[SQUEAKING]

[RUSTLING]

[CLICKING]

IAN BALL: So today I want to talk about-- we're going to move to the final phase of the course, which is going to be on dynamic Bayesian games. But I want to start with maybe one example of a VCG mechanism, because I think, yesterday, that was maybe a bit abstract and a bit hard to follow. So let's start with one final example. And it's going to be a VCG mechanism, or maybe we'll call it a VCG auction for two items.

So a lot of the classical work that we did so far in the course on auctions, we assumed there was just one item being auctioned off. Then we looked at the ad auction setting, which had this special structure where there was this ranking between the items. Here, we're just going to have two abstract items.

So I'm going to call these maybe item A and item B. So what the auction needs to do is it needs to solicit bids. Maybe I'll write what it's going to do. It's going to collect bids from the bidders.

And then, based upon these bids, as usual, we have to decide two things. We have to decide how we allocate the goods and how much everyone pays. So then it specifies an allocation, which, in this case, it says who gets item A if someone gets it and who gets item B if anyone gets it.

And then, in addition to the allocation, we have to specify a payment by each bidder. Now, I put bids in quotes because the kind of surprising or maybe somewhat confusing aspect of a VCG auction is that a bid is essentially a report or claim about how much you value the items. So you think of my bid as the auctioneer saying, tell me how much you value everything.

And then the bidders say, I value it this much. And that may seem strange, but if you think about a second-price auction, that's exactly what happens because, in a second-price auction, people are bidding their true valuation. So it's as if the auctioneer is saying, tell me your valuation. And then the bidders are submitting their valuation.

So let's put a little more structure on this. So let's imagine that we have two bidders. And it's a little complicated to specify the whole auction because we have to specify the allocation and payment for any vector of bids that's submitted. So let's just go through one example. Let's look at some vector of bids that could be submitted.

So, well, we first have to say, well, what is a valuation? What do I need to tell the auctioneer if the auctioneer is asking me how much I value things? Well, I actually have to say three things.

I have to say, how much do I value item A if I just get item A? How much do I value item B If I just get item B? And then how much do I value getting both items together?

And depending on these items, it may not be the value of the bundle of the two items is exactly the same as some of your valuations for the two items. If it's shoes, a left shoe and a right shoe together are much more valuable than the two shoes separately whereas other items that are substitutes, the bundle may be less valuable than the sum of its parts, if there's some duplication between the two parts. Maybe 2 apples is not twice as valuable as 1 apple.

So we're going to think of it as a table where my valuation is going to say, if I get item A alone, what's my value? If I get item B alone, what's my value? And if I get the bundle, what's my value?

So bidder 1 is going to make some report. And bidder 2 is going to make some report. I've made the numbers here so it works out.

So let's suppose that the report beta 1 makes is they say, if I get only item A, I value that at 5. If I get only item B, I valued at 3, and if I get both items, I valued at 8. So for bidder 1, it turns out, their value of the bundle is exactly the sum of the value of the two parts, but it doesn't have to be that way.

Now let's look at bidder 2. Let's suppose bidder 2 says, I value item A alone at 2. I value B alone at 6, but I only value the bundle at 7.

So notice for bidder 2, the bundle is not worth as much to them as the sum of its parts. If they just got A and if they just got B, that's different from if they get the bundle together because, for them, these two items, maybe if item B is a better version of item A, then getting both of them together is basically just as good as getting item B.

So let's suppose these are the reported valuations for the two bidders. And now we have to specify who gets and what the payments are. So let's go through this. Maybe we'll do allocation over here.

So remember, the rule of the VCG mechanism is that we always allocate things efficiently. So that is, we want to maximize the sum of the players' utilities. So we need to think, who should item A go to? And who should I be go to? Any thoughts here on what would be the efficient way of allocating things, given these reports?

Remember, when we say we allocated efficiently, what we mean is, we collect the reports, and we allocate in a way that would be efficient if these were the players true valuations. So taking this as the truth, let's think about the efficient way of allocating it? So how would you do this? Yeah.

AUDIENCE: You could [INAUDIBLE] person 1.

IAN BALL: Right, that seems like a good way of doing it. So let's see. If we give item A to person 1, they get a valuation of 5.

If we give item B to person 2, they get evaluation of 6, which means the sum is 11. That sounds pretty good. If we did the opposite, we would only get a sum of 5.

And if we gave both items to the same player, we'd either get 8 or 7. But that's still less than 11. So this is going to be the efficient allocation. This is going to be the allocation that maximizes the sum of the reported utilities.

But now we have to do the payments. And I think the payments part is a bit trickier. So we have to specify two payments, t1 and t2.

And maybe if I'm more precise, I would say this is t1 of v1 hat, v2 hat. And this is t2 of v1 hat, v2 hat. So what the mechanism is saying is if these are the submitted valuation reports that the two players submit, this is how much I'm going to ask player 1 to pay. And if these are the valuation reports that are submitted, this is how much I'm going to ask player 2 to pay.

And the full mechanism is going to have to specify a function that says what these payments are for any submitted reports, but writing out that full function's a bit complicated. So we'll just do this example. Now, what we said in class last time is what bidder 1 is supposed to pay is the externality that they impose on the other bidders.

So we have to compute two things. We have to say if we just cared about the other bidders and we allocated the items to maximize the utility of the other bidders, how well would they do? And then we have to subtract from that how well the bidders actually do under this allocation.

So let's start with the first part. And this is how well the other bidders could do. So let's say I'm bidder 1. These are the submitted valuations.

If we gave bidder 2, if we just allocated things in the best way for bidder 2, what would bidder 2's valuation be? Well, what we would do is we would give both items to bidder 2. And if we gave both items to bidder 2, then bidder 2's valuation would be 7.

So we're going to first start with a 7 here because this is how well the other bidders could do optimally if bidder 1 weren't here, or in other words, if we didn't care about bidder 1 and we only tried to maximize bidder 2's payoff. Now we want to subtract from that how well bidder 2 actually does. How well does bidder 2 actually do under the given allocation? What's their actual valuation? 6, right?

So we subtract 6, and we get 1. So this is the payment that bidder 1 is asked to make. Why is it 1? Well, because what is the effect that they're having on the other bidders?

The other bidder, bidder 2, could've gotten the bundle A, B, but because of the presence of bidder 1, bidder 2 is only getting the item B by itself. And the difference in utility for bidder 2, between the bundle and getting only B by itself is exactly 1. And that's why we get this.

Now, let's do it for bidder 2. So if I'm bidder 2, we first need to compute how well bidder 1 would do if bidder 2 weren't here. If bidder 2 weren't here, bidder 1 would get everything, and their utility would be 8.

How well does bidder 1 actually do? Well, bidder 1 actually gets item A, which gives them a utility of 5. So we get 8 minus 5. And we see the payment is 3.

Why does bidder 2 have to pay more? Because bidder 2 has a bigger externality on bidder 1. Bidder 1 would've been able to get everything together A and B, but because of the presence of bidder 2, they only get item A by itself, and the difference in utility between 8 and 5 is 3, which is a larger number. And this is going to-- this completes the specification of VCG for this particular profile [INAUDIBLE].

And the result that I didn't prove but that I claimed yesterday is that if you use this kind of rule, in a VCG auction, it's going to be optimal for each bidder to report truthfully. So I should maybe make a note on this. There's an optimal strategy for each bidder. And what is that strategy? Well, it's reporting truthfully.

And we might call that vi hat equals vi. So whatever my true valuation is, I look, and I say, this is how much I really value A, this is how much I really value B, and this is how much I really value the bundle. It's optimal for me to tell the designer exactly how I really feel and be honest about it.

And I can't benefit by misreporting my valuation to the designer. All right, any questions on that? Yes.

AUDIENCE: Will you ever have a situation where your externalities are going to be less than the valuation that you have for the good? That payment would be negative in essence?

IAN BALL: Good question. So you're never going to have a case where the payments are negative. And this follows from the way that we define an externality. So, yeah, we have to be a little careful about this.

So it is true that, more broadly, in economics, when we talk about externalities, some things do have positive externalities. There are things I can do that make activities I can do that make other people better off. So I guess I should be maybe a little careful about the terminology of externality.

So here, what we mean in this context by externality is, well, we're comparing how well the other bidder would do if we just cared about maximizing that bidder's utility to how well that bidder actually does under this allocation. And there's no way that the current allocation can give more to the other bidder than if we just maximize that bidder's utility. So for that reason, this number is always going to be nonnegative.

And you're right, if you're thinking of other externality context, this is a bit more special. So it does turn out here that the payments will always be nonnegative. They could be 0, but they would never be negative because, in order for them to be negative, it would have to be that the allocation we're choosing makes the other bidder strictly better off than the best allocation for that bidder. And that goes against the definition of best. Yeah.

AUDIENCE: 7 is maximizing their utility when everyone besides myself is in the game? Or is it just their utility alone?

IAN BALL: Good question. So the 7 here is-- what we're maximizing is the sum of everyone else's utility, and we're ignoring my own utility. So in this case, because there's only two bidders, the sum of everyone else's utility is just maximizing the other bidder's utility.

But if we had three bidders, then we would think about how could we allocate the goods in a way to maximize the sum of, say, bidders 2 and 3's utility, ignoring the effect of the presence of bidder one. Yeah, great. Any other questions on this?

So now let's change tacks a little bit. And we're now going to move to the final phase of the course, which is called dynamic Bayesian games. And I think it's good at this point to pause and see where we've come. So we can think of the course as organized into four parts. We have looked at static games and dynamic games.

And we've also looked at games of complete information and games of incomplete information. And basically, so far, we've filled in three of these boxes, and now we're going to fill in the fourth box. And for each box, each class of games, we have a different solution concept, a different equilibrium notion. So let's fill this in.

So in the first part of the course, we started here. This would just be your basic strategic-form game with complete information. And our solution concept there was just Nash equilibrium.

Next, we said, well-- we looked at extensive-form games where players make moves over time. And we said, one problem of the Nash equilibrium is it allows players to make non-credible threats. And we said we wanted to refine Nash equilibrium and impose a stronger solution concept that rules out threats that are not credible. And that solution concept was called subgame-perfect Nash equilibrium.

So that was a stronger notion than Nash that ruled out some Nash equilibria that we didn't find compelling because they involved threats that were not credible at nodes that were not reached in the game.

Then we stepped back, and we said, wait a second, our complete information seems like a really strong assumption. A lot of times when people come to the game and they start the game, they know things that other players don't. They know things about their preferences that other players in the game don't.

So we introduced Bayesian games, that is, games that have incomplete information, and we introduce a solution concept for that called Bayes-Nash equilibrium or Bayesian Nash equilibrium.

Now this dealt with the issue of incomplete information. But it turns out that Bayesian Nash equilibrium doesn't deal with the issue of non-credible threats. And if we want to look at dynamic games with incomplete information, we have to both take into account the fact that players have incomplete information.

But we also may want to rule out equilibria that involve non-credible threats. So we're going to introduce today our final solution concept, which it's not very creative. We basically combine the perfect from here and the Bayes from here, and it's just called perfect Bayesian equilibrium.

I guess maybe I'll give somewhat of a Warning about this. These first three concepts are very-- are well-agreed upon. Everyone agrees what's the right notion. Once we get to dynamic games of incomplete information, there's been a lot of recent work arguing about what's the right notion of equilibrium.

So we're going to present a particular notion called perfect Bayesian equilibrium. But there's 10 other equilibrium notions that could fill this box that are more exotic and involve more constraints and people are arguing to this day about what's the right way to define equilibrium. So I'll just warn you that things are a little maybe trickier or less clear cut in this case, but we'll try to focus on well-trodden ground in this class.

OK, so I want to first, today, before defining perfect Bayesian equilibrium, I want to give a few examples that try to motivate why we need a new definition, why our existing equilibrium notions are insufficient. And then we'll go over how we can actually solve for perfect Bayesian equilibrium and go over the formal definition. So let's start with maybe a reminder about non-credible threats and a weakness of Nash equilibrium with complete information, and then we'll build on that.

So let's go back to maybe our classic example where we had two firms. We had firm 1, and we had firm 2. And we'll think of firm 1 as a potential entrant. And we'll think of firm 2 as an incumbent.

So firm 2's already in the market. Firm 1 is deciding whether to enter the market or not. So they're called an entrant, but really, they're a potential entrant. They may or may not enter.

And firm 1 is choosing whether to enter the market-- we'll call that E-- or not enter the market. We'll call that N. And this is an example we've done before.

Firm 2 is the incumbent. If firm 1 enters, firm 2 is the incumbent, can choose whether to accommodate their entry or to fight them by starting some price war or trying to cut prices and doing something that hurts both players. So firm 2 is choosing A or F, a for accommodate the new entrant and f for fight against the new entrant.

So let's write this as an extensive-form game. We have firm 1 is here. They're choosing whether to-- maybe I'll put N here, not enter or enter.

And then if they do enter, we have firm 2. And they're choosing to either accommodate the entrant or fight. Let me make sure I get the numbers right. So what we did here, let's put 0, 5, 1, 1, and negative 10, 10 to make this easy.

So what's the idea? If firm 1 does not enter, well, that's great for firm 2, the incumbent. They're the monopolist.

They're all by themselves. So they get a payoff of 5. Firm 1 isn't entering, so they get a payoff of 0.

If firm 1 does enter, well, then the payoffs of the firms depend on whether to accommodate the entrant or fights. If they accommodate, then they both do pretty well. They both get 1, but of course, not as well as firm 2 would've gotten if firm 2 were by himself as monopolist.

And then if firm 2 fights, that's really bad for both sides. They a price war. Prices go, say, really low. And this is bad for everyone.

We're thinking of fighting in an abstract sense. It could literally involve violence, or it could be cutting prices, or it could be flooding the market. There's a lot of things it could be.

So let's now just write this as a bit of a review today of our standard strategic form game box. Player 1 has two strategies, N, E. Player 2 has two strategies, A, F. N, E to keep the order consistent.

And let's see. If player 1 doesn't enter, then they both get 0, 5 either way. If player 1 does enter, then the payoffs depend on what player 2 does. And we get 1, 1 here and negative 10, negative 10 here.

Well, now let's just do our standard approach to try to find the Nash equilibria of this. This is a good little review. The final exam is going to be cumulative. So you do have to remember how to solve the basic games.

So let's see. We go down here. If firm 2 is going to accommodate, then it's better for firm 1 to enter. If firm 2 is going to fight, it's better for firm 1 to not enter.

And then we can do things the other way. If firm 1 doesn't enter-- actually, it doesn't matter what firm 2 does. If firm 1 enters, then firm 2 would rather accommodate. So we get two equilibria. We get EA and NF, which we can see on our graph here.

Now, both of these are Nash equilibria. But we precisely introduced subgame-perfect Nash equilibria as a more demanding solution concept to rule out one of these equilibria. So which of these equilibria is not a subgame-perfect Nash equilibrium? Which one did we-- does subgame perfection rule out? Yeah.

AUDIENCE: [INAUDIBLE] because you can't really--

IAN BALL: Yeah, sorry, NF or EA. It's one of these two. It's EA and NF.

AUDIENCE: Yeah.

IAN BALL: Maybe, yeah.

AUDIENCE: NF.

IAN BALL: NF, I agree, OK. And why?

AUDIENCE: Because if the entrant enters, it doesn't make sense for the incumbent to fight.

IAN BALL: Exactly right. So both of these are Nash equilibria. Only EA is a subgame-perfect Nash equilibrium. As you said, NF is not subgame-perfect.

And it's not subgame-perfect because it involves a threat by firm 2 that is not credible. Firm 2 says, if you enter, I'm going to fight. But if firm 1 actually enters, well, firm 2 is here, and they look, well, wait a second.

If I fight, I get negative 10. If I accommodate, I get 1. I'm much better off accommodating. So it's not optimal for firm 2 to actually carry out this punishment or this threat that they threatened to carry out.

Formally, well, we have to think of the definition of subgame-perfect Nash equilibrium. In the definition, we have to look at every subgame. We have the full game here, but we also have subgame that starts here.

So remember, this is one subgame of the full game. And the problem with NF is that fighting is not an equilibrium of this subgame. Now, it may be weird to think of this as subgame because there's only one player moving.

But remember, we can have a game with a single player, and an equilibrium of a single-player game just means that player's behaving optimally. So here F is not a Nash equilibrium of this subgame because F is strictly worse for player 2 than A. So this doesn't satisfy our definition of subgame-perfect Nash equilibrium.

But let's maybe change the game slightly. So what if we change the game and say, well, actually, firm 1 has two different ways of entering. They can enter the market and produce one kind of good, or they can enter the market and produce a slightly different kind of good. So we're now going to say that firm 1 is choosing between not entering or I'll call it E1 or E2. These are just two different ways they can enter and maybe use a slightly different marketing strategy or slightly different pricing strategy, but they're effectively entering either way.

And now let's write out how this looks. So we have firm 1 here. Firm 1 is now choosing enter, not enter-- sorry, not enter, E1 or E2.

Again, if they don't enter, the payoff is 0, 5. If they enter in this way, firm 2 chooses to accommodate a fight. They enter in this way. Firm 2 chooses to commentator or fight. This is still firm 2, firm 2.

And maybe the exact pricing strategy they use makes a bit of a difference. But just to make things simpler, let's say it doesn't really affect their payoffs whether they sell them in one way or a different way or whether they use this advertising firm or this other advertising firm. So let's say the payoffs down here are going to be 1, 1, negative 10, negative 10, 1, 1, negative 10, negative 10.

So these are different ways of entering. But the effect on the payoffs is actually 0. It doesn't actually change the way the payoffs are. And we could change these slightly if we wanted, but it doesn't matter.

And notice, the way I'm writing this, just let me slightly modify this. We have to make a choice, does firm 2 observe the advertising strategy that firm 1 is using? And we're going to assume no.

So we're going to put this in the same information set. So as firm 2, all I see is that firm 1 has entered. I don't really know under the Hood exactly what, say, production process they're using or what advertising firm they are.

And now let's think of the strategies in this game. It turns out the strategies in this game are exactly the same as the strategies over here. Firm 1 is still choosing-- sorry, for firm 1, it's slightly different.

So firm 1 is choosing between N, E1, and E2. And firm 2 is still choosing accommodate or fight. It's the same for firm 2, but it's slightly different from firm 1.

Let's try to compute the Nash-- let's try to find the Nash equilibria of this game. And the idea is what are some Nash equilibria of this game? It looks really similar. So I think it shouldn't be too hard to figure this out.

Well, let's organize it by what firm 2's doing. Let's say firm 2 is accommodating. And notice firm 2 makes a single choice-- accommodate or fight. It's a single choice because these two nodes are in the same information set. So I can't choose to accommodate E1 and fight against E2 because I can't distinguish those.

So if firm 2 accommodates, what does firm 1 want to do? Yeah.

AUDIENCE: E1 or E2.

IAN BALL: E1 and E2 is going to be good for them. They want to enter, and they're happy to do either one. So we're going to get E1A and E2A are both going to be Nash equilibrium.

What if firm 2's going to fight? What does firm 1 want to do? Yeah.

AUDIENCE: Not enter.

IAN BALL: Not enter. So now it looks similar to before. But now we actually have three equilibria. Same idea, if firm 1 is going to accommodate, then-- sorry.

If firm 2 is going to accommodate, then firm 1 wants to enter. They're happy to enter either way. If firm 2's going to fight, then firm 1 doesn't want to enter.

Now, I think that NF still feels like a non-credible threat. For exactly the same reason as before. Firm 2 is fighting.

They know that, by fighting, they get a payoff of negative 10, and if they didn't fight, they would get a payoff of 1. So to me, this feels like a non-credible threat. But the question is, is NF a subgame-perfect Nash equilibrium of this game?

What do people think? Well, the first question we always have to ask ourselves about subgame-perfect Nash is, first, we have to identify the subgames. Remember, maybe we're going a little bit back. We need to review a little bit. So if we go all the way back, what are the subgames of this game here?

AUDIENCE: Only if the entire game is a subgame because the decision point for player 2 is not a singleton node.

IAN BALL: Exactly right. So we can't start a subgame here because we would break this information set. We can't start here because we would break this information set. And we can't include both of them as part of our subgame because we had a rule that a subgame has to start with a single node.

So the only subgame of the full game is just the full game itself. And if there's no proper subgames, no subgames other than the game itself, then subgame-perfect Nash equilibrium doesn't have any additional bite because the additional strength of subgame-perfect Nash is it says, in every subgame, we must be playing a Nash equilibrium. But if there's no other subgames, well, it doesn't have any bite. So Nash equilibria and all of these-- all three are subgame-perfect.

So this seems like maybe a problem. It shows that our notion of subgame perfection doesn't have the bite to rule out non-credible threats in games like this where we have multiple nodes in the same information set. And we're going to introduce our notion of perfect Bayesian equilibrium to rule out these kinds of non-credible threats, even in situations like this. So let's maybe go through a more involved example.

So let's think of, maybe we'll call this, an application game. People have probably applied to internships or jobs or schools like this. And sometimes, there's an optional essay or an optional interview or something optional.

And often, people feel it's not really optional. If I don't do it, I'm not going to get the job. I'm not going to get in. Well, let's try to explore that in this game.

So let's say, well, it's an application game. We have an applicant. And we could call it a firm or a college or a school. Let's say school. Let's use school just in this example, so we have a school.

And the applicant, let's assume, has private information. In this example, it could be private information about how strong of a candidate they are. But maybe, here, just to make it a little nicer, we'll say the applicant privately knows how excited they are about going to this school.

So the applicant has two types. We'll call them types. Maybe t excited or passionate, let's say P, tP or tN. So tP says, I'm really passionate about going to this school about the thing that this school does.

I'm I want to go. And tN doesn't really-- they're not quite as excited. They still want to go, but they're not quite as excited.

And then we have the school and the school-- let's just say in words, what they do is they want to accept tP but not tN. So there's two kinds of students, two kinds of applicants. I'll be more precise about what the utilities are, but just at a high level to understand, if someone's really passionate and excited about coming, they want to accept that person. If they're not excited, they don't want to accept them.

So let's try to-- let's draw out a game tree here. So the applicant has private information at the beginning of the game about whether they're excited or not about going to the school. And the way that we model this is we start with a move by nature.

We first assume that nature makes a move and draws this applicant's type. And we need to specify some probabilities. So let's say that they're passionate with probability 1/2, and they're not passionate with probability 1/2, just to make the math a little bit easier.

So I need a new board just to-- this is going to get a little big, this game. Let's start here. So we start with nature.

And nature is first going to draw whether this person is passionate or not. So we're going to have nature making this first move, probability 1/2, and probability 1/2. And this is going to be the passionate student. This is going to be the non-passionate student.

And then the student or the applicant-- I'm calling them the applicant, so I'll say A-- is going to make a choice. They can either submit the optional essay or not. They're applying to this job-- or this school.

There's some optional essay. So let's write here, let's say I can include the essay or not. Maybe I'll use these words. So they're applying either way, but they're just choosing whether to submit the essay.

And now we have the school is making a choice whether to accept or reject the candidate. Let's say accept, reject, and let's save some space over here, A, R. OK.

Now, we said the school would rather accept someone who's passionate than someone who's not. And let's say that a passionate candidate gives the school a utility of 3, but someone who's not excited gives the school a payoff of negative 2, if they accept them. And if they don't accept them, they just get a payoff of 0.

So let's go and fill that in here. On this side of the tree, the student is excited. They're passionate. So if the school accepts them, the school is going to get a payoff of 3.

And if they reject them, the school's going to get a payoff of 0. I'm writing a comma here because we're going to fill in the applicants payoffs next. Yeah.

AUDIENCE: Is there a reason why there are no-- [INAUDIBLE]?

IAN BALL: Yes, I going to fill it out, but maybe I'll-- let me fill out the payoffs, and then I'll do that. So I'm going to fill in some information sets next. And then if we go over here, if we accept the not very excited student, then we get negative 2, negative 2, and then 0, 0.

But as you pointed out, the school only observes whether the applicant has submitted the optional essay. They don't directly observe whether the student is excited or not. So we're going to need to define our information sets like this. Here. And like this.

So the school has two information sets. At this information set, I know that you've not submitted the essay, but I don't know your true enthusiasm. At this information set, I know you've submitted the essay, but again, I don't know your true enthusiasm.

Now we need to specify the student's payoff. And let's start, when the student doesn't submit the essay, let's suppose that if they're rejected, they get a payoff of 0. And actually, let's assume that everywhere.

So all I've said here is whenever the student is rejected, they get a payoff of 0 because they want to be accepted. Now, if they do get accepted, let's suppose that-- so let's talk about the value of acceptance. Well it's 2 to the passionate student. It's only 1 to the less passionate student.

And let's assume that writing the essay is costly. So writing the essay also has a cost of c. And I'm just using c so I can vary this over the game. But c is a known parameter. We all know c is just like a number like 1 and 2, but I just want to have some flexibility over what this is.

So let's see if we can fill this in. So down here, the student is passionate, and they don't submit the essay, and then they get accepted. Yeah. Question.

AUDIENCE: I guess we'll modify it, but for the one where the student writes the essay and is rejected, would it be [INAUDIBLE]?

IAN BALL: Good, good, yes, we should do that. Good point. They still had to write it. Yeah, great point. Let's do that.

Good point. Thank you. OK, thanks. So let's have that in here.

And now we go over here. So let's see. The student is passionate. They don't write the essay, and then they get accepted.

So they get 2 for being passionate, and that's it. Over here, the student is passionate, but they have to write the essay, and then they're accepted. So we get 2 minus c.

Then we go over here. The student is not passionate. They don't write the essay, and they're accepted. So they get 1.

And then, over here, the student is not passionate. And they do write the essay so they get 1 minus. It's maybe getting a little messy, but I think we have everything. Maybe I'll put parentheses around these things just to make it a little clearer.

OK, so here's our game. Let's go back over here. Start analyzing.

So we're going to look at a few different versions of the game and a few different kinds of equilibria. But let's first assume that c is less than 1. so this means even if I'm not passionate, if I'm going to get accepted, it's still worth writing the essay. The benefit of being accepted outweighs the essay cost.

And let's see if we can look for an equilibrium where every student applies and writes the essay because I think this is often what we see in practice. Even the students who aren't that excited, they write some essay about how excited they are to go to the school. So let's look for an equilibrium. In which both types of students submit the essay.

So formally, let's look at the student strategy. Remember, a student strategy has to say what the student does as a function of their type, as a function of their private information. So we have the passionate student and the non-passionate student, and they both choose the essay. It's pretty easy.

And now we have to go to the school and see what they do. So let's put ourselves in the position of the school. And let's look at the information set where they have received an essay. So remember, for the school, their information sets are essay and no essay.

So the school needs to decide what to do if they see that the student has submitted an essay and what to do if they see the student has not submitted an essay. And if we think about our experience with how these things work, what is a guess about what the school strategy might be if we want to try to look for this equilibrium where all the students are writing the essay. What's your guess about what the school might do?

If they see that you submitted the essay, what are they going to do? They're probably going to accept you. And if they see you didn't write the essay, they're going to reject you.

Let's see how this works out. So first, we need to say, well, let's suppose they see that you've written the essay. What is the school believe about your passion for the position. Do they think you're passionate or not passionate? What probability do they assign?

1/2, 1/2, so technically, we're using Bayes' rule, but here it's pretty simple because the prior is 1/2, 1/2. 1/2 the students are passionate. 1/2 the students are not passionate. If all the students submit the essay, well, then the students who submit the essay are also 1/2 passionate, 1/2 not passionate. Nothing has changed.

But what does the school believe if they see that you didn't submit the essay? I don't know, what do you think. This is a bit trickier.

Should it be 1/2, 1/2? Why should it be 1/2, 1/2? What do you think?

There's not one right answer to this. So I think it's unclear, but, Amy, you said 1/2, 1/2, why?

AUDIENCE: I guess the idea is symmetric to the other one where everyone's-- no one [INAUDIBLE] essay, but-- wait. So if everyone submits essay [INAUDIBLE] 1/2, 1/2. And then that's like, you have a student that you don't receive an essay from would also be 1/2, 1/2 because, basically, the way that Bayes' rule plays out, they're all the same per person for other cases.

IAN BALL: OK, so let's maybe write out-- let's get another thought, and then let's write out Bayes' rule and see what happens here. Were there any other thoughts on what we do here? Yeah.

AUDIENCE: I mean, just like qualitatively, then you assume that there's a greater chance that they're not passionate about [INAUDIBLE].

IAN BALL: That also seems reasonable. I think there's a lot of reasonable answers to this, but let's see how Bayes' rule works. So to apply Bayes' rule--

AUDIENCE: [INAUDIBLE]

IAN BALL: Yeah, so I think we're going to have an issue when we apply Bayes' rule. So let's see what happens. So remember, just a quick review of Bayes' rule.

Bayes' rule allows us to calculate conditional probabilities. So it allows us to say something like if we have two events, A and B, what is the probability that event A occurs conditional on B occurring. And the rule for Bayes' rule is that this is the probability of A and B occurring divided by the probability of B occurring.

And this basic picture that we give, this is A, and this is B. Well, what's the probability of A occurring given B? Well, if we know that B occurred, we know we're somewhere in this circle.

The probability that A occurs given B is saying what's given that we're in this circle B, what's the likelihood that we're in the shaded region. And that's exactly what's being pictured here, P of A intersect B is the probability that we're in this shaded region. And probability B is just the probability that we're in this outer circle here.

So let's try to apply Bayes' rule when we see that the student writes the essay. So what we're interested in is, given essay, we want to say, what is the probability that the student is passionate, tP, given that they write the essay? And Bayes' rule tells us this is equal to the probability that they're passionate and they write the essay divided by the probability that they write the essay.

Now, the probability that they write the essay is a little hard to calculate. Let's break it down. So write this one more time-- maybe I'll write it over here to give myself more space. This is the probability that they're passionate and they write the essay divided by-- well, there's two ways they could write the essay.

They could be passionate and then write the essay, or they could be not passionate and write the essay. So let me break it down like that. We have the probability of tP and essay plus the probability of tN.

So let's see if we can fill this in. What is the probability that they're passionate, and they write the essay? Well, they always write the essay. So this is just the probability they're passionate, and that's 1/2.

And on the bottom, again, we have the 1/2 plus we're going to get a second 1/2 because the probability that they're not passionate and they write the essay, well, if they're not passionate, they always write the essay. So this is just the probability that they're not passionate, which is also 1/2. So we get 1/2 divided by 1/2 plus 1/2, and that's going to be 1/2.

I think it's good to go over this carefully once. And then I think people get the hang of it. Any questions on this? Two questions, OK, great. Yeah.

AUDIENCE: [INAUDIBLE]

IAN BALL: This is just confirming this. Yeah, yes.

AUDIENCE: So then in the other case--

IAN BALL: So let's go to the other case, yeah, yeah. So maybe I'll erase our little review and go to the other case. So now we want to say, given that we see a student-- we didn't expect to see this. We thought everyone was supposed to write an essay. That's what they're supposed to do under their equilibrium strategy.

But now we see that someone submits an application with no essay. And we say, well, what belief should we form about this student? Well, let's apply Bayes' rule.

We say, what's the probability that they're actually passionate, that I want to accept them, given no essay? And that's exactly equal to the probability that they're passionate and no essay divided by the probability of no essay. And what is this? Yeah.

AUDIENCE: 0 over 0.

IAN BALL: This is 0 over 0. Because I'm never-- no one is writing no essay. Everyone is writing an essay.

So the probability that you're passionate and you don't write an essay is 0 because if you're passionate, you always write an essay. The probability that you don't submit an essay, well, if you're passionate, you always submit an essay. If you're not passionate, you always submit an essay. So we get a 0 here as well.

So what we see is that there's a fundamental problem with Bayes' rule, that Bayes' rule only allows us to define conditional probabilities conditional on events that have positive probability. When we try to define conditional probabilities conditioned on 0 probability events, Bayes' rule doesn't really apply, and we run into problems. So I'll say not well defined.

And I think this is exactly why we got a few different answers here. What should we conclude when someone doesn't write an essay? I think there's a few reasonable choices.

You might say, well, this isn't supposed to happen. So let me just go back to my prior and say it's 1/2, 1/2. You could say, well, the fact that they don't want to write an essay, that must mean they're not really passionate about the program.

I'm going to assume-- I'm going to say that they're not passionate. Or maybe you think they are passionate, and they just-- I don't know, you could tell a lot of stories. But basically, we don't really pin down what the belief is.

But now this creates a problem because now the question is-- so this equilibrium seems to fit. Let's see if this makes sense. If I'm the student, if I believe that the school is going to accept me if I write an essay and not accept me if I don't write the essay, then it's optimal for me to write the essay. That makes sense.

And given that all the students are writing the essay, well, the school is happy to accept you because the school gets a payoff of 3 from accepting a passionate student, negative 2 from accepting a not passionate student, And the average of 3 and negative 2 is strictly positive. So everything looks good here. The only question is, is this a credible threat?

But there's a question here. Yeah.

AUDIENCE: I was just curious. If you were to get someone who submits an essay, could that also imply that maybe there's a third type of student that you don't know of?

IAN BALL: That's a good point. So maybe you want to revise your model of the world. And maybe our types, we've only represented types by how excited they are about the program, but maybe there's a laziness component that we also need to model.

Or-- and we'll talk about this later-- it could be that this cost c varies in the population. Again, talking about essay costs, I feel like it's 0 for everyone in the world of GPT, but in the old days, there was a cost. And some people were better at writing essays and the cost for them was lower, and we're going to talk about that later. Yeah, exactly.

So I guess the basic question here is, is this a credible threat? Is it credible for the school to say, as they do in this equilibrium, if you don't submit an essay, we're going to reject you? Yeah.

AUDIENCE: Well, isn't it kind of murky because, as you were just talking about, there are some ways you could just say, if you don't write an essay that you're not passionate, so we don't want you. But what if the school has a really different way of thinking like, oh, he didn't write the essay. That's because you were spending more time studying for an exam. So then we're actually more excited, so I think because you don't know exactly what the school will conclude, the threat is-- it's not necessarily--

IAN BALL: I think that's a great answer. So it's definitely murky. And that's exactly what I'm trying to convey with this example, that it is murky.

But I think the key point, which you hinted at, was whether this threat is credible depends on what the school believes when they don't see an essay. So is this threat credible? Well, it depends on the belief.

If the school believes that everyone who doesn't submit an essay is passionate, then it's not credible for them to say, we're going to reject those students because they'd rather accept passionate students than not. But if their belief, after not seeing an essay, is that the student is not passionate, then it is credible for them to reject that student because that's exactly what's optimal for them, if they see that a student-- if they believe that a student is not passionate.

So how do we get around this? Well, beliefs aren't well defined, but whether something is credible depends on the belief. The way we get at this is our equilibrium notion is going to need to explicitly specify what beliefs the school forms, even at this information set. So as a result, beliefs must be part of the equilibrium.

Why? Well, specifying strategies alone is not enough to tell us what people believe. But our whole equilibrium notion is supposed to capture whether these threats are credible. And therefore, we're going to need to explicitly specify what the belief is.

And what we'll see here is that, here, there's no flexibility in what belief we specify here. In this case, the belief is pinned down by Bayes' rule. Bayes' rule tells us what the belief has to be.

Here, we can choose whatever belief we want. There's no further restriction imposed by Bayes' rule on what the school's belief is here. Now, I said before, there's all this big debate in the literature about how to define it.

People came along and made all these arguments about what belief should be at this kind of information set. What should you believe here? For now, we're not going to impose any restrictions on what those beliefs are. And now I'm going to go into the formal definition. Let's go back over here.

Let's define perfect Bayesian equilibrium. So the first key point is that one thing I've harped on throughout the semester is no longer true. So throughout the semester, I've said, an equilibrium is a strategy profile. And I said this over and over again.

Now, that's no longer true because equilibrium is going to be a strategy profile together with beliefs. So we're going to have to have this new object, which is going to be called an assessment. I think this is a silly word, but this is the word that's used.

And that specifies two things. It specifies a strategy profile. And also beliefs, sometimes it's called a belief system.

So strategy profile, we already know what that is. It specifies a strategy for each player. And a strategy in this game and in any game is a mapping from information sets to actions.

It's a complete contingent plan that says that my information set and any of my information sets, what am I going to do? So in this game, it's exactly what we wrote down over here. The student has two information sets tP and tN.

So a strategy for the student specifies what they do at tP and what they do at tN. Similarly, the school has two information sets. The information set where they observe the student wrote an essay and the information set where they observed the student didn't write an essay, and a strategy specifies what action they take at each of those informations. So I said that in words.

Now, a belief system is going to specify-- specifies a belief at every non-singleton information set. So a lot of the information sets are singletons. I don't need to specify a belief here. I only need to specify a belief at an information set that contains multiple nodes.

And what is a belief? A belief is going to be a probability distribution over the nodes in that information set. And that's exactly what we did here. We said if the student submitted an essay, we had two nodes in that information set.

And we had to form belief about what was the relative likelihood of those two nodes. And similarly, at this information set, no essay-- or maybe I got it backwards. At the essay information set, we have to form a belief about the two nodes within that information set.

So a belief at every non-singleton information set. What is a belief? i.e., a probability distribution over-- well, it can't be over all nodes.

I certainly can't assign positive probability to nodes that are not in that information set. It's only a distribution over the nodes that are within that information set. So it's a probability distribution over nodes in the information set.

So now we've defined an assessment. Let's formally define a perfect Bayesian equilibrium. So definition, an assessment is a PBE if it satisfies two conditions.

The first condition is sequential rationality. And the second is belief consistency. Let's start with belief consistency.

This is what we're getting at here. We want to say that the school's belief here must be 1/2, 1/2. But the school's belief here is totally flexible.

So what belief consistency says is, well, the beliefs are consistent. The beliefs are consistent with Bayes' rule.

Well, they can't be consistent with Bayes' rule everywhere. Sometimes we use this vague phrase wherever possible. There's often different ways of defining exactly what wherever possible means. What I mean by wherever possible is at any information set that's reached with positive probability, so i.e.

Now, let's be careful-- the probability that we reach an information set depends on the strategy profile. So what we're saying is we have an assessment, looking at the strategy profile that's part of that assessment, I can go every information set. And I can compute the probability that we actually reach that information set under the strategy profile.

And I can split those information sets in two. I can say there's some information sets that are reached with positive probability. And there are some that are never reached.

And the ones that are reached with positive probability, the beliefs that those information sets must satisfy Bayes' rule. But at the other information sets that are not reached with positive probability, I have total flexibility over those beliefs. There's no restriction on the beliefs there.

So now sequential rationality. We're close. Well, the whole point of introducing beliefs was to determine whether these threats were credible.

And what credible meant, precisely, was optimal. Optimal given my beliefs and given the strategy of others. So sequential rationality means, at every information set, the action under the strategy profile is optimal, given beliefs and the rest of the strategy profile. And let me say the rest.

Put this in quotes. So what do I mean here? Whenever we get to an information set, a strategy is going to specify what action the player is supposed to take at that information set. And that action that they're supposed to take better be optimal.

Well, how do we define that it's optimal? Well, we have to check that it's optimal given the beliefs that the player has there and how everyone else is going to play, which is the "rest of the strategy profile." So what's optimal for me at this node depends on-- or at this information set depends on my beliefs over which node I'm actually at and also, how people are going to subsequently play in the game.

And the way they subsequently play is pinned down by the "rest of the strategy profile." I could write all this in math, and that's what the notes do, but I think that maybe just makes things a little more confusing. So I'm trying to write it in words, but if anything's unclear, let me know.

Questions? So let's go back to this game and see if we can find some more PBE. So let's look here.

So let's try to finish off our construction of a perfect Bayesian equilibrium over here. We already have a strategy for the student. We already have a strategy for the school.

The next step is we have to specify beliefs. We've already specified the 1/2, 1/2 belief. But we have flexibility over the p1 minus p belief. So let's emphasize here that this is unconstrained by Bayes' rule.

Now one confusion I often see on problem sets or exams is people say, oh, here's my equilibrium. You can take p whatever you want. Just because it's unconstrained by Bayes' rule, to write down an equilibrium, we still have to make a choice about what this is.

We have a lot of flexibility in what that choice is, but you still have to make a concrete choice. You can't say the equilibrium is whatever p you want. You have to choose a particular P. And that's what we're going to do here.

So let's see what p we can choose. So this p is unconstrained by Bayes' rule, but it is constrained by something else. It is constrained by Sequential Rationality. Say Sr.

Why? Why am I not allowed to choose whatever p I want here? So I want to make this construct here a perfect Bayesian equilibrium. But you might think, oh, I can just choose whatever P I want, but I think there's an issue. Yeah.

AUDIENCE: I mean, if you were to accept-- or to assume that it was more passionate and not passionate [INAUDIBLE] since you would be rejecting more [INAUDIBLE] likely would've been a higher utility [INAUDIBLE].

IAN BALL: Exactly right. So let's check whether rejecting is sequentially rational given this belief. So my payoff as the school is always 0 if I reject. So my payoff from rejecting is 0.

What's my expected payoff from accepting given this belief? My expected payoff of accepting depends on the belief. So let's write it down. So we have-- well, p is the probability that the student is passionate.

And if they're passionate, I think we said I get a payoff of 3. And if they're not passionate, I get a payoff of negative 2. So in order for rejecting to be sequentially rational at this node, it must be that the expected payoff from rejecting is weakly higher than the expected payoff from accepting.

So I need 0 to be greater than or equal to this. And now let's simplify this a little bit. We get 3p, and then we get 5p minus 2. That's right.

So we need 0 greater than or equal to 5p minus 2. Let's put the 2 over here and then divide by 5. So we must get p less than or equal to 2/5.

So we've actually constructed a family of perfect Bayesian equilibrium. For any p that you give me, as long as that p is smaller than 2/5, we have a perfect Bayesian equilibria where the student always submits the essay, the school accepts if you submit an essay, they reject if you don't submit an essay, and if they see that you don't submit an essay, they assign belief p to you being a passionate student. And that belief p must be small enough that it's credible for them to actually reject. Yes.

AUDIENCE: So does that mean that, given different students, as long as they can assign one person 3 and one person-- like, different students who don't submit essays can have different values of p as long as it just doesn't make sense.

IAN BALL: So I think you're describing a slightly different game. Yeah, here, we've kind of made a simplification here where we've basically imagined that there's really only one student. So here what we're imagining is a student is drawn from the population and applies.

I think this is a reasonable model if-- in reality, what happens is many people apply, but the school adopts a uniform rule, they either accept students with essays or they don't, and they either accept students who don't have essays or they don't. If they wanted to do something more complicated where there's five applications and they do different things to different applications, then we'd have to explicitly model the number of applications.

And here, look at the information set. All I see is I see 1 applicant, and it's either an essay or not an essay. And if we want to capture what you're describing, which I think is important for applications, we need to enrich the game.

AUDIENCE: [INAUDIBLE]

IAN BALL: We're just looking at one student. Or the interpretation is a uniform policy that treats all students equally. If we wanted to treat students differently, then we'd have to enrich the game.

Great, so this is what's called a-- let's put this in words. This is what's called a pooling equilibrium. It's called a pooling equilibrium because both types of students, the passionate and the non-passionate students, choose the same action. They pool together, and all submit essays. And as a result, their choice about whether to submit an essay doesn't reveal that much information about whether they're passionate or not.

So next, let's see if we can compute what's called a separating equilibrium where different types of students make different choices. Let's go down here. This should be the last thing we do.

So earlier, I had assumed that c-- remember c is the cost of submitting an essay-- was less than 1. Let's now relax that assumption. And let's let c be arbitrary. So let c be arbitrary.

And now what we'd like to do is see, for which values of c, can we find a separating equilibrium? And this is actually an exam question I often like to ask, where you try to find the set of parameter values such that something is in equilibrium. So let's see if we can work through this.

So we want to look for an equilibrium where the essay distinguishes the types of students. I think this is exactly why schools offer optional essays. What they're hoping is that people who really want to get in and are excited write the essay, and the people who don't really want to get in don't write the essay.

I think it often doesn't work because we end up in the other equilibrium I described where everyone submits the essay even if they don't want to go. But this is the idea. And if essays are costly enough, this is what might work.

So we're looking for an equilibrium where what is the student doing? Well, if they're passionate, they submit the essay. And if they're not passionate, they don't submit the essay.

And then what does the school do? Well, let's first specify the school's beliefs. In this case, both information sets are reached with positive probability.

There's a positive probability that a student submits an essay and a positive probability that student does not submit an essay. So in this case, we can apply Bayes' rule at both information sets. We don't run into this issue of dividing by 0. And these beliefs are going to be well defined.

So if I see that a student submits an essay given that this is the strategy, what does Bayes' rule tell me as a school? What is the probability that they're passionate? And what is the probability that they're not passionate? Yeah.

AUDIENCE: 1?

IAN BALL: Yeah, it's just 1. And I think the way I'm writing it, let's write this 1, 0. So I say there's a probability of 1 that they're passionate and 0 that they're not passionate.

What if I see that there's no essay? Well, now it's the opposite. The only students who don't submit essays are the ones who are not passionate. So I'm certain that they must be not passionate. And I get 0, 1.

OK. Well, now it's pretty clear, if I'm certain that students who submit the essay are passionate, I certainly want to accept them. And if I'm certain that students who don't submit the essay are not passionate, then I certainly want to reject them.

So here's our candidate equilibrium. And let's just go through the conditions. And I think this is always a good way to check.

So does this equilibrium satisfy belief consistency? Yes. We just checked Bayes' rule, so we're good here.

And just to remind you, when I say equilibrium, notice I've not just specified the strategies. I've also specified the beliefs. And that's because this is an assessment, not just a strategy profile.

And so I have belief consistency. Now we move to sequential rationality. Well, let me separately check it for the student and the school.

Sequential rationality for the school is pretty clear. This is what I already said. They're accepting students who they think are likely to be passionate, and they're rejecting students they think are likely to be not passionate. So we're good here.

But what about sequential rationality for the student, this is going to be a little trickier. And this is going to depend on c. Yeah.

AUDIENCE: Yeah, [INAUDIBLE] the expected value for passionate students have to be like, negative [INAUDIBLE].

IAN BALL: Right, so let's exactly check it out. So there's two things we have to check. We have to check that if I'm a passionate student, I prefer writing an essay to not. And if I'm a not-passionate student, I prefer not writing an essay to writing an essay. So let's just write that down in math, but in words, that's exactly what we want.

So let's suppose I'm a student who's passionate. And if I write an essay, what is my payoff? And if I don't write an essay, what is my payoff?

Well, if I'm passionate and I write an essay, what happens? What's my payoff here? 2 minus c. OK.

And if I don't write an essay, well, I'm not accepted, and I don't write the essay, so my payoff is just 0. Now, notice a really crucial point here. The school's strategy is that if you don't submit an essay, you're rejected.

When I, as a passionate student, contemplate deviating and not submitting an essay, when I don't submit an essay, the school's strategy is fixed, and therefore, they reject me. People often get confused here because you might say, well, the school's belief is wrong. Their belief is that only non-passionate students don't submit essays.

And now a passionate student's submitting an essay. But the school can't tell the difference. If the school's belief is that if you don't submit an essay, you can't be passionate.

And if a passionate student deviates and doesn't send an essay, they're not able to change the beliefs that the school has. The school has their beliefs fixed, and therefore, what happens to them when they don't write an essay is determined by the strategy of the school. It doesn't change.

Some people are tempted to say, oh, they're passionate, so now the school's going to accept them. No, it's a unilateral deviation. The school strategy is fixed. And we're considering what happens when a passionate student deviates by not submitting this.

OK, now let's consider a non-passionate student. They're not submitting an essay. And again, they're getting a payoff of 0.

What if they submitted an essay? What would their payoff be? Yeah.

1 minus C. Again, they're able to trick the school. If they submit the essay, they successfully trick the school, and the school's going to accept them even though the school doesn't want to accept non-passionate people, but they have to pay the essay cost of c.

So now let's check the equilibrium conditions. A passionate student is supposed to submit the essay, so we need 2 minus c to be greater than or equal to 0. But a non-passionate student is not supposed to submit the essay. So we need 0 greater than or equal to 1 minus c.

So what's our solution? Well, this is going to be an equilibrium if both of these inequalities are satisfied. And that's going to be true if 1 is less than or equal to c less than or equal to 2.

And let's try to understand the role of this. Why do we have these two assumptions? Well, we want the essay to separate the good students from the bad students.

What would go wrong if c was too big? If c was greater than 2? Yeah.

AUDIENCE: Even the good students--

IAN BALL: Yeah, it's so costly-- if you say you have to write 100 essays, well, no one's going to apply to your school-- again, in the pre-GPT world. So we're not able to separate the students. If c is too low, what goes wrong?

If c is less than 1, well, even the non-passionate students say, well, it's a one-paragraph essay, I can do that. It's worth it for me to do that. So if you think of the school as-- this is a broader question about the game, you could think of the design question, if the school can choose c by choosing how hard the essay is, they might want to choose the essay to be a moderate difficulty.

It's hard enough to deter the uninspired students from applying, but easy enough that the inspired students would still apply. And I think that's exactly what admissions offices and jobs think about. OK, let me stop there unless there's any questions. And I'll see everyone next week.