PRINCIPLES OF BEHAVIORAL ECONOMICS

Prof. Sujata Kar

Department of Management Studies
IIT Roorkee

Week 54

Lecture 54

Hi, welcome back to the course on Principles of Behavioral Economics. This is lecture 54, and we are now going to talk about the problems with classical game theory. In the last three modules, we basically introduced analytical or classical game theory. They are alternatively referred to as classical or analytical game theory.

So, having talked about them, now I am going to discuss the problems that exist with classical game theory, as observed by behavioralists through a large number of experiments. Classical game theory has long provided powerful tools for analyzing strategic human choices. However, an extensive and growing body of experimental and field evidence challenges its predictions. These empirical findings highlight systematic deviations between theoretical expectations and actual human behavior in strategic settings. As researchers conduct more rigorous tests, it becomes increasingly evident that classical assumptions about rationality, strategic reasoning,

and equilibrium selection do not always hold in practice. One of the key motivations for testing classical game theory is to understand how real people behave in strategic situations, as opposed to how perfectly rational agents are assumed to behave. The empirical evidence indicates that human behavior is influenced by factors such as bounded rationality, social preferences, framing effects, assurance, resentment, and reciprocity. These behavioral aspects can significantly alter outcomes in strategy games, often leading to systematic departures from what classical game theory predicts. Consequently, the scope of classical game theory is

Severely restricted if it relies solely on assumptions of perfect rationality and common knowledge of rationality among players. So again, you can see that throughout the developments of various concepts under behavioral economics—like prospect theory, mental accounting, and intertemporal choices—we have repeatedly questioned the validity of rationality as assumed under the neoclassical tradition. And there is no exception to this when it comes to strategic interactions or classical game theory. Again, we are questioning the rationality, which is the basis.

Again, we are questioning the rationality of individuals, which is one of the most important bases of classical game theory. The empirical challenges to classical game theory have spurred the development of an emerging field known as behavioral game theory. This new approach integrates insights from psychology, experimental economics, and empirical observations to build models that better capture actual human decision-making in strategic contexts. Now we talk about something called iterated deletion of dominant strategies.

This is also one method. A fundamental prediction of classical game theory is that players can solve games by iteratively deleting dominated strategies. The method assumes that players are rational and believe their opponents are rational as well, leading to common knowledge of rationality. However, empirical evidence reveals that players are typically able to engage in only a few steps of iterated dominance, usually one or two, and higher levels of reasoning are actually very rare.

Suppose that the unique outcome of a game can be found by successively deleting dominated strategies in J steps. Then the game is said to be solvable with J steps of iterated deletion of dominated strategies. In classical game theory, J can be any finite number. For J equal to 1, the predictions of classical game theory usually hold, which implies that people are able to identify dominated strategies and delete them. If there is one move where J equals 1 (one step), however, when J equals 2, a majority of the players violate the predictions, so forget about having a large number or infinite number of steps.

Even when we have just two steps, it is observed that a large number of individuals are not able to identify the dominated strategy and delete it. It is quite possible that they are able to identify the dominated strategy. Nevertheless, they prefer not to delete it. Now, there could be various situations.

We are going to discuss those aspects. One possibility is that as J increases beyond 1, players may have less confidence in the rationality of others to engage in iterative deletion of dominated strategies. There are two main problems with such a priori arguments. So, first of all, it is interesting to see that what we are claiming is that player one or players may have less confidence in the rationality of others to engage in iterated deletion. So, in this sense, even if I am willing to go for iterated deletion, I really do not trust the other player.

First, if a player is rational, why does he believe that his opponent is irrational? Second, if opponents are believed to be irrational in such exceedingly simple games, then there are hardly any games where we can invoke common or even mutual knowledge of rationality. So basically, two steps are pretty simple games. Still, if people do not think that their opponents are going to be rational, then this implies that rationality is actually not very common. It's not a common occurrence.

Let us take an example to see to what extent people actually behave rationally and go for iterated deletion of dominated strategies. These are the words of J.M. Keynes when he writes that professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from 100 photographs. The prize is awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole. So that each competitor has to pick not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors

all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest. Not even those which average opinion genuinely thinks the prettiest. We have reached the third degree, where we devote our intelligence to anticipating what average opinion expects the average opinion to be. So this is called the beauty contest game or the P-beauty contest game more commonly.

Let us try to understand what the game tries to tell us. Suppose there are n players simultaneously. They have to pick a number between 0 and 100, both inclusive. So, between 0 and 100, one has to pick a number. The the person whose number is closer to a fraction say two-thirds of the average number wins a fixed price we can assume that there are 100 individuals who would be picking numbers between 0 and 100. Suppose everyone picked 100, the highest number available.

In that case, two-thirds of 100 would actually be approximately equal to 67. So, the winning number would be 67. Now, if I know that most people would be picking 100, so the winning number would be 67, then why should I pick 100? I'll be picking 67. And then, most people like me would also be picking 67.

So, no rational person would ever pick a number greater than 67. So, if nobody picks a number greater than 67—say, we all pick 67—then the winning number cannot be two-thirds of 67, which is approximately equal to 44. So, this is the second stage of iteration.

yet if nobody picks a number greater than 44 so if I can get to the second stage of iteration that it cannot be greater than 44 and others also would think along the same line and then nobody would be picking up a number greater than 44 so if nobody picks up a number greater than 44 and we assume that everybody picks up 44

Or majority picks up 44, then two-thirds of 44 would be roughly equal to 30, so the winning number cannot be greater than about 30, and so on all the way down to 0. So the Nash equilibrium in this game is 0. Now the point is that if this game is played by individuals empirically, do they actually follow so many levels of iterations and give a 0 answer? The average choice was about 35, with frequency spikes at 33 and 22. That is what was observed by Nagel in 1995.

The players performed only one or two steps of iteration. That was observed by Ho, Camerer, and Weigelt. The result suggests that most humans are level 1 or level 2 players. That is, we can go up to the first level of iteration, choosing 67, or the second level of iteration, choosing 45. Not much beyond that, and unable to iterate beyond a couple of steps.

Or they do not believe that other people are capable of doing so. So even if I could think that, OK, it should be 30, and then it should be two-thirds of 30, and so on. I would most often think, or I might think that even if I'm thinking like this, others don't think along those lines. Or players may not be willing to play the Nash equilibrium strategy even if they know it. So this severely restricts the scope of classical game theory.

We just observed that iterated deletion of dominated strategies, the way classical game theory conceptualizes it, may not work all the time empirically. A range of other factors, such as resentment, assurance, and reciprocity, may also be important in such games. But we will be talking about those games much later, or applications of these factors much later, specifically while discussing public goods games. When J equals 2, that is, we are at the second step of iteration. Extensive-form games and their equivalent normal-form representations elicit different behavioral responses.

Consider a game in its extensive form as well as its strategically equivalent normal-form game. So this is the extensive-form game. What is the equilibrium following the rule of backward induction? So Player 2 would see that here he is getting 3, here he is getting 4.75. Here he is getting 5.

So obviously, Player 2 would play R. If Player 2 plays R, then Player 1 would check between 9.75 and 10. And, of course, R is preferred. So RR is basically the Nash equilibrium or unique outcome. Two steps of iterated dominance are needed to obtain the backward induction outcome. First, strategy R weakly dominates L. So, strategy R weakly dominates L. Of course, 3 is also less than 10.

Mutual knowledge of rationality allows Player 1 to delete strategy L altogether. Rolling back the game tree, strategy R is the dominant strategy for Player 1. Now, when I write it in normal-form game, then what happens? When Player 1 plays L, Player 2 is actually indifferent. When Player 1 plays R, then Player 2 plays R. When Player 2 plays L, then Player 1 plays L. So, 9.75.

When Player 2 plays R, then Player 1 plays R. So, what is happening here is that besides RR, LL is also another game. But the normal form shows that there is another Nash equilibrium, which is LL. One of the earliest and simplest demonstrations of the failure of iterated dominance in this game was provided by Beard and Beil in 1994. In their experimental setup, players faced a two-player sequential game requiring two steps of iterated deletion of dominated strategies to reach the classical prediction. Exactly the game that we showed in the previous slide was given.

Despite the simplicity of the game, 66% of subjects in the role of Player 1 failed to choose the strategy predicted by classical game theory, that is, strategy R. This suggests that either players are unable to perform two steps of iterated reasoning or they lack confidence in the rationality of their opponents. Beard and Beil's experiment also examined whether players' beliefs about the rationality of others could explain these deviations. Suppose that Player 1 assigns probability P that Player 2 will play R. So this has a probability of P. Therefore, Player 1 will be indifferent between L and R if and only if 9.75 is equal to 3 into 1 minus P plus 10 into P. Solving, we will get P equals 0.97.

Suppose Q denotes the actual fraction of player 2s who responded to the play of R with small r. So, that is when player 1 played capital R. Then they played small r. In the experiment, the fraction of player 2s who responded to the play of R by player 1s with the dominant action small r is $q = 0.83$, and this is lower than 0.97. Therefore, we cannot necessarily infer that the 66% of player 1s who played L were irrational. Hence, the observed deviations could reflect doubts about mutual rationality rather than outright irrationality. So, basically, even though a large number of player 2s could actually play R—if you remember, R was the subgame-perfect Nash Equilibrium—

giving both players the maximum payoff, but only 83% played that. While ideally, 97% should play it in order for player 1 to be indifferent between L and R. So, since only 83% played R and still 17% played small L—that is, 17% of player 2s played small L, where player 1 had the lowest payoff at 3. If you remember, it was 3 and 4.75. So, there is a 17% chance that player 2 will play L.

And player 1 will end up having a very low payoff. So, given that percentage, what is safer for player 1 to do is to straight away play L here, which implies playing L where the payoff is 9.75, which is marginally smaller than 10, and as a result, the 66% who played L cannot be called irrational. Hence, the observed deviations could reflect doubts about mutual rationality rather than outright irrationality. To further investigate these deviations, Beard and Beil in 1994 conducted a range of experimental treatments, altering payoffs to test the influence of factors.

Such as risk assurance, resentment, and reciprocity. Despite these modifications, which theoretically should have led to the predicted play, a substantial proportion of players continued to deviate from the classical prediction. For example, even when player 2's response was known to be fully rational, That is, q was equal to 1, many players still chose actions inconsistent with iterated dominance. Next, we talk about a dynamic game of perfect information. These results challenge the assumption that players engage in even two steps of iterated deletion.

So again, the same problem continues. Other researchers used a sequential bargaining game to examine the limits of iterated reasoning. In the experiment, a pie has to be divided, which shrinks in size from $5 to $2 from the first stage to the second stage. The reservation amount or utility for each player in the experiment is 0. In the first stage, player 1 moves first to propose a split of $5. If accepted by player 2, the proposed split is implemented, and the game ends.

So, player 1 basically proposes a split of $5, say x and 1 minus x. If it is accepted, the game ends there. If player 2 does not accept the split of 1 minus x, then player 2 can make a counteroffer based on a pie size of $2. So next, the game starts with a pie size of $2, where player 2 makes the move. Player 1 can either accept the counteroffer or reject it, in which case both players get 0. So now, player 2 is making offers like Y and 1 minus Y. If it is accepted, then of course they would be getting these amounts.

If it is rejected, then both of them get 0. The unique backward induction outcome is for player 1 to keep $3, and offer $2 to player 2 at stage 1, which should be immediately accepted. Why is this so? Because the minimum amount that player 2 can offer, since the pie size is $2, is 1 cent—so $0.01 or 10 cents—and keeps $1.99 for themselves.

Now, $1.99 is less than $2. So, if player 1 keeps $3 for themselves and offers $2 to player 2, this should be acceptable to player 2, and the game should end in the first stage. Notice that the equilibrium offer of player 1 is the size of the initial pie minus the amount by which the pie shrinks—that is, $3 minus $2. Thus, if the pie shrinks from $5 to $0.50, then in equilibrium, player 1 keeps $4.50 and offers $0.50 to player 2, which should be immediately accepted by any rational individual because player 2 will begin with a pie size of $0.50.

And then, whatever he or she offers to player 1 should be less than $0.50, so accepting $0.50 is the best possible option. The experiments conducted by Goeree and Holt when the pie shrank from $5 to $2, the average amount kept by player 1 was $2.83, not far from the equilibrium prediction of $3. But when the pie shrank from $5 to $0.5, the average amount kept by player 1 was $3.38, which was substantially lower than the equilibrium prediction of $4.5. Furthermore, rejections by player 2 were common, though by rejecting the offer of player 1, player 2 could get no more than $0.5, which is not rational. These results challenge the assumption that players engage in even two steps of iterative deletion.

Next, we talk about the impact of framing effects in strategic reasoning. In classical game theory, one may expect the equilibrium outcomes to be identical if we consider games in extensive form or in their equivalent normal form representation. So, you see, we are again coming back to framing effects. Framing, we talked about in the context of prospect theory, in the context of mental accounting. And then again, we say that framing is important even in the context of classical or analytical game theory.

Schotter in 1995 asked subjects to play the same strategic game presented in two formats, the extensive form and a strategically equivalent normal form. However, spotting dominance may be relatively easier in one representation rather than the other. So, this is the game in sequential form or extensive form. You can see that first, going by backward induction, player moves. Player 2 will see he is getting 4 here, he is getting 1 here, and he is getting 3 here.

So, R is his strategy, and then if he is choosing R, then it, of course, makes sense for player 1 to choose R. So again, R-R is the unique equilibrium outcome. Now, if we represent it in the normal form game, then what is happening? When player 1 plays L, then player 2 is indifferent. When player 1 plays R, then player 2 plays R. Player 2 plays L, player 1 plays

L, player 2 plays R, and player 1 plays R. So here, we have two equilibria: 4-4 or L-L and R-R.

For the games given, experiments showed that in the normal form version, 43% of player 1's chose R because even if LL and RR both are Nash equilibrium you can see that here player 1 is benefited substantially. Still, only 43% of player 1's chose R, while in the extensive form version, 92% chose R. And of course, when he will choose R, player 2 will choose small r. The reason could be a lack of confidence in player 2 to respect dominance in normal form.

In fact, In the normal form version, 80% of player 2s choose the dominant strategy, that is R, while in the extensive form version, this increased to 98%. So extensive forms are considered to be frames, which makes it more evident which one is the dominant strategy or which are the dominated strategies. These results indicate that framing effects significantly affect players' strategic reasoning and their confidence in opponents' rationality. One interpretation of these results is that the extensive form presentation makes it cognitively easier for players to spot dominated strategies.

The visual structure of the game tree, which explicitly illustrates sequential moves and outcomes, aids players in reasoning backward through the game. Beyond simple two-steps reasoning, shorter-tested these players could successfully perform three steps of iterated deletion of dominated strategies. The results were sobering. Only 16% of players once and 26% of players choose the equilibrium strategy requiring three steps of reasoning. So you can see that when it comes to three steps of iterated deletion of dominated strategy, then people perform very poorly.

Only 16% of players chose once, and 26% of players chose. Most players opted for safer strategies, avoiding deeper levels of iterative thinking. Furthermore, when games were presented through verbal descriptions rather than formal diagrams, players' mental representations were in neither purely extensive form nor normal form, complicating their strategic reasoning. These findings highlight that how a game is framed visually or verbally can fundamentally alter players' ability to reason through dominance and backward induction.

These deviations highlight the cognitive limitations of players as well as the influence of fairness and relative payoffs, which are not accounted for in classical game theory. Players may possess other-regarding preferences, placing importance on relative payoffs rather than absolute gains. For instance, Player 2s may reject offers perceived as unfair, even at a personal cost. We will talk about such games much later while discussing public goods games.

Similarly, players may use simple heuristics instead of engaging in complex iterative reasoning. So, with this, I close the current module discussing problems with analytical or classical game theory, but we will continue with more observations in the coming modules. These are the references used. Thank you.