This is page 1
Printer: Opac

16

Strategic Bargaining

The strength of the framework we have developed so far, be it normal form or extensive form games, is that almost any well structured game can be presented in this fundamental language, and formal analysis based on rationality and strategic interaction could be performed. The repeated game was especially useful to depict a strategic situation that repeats over time, and the infinite horizon structure was helpful in setting a formal framework for repeated games with long horizons, in which the future is always important to a certain degree.

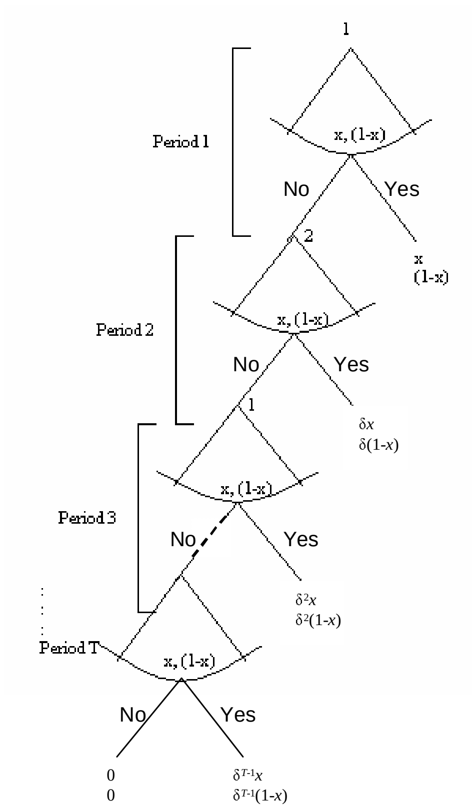
As it turns out, there is a large and important set of strategic situations that differ from the type of repeated game model we have developed, yet they do share the feature that a particular situation seems to repeat itself over time. These are the family of *Bargaining games*. This games generally include players who need to reach some agreement over a decision that affects them all, and failing to do so results in an inferior status quo.

Clearly, this is a very important set of situations of strategic interaction, since bargaining is regarded primarily as a situation in which a small number of parties, often two, try to split some gains from interaction. Be it a firm and a union that are bargaining over wages and benefits, a local municipality bargaining with a local utility provider over the terms of service, the head of a political party bargaining with other party members on the campaign issues, and even the mundane bargaining between a client and a merchant at a street bazaar.

How do we model bargaining? As argues above, the issue is often about a surplus that has to be split, and involves the parties making proposals, responding to them, and trying to settle on an agreement. One particular stylized way of modelling such a situation would be to consider two players that need to split a “pie” (the surplus from an agreement) of total value 1, and bargain according to a pre-specified procedure that evolves as follows:

- In the first round:
  - player 1 offers shares $(x, 1-x)$ where player 1 gets x and 2 gets $1-x$.
  - Player 2 then chooses between “accept”, causing payoffs $u_1 = x$, $u_2 = 1-x$, and the game ends, or “reject”, causing the game to move to the second round.
- In the second round a share $1-\delta$ of the pie “evaporates” (discounting):
  - Player 2 offers shares $(x, 1-x)$ where player 1 gets x and 2 gets $1-x$.
  - Player 1 chooses between “accept”, causing payoffs $u_1 = \delta x$, $u_2 = \delta(1-x)$, and the game ends, or “reject”, causing the game to move to the third round.
- The game continues in this way where following rejection in an odd period player 2 gets to offer in the next even period, and vice versa. Each period has further discounting of the surplus so that in period t the total “pie” is worth $\delta^{t-1} \cdot 1$.

From the specification of the bargaining procedure, the game may end with an agreement, or it may continue indefinitely. For starters, assume that the game ends in some pre-specified final period T that is reached if the players did not settle previously, and if they do not reach an agreement in that period then both receive a payoff of zero. This, for example, can be the case of a fisherman bargaining with a restaurateur over a fresh catch. If they wait more than a few hours, the fish will go bad and there will no longer be gains from trade. The rough extensive form of this game is:



Notice that this game has an interesting structure that is different from what we have seen so far. On one hand, the game has some features of a finitely repeated game as follows: If we think of an "odd" round as one where player 1 proposes and player 2 responds, and an "even" round as the converse, then we can treat each *pair of rounds* as *one stage* that repeats itself as long as an agreement is not reached. Furthermore, as rounds proceed the value of the pie that these players need to split is shrinking according to the discount factor δ. On the other hand, there are two features that are not part of the repeated game structure we saw above. First, the game can end during any round if the proposal is accepted, and second, payoffs are obtained only when the game actually ends and not as a flow of stage-game payoffs.

Is this general game of bargaining a reasonable caricature of reality? We need to convince ourselves that this game itself is, at least at some level, representative of what we believe bargaining is about. To some extent, sequences of offers and counter-offers are natural components for any type of bargaining, and clearly, once agreements are reached the bargaining stage is over. However, the pre-fixed end date T, at which an all-or-nothing agreement is reached is not extremely appealing. For this reason we will first analyze the game with such a finite end date, and later see what happens when we eliminate this artificial termination stage.

16.1 The Simplest Game: One Round

Let's start with an almost trivial case where there is only one round, so that $T = 1$. Since this is a game of perfect information (player 2 sees the offer of player 1 before he needs to accept or reject), we know that we can apply backward induction to find the subgame perfect equilibria. We start, instead, by analyzing the game with the concept of Nash equilibrium: what can be supported without requiring sequential rationality?

Claim 1: In the bargaining game with round **any division of surplus** $x^* \in [0, 1]$, $(u_1, u_2) = (x^*, 1 - x^*)$, can be supported as a Nash equilibrium.

proof: We will construct a pair of strategies that are mutual best responses, and that lead to $(x^*, 1 - x^*)$ as the division of surplus. Let player 1's strategy be "I propose x^*", and let player 2's strategy be "I accept any offer $x \leq x^*$ and reject any offer $x > x^*$." It is easy to see that these two strategies are mutual best responses independent of the value of $x^* \in [0, 1]$. ■

This observation is telling us one simple thing: the concept of Nash equilibrium has no bite for this simple game, and it can rationalize *any division of surplus*. In other words, we cannot predict in any meaningful way what the outcome of such a game will be if we only require people to respect mutual best responses.

A quick observation of the strategies that support this division of surplus should immediately scream out that strategic rationality, or as we now call it subgame perfection, is violated: Player 2 is saying "there is a minimum that I am willing to accept.” But what if player 1 offers him less? In particular, what if she offers him $(1-x) = \varepsilon > 0$? If player 2 rejects this offer he will get zero, while if he accepts it he will get a payoff of $\varepsilon > 0$, implying that his best response is to accept *any strictly positive payoff.* Anticipating this, player 1 should offer player 2 the smallest possible amount. This leads to the following result,

Claim 2: In the bargaining game with round, player 1 offering $x = 1$ and player 2 accepting any offer $x \leq 1$ is the unique subgame perfect equilibrium.

proof: We have established that player 2 must accept any positive share $(x < 1)$ and he is indifferent to accepting $x = 1$ so the proposed strategy is sequentially optimal and player 1 is playing a best response. The only other sequentially rational strategy for player 2 is to accept any strictly positive share $(x < 1)$ and reject getting zero $(x = 1)$. But player 1 has no best response to this strategy and thus it cannot be part of a SPE.■

This result is very stark, especially when compared with claim 1 above: simultaneous rationality can justify any outcome, but sequential rationality predicts only one outcome: player 2 should accept anything, and thus player 1 has an extreme form of a “take-it-or-leave-it” advantage giving her the whole pie of surplus.

Remark 8 *There have been many experiments that were run by researchers that basically tested behavior for exactly this type of game, also called the “ultimatum game”, in which a player 1 can offer to split a sum of money with player 2. If player 2 accepts then the split is realized and both players receive the proposal of player 1. If player 2 rejects then the players receive nothing. Contrary to the theory, and maybe not too surprising, the experiments show that those players in the role of player 1 offer significant shares to those in the role of player 2, and the latter will reject offers that are rather small. (expand, references)*

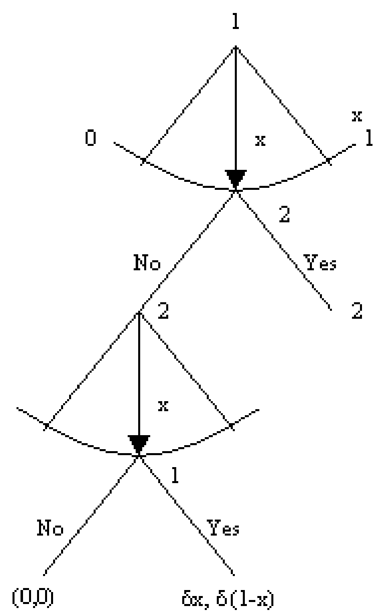
L12

16.2 The finite game: $T < \infty$

When we consider longer horizons, but still stick to the finite game case with some end date $T < \infty$, the fact that the Nash concept has no bite continues to hold. In particular, we can easily generalize the strategies in the proof of claim 1 above to *any* horizon, including an infinite one. To support a division of surplus $x^* \in [0,1]$ we can use the following strategies: In every odd round in which player 1 proposes, player 1 will propose x^* and player 2 will accept proposals of $x \leq x^*$. In every even round in which player 2 proposes, player 2 will propose x^* and player 1 will accept proposals of $x \geq x^*$. It is easy to see that these strategies are mutual best responses, and an agreement is reached in the first round with the division of surplus $(x^*, 1 - x^*)$.

It is not surprising that the uniqueness of the SPE will also apply to the finite bargaining game since backward induction will apply in a similar way that it did for the one round case. That is, in the last round, the proposer will offer to keep the entire pie to himself, and the responder will agree.

We saw that with one round, player 1 gets all the surplus because it is by construction the last round. What will happen with two rounds?



The second round is a one round game, so *if it is reached* player 2 gets everything. Continuing with backward induction, and with $\delta \leq 1$, we can find the subgame perfect equilibrium for the two-round game as follows. In the first round player 2 realizes that in the next (and last) round he will receive the whole pie, which will be worth only δ due to the discounting. Thus, if he is offered less than δ for himself (player 1 offers $x > 1 - \delta$) he should reject the offer and get the whole pie in the next period. Going then to the stage where player 1 makes her first offer, she realizes that any offer $x > 1 - \delta$ will be rejected, and thus she offers $x = 1 - \delta$, leading to a division of surplus equal to $(u_1, u_2) = (1 - \delta,\ \delta)$. For $\delta = 1$ we have player 2 getting all the pie since he then has the ultimate form of a "take-it-or-leave-it" advantage, and with $\delta < 1$ player 1 has a bit of a first-mover advantage due to the discounting.

Essentially, the backward induction argument that we have just demonstrated for the two round bargaining model will generalize to any finite round game, and there will be an inter-play between the ability to make a last "take-it-or-leave-it offer", and the first mover advantage due to the discounting of $\delta < 1$. Another aspect is who is the last mover, player 1 or player 2, which will artificially depend on whether the game has an odd number or even number of periods.

Consider the case with an odd number of rounds $T < \infty$, implying that player 1 has both the fist mover and last mover advantages. The following backward induction argument applies:

- In period T, player 2 accepts anything, so player 1 offers $x = 1$ and payoffs are $u_1 = 1 \cdot \delta^{T-1}$; $u_2 = 0$

- In period $T-1$ (even period – player 2 offers), by backward induction player 1 should accept anything giving $u_1 \geq \delta^{T-1}$. If player 1 gets x, then $u_1 = x \cdot \delta^{T-2}$; so player 2 should offer $x = \delta$, which gives player 1 $u_1 = \delta \cdot \delta^{T-2} = \delta^{T-1}$ and player 1 will accept any $x \geq \delta$. Payoffs are then $u_1 = \delta^{T-1}$; $u_2 = (1-\delta) \cdot \delta^{T-2}$.

- In period $T - 2$ (odd period), conditional on what we solved for $T - 1$, player 2's best response is to accept any x that gives him $(1 - x) \cdot \delta^{T-3} \geq (1 - \delta) \cdot \delta^{T-2}$. Player 1's best response to this is to offer the largest x that satisfies this inequality, and solving it with equality then yields player 1's best response: $x = 1 - \delta + \delta^2$. This offer followed by 2's acceptance yields $u_1 = x\delta^{T-3} = \delta^{T-3} - \delta^{T-2} + \delta^{T-1}$ and $u_2 = \delta^{T-2} - \delta^{T-1}$.

- In period $T - 3$ (even period), using the same backward induction argument, player 1's best response is to accept any x that satisfies $x\delta^{T-4} \geq \delta^{T-3} -$

$\delta^{T-2}+\delta^{T-1}$, and player 2 then offers an offer that satisfies this with equality, $x=\delta-\delta^2+\delta^3$, yielding $u_1=\delta^{T-3}-\delta^{T-2}+\delta^{T-1}$ and $u_2=(1-x)\delta^{T-4}=\delta^{T-4}-\delta^{T-3}+\delta^{T-2}-\delta^{T-1}$

We can continue with this tedious exercise only to realize that a simple pattern emerges. If we look at the solution for an odd period $T-s$ (s being even) then the backward induction argument leads to,

$$x_{T-s}=1-\delta+\delta^2-\delta^3\cdots+\delta^S\ ,$$

and for an even period $T-s$ (s being odd) then the backward induction argument leads to,

$$x_{T-S}=\delta-\delta^2+\delta^3-\delta^4...+\delta^S.$$

Now we can use this pattern to solve for the offer in the first period, x_1, that by backward induction is accepted by player 2, and it is

$$\begin{aligned}
x_1 &= 1-\delta+\delta^2-\delta^3+\delta^4\cdots+\delta^{T-1}= \\
&= (1+\delta^2+\delta^4+\cdots+\delta^{T-1})-(\delta+\delta^3+\delta^5+\cdots+\delta^{T-2}) \\
&= \frac{1-\delta^{T+1}}{1-\delta^2}-\frac{\delta-\delta^T}{1-\delta^2} \\
&= \frac{(1-\delta)(1+\delta^T)}{(1+\delta)(1-\delta)} \\
&= \frac{1+\delta^T}{1+\delta}\ ,
\end{aligned}$$

and this in turn implies that

$$\begin{aligned}
u_1 &= x_1=\frac{1+\delta^T}{1+\delta}\ , \\
u_2 &= (1-x_1)=\frac{\delta-\delta^T}{1+\delta}\ .
\end{aligned}$$

We can now offer some insights into this solution, which turns out to have some very appealing features. First, any SPE must have the players reach an agreement in the first round. The reason is that if agreement is reached in a later round with payoffs (u_1',u_2'), then discounting implies that part of the surplus is waisted and

$u_1' + u_2' < 1$. Sequential rationality would then imply that player 1 could deviate and offer $x = 1 - u_2'$, which guarantees player 2 the payoff u_2' immediately, and leaving herself more than u_1'.

Second, since player 1 has both the last mover "take-it-or-leave-it" advantage, and the first mover "discounting" advantage, then we have $u_1^* > u_2^*$ for any discount factor $\delta \in [0, 1]$. Now imagine that T is fixed and we are trying to see how patience affects the payoffs. Clearly, if players are very impatient and $\delta = 0$, then this is basically equivalent to a one round game in which player 1 gets the whole surplus. At the other extreme, if we take the limit of $\delta = 1$, then $u_1^* = 1$ and $u_2^* = 0$, and this is independent on the length of the game T.

This observation is somewhat disturbing: if players are very patient then the last mover "take-it-or-leave-it" advantage just flows through no matter how long the game is! The reason we get this rather counter-intuitive result is precisely due to the artificial stopping period T. To see this we fix the discount factor δ, and look at what happens to the equilibrium payoffs as the game gets longer and longer. We get,

$$\lim_{T\to\infty} u_1^* = \lim_{T\to\infty} \frac{1+\delta^T}{1+\delta} = \frac{1}{1+\delta} ,$$

and

$$\lim_{T\to\infty} u_2^* = \lim_{T\to\infty} \frac{\delta-\delta^T}{1+\delta} = \frac{\delta}{1+\delta} .$$

Now, looking at the equilibrium payoffs for the limit game where $T \to \infty$, we can ask what happens as the patience of the players changes without the artificial end period T having such strong bite. For the extreme case of impatience with $\delta = 0$ we get the same result: the game is equivalent in nature to a one round game. What happens when people become very patient? We now have,

$$\lim_{\delta\to 1} \lim_{T\to\infty} u_1^* = \lim_{\delta\to 1} \frac{1}{1+\delta} = \frac{1}{2} ,$$

and

$$\lim_{\delta\to 1} \lim_{T\to\infty} u_2^* = \lim_{\delta\to 1} \frac{\delta}{1+\delta} = \frac{1}{2} ,$$

which is a rather appealing result for very long horizons with very patient players: the long (potentially infinite) horizon eliminates the last mover's "take-it-or-leave-it" advantage, and the high level of patience takes away the first mover's "discounting" advantage.[1]

As an exercise, you can solve for the subgame perfect equilibrium when the number of periods is even, and see that at the limit we get the same result.

16.3 The Infinite game: $T = \infty$ ♠♠

When we consider the possibly infinite horizon game (if an agreement is never reached) then the assumption of disagreement leading to zero payoffs comes up naturally from the discounting: if we disagree forever then nothing is left to agree upon. However, there is a crucial difference with the analysis not being the limit of a finite horizon game, and that is that the path of perpetual disagreement has an infinite length, and we cannot therefore apply a backward induction argument.

There is, however, an interesting feature of the infinite game's structure, and it is the *stationary structure of the game following disagreement.* Namely, every odd period is the same, with player 1 making an offer, and the continuation game has a potential infinite horizon. Similarly, every even period is the same with player 2 making the offer. This allows us to apply a rather appealing, and not too difficult logic, to solve for the unique subgame perfect equilibrium of the game.[2]

Loosely speaking, the argument goes as follows. First notice that sequential rationality implies that an agreement must be reached in the first period, following the same logic we introduced earlier: waste will not be tolerated. Now imagine that there were possibly many SPE. This implies that in period 1 where player 1 makes an offer, from player 1's perspective there will be a *best* SPE yielding her a value of $\overline{v}_1$, and a *worst* SPE yielding her a value of $\underline{v}_1$. Similarly, in period 2 where player 2 makes an offer after rejecting the first offer from player 1, there will be a *best* SPE yielding player 2 a value of $\overline{v}_2$, and a *worst* SPE yielding him a value of $\underline{v}_2$. The stationary structure of the game implies that $\overline{v}_1 = \overline{v}_2 = \overline{v}$, and $\underline{v}_1 = \underline{v}_2 = \underline{v}$.

[1] Note that to reach this appealing conclusion the order of limits matters. If we reversed the order and did $\delta \to 1$ first followed by $T \to \infty$ then $\lim u_1^* = 1$ and $\lim u_2^* = 0$, preserving the artificial "take-it-or-leave-it" advantage.

[2] The idea comes from a paper by Shaked and Sutton (1984), but the infinite model was proposed, and the unique solution was identified in a seminal paper by Rubinstein (1982). The finite version is due to Stahl (1972).

It must then be that the lowest SPE payoff for player 1 in period 1 is obtained from the SPE that gives player 2 his highest payoff following rejection in the first round, implying that player 1 will offer player 2 $\delta\overline{v}$, and her payoff is then

$$\underline{v} = 1 - \delta\overline{v}. \tag{16.1}$$

Similarly, the highest SPE payoff for player 1 in period 1 is obtained from the SPE that gives player 2 his lowest payoff following rejection in the first round, implying that

$$\overline{v} = 1 - \delta\underline{v}. \tag{16.2}$$

Taking (16.1) and (16.2) we obtain that

$$\overline{v} = \underline{v} = \frac{1}{1+\delta} \ .$$

What does this mean? First, it means that player 1's SPE payoff at the beginning of the game is a uniquely determined by this equality. Second, since we know that agreement is reached in the first round this also implies what the SPE strategies are: in each odd round, player 1 offers $x = \frac{1}{1+\delta}$ and player 2 accepts any $x \leq \frac{1}{1+\delta}$. In each even round player 2 offers $x = \frac{\delta}{1+\delta}$ and player 1 accepts any $x \geq \frac{\delta}{1+\delta}$. The appealing feature of this unique SPE is that it coincides with the unique limit of the SPE for the finite horizon bargaining game.

16.4 Additional Bargaining Material:

- fixed cost of offers
- delay in bargaining
- axiomatic/cooperative frameworks.