1 Introduction

In previous notes and lectures, we have discussed various social welfare functions and mechanisms. These discussions have spanned not only environments where agents only know their preferences but also environments where agents know their utility. The mechanism design problem was to set up the rules of the game so that socially “good” things were either equilibria of the game or were strategically dominant solutions of the game.

In this section, we will continue to study the problem of reaching a social choice by designing mechanisms. We will study the bargaining problem. Approaches to bargaining fall into two divisions: axiomatic bargaining and strategic bargaining. Axiomatic bargaining identifies some desirable properties about the outcome of the bargaining process and then identifies process rules that produce this outcome. I think about the axiomatic bargaining problem similar to the way that I think of social welfare function design: identify what traits I want the bargaining solution to posses, and then find a rule that produces this solution when agents are honest about their feelings.

Strategic bargaining, by contrast, seems to be more similar to the mechanism design problem. It treats the bargaining problem as a game, assumes that the desirable property of the bargaining process is that the solution be an equilibrium solution, and then identifies what strategies will be followed by perfectly rational agents so that an equilibrium is reached. By establishing the rules of engagement but granting agents the power to reach an agreement (possibly through repeated interactions), strategic bargaining is a fairly unconstrained type of social choice mechanism.

If we had a lot more time, we would discuss some approaches to both axiomatic and strategic bargaining. Since we do not have this free time, we will focus our attention on an axiomatic solution that illustrates some of the key aspects of bargaining.

In words, we want to determine what properties will be satisfied when our design goals are to (a) impose constraints on what kind of outcomes we think are desirable and then try to (b) find a rule that produces this outcome. To deal with the incentive to lie, we will assume that there is an external agent, such as a benevolent court system, that can enforce the agreement reached by the agents.

2 What is Bargaining?

Bargaining, loosely speaking, is the process of figuring out how to divide up a shared resource between two or more agents. This resource can be money, utility units, cake, etc. In class, remind me to talk about the Ultimatum Game.
3 A Fair Bargaining Solution

The Nash bargaining solution is the unique formula that satisfies a set of axioms for a “fair” bargain between two agents. As with Arrow’s axioms of fairness, Nash introduces a set of axioms for a fair bargain. These axioms are:

- Invariance to equivalent utility representations,
- Symmetry (anonymity),
- Independence of irrelevant alternatives, and
- Pareto efficiency.

We will precisely define each of these, but it is first necessary to introduce some notation.

3.1 Formalism

Let $G$ be a set of goods that are to be divided between agents, let $g$ be a division of goods between agents, and let $u_i(g)$ be the utility received by agent $i$ for the division.

Example. Let $G$ be a pie that must be divided. Let $g$ be the division of the pie between agents, so that $g = [g_1, g_2]$. For example, $g_1 = g_2 = \frac{1}{2}$ indicates that the pie is divided precisely in half. Associated with each possible division, $g$, are utilities for the agents. For example, if the pie is worth $5.00 to each agent, then half the pie is worth $2.50 to each agent. Thus, the utility is $u_1(g = \left[\frac{1}{2}, \frac{1}{2}\right]) = u_2(g) = 2.5$. Other divisions are also possible so that, for example, each agent could get one-fourth of the pie and the remaining half of the pie goes to waste.

There is a special outcome that can result if bargaining breaks down. Call this outcome the fall-back outcome, and denote it by $g_{fb}$. Associated with this fall-back outcome is the utility pair $[u_1(g_{fb}), u_2(g_{fb})]$.

Example continued ... If negotiations for the pie break down, assume that the two agents start trying to wrestle the pie out of each other’s hands. The pie falls and is ruined, but the two agents get to lick pie filling off of their fingers resulting in $g_{fb} = [0.05, 0.05]$ and $[u_1(g_{fb}), u_2(g_{fb})] = [0.05, 0.05]$.

We assume, for the Nash bargaining solution, that the set of possible utility pairs is convex. This means that for any two utility pairs, $[u_1(g), u_2(g)]$ and $[u_1(g'), u_2(g')]$, there exists a division of goods $g''$ which has a utility “between” $[u_1(g), u_2(g)]$ and $[u_1(g'), u_2(g')]$. This occurs, for example, if outcomes include all possible lotteries over actual division of goods.

A convex and non-convex set are shown in Figure 1. In the figure, the horizontal axis represents the utility to player 1 and the vertical axis represents the utility to player 2. The set enclosed by the boundary points on the right is convex because there is no point “in between” other points that is not also in the set. The set on the left is not convex because there are several utility pairs “in between” the points on the right side of the set.

To help reinforce these ideas, think about the payoff structure for the battle of the sexes game. The corner points of this game are the set of utilities in the payoff matrix: (4,3), (3,4), (2,2), and (1,1). If we include mixed strategies, then we create a closed set, as shown in the online notes, but this set is not convex.
not convex; there are no points in this set between (4,3) and (3,4). However, if we include turn taking and use average utilities, then it is possible to get points between (4,3) and (3,4). For example, to get \((3.5, 3.5) = \frac{1}{2}(4, 3) + \frac{1}{2}(3, 4)\), we can let players alternate between going fishing and going to the ballet.

For the battle of the sexes to be treated as a bargaining problem, the goods that we will divide are various configurations of taking turns, and the utilities to the agents are the average payoffs for taking turns.

**Example continued** ... The joint payoff matrix for the divide-the-pie game is shown in Figure 2. The

![Figure 2](image)
This function is a bit different from most of the functions that we study in this class. Rather than returning some product of the argument (in this case $U$), this function returns an index, $g^*$, that satisfies a given set of properties. Given the index, we can identify the corresponding values within $U$ as

$$(u_1^*, u_2^*) = (u_1(g^*), u_2(g^*)).$$

We now turn attention to formalizing the axioms of fairness. Once we have formalized these axioms, we will define the function $f_{\text{Nash}}$ that is the unique bargaining solution that satisfies these axioms.

### 3.2 Invariance to Equivalent Utility Representations

Recall that utilities are unique only up to a positive affine transformation. This means that if a function $u(\cdot)$ satisfies the axioms of preference for a particular agent, then so does $\alpha u(\cdot) + \beta$ for $\alpha > 0$ and $\beta \in \mathbb{R}$. The invariance axioms says that the good chosen by the bargaining solution, $g^*$, should not depend on which of these equivalent utility representations is used.

To formalize, let $u_1(\cdot)$ and $u_2(\cdot)$ denote utility functions for players 1 and 2, respectively, let $T_1 u_1(\cdot) = \alpha_1 u_1(\cdot) + \beta_1$ denote a transformed version of $u_1(\cdot)$, and let $T_2 u_2(\cdot) = \alpha_2 u_2(\cdot) + \beta_2$ denote a transformed version of $u_2(\cdot)$. Let $U' = \{(T_1 u_1(g), T_2 u_2(g)) : g \in G\}$ denote the the utility set in the transformed spaces. Invariance to equivalent utility representations means that

$$g^* = f(U, g_{fb}) = f(U', g_{fb}).$$

In other words, the good that is chosen does not depend precisely on how the utility is represented.

**Example continued ...** Suppose that both players assign a utility unit for every fraction of the pie that he or she receives, $u_1(g) = u_2(g) = g$. Suppose that the bargaining solution is $g^* = f(U, g_{fb}) = [\frac{1}{2}, \frac{1}{2}]$; both players get half of the pie. Suppose now that player two expresses his or her utilities in terms of milliliters of pie filling received instead of pie fractions: $u_2'(\cdot) = T_2 u_2$. A fair solution should still choose $g^* = [\frac{1}{2}, \frac{1}{2}]$.

### 3.3 Symmetry (Anonymity)

Symmetry means that if the players’ utility space is symmetric and if the fall-back position is symmetric then a fair bargain produces a $g^*$ such that is equally valued by both players. We will formalize what these symmetric properties mean, and then talk about what a symmetric bargaining solution means.

Formally, the fall-back position is symmetric if $u_1(g_{fb}) = u_2(g_{fb})$.

Formally, the utility space is symmetric if, for every $g$ for which $(u_1(g), u_2(g)) \in U$ then there exists a $g'$ for which $(u_1(g'), u_2(g')) \in U$ and $u_1(g) = u_2(g')$ and $u_2(g) = u_1(g')$. In other words, if one pair of utilities, $(x, y)$ is in $U$ then the symmetric pair $(y, x)$ is also in $U$.

If the fall-back position is symmetric and if the utility space is symmetric, then it makes sense that players should receive equal utilities. Simply put, if the space of possible bargains is the same for player 1 and player 2, and if the fall-back positions are the same the players have equal bargaining power so the fair solution should award them equally. Thus, $u_1(g^*) = u_2(g^*)$.

This property is sometimes called anonymity because, if payoffs are symmetric, it does not seem fair for one player to receive more just because the bargaining mechanism “knows” that player. The solutions should treat players as if they are anonymous.
Example continued... In the divide-the-pie game, suppose that both players express their utilities as the fraction of the pie that they receive. Then, the utility space is symmetric and the fall-back position is symmetric. It stands to reason that players should receive equal portions of the pie if the solution is fair.

Since independence of irrelevant alternatives is the most complicated, we will discuss Pareto efficiency first.

3.4 Pareto Efficiency

A fair bargain should give the maximum allocation of utility to the players, in some sense. Said in another way, a fair solution will be one for which no alternative solution is better for both players.

Formally, if \( g^* = f(U, g_{fb}) \) then there should not exist a \( g' \) such that \( u_1(g') > u_1(g^*) \) and \( u_2(g') \geq u_2(g^*) \), or \( u_1(g') \geq u_1(g^*) \) and \( u_2(g') > u_2(g^*) \).

Example continued ... In the divide-the-pie game, it seems stupid for a fair solution to be one where some pie is left over after players reach a solution. Under such conditions, both players would be better off if the remaining pie were split in two and given to the players too. This means that the fair solution must be on the Pareto frontier\(^2\) which, for this problem, is the line between \((1,0)\) and \((0,1)\) in the utility space.

3.5 Independence of Irrelevant Alternatives

We have already discussed, in the context of Arrow’s impossibility theorem, the notion of an irrelevant alternative. In that context, an irrelevant alternative was a choice not considered in ordering a pair of different alternatives. The irrelevant alternative was problematic because it gave some information about strength of preference, but not enough to resolve ambiguity.

It seems strange to talk about irrelevant alternatives when we are using utilities, since utilities already encode preference strengths. However, these utilities alone do not tell us the strength of a bargaining position. Instead, the entire set of utilities and the fall-back position tell us about how strong a bargaining position is for an agent.

The independence of irrelevant alternatives axioms invites us to consider what would happen if we remove some goods that decrease the utility space, but still keep the fall-back position around. These removed goods should be irrelevant, meaning that they should not affect the outcome of the bargaining process.

Formally, let \( U = \{ (u_1(g), u_2(g)) : g \in G \} \) and let \( U' \subset U \) be some subset of the original utility space satisfying \( U \supset U'U' = \{ (u_1(g), u_2(g)) : g \in G - H \} \) and \( g^* \in G - H \). In other words, we have created a new utility space that is a subset of the old, but the new space still contains the bargaining solution. Moreover, suppose that \( U', g_{fb} \) is in set \( G - H \). Thus, we have a new utility space that still contains the fall-back position and the solution returned from the original utility space.

Consider the solution to the first problem, \( g^* = f(U, g_{fb}) \). The independence of irrelevant alternatives says that since \( g^* \in G - H \) the \( g^* \) should still be a solution to the second problem, \( g^* = f(U', g_{fb}) \).

Example continued ... In the divide-the-pie problem, suppose that we eliminate any alternative where player 2 gets more than two-thirds of the pie. Independence or irrelevant alternatives implies that the fair solution, the one where each player gets half of the pie, does not change simply because we constrain the search to a smaller utility space.

---

\(^2\)As me in class to define the term Pareto frontier.
The Nash Bargaining Solution

The Nash bargaining solution is the good that satisfies the following equation:

\[
g^* = \arg \max_{g \in GP} [(u_1(g) - u_1(g_{fb}))(u_2(g) - u_2(g_{fb}))] \tag{1}
\]

The notation \( GP \) indicates those goods that are better than the fall-back position, \( g^P = \{ g : u_1(g) \geq u_1(g_{fb}) \text{ and } u_2(g) \geq u_2(g_{fb}) \} \). It is necessary to restrict the search to these goods so that things that are much worse than the fall-back positions for both agents (meaning that both \( u_1(g) - u_1(g_{fb}) \) is less than zero and \( u_2(g) - u_2(g_{fb}) \) is less than zero) do not produce the maximum solution.

**Example continued ...** Consider the divide-the-pie problem. We want to search over all pie divisions to find the one that maximizes the product of the differences between an agent’s utility and its fall-back position. Since we will restrict attention to the Pareto frontier and since the solution is invariant to utility representations, we can restrict our search to those points on the line between \((1,0)\) and \((0,1)\). For simplicity, and since we are on the Pareto frontier, we can specify goods as the fraction of the entire pie that goes to player 1 knowing that the rest of the pie will go to player 2. Thus, we can let \( g \in [0,1] \) denote the portion of the pie going to player 1. This means that \( u_1(g) = g \). Moreover, the line between \((1,0)\) and \((0,1)\) is written as \( u_2 = 1 - u_1 = 1 - g \).

We are now in a position to solve the extrema problem posed by Equation (1). We can rewrite the equation by substituting \( u_1 \) and \( u_2 \) in terms of \( g \)’s,

\[
g^* = f(U, g_{fb}) = f(U, [0.05, 0.05])
\]

\[
= \arg \max_{g \in G^P} [(u_1(g) - u_1(g_{fb}))(u_2(g) - u_2(g_{fb}))]
\]

\[
= \arg \max_{g \in [0,1]} [g - 0.05][(1 - g) - 0.05]
\]

\[
= \arg \max_{g \in [0,1]} [g - 0.05][0.95 - g]
\]

\[
= \arg \max_{g \in [0,1]} -g^2 + g - (0.95)(0.05).
\]

We can solve this by taking the derivative with respect to \( g \) and setting it equal to zero.

\[
0 = \frac{df(g)}{dg} = \frac{d}{dg} - g^2 + g - (0.95)(0.05)
\]

\[
= -2g + 1.
\]

Solving yields \( g = \frac{1}{2} \). We must now check to see if \( g \) is in the feasible set, and since \( g = \frac{1}{2} \in [0, 1] \) we are satisfied.

Thus, the Nash bargaining solution is for each player to receive half of the pie. This solution seems to make intuitive sense. What would happen if one player was stronger than the other and could guarantee more pie filling if they end up wrestling over the pie? This would change the fall-back position in favor of the stronger player and would therefore change the bargain that is reached. This change occurs because a shift in fall-back position is a shift in strength in the bargaining process — if one player has more to lose than the other, then the player who stands to lose more has less influence over the ultimate outcome than the other.
5 The Proof

It is rather remarkable that Equation (1) is the unique function that satisfies all of the axioms of fairness. Can we prove this? Of course we can.

First, we will prove that Equation (1) does indeed satisfy the axioms. Second, we will show that it is the only function that satisfies it.

5.1 Step 1

Does the solution satisfy invariance with respect to equivalent utility representations? Of course it does. To see this, consider Equation (1) when utilities are transformed.

\[ g' = \arg\max_{g \in G^P} [(u'_1(g) - u'_1(g_{fb}))(u'_2(g) - u'_2(g_{fb}))] \]

\[ = \arg\max_{g \in G^P} [(\alpha_1 u_1(g) + \beta_1 - \alpha_1 u_1(g_{fb}) - \beta_1)(\alpha_2 u_2(g) + \beta_2 - \alpha_2 u_2(g_{fb}) - \beta_2)] \]

\[ = \arg\max_{g \in G^P} [(\alpha_1 u_1(g) - \alpha_1 u_1(g_{fb}))(\alpha_2 u_2(g) - \alpha_2 u_2(g_{fb})]] \]

\[ = \arg\max_{g \in G^P} [(u_1(g) - u_1(g_{fb}))(u_2(g) - u_2(g_{fb})]] \]

\[ = g^*. \]

Does the solution satisfy Pareto efficiency? The answer is clearly yes since if another good had higher utility for both players, then this good would produce a larger value for \((u_1(g) - u_1(g_{fb}))(u_2(g) - u_2(g_{fb}))\) and hence be the argument that maximizes this function.

Since the function \((u_1(g) - u_1(g_{fb}))(u_2(g) - u_2(g_{fb}))\) is symmetric, then whenever \(u_1(g_{fb}) = u_2(g_{fb})\) and \(U\) is symmetric, then \(g^*\) produces symmetric utilities. Thus, the solution satisfies the symmetry axiom.

Finally, consider decreasing \(U\) by removing irrelevant alternatives but keeping \((u_1(g^*), u_2(g^*))\) in \(U\). Since the fall-back points with and without the irrelevant alternatives are the same, it follows that the product is still maximal at \(g^*\). Simply put, if \(g^*\) maximizes the function, then decreasing the search space to a smaller set of goods but keeping \(g^*\) and \(g_{fb}\) in the set will not change the maximizing argument.

Thus, the Nash bargaining solution satisfies all four axioms.

5.2 Step 2: Uniqueness

Suppose that another bargaining solution, denoted by the function \(f\), satisfies the axioms. If we can show that \(f\) must equal \(f_{\text{Nash}}\), then we will know that \(f_{\text{Nash}}\) is unique.

To begin, we can invoke the invariance of equivalent utility representations and shift the utility space such that the origin is the fall-back position without altering the solution produced by \(f\). We do this by transforming \(u_i(\cdot)\) into \(u'_i(\cdot) = \alpha_i u_i(c) + \beta_i\). Transforming the fall-back position to the origin requires \(\beta_1 = -u_1(g_{fb})\) and \(\beta_2 = -u_2(g_{fb})\).

By the same axiom, we can scale the utilities such that the solution produced by \(f_{\text{Nash}}, g^* = f_{\text{Nash}}(U, g_{fb})\), falls on \(u_1(g^*) = u_2(g^*) = 1\). We do this by setting \(\alpha_1 = \frac{1}{u_1(g^*) - u_1(g_{fb})}\) and \(\alpha_2 = \frac{1}{u_2(g^*) - u_2(g_{fb})}\).

Since \(f\) satisfies the invariance of equivalent utility representations, \(g' = f(U', g_{fb})\), where \(U'\) is the utility space transformed as above, is the same as \(g' = f(U, g_{fb})\). We will exploit this fact and show that \(f\) chooses a \(g'\) for which \(u'_1(g') = u'_2(g') = 1\); in other words, \(f\) chooses the same point as \(f_{\text{Nash}}\).
By independence or irrelevant alternatives, we can remove elements of $U'$ until the utility space is symmetric. We do this by doing a search of all $(x, y) \in U'$ and checking to see if $(y, x) \in U'$. If both $(y, x) \in U'$ and $(x, y) \in U'$, then we allow both to be in the new utility space. Denote this new utility space as $U''$. Clearly, $(u'_1(g^*), u'_2(g^*)) = (1, 1)$ is an element of $U''$. Also, it is clear that $(u'_1(g_{fb}), u'_2(g_{fb})) = (0, 0)$ is also in $U''$. Finally, note that since $U$ is convex, it follows that $U'$ and $U''$ must also be convex.

Since $f_{Nash}$ satisfies all of the axioms of fairness, we know that the point $u'_1(g^*) = u'_2(g^*) = 1$ must be on the Pareto frontier in $U''$.

5.3 Step 2: Interlude

Before we conclude, it is useful to review what we have done. First, we shifted and scaled the utility space $U$ to a new space $U'$. The Nash equilibrium solution, $g^*$, from the original space was still the bargaining solution in the shifted and scaled space.

We then created a new, symmetric, utility space, $U''$, by eliminating all $(x, y) \in U'$ for which $(y, x) \not\in U'$. Since $g^*$ and $g_{fb}$ corresponded to symmetric utilities, they stayed in $U''$. By the independence of irrelevant alternatives, $g^*$ must be the bargaining solution in $U''$.

We then observed that $g^*$ was on the Pareto frontier in $U''$.

5.4 Step 2: Exploiting Symmetry

We can now apply symmetry which states that if $U''$ is symmetric and if $u'_1(g_{fb}) = u'_2(g_{fb})$ then the bargaining solution, $g_B$, on the new space must have symmetric utilities, that is $u'_1(g^B) = u'_2(g^B)$. But, the only point on the Pareto frontier of the new space for which $u'_1(g^B) = u'_2(g^B)$ is when $u'_1(g) = u'_2(g) = 1$ which occurs when $g = g^*$. Thus, the only solution that satisfies all of the axioms of fairness in this shifted, scaled, and symmetrized space is $g^*$.

When we pop back from $U''$ to $U'$, $g^*$ must still be the unique bargaining solution. When we pop back from $U'$ to $U$, $g^*$ must still be the unique bargaining solution. But this means that $g' = f(U, g_{fb})$ is the same as $g^* = f_{Nash}(U, g_{fb})$. And this means that $f$ is the same function as $f_{Nash}$ since it produces the same output for a given input.