Chapter 3

THE HARSANYI SOLUTION TO THE BARGAINING PROBLEM

3.1 Introduction

In the present chapter we introduce several generalizations to the formulation and solution of the bargaining problem. Until the early 1960s, the study of social power (including political power) had been carried out exclusively by social scholars. That changed in 1962, when Harsanyi managed to introduce the rigor of mathematical reasoning into the theory of social power. He achieved this by applying the analytical tools and models of game theory. In addition to certain significant contributions, Harsanyi provided a clear interpretation which eliminated much of the ambiguity that had plagued the theory of social power prior to the publication of Harsanyi’s works on the subject (Harsanyi 1962a, 1962b).

Chapter 2 focused on Nash’s two-person bargaining problem which treats the bargaining parties’ conflict payoffs as given. This formulation suffers from two important drawbacks in the context of political policy formation. First, in many real world cases, the threats can actually be decided by the parties themselves. Nash himself recognized this and in a 1953 paper presented a model in which the determination of the threat payoff is the reflection of the parties’ strategic choices. In Section 3.2 we describe this more general model.

Second, Nash’s original formulation of the bargaining problem dealt exclusively with the two-person bargaining problem. However, more often than not, the number of bargaining
parties exceeds two. This is particularly true of policy negotiation settings. When there are more than two parties involved in the bargaining, coalition forming is always possible. Many political economic models take the coalitions as given. In his seminal 1983 paper, Gary Becker took this approach but noted that "an explicit modeling of coalition formation would surely add to the power of the approach" (Becker, 1983).

This explicit modeling is exactly what Harsanyi did in his 1963 formulation and proposed solution of the generalized \( n \)-person bargaining problem. Of necessity the bargaining problem is far more complicated in the \( n \)-party case. Furthermore, both the formulation and the solution concept of the \( n \)-person bargaining problem reflect both strategic and cooperative conceptualization. Section 3.3 is devoted to describing and explaining the Harsanyi model.

Finally, our political-power theoretic approach adopted in later chapters relies heavily on the model of political power proposed in Harsanyi (1962a, 1962b). To lay the groundwork, we present Harsanyi’s model of the social power relationship in Section 3.4. The presentation is rather brief and concise, but it is sufficiently comprehensive to be self-contained.

### 3.2 Endogenous Disagreement Payoffs

In the bargaining game discussed in Chapter 2 it was assumed that the disagreement (conflict) payoffs, \( t = (t_1, t_2) \), were predetermined by the rules of the game as given fixed values. Such a relationship is not unheard of in real-world bargaining situations. Nevertheless, quite often the bargaining parties are capable of influencing the disagreement payoffs, thus affecting the solutions payoffs of the corresponding bargaining game. This suggests extending the "simple" bargaining game by turning it into a stage of a more complex game. It is assumed that in the first stage of this game each player announces the conflict strategies that will be employed by him should the players fail to agree in the second stage. It is also assumed that the announced strategies are binding and will be carried out by both players. Let \( \Sigma_i \) be the sets of conflict strategies available to player \( i \). Similarly, let \( \theta_i \) be player \( i \)’s
disagreement strategy choice. That is, $\theta_i \in \Sigma_i$. We begin by considering a two person game where $i \in \{1, 2\}$. In this setting, we let $U_i (\theta_i, \theta_j)$ denote player $i$’s payoff as a function of player $i$’s and player $j$’s conflict strategies. This means that player $i$’s disagreement payoff can be written $t_i = U_i (\theta_i, \theta_j)$.

In Chapter 2, we defined $H$ as the set of payoff vectors in the payoff space, $P$, not dominated, even weakly, by any other payoff vector in $P$. We also defined

$$P^* = \{u \in P: u_1 \geq t_1, u_2 \geq t_2\},$$

the set of payoff vectors in $P$ that leave both parties at least as well off as in the conflict outcome. We denoted the upper-right boundary of $P^*$ by $H^* \subseteq H$. The bargaining problem in Chapter 2 was then: Given $P^*$ and $(i = 1, 2)$, what solution will the bargaining parties eventually reach? In this chapter, $t$ is made endogenous and the bargaining problem becomes: Given $P$ and a set of possible conflict strategies $\Sigma_i (i = 1, 2)$, what will be the solution $\bar{u}$? To reach the solution, it is assumed that once the disagreement payoffs have been determined in the first, noncooperative stage of the bargaining game, the solution to the second stage is the plain Nash bargaining solution, i.e. the point $\bar{u} = (\bar{u}_1, \bar{u}_2) \in H^*$ satisfying

$$(3.1a) \quad (\bar{u}_1 - t_1) (\bar{u}_2 - t_2) = \max_{u \in P} [(u_1 - t_1) (u_2 - t_2)]$$

such that

$$(3.1b) \quad u_i \geq t_i \quad (i = 1, 2)$$

Formally, let $H(u_1, u_2) = 0$ be the equation of $H^*$ and assume that the partial derivatives, $H_1 (\bar{u}_1, \bar{u}_2) = a_1$ and $H_2 (\bar{u}_1, \bar{u}_2) = a_2$, are nonzero. We know that along the upper-right boundary of the (convex) payoff set $P$, an increase in player $i$’s payoff must decrease player $j$’s payoff or that $\frac{du_j}{du_i} \bigg|_{H=0} \leq 0$. But $\frac{du_i}{du_j} = -\frac{\partial H/\partial u_i}{\partial H/\partial u_j}$, so we know that $H_1$ and $H_2$ have the
same sign. Without loss of generality, we can consider the case where \( H_1, H_2 > 0 \). The Kuhn-Tucker conditions for maximizing \((u_1 - t_1)(u_2 - t_2)\) with respect to \( u_1 \) and \( u_2 \) subject to \( H(u_1, u_2) = 0 \) are

\[
(3.2a) \quad a_1(\bar{u}_1 - t_1) = a_2(\bar{u}_2 - t_2)
\]

where

\[
(3.2b) \quad a_i = H_i(\bar{u}_1, \bar{u}_2) \equiv \frac{\partial H}{\partial u_i}(\bar{u}_1, \bar{u}_2) \quad (i = 1, 2)
\]

These conditions imply that for a given payoff space \( P^* \), any pair of disagreement payoffs \((t_1, t_2)\) that satisfy (3.2a) yield the solution \( \bar{u} \). This situation is demonstrated in Figure 3.1 where both \( t \) and \( t' \) satisfy (3.2a) and both yield the solution \( \bar{u} \).

Let \( C(\bar{u}) \) denote the line in \( P \) whose equation is \( a_1 t_1 - a_2 t_2 = a_1 \bar{u}_1 - a_2 \bar{u}_2 \). Note that \( C(\bar{u}) \) is then the set of disagreement payoffs that satisfy (3.2a) and thus yield the solution \( \bar{u} \). As we move up and to the left along \( H^* \), the outcome becomes more favorable for player 2. This movement increases the point at which \( C(\bar{u}) \) hits the \( u_2 \) axis and simultaneously decreases the slope of the line \( C(\bar{u}) \). Because a given point \((t_1, t_2)\) has a unique solution, no two lines \( C(\bar{u}) \) and \( C(\bar{u}^*) \) can intersect. As a result, if \( C(\bar{u}^*) \) lies above and to the left of \( C(\bar{u}) \), \( \bar{u}^* \) is more favorable to player 2 than \( \bar{u} \). It can also be shown that if \( \bar{u} \) is more favorable to player 1 than \( \bar{u}^* \) (i.e. \( \bar{u}_1 > \bar{u}_1^* \)), then

\[
(3.3) \quad a_1 \bar{u}_1 - a_2 \bar{u}_2 > a_1 \bar{u}_1^* - a_2 \bar{u}_2^*
\]
Figure 3.1: Nash solution to the two-person bargaining game in the $(u_1, u_2)$ space.
Similarly, if $\bar{u}^*$ is more favorable to player 2 than $\bar{u}$ (i.e. $\bar{u}_2^* > \bar{u}_2$), then

$$a_1 \bar{u}_1 - a_2 \bar{u}_2 < a_1 \bar{u}_1^* - a_2 \bar{u}_2^*$$

Accordingly, it is in the interest of player 1 to maximize $a_1 t_1 - a_2 t_2$, while player 2’s interest is to minimize the same expression. Consequently, the chosen conflict strategies are given by

$$a_1 t_1 - a_2 t_2 = a_1 U_1 (\theta_1^0, \theta_2^0) - a_2 U_2 (\theta_1^0, \theta_2^0) = \max_{\theta_1 \in \Sigma_1} \min_{\theta_2 \in \Sigma_2} \left[ a_1 U_1 (\theta_1^0, \theta_2^0) - a_2 U_2 (\theta_1^0, \theta_2^0) \right]$$

$\theta_1^0$ and $\theta_2^0$ are referred to as the mutually optimal conflict strategies. Note again that the selected conflict strategies affect the ultimate bargaining outcome through their mutual effects on the disagreement payoffs, $t$.

We can therefore characterize the solution to the full two-person bargaining model by the set of necessary and sufficient conditions given by

$$H (u_1, u_2) = 0$$

$$a_i = H_i (u_1, u_2)$$

$$a_1 (u_1 - t_1) = a_2 (u_2 - t_2)$$

$$t_i = U_i (\theta_1^0, \theta_2^0)$$

$$a_1 t_1 - a_2 t_2 = \max_{\theta_1 \in \Sigma_1} \min_{\theta_2 \in \Sigma_2} \left[ a_1 U_1 (\theta_1^0, \theta_2^0) - a_2 U_2 (\theta_1^0, \theta_2^0) \right]$$

### 3.3 The $n$-Person Bargaining Game

In contrast to the two-person bargaining game with endogenous disagreement payoffs, it is natural to assume that in the $n$-person case ($n > 2$) parties will be able to form coalitions. We focus here on Harsanyi’s solution to the resulting $n$-person cooperative bargaining game (Harsanyi 1963). In formulating the relevant model, Harsanyi stated:
"the final payoffs of the game are determined by a whole network of various agreements among the players and we shall try to define the equilibrium conditions of mutual consistency and interdependence that these various agreements have to satisfy. That is, we shall assume that each such agreement between players will represent a bargaining equilibrium situation between the participants if all other agreements between the players are regarded as given.” (Harsanyi, 1963: 205)

Essentially, Harsanyi’s solution results from considering the entire bargaining game to be a complex nest of two-party bargaining games between players and coalitions where the solution to any given bargaining is the solution to the two-party game resulting from taking all other payoffs as given.

To discuss Harsanyi’s solution, we must first define his notation. Let $N$ represent the set of all $n$ players involved in the bargaining game. Each player $i \in N$ can belong to many coalitions $S \subset N$. From each of these, he receives a secure payoff $w^S_i$. Note that in general this payoff is different for various members of the coalition because they have different bargaining positions. The total amount paid out by $S$ is determined by bilateral bargaining between the coalition $S$ and its complimentary coalition $\bar{S}$ in $N$ (i.e., $\bar{S} = N - S$). The final payoff, $u_i$, of player $i$ from the entire game is the sum of his payoffs from each coalition he belongs to. That is,

\begin{equation}
    u_i = \sum_{S \subseteq N : i \in S} w^S_i \text{ for all } i \in N
\end{equation}

where the notation $S \subseteq N : i \in S$ refers to summing over all subsets of $N$ that contain individual $i$. (Similarly, the notation $S \subseteq N : i \notin S$, used later, refers to all subsets of $N$ that do not contain individual $i$.)
It is, of course, required that the final solution be on the upper-right boundary of the payoff space, that is

\[(3.8) \quad u = (u_1, u_2, \ldots, u_n) \in H.\]

As noted in Chapter 2, this is equivalent to requiring that the final solution be Pareto optimal.

Each coalition \(S\) guarantees its members’ payoffs by announcing a conflict strategy \(\theta^S\). The conflict strategies determine the conflict payoffs, \(u^S_i\) and \(u^S_j\), that the members of \(S\) and \(\bar{S}\) receive. Thus we can write

\[(3.9) \quad u^S_i = U_i \left(\theta^S, \theta^{\bar{S}}\right) \text{ for each } S \subset N, i \in S.\]

The set \(S\) may contain several subsets \(R\) that also contain \(i\). It is logical to conclude that

\[(3.10) \quad u^S_i = \sum_{R \subseteq S : i \in R} w^R_i \text{ for all } S \subset N.\]

This condition simply implies that the subsets of \(S\) cannot guarantee its members more in the aggregate than the coalition \(S\) could secure by bargaining with \(\bar{S}\) and that all payoffs secured by the coalition \(S\) will be distributed by its member coalitions. The payoff from the entire game is given by

\[(3.11) \quad u_i = u^N_i = \sum_{R \subseteq N : i \in R} w^R_i\]

Now, suppose we assume that no payoffs are negative.\(^1\) The distribution between players \(i\) and \(j\) of payoffs from a coalition \(S\) they both belong to is determined in a bargaining game between \(i\) and \(j\) where the following payments are taken as given: (1) the payoffs to other

\(^1\)Negative dividends are possible; however, we shall ignore this case.
players \((u^S_k \text{ for all } k \in S \text{ where } k \neq i, j)\), (2) player \(i\)’s payoffs from all coalitions \(j\) is not a member of \((u^S_j \text{ for all } S \subset N \text{ where } j \notin S)\), and (3) player \(j\)’s payoffs from all coalitions that \(i\) is not a member of \((u^S_j \text{ for all } S \subset N \text{ where } i \notin S)\). Denote this bargaining game \(G^S_{ij}\).

If players \(i\) and \(j\) fail to agree on payoffs, the coalition \(S\) will be unable to operate since it requires unanimous agreement from its members. Therefore, the disagreement payoffs are given by what players \(i\) and \(j\) can secure from other coalitions, i.e.

\[
(3.12) \quad t^S_i = \sum_{R \subset S: i \in R} w^R_i \quad \text{and} \quad t^S_j = \sum_{R \subset S: j \in R} w^R_j
\]

This relationship implies that

\[
(3.13) \quad t^S_i = u^S_i - w^S_i.
\]

In similar fashion, we can express the disagreement payoffs for the entire game as

\[
(3.14) \quad t^N_i = u^N_i - w^N_i \quad \text{and} \quad t^N_j = u^N_j - w^N_j.
\]

Harsanyi then expresses the disagreement payoffs in alternative form, which will be useful later. Note that the inverse of the relationship given in (3.10) and (3.11) is

\[
(3.15) \quad w^S_i = \sum_{R \subseteq S: i \in R} (-1)^{s-r} u^R_i
\]

where \(s = |S|\) and \(r = |R|\) denote the number of players in coalitions \(S\) and \(R\), respectively. Since this relationship is not immediately obvious, we present a simple example to demonstrate. Consider a game where \(N = S = \{1, 2, 3\}\). There are three subsets of \(S\) which contain player 1: \(R_1 = \{1\}, R_2 = \{1, 2\}, \text{ and } R_3 = \{1, 3\}\). By (3.10)

\[
(3.16) \quad u^S_i = u^S_i + u^R_1 + u^R_2 + u^R_3.
\]
Similarly, since $R_1 \subset R_2, R_3$, we can write

$$u_{1}^{R_2} = w_{1}^{R_2} + w_{1}^{R_1} \Rightarrow w_{1}^{R_2} = u_{1}^{R_2} - w_{1}^{R} \quad (3.17)$$

$$u_{1}^{R_3} = w_{1}^{R_3} + w_{1}^{R_1} \Rightarrow w_{1}^{R_3} = u_{1}^{R_3} - w_{1}^{R} \quad (3.18)$$

Finally, since there are no subsets of $R_1$, we have

$$u_{1}^{R_1} = w_{1}^{R_1} \quad (3.19)$$

Combining (3.16), (3.17), (3.18), and (3.19) and solving for $w_{1}^{S}$ gives

$$w_{1}^{S} = u_{1}^{S} - u_{1}^{R_2} - u_{1}^{R_3} + u_{1}^{R_1} \quad (3.20)$$

The sign pattern here is consistent with (3.15) since the coalitions with two members ($s - r = 1$) have negative signs, the entire group has a positive sign ($s - r = 0$), and the individual member has a positive sign ($s - r = 2$).

Substituting (3.15) in (3.13) gives us

$$t_{i}^{S} = u_{i}^{S} - \sum_{R \subseteq S : i \in R} (-1)^{s-r} u_{i}^{R} = \sum_{R \subseteq S : i \in R} (-1)^{s-r+1} u_{i}^{R} \quad (3.21)$$

which is the form used by Harsanyi’s in defining the solution to the problem.

The entire game can be thought of as a series of two-party bargaining games. Hence, the final payoffs, $\bar{u} = (\bar{u}_1, ..., \bar{u}_n)$, must satisfy

$$H (\bar{u}_1, ..., \bar{u}_n) = 0 \quad (3.22a)$$

$$a_i (\bar{u}_i - t_i^N) = a_j (\bar{u}_j - t_j^N) \quad (3.22b)$$
where

\[(3.22c) \quad a_m = H_m (\bar{u}_1, \bar{u}_2, ..., \bar{u}_n) \equiv \frac{\partial H (\bar{u})}{\partial u_m} \text{ for all } m\]

\[(3.22d) \quad u^S_i = U \left( \theta^S, \theta^S \right) \text{ for all } i \in S, S \subset N\]

\[(3.22e) \quad t^S_i = \sum_{R \subset S : i \in R} (-1)^{s-r+1} u^R_i \text{ for all } i \in S, S \subseteq N : s > 1\]

and

\[(3.22f) \quad \sum_{i \in S} a_i u^S_i - \sum_{j \in S} a_j u^S_j = \max \min_{\theta^S \in \Sigma_S, \theta^\bar{S} \in \Sigma_{\bar{S}}} \left[ \sum_{i \in S} a_i U_i \left( \theta^S, \theta^S \right) - \sum_{j \in S} a_j U_j \left( \theta^S, \theta^S \right) \right]\]

subject to

\[(3.22g) \quad \begin{cases} a_i \left( u^S_i - t_i^S \right) = a_k \left( u^S_k - t_k^S \right) & \text{for all } i, k \in S \\ a_j \left( u^S_j - t_j^S \right) = a_m \left( u^S_m - t_m^S \right) & \text{for all } j, m \in \bar{S} \end{cases}\]

While they may appear complicated due to the notation, these conditions follow quite naturally from the previous discussion. (3.22a) requires that the solution be on the efficiency frontier and (3.22b) is the familiar first order condition derived by Nash. (3.22c) through (3.22e) formally define the quantities \(a_m, u^S_i, \text{ and } t^S_i\) as we have defined them in the preceding text. Finally, conditions (3.22f) and (3.22g) require that the solution payoffs for the bargaining within coalitions also satisfy the Nash first order conditions.

As written, these conditions in (3.22a) through (3.22g) require the existence of the partial derivative of \(H\) at the solution value \(\bar{u}\). However, we can replace (3.22a) and (3.22c) with

\[a_i \geq 0\]

and

\[\sum_{i \in N} a_i \bar{u}_i = \max_{u \in \bar{P}} \sum_{i \in N} a_i u_i\]
to attain the general conditions. The last two conditions imply that, given constants \(a_1, ..., a_n\) \((a_i \geq 0 \text{ for all } i \in N)\), the bargaining solution, \(\bar{u}\), is the payoff vector maximizing the weighted sum \(\sum a_i U_i\) over the payoff space, \(P\). Since \(P\) is compact and convex, \(\bar{u}\) exists and is a unique element of \(H\), or \(H(\bar{u}) = 0\).

Harsanyi (1963) also demonstrated that the solution payoff is given by what he calls a "generalized Shapley value." To do so, he defined the quantity \(Z^S = \sum_{i \in S} a_i u^S_i\), which is a weighted sum of the utility gained by the individual members of \(S\). If, as before, \(s = |S|\) is the number of members in coalition \(S\) and \(r\) is the number of players in coalition \(R\), Harsanyi proved that

\[
(3.23) \quad a_i u^S_i = \sum_{R \subseteq S, i \in R} \frac{(r-1)! (s-r)!}{s!} \left[Z^R - Z^{R|i}\right] \text{ for all } i \in N.
\]

This alternative method for specifying the solution plays an important role in the measurement of power described in Section 3.4.

### 3.4 Reciprocal Power Relations

In his analyses, Harsanyi distinguished two principal modes of social influence broadly conceived as situations where the intervention of one actor, \(A\), causes another actor, \(B\), to alter his/her behavior in a manner that actor \(B\) would not otherwise do. In the first formulation—the unilateral power relationship—actor \(A\) unilaterally determines \(B\)'s behavior by setting the values of certain variables under \(A\)'s control. As \(B\)'s reaction is a function of \(A\)'s actions, \(A\) thereby influences \(B\)'s reaction. (Thus, \(A\) is a Stackelberg leader, while \(B\) is a Stackelberg follower.) Alternatively, \(A\) may irrevocably set certain conditional incentives in the form of sanctions for noncompliance and rewards for compliant behavior which leads \(B\) to comply with \(A\)'s demands (ultimatum game). Both the Stackelberg relation and ultimatum games are unilateral power relations.

According to the second distinct principal mode—the reciprocal power relationship—both
parties possess power over each other. As demonstrated by Harsanyi, this mode of power relationship is best modeled as a bargaining game. To simplify the exposition and render the relevant underlying principles more visible, Harsanyi focused first on the two-person bargaining game as a model of a reciprocal bilateral power relationship. Prior to Harsanyi, the existing literature focused on five dimensions of A’s power over B: (1) the base of power or the resources A can use to influence B; (2) the means of power or the specific actions A takes; (3) the scope of power or the actions A can induce B to perform; (4) the amount of power or the increase in the probability of B performing the action A wants; and (5) the extension to the set of individuals over whom A has power.

Harsanyi argued that two additional dimensions are required for a meaningful theoretical model of the social power relationship. In particular, modelers should always include the "cost and strength of A’s power over B” in their models. The "cost dimension” refers to the cost borne by A in influencing B’s behavior. The "cost of power”, whether a direct or opportunity cost, comprises the cost to A of rewarding B for compliance or the cost to A of penalizing B when the latter ignores A’s demands. The other needed dimension is the "strength of A’s power over B.” This term refers to B’s utility loss when not complying with A’s demands. It is the sum of B’s utility losses due to unrealized rewards and the subjective cost caused by A’s sanctions under noncompliance.

### 3.4.1 Reciprocal Power in Two-Party Games

As we did with the bargaining model, we begin by considering the measurement of power in games involving only two parties. Our treatment adopts the scenario offered in Harsanyi (1962a). Suppose A wants B to perform X with probability $p_2$ when, in the absence of A’s intervention, B would perform X with probability $p_1$ only ($p_1 < p_2$). If B completely refused to do X he would obtain a utility of $u_0$ and A would obtain a utility of $u_0^*$. A offers B a reward of $R$ if B complies with A’s demand and threatens B with a penalty of $T$ if B elects to perform X with only probability $p_1$. Suppose the value of the reward $R$ to B is $r$ utility
units and the cost of \( R \) to \( A \) is \( r^* \) utility units. Similarly, the sanction \( T \) entails a loss of \( t \) units to \( B \) and a cost of \( t^* \) units to \( A \). If \( B \) performs \( X \) he loses \( x \) utility units while \( A \) enjoys a gain of \( x^* \) units. Then, if \( B \) complies with \( A \)'s demand and performs \( X \) with probability \( p_2 \), the parties’ expected utilities are

\[(3.24a) \quad u_A = u_0^* - r^* + p_2 x^* \]

and

\[(3.24b) \quad u_B = u_0 + r - p_2 x. \]

If \( B \) does not comply with \( A \)'s demand and a conflict situation arises, then the parties’ expected utilities are

\[(3.25a) \quad \tilde{u}_A = u_0^* - t^* + p_1 x^* \]

and

\[(3.25b) \quad \tilde{u}_B = u_0 - t - p_1 x. \]

In this context, a bargaining problem emerges where the parties bargain over how much influence \( A \) can exert on \( B \) (i.e. on the value of \( p_2 \)). The agreement payoffs are given by \( u_A \) and \( u_B \) while the disagreement payoffs are given by \( \tilde{u}_A \) and \( \tilde{u}_B \). From Section 3.2, we know that taking \( T \) and \( R \) as given, the Nash solution solves

\[(3.26) \quad \max_{p_2: p_1 < p_2 < 1} (-r^* + p_2 x^* + t^* - p_1 x^*) \left( r - p_2 x + t + p_1 x \right). \]
The solution is

\[(3.27) \quad p_2^0 = p_1 + \frac{1}{2} \left[ \frac{r^* - t^*}{x^*} + \frac{t + r}{x} \right].\]

The theory of optimal threat strategies presented in Section 3.2 tells us that A’s optimal strategy is to select the value of \( T \) that maximizes \( [(t/x) - (t^*/x^*)] \) and a value of \( R \) that minimizes \( [(r^*/x^*) - (r/x)].\)

Harsanyi (1962a) then considered Dahl’s measure of the quantity of A’s power over B (Dahl 1957). This dimension of power was defined by Dahl as the probability difference \( \Delta p = p_2^0 - p_1 \). It follows immediately from Equation (3.27) that

\[(3.28) \quad \Delta p_2^0 = p_2^0 - p_1 = \frac{1}{2} \left[ \frac{r + t}{x} - \frac{t^* - r^*}{x^*} \right].\]

According to Harsanyi, \( r + t \) is the cost to B of being in conflict with A instead of cooperating and thus “measures the (gross) absolute strength of A’s power over B.” The quantity \( (r + t)/x \) can then be thought of as “the gross relative strength of A’s power over B with respect to action \( X \)” (Harsanyi, 1962a, 77). The complementary action \( X^* \) denotes A tolerating B’s noncompliance, which can be thought of as B inducing A to take an action he/she would not otherwise take. Therefore, \( t^* - r^* \) is the absolute strength of B’s power over A while \( (t^* - r^*)/x^* \) is the relative strength of that power. The difference between the gross relative strengths, \( \frac{r+t}{x} - \frac{t^*-r^*}{x^*} \), measures ”the net strength of A’s power over B with respect to action \( X \)” (Harsanyi 1962a: 78). Therefore, Dahl’s quantity-of-power measure in a reciprocal bilateral power relationship is equal to one half of the net strength of A’s power over B with respect to the action \( X \). The foregoing definitions and derivations reflect the reciprocal power relationship inherent in the bargaining model of social power.

\[\text{Recall that (by assumption) by selecting } T, \ A \text{ uniquely determines the values of } t(T) \text{ and } t^*(T); \text{ by selecting } R, \ A \text{ uniquely determines the values of } r(R) \text{ and } r^*(R).\]
3.4.2 Reciprocal Power in $n$-Party Games

Just as the bargaining model becomes considerably more complex with more than two parties, measuring power in $n$-party games is more difficult. Harsanyi’s method for measuring this power draws heavily on the bargaining framework we reviewed in Section 3.3. We present here some of the highlights of Harsanyi’s formulation.\(^3\)

Harsanyi defines several types of power that exist when a group of individuals has to choose among several alternative policies. The first is called specific power and is the probability that an individual’s most-preferred policy is adopted. If players have preferences between other policies, however, this is not an adequate measure of an individual’s power. Instead, Harsanyi defines generic power as the sum of the utility the individual gains from each alternative, weighted by the probability of that policy being adopted by the group. It is thus a measure of ”$i$’s power to get the group to adopt some policy reasonably satisfactory to him.” He then decomposes an individual’s generic power into two components. Independent power measures the ability of individuals to secure good outcomes without cooperation from other parties and incentive power measures their ability to use rewards and penalties to convince other groups to accept outcomes the individuals prefer.

Suppose there are $N$ players considering $M$ alternative policies, one of which will be adopted. Let $x_{ij}$ represent the utility player $i$ gains from alternative $j$, and let $p_j$ denote the probability that alternative $j$ is chosen. The probability $p_i$ measures individual $i$’s specific power and the quantity

$$\tilde{p}_i = \frac{\sum_{j \in M} p_j x_{ij}}{x_{ii}}$$

measures individual $i$’s generic power. Using his bargaining framework, Harsanyi is able to predict the value of $\tilde{p}_i$ for any given game. In particular, it can be derived using the generalized Shapley values discussed at the end of Section 3.3. In the case of conflict between

\(^3\)For further details, see Harsanyi (1962b).
coalitions $S$ and $\bar{S}$, let the quantities $P^S$ and $P^\bar{S}$ represent the weighted total utility gained by members of coalitions. Similarly, the quantities $T^S$ and $T^\bar{S}$ represent the weighted total utility losses caused by the punishment strategies pursued by these coalitions. The quantity $P^S - T^S$ is thus equivalent to the quantity $Z^S$ from Section 3.3. If a cooperative solution between all the parties is reached, the players reward one another. Let $R$ represent the weighted sum of the net utility gained by each individual from these reward strategies. If we adopt the normalization $P^N = 1$, $Z^N = 1 + R$. In this setting, (3.23) implies that the solution is given by

$$a_i u_i = \frac{(1 + R)}{n} + \sum_{S \subseteq N; i \in S} \frac{(s - 1)! (n - s)!}{n!} \left[ (P^S - T^S) - (P^\bar{S} - T^\bar{S}) \right]$$

By definition, $u_i = \sum_{j \in M} p_j x_{ij} + r_i$. Substituting this value into (3.30) and re-arranging yields the expression

$$\tilde{p}_i = \frac{1}{a_i x_{ii}} \left[ \frac{1}{n} + \sum_{S \subseteq N; i \in S} \frac{(s - 1)! (n - s)!}{n!} \left( P^S - P^\bar{S} \right) \right]$$

$$+ \frac{1}{a_i x_{ii}} \left[ \frac{1}{n} (R - a_i r_i) - \sum_{S \subseteq N; i \in S} \frac{(s - 1)! (n - s)!}{n!} (T^S - T^\bar{S}) \right]$$

As noted previously, this term measures player $i$’s *generic power*. The two terms in this expression represent the decomposition of this power into its two components. The quantity $P^S - P^\bar{S}$ represents the ability of coalition $S$ to get its agenda adopted without the cooperation of the rest of the group. Hence, the first term measures the *independent power* of $i$ (and his potential allies). In contrast, the second term measures difference between the rewards gained by the group, $R - a_i r_i$, if they cooperate with individual $i$ and the penalties they face, $T^S - T^\bar{S}$, if they fail to reach agreement with $i$ (and his potential allies).

Harsanyi also emphasized that this power measure incorporates “the effects of alliances and party alignments among participants.” These impacts are reflected by the summation

17
over all potential coalitions $S$. Furthermore, the approach reflects the "improvements in all participants’ power positions when suitable compromise policies are discovered."

3.5 Conclusion

In Chapters 2 and 3, we have described the Nash-Zeuthen-Harsanyi bargaining theory. Essentially, Harsanyi’s imaginative model of social power as a bargaining relation provides the basic conceptual formulation. In later chapters we will use this as the foundation for our political power theory of endogenous policy formation. Chapter 4 will be dedicated to the detailed description and formulation of the political process in terms of Harsanyi’s underlying framework. In the rest of this book we shall outline the many ramifications of the basic idea, thus providing a rather broad view of the impressive tree that grew out of the seed planted by Harsanyi in the early 1960s.