Indian Statistical Institute, Delhi Centre

Instructor:

Prof. Arunava Sen[1]

Summary of Readings on Coalition Formation and Allocation Mechanisms

Dev Kinger[2]

October, 2025

[1]Professor, Economics and Planning Unit, Indian Statistical Institute, Delhi Centre

[2]MSQE_II student, Indian Statistical Institute, Delhi Centre

1 Introduction

We began our exploration with a review of the literature on coalition formation, which serves as a fundamental framework for the analysis of cooperative behavior in game theory. Coalition formation provides a structured way to model situations where groups of agents coordinate their actions to achieve outcomes that would be unattainable individually. In contrast to non-cooperative frameworks—where each player independently seeks to maximize their own utility—the cooperative setting emphasizes collective optimization, wherein members of a coalition jointly determine strategies that maximize the overall surplus of the group. Formation of coalitions is the most natural way to extend the ideas of cooperative games.

The theoretical inquiry within this domain centers around two key questions. The first concerns the **existence and stability of coalition structures**: under what conditions will agents find it optimal to form and stay within a given coalition? The second concerns the **allocation of the coalition's output or surplus**: once cooperation is established, how should the collective gains be distributed among members in a manner that preserves notions like fairness and stability? Together, these questions form the foundation of cooperative game theory and motivate a large body of research seeking to formalize the principles underlying cooperation and collective rationality.

We began with the problem of allocating coalition output, for which the Myerson value provides a foundational framework. Myerson (1977) introduced graph-theoretic ideas to model cooperation, viewing the game as a communication network where players are connected through links that determine feasible interactions. Instead of focusing on all players collectively, cooperation occurs within connected components formed by these links, which define coalitions. The Myerson value is built on two principles—equity and efficiency—and extends the Shapley value to settings where communication is constrained. In doing so, it captures how the structure of connections influences both the possibility and the payoff of cooperation.

We also looked at ideas of stability of a coalition structure. First, it seems that the notion of stability is a simple affair: for a coalition structure to be stable all we need is that there does not exist a subset of players of existing coalition who can simultaneously deviate and make themselves better off. This thinking led to the idea of core allocations: *The core is the set of all allocations which are unblocked by any subcoalition.* However, this myopic or short-sighted idea has received a fair amount of criticism (Aumann and Myerson (1988))):

> When a player considers forming a link with another one, he does not simply ask himself whether he may expect to be better off with this link than without it, given the previously existing structure. Rather, he looks ahead and asks himself, 'Suppose we form this new link, will other players be motivated to form further new links that were not worthwhile for them before? Where will it all lead? Is the end result good or bad for me?

A natural next step is to formalize the notion of stability within coalition structures—where players account not only for immediate gains but also for the future repercussions of their actions on other coalitions. This perspective gave rise to the concept of far-sightedness (Chwe (1994)), in which players anticipate the sequence of potential reactions that might follow any deviation. When a subset of a coalition deviates, the system transitions from one outcome (say, a) to another (b), which may in turn induce further deviations leading to a new outcome (c). An outcome is deemed stable if it survives all possible deviations—meaning that no subset of players finds it profitable to deviate, given their understanding that any resulting sequence of moves would ultimately leave them worse off.

The concept of far-sightedness in coalition formation highlights that an individual's payoff is closely tied to the stability of the coalition they belong to. This perspective motivates the analysis of coalition stability under different allocation mechanisms, such as the Myerson value. A coalition structure is said to be stable with respect to an allocation rule (e.g., the Myerson or Equal Share value) when each player recognizes that deviating from their assigned outcome would lead to a lower individual payoff, thereby aligning individual incentives with collective stability.

To summarize, the stability of coalition structures and the allocation of output within coalitions constitute the two fundamental requirements that enable sustained cooperation in a strategic setting. Most formulations of stability integrate both myopic and far-sighted considerations. Importantly, these notions are often imposed within an externally defined structure—such as a communication network, power relations, institutional framework, or information constraint—that shapes the stable coalitions and the flow of cooperation among agents. Together, these theoretical components provide a coherent foundation for understanding how cooperative behavior can emerge, persist, and evolve across different strategic environments.

2 Review of Select papers

This section provides a summary of the main results of the papers that were discussed to get a formal idea of coalition stability, far-sightedness in coalitions, Myerson value and papers that lately use these concepts to define stability of coalitions.

2.1 Core allocations and stability in graph restricted games

Let there be set N of n players. Begin with a characteristic function $\mathcal{U}$ defined on $S \subset N$; it specifies for every coalition S a set of payoff vectors $\mathcal{U}(S) \subset R^S$ for that coalition. If for each coalition S there is a number $v(S)$, describing the overall worth of that coalition, such that

$$\mathcal{U}(S) = \{\mathbf{u} \in R^S | \sum_{i \in S} u_i \leq v(S)\}$$

Simply put, $\mathcal{U}(S)$ is a set of payoff vectors $\mathbf{u}$ that are feasible for coalition S. Let $\mathbf{u}$ be an allocation for the set of players S and $T \subset S$ a subcoalition of S. $(T, \mathbf{u}')$ blocks $\mathbf{u}$ if $\mathbf{u}' > \mathbf{u}$.

> Definition 1: The set of all feasible allocation for a coalition that are unblocked is the *core of a coalition.* Formally, a coalition structure S belongs to the core, if $\nexists$ $(T, \mathbf{u}')$ where $T \subset S$ which blocks $(S, \mathbf{u})$.

Let us look at stability concept provided by Myerson, which just an extension of the above idea in a graph restricted setting. Let N be a nonempty finite set of n players. Let $\mathcal{C}$

be the set of all possible coalitions of players in N:

$$\mathcal{C} = \{S | S \subset N, S \neq \emptyset\}$$

A link between two players, n and m, is denoted by $\underline{n, m}$. Let g^N be the complete graph of all such links in the set N:

$$g^N = \{\underline{n, m} | n \in N, m \in N, n \neq m\}$$

Let GR be the set of all graphs of N:

$$GR = \{g | g \subset g^N\}$$

Finally, let $S \setminus n$ denote the removal of a member from a set S.

Definition 2: We say an allocation rule $Y : GR \to R^N$ is stable if and only if

$$\forall n \in N, \forall S \subset N \setminus n, \quad Y_n(g^N) \geq Y_n(\{\underline{i, j} \in g^N, \underline{i, j} \notin \{\underline{n, m} | m \in S\}\})$$

By definition, the payoff of a player in a complete graph is at least as high as the payoff the player would receive in any graph where they are permitted to omit links with a subset of other players. Clearly, if the complete graph is not stable, one can examine the remaining subgraphs to identify a stable coalition structure. The two definitions discussed are closely related; however, the latter explicitly incorporates the links between players, allowing for the analysis of stability within a graph-restricted game.

2.2 Farsighted stability

As discussed in the previous section, the natural extension of this definition in cooperative setting is far-sightedness. To incorporate this notion formally, we alter the structure of the game. The game $\mathcal{T}$ is defined as $\mathcal{T} = (N, Z, \{\prec_i\}_{i \in N}, \{\to_S\}_{S \subset N})$, where N is set of players, Z is the set of 'all' possible outcomes, and $\{\prec_i\}_{i \in N}$ defines a set of strict preference for all individuals in the set N

$\to_S$ is new addition within a game which now represents what a coalition S can do. If $a \to_S b$, the coalition S can move the status quo outcome from a to b. This however does not mean the coalition S can enforce the outcome b. No restrictions are placed on the ability of coalition to move their status quo outcome. Specifically, the set $\to_S$ may be empty, $a \to_S a$ is possible and $a \to_S b$ does not imply $b \to_S a$. We have now formally written down a structure which allows coalitions to look at the action of their repurcussions.

The formal definition of stability requires us to define largest consistent set of a game, $lcs(\mathcal{T})$. Let us define two notions based on which lcs is defined. If $a \succ_i b$ for all $i \in S$, then a is directly dominates b and we say $a \succ_S b$. Here is the definition of indirect dominance:

Definition 3: We say a indirectly dominates b, or $a >> b$, if there exist $b, a_1, a_2, \ldots, a_{n-1}, a$ and $S_0, S_1, \ldots, S_{m-1}$ such that $a_i \to_{S_i} a_{i+1}$ and $a_i \succ_{S_i} a_{i+1}$ for $i = 1, \ldots, m-1$.

In other words, outcome a *indirectly dominates* outcome b if there exists a sequence of coalitional deviations through which successive groups of players can move from b to a, with each deviation being both feasible and strictly preferred by the deviating coalition.

The explicit idea of far-sighted stability is: outcome a is said to be *stable* if no coalition S has an incentive to deviate from a to another outcome b, recognizing that such a deviation could trigger subsequent adjustments by other coalitions leading to a new outcome c, under which not all members of S would be better off. Formally, the largest consistent set (of stable outcomes) is:

> Definition 4: A set $Y \subset Z$ is consistent, that is $a \in Y$ if and only if $\forall\ b, S$ such that $a \rightarrow_S b$, $\exists\ c \in Y$, where $b = c$ or $c >> b$, such that $c \not\succ_S a$

Chwe (1993) established the existence and uniqueness of the largest set of consistent outcomes for any game within the general framework, thereby providing a foundational characterization of farsighted stability. In later section, we see how farsightedness can be applied to Farell and Scotchmer equal sharing partnership.

2.3 Coalition formation under power relations

The model proposed by *Piccione and Razin (2009)* studies a society driven by an exogenous **power relation** defined over all possible coalitions of agents. Each agent seeks to maximize their position in a social ranking that is determined jointly by individual and group power. A *social order* is represented as a partition of agents into coalitions, ranked according to their collective power.

Within this framework, the authors define **stability** of a social order using two key ideas: the **core** and **farsighted (durable) deviations** as discussed in the previous two sections. A coalition may deviate from the existing order if doing so improves the members' relative positions, but such deviations are subject to farsighted reasoning about possible 'counterresponses'. Formally, stability is defined through the following conditions:

1. A deviation by a coalition is **profitable** if every member of the coalition strictly improves their rank in the new social order.

2. A deviation is **durable** if no disjoint coalition can subsequently counter-deviate in a way that makes any of the original deviators worse off than before.

A **stable social order** is one for which no durable deviation exists. This concept extends the traditional core by incorporating farsightedness: agents anticipate the reactions of others before deviating. A stronger notion, the **strongly stable social order**, requires stability under all possible power relations, ensuring robustness of the social structure to any configuration of power.

We can clearly see how the concepts discussed in the previous two sections can be applied to different settings to define the idea of stability.

3 Farsighted Stability in Farrell–Scotchmer Partnerships

Farrell and Scotchmer (1988) tries to answer what would happen when groups were obliged to share the output equally. They call this 'equal-sharing' contract a *partnership*. The paper suggests a simple model where each individual has some inherent ability e_i. The group's payoff depends upon the ability level and the number of people within a group. Within a group, the total payoff is shared equally. The individual payoff and the group's payoff is given by:

$$\mathcal{G}(|S|, e_i) = t(|S|) \sum_S e_i$$

$$v(|S|, e_i) = \frac{t(|S|)}{|S|} \sum_S e_i$$

where, $|S|$ is the number of people in a group, $\mathcal{G}(|S|, e_i)$ is the group's output and $v(|S|, e_i)$ is the individual output.

They make two assumptions to get their results: $t'(|S|) > 0$ and $a'(|S|) < 0$ for some $|S|$ large enough. This assumption states that adding more people always increases the group's output but as more and more people are added, the individual output falls. Note that each individual can rank the potential outcomes corresponding to all the possible groups or coalitions to which they may belong. Farrell and Scotchmer result states that:

Theorem 1. *F-S Theorem: There is partition of the players into coalitions such that no new coalition could form and make all its members strictly better off and the partition is generically unique*

Using the structure from Section 2.2, we identify the largest consistent set within the framework of Farrell–Scotchmer partnerships. For this, we express the game in its transformed form and specify all its components to analyze it from a farsighted perspective.

Here are the components of the game:

- Let $N = \{1, 2, 3, \ldots, n\}$ be a finite set of n players each having some inherent ability e_i for $i = 1, 2, 3, \ldots, n$

- Each individual can rank the potential outcomes corresponding to all the possible coalitions to which they may belong. Note that preferences are strict and no player could derive the same output from two different groups.

- To define the set of all possible outcomes, let us define the possible coalitions that can form. Let $\Pi(N)$ denote the set of possible partition/coalitions of N. The total space of feasible outcomes can be written as:

$$Z = \{(\pi, u) : \pi \in \Pi(N); u : \pi \to R\}$$

- To say that a coalition $S \subset N$ changes the outcome from a to b, now means a coalition S forms their own coalition and changes the partition $\pi \to_S \pi'$ where $\pi, \pi' \in \Pi(N)$.

Here, π' is formed by creating a new partition that contains the coalition S, while the remaining partitions stay unchanged except that the members of S are removed from them.

This gives us all the components of the game as required in the Chwe's framework. We state and prove the following result:

Theorem 2. *Farsighted F-S Theorem: The largest consistent set is unique and equals the Farrell-Scotchmer partition*

Proof. We prove the result in two steps. First the the F-S partition π^* lies in the largest consistent set and second no other outcome lies in the largest consistent set.

To prove the first claim note that Farrell and Scotchmer partition, say π^* is obtained in the following manner:

- Let $\mathcal{R}_0 := N$. Choose $S_1 \subset \mathcal{R}_0$ that maximizes $v(S_1)$ over all non-empty coalitions of $\mathcal{R}_0$. Such an outcome exist because of assumptions over $t(|S|)$ and genericity of v
- Let $\mathcal{R}_1 = N \setminus S_1$. Choose $S_2 \subset \mathcal{R}_1$ that maximizes $v(S_2)$ over coalitions of $\mathcal{R}_1$
- Let $\mathcal{R}_2 = N \setminus S_1 \cup S_2$. Choose $S_3 \subset \mathcal{R}_2$ that maximizes $v(S_3)$ over coalitions of $\mathcal{R}_2$
- Continue until $\mathcal{R}_k = \emptyset$

The outcome so obtained is the (generically) unique Farrell and Scotchmer partition $\pi^* = \{S_1, S_2, ..., S_k\}$ where S_i is obtained by forming a coalition of members from $\mathcal{R}_{i-1}$.

Claim 1: No coalition $S \subset N$ can form from π^* and make all of its member strictly better off.

This follows because for any coalition $T \neq S_j$ formed by the members of $\mathcal{R}_{j-1}$ must necessarily have $v(T) < v(S_j)$. Hence no coalition T exists that can make all the members strictly better off starting from π^*. In the notation of definition 4 from section 2.2, π^* belongs to the consistent set as $\nexists\, T_1 \neq S_1$ such that members of $T_1 \succ S_1$. Similarly, $\nexists\, T_2 \neq S_2$ such that 'all' members of $T_2 \succ S_2$ and so on.

Claim 2: For any partition $P \neq \pi^*$, $\pi^* >> P$ i.e. π^* indirectly dominates P

Let j denote the *smallest index* such that the coalition S_j is not present as a coalition in the partition π^*. Such a coalition must exist since $P \neq \pi^*$.

Note that all coalitions $S_1, S_2, \ldots, S_{j-1}$ are identical in both P and π^*, because j is the smallest index at which the two partitions differ. Therefore, the members of S_j must belong to some other coalition in P, formed by members within $\mathcal{R}_{j-1}$.

Consequently, coalition S_j has an incentive to deviate, changing the partition from

$$P \to_{S_j} P_1$$

In the new partition P_1, the next smallest index where P_1 differs from π^* is $j+1$, corresponding to the coalition S_{j+1}, which is not present as a coalition in π^*. Hence, members of S_{j+1} deviate from

$$P_1 \to_{S_{j+1}} P_2$$

Proceeding in this manner, the coalitions $S_{j+1}, S_{j+2}, \ldots, S_k$ successively deviate through

$$P_i \rightarrow_{S_{j+i}} P_{i+1}$$

This continues till $P_k = \pi^*$. By definition 3 in Section 2.2, this chain of deviations shows that

$$\pi^* \gg P$$

Together, claim 1 and claim 2 proves that the largest consistent set is unique and equals the Farrell–Scotchmer partition. $\square$

4 Conclusion

The readings during this term helped in understanding how cooperation and stability can arise when individuals or groups come together to form coalitions. Starting from the basic idea of the core, which ensures that no group of players can gain by breaking away, we moved towards farsighted stability, where players think ahead about possible future reactions before deciding to deviate. This showed that stable outcomes depend not only on immediate benefits but also on how players anticipate the behavior of others in response to their actions.

We reviewed key papers that build on these ideas—beginning with the concept of the core and stability in graph-restricted games, moving to farsighted stability, and then to models such as Piccione and Razin (2009) that connect stability to power relations. Finally, we applied these concepts to the Farrell–Scotchmer model and proved that the largest consistent set is unique and coincides with the Farrell–Scotchmer partition. Together, these results show how different models of cooperation use the same underlying ideas to explain why some coalitions remain stable over time.

References

[1] Debraj Ray. *A Game-Theoretic Perspective on Coalition Formation*. Oxford University Press, 2007.

[2] Roger B. Myerson. Graphs and cooperation in games. *Mathematics of Operations Research*, 2(3):225–229, 1977.

[3] Michael Suk-Young Chwe. Farsighted coalitional stability. *Journal of Economic Theory*, 63(2):299–325, 1994.

[4] Joseph Farrell and Suzanne Scotchmer. Partnerships. *The Quarterly Journal of Economics*, 103(2):279–297, 1988.

[5] Michele Piccione and Ronny Razin. Coalition formation under power relations. *Theoretical Economics*, 4(1):1–15, 2009.

[6] Elena Parilina and Artem Sedakov. Stable cooperation in graph-restricted games. In Leon A. Petrosyan and Nikolay A. Zenkevich, editors, *Contributions to Game Theory and Management*, pages 271–281. Saint Petersburg State University, 2014.