POWER, SHARING RULES, AND STABILITY IN COALITION FORMATION: THEORY AND EXPERIMENT

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ECONOMICS

MAY 2016

By

Karl Robert L. Jandoc

Dissertation Committee:

Ruben Juarez, Co-Chairperson
James Roumasset, Co-Chairperson
Ekaterina Sherstyuk
Nori Tarui
Gurdal Arslan, University Representative

Keywords: Coalition Formation, Self-enforcement, Sharing Rules, Experiment

DEDICATION PAGE

To Giezl and Alon–members of my strongest and most stable coalition.

ACKNOWLEDGMENTS

There is an old Filipino saying: *Ang hindi marunong lumingon sa pinanggalingan ay hindi makakararating sa paroroonan.* He who doesn't look back will never reach his destination. The saying conveys a sense of groundedness, that one must acknowledge the past–the struggles, the right connections with people, the favorable circumstances–in order to move forward to greater achievements.

Looking back these past few years, I realized that finishing my PhD required not only my best effort but also needed a happy confluence of events, people and opportunities. In the following, I shall try my best to enumerate these but, alas, the space allotted is not enough for a full accounting.

First, my sincerest gratitude goes out to my advisers Ruben Juarez and Jim Roumasset. Ruben hired me for my first project, the product of which is the chapters of this dissertation. I never thought I had the smarts, nor the grit, for formal economic modelling, a special area reserved for the mathematical and logical gods and not for mere mortals like me. But brick by brick he imparted to me the tricks of the trade until I developed my self-confidence. Ruben's painstaking patience with me is legendary, and truly deserves an award. Jim, on the other hand, is the mentor who always encouraged me to look at the big picture–a model is only good if it contributes to understanding a complex phenomenon. He is a superb theorist with his ears on the ground, eschewing fads and fancies for more fundamental explanations of economic reality. His favorite maxim is, "it is natural in academic life to lean against the wind", attributed purportedly to Ken Arrow (still, after all these years, I can't find the source nor the particular paper containing this passage). Notwithstanding that, this is characteristic of Jim: he may ruffle some feathers but still his disdain for the popular is the only way to advance knowledge.[1]

I also thank my committee members Katya Sherstyuk, Nori Tarui and Gurdal Arslan. They provided critical but valuable comments on the different incarnations of the chapters. My deep appreciation and thanks goes especially to Katya, who sacrificed a lot of her time helping me improve Chapter 4 of this dissertation.

[1] An aside: a big multilateral institution asked Jim to write something on a contentious Philippine policy. He asked me to collaborate on it. After we submitted our paper, we never heard back from the institution. Our suspicion is that our findings did not conform to their taste.

I would also like to express my thanks to those who partnered with me in this struggle. My friends at UH Pinoy, UP Alumni Association and IBF who shared their time, money and energy–whether picking me up at the airport, lending me money during periods of cash flow problems, or just plain laughing at the world around us. Because of them my life became a little richer despite being paid just a little above the poverty guideline for Hawaii.

Speaking of poverty guidelines, I would also like to thank those who provided additional funding that enabled me to escape penury during my stay at UH: Michael Roberts, UHERO, the Air Force Office of Scientific Research (via Ruben Juarez), the UH GSO, the UH CSS, SeaGrant, the Seiji Naya Outstanding First-Year Student Award, the Mok-Mak Scholarship Award, and Jim Roumasset.

Finally, I would like to thank my wife Giezl. My bedrock. My constant. She made sure that I was whole physically, mentally and spiritually. What a wife.

And above all, everything is for God's greater glory.

ABSTRACT

The chapters included in this dissertation represents an attempt to incorporate the notion of power in examining stability in coalition formation games. Power can emanate from different sources: wealth, military might, or political influence. Whatever the source, power is used by agents to impose their preferences on the rest of society. In the coalition formation games presented in this study, agents who are endowed with power compete for a divisible resource by forming coalitions with other agents. The coalition with the greatest power wins the resource and divides it among its members.

Over time, coalitions form and disintegrate and different coalitions yield different payoffs for agents. Which coalitions will ultimately form when agents have heterogenous power? Can we find rules to select coalitions that are stable once formed and at the same time gives its members their highest payoff? The chapters in this study employ an axiomatic approach of searching for these rules that satisfy two main desirable properties: *self-enforcement*, which requires that no further deviation happens after a coalition has formed, and *rationality*, which requires that agents pick the coalition that gives them their highest payoff. The existence of these rules may be sensitive to the different features introduced in our coalition formation game.

This dissertation and its key results are organized as follows. Chapter 1 enumerates the various applications of coalition formation in some key areas such as industrial organization, trade and international economics, public economics, environmental and resource economics, and political economy. It also introduces the two main approaches that is used in modelling how agents form coalitions. Chapter 2 investigates the implication of power dynamics, where power changes according to the resource share. The main findings show that self-enforcement may not be satisfied in some specific cases, and restrictions on the domain of power or the types of coalitions that can form are needed to satisfy self-enforcement and rationality simultaneously. Chapter 3

disallows power accumulation and investigates the sensitivity of these rules to different ways of dividing the prize (sharing rules). The main findings reveal that, in general, sharing rules may cause disagreements among agents on which coalitions should form. The chapter provides the sufficient and necessary conditions of the sharing rule parameter under which these disagreements can be ruled out. Chapter 4 conducts an economic experiment using a simplified version of a coalition formation game to investigate possible behavioral factors that may explain deviations from theoretical predictions. The main findings show that agents display rational behavior when forming coalitions, especially when they know that a large proportion of their opponents play myopic strategies in the beginning. Over time, however, agents learn to behave more strategically and even more rationally, and thus enables agents to display more of the behavior predicted by our model. Chapter 5 summarizes the key findings of the previous chapters and provides possible further directions for research on coalition formation.

The models presented in this dissertation illustrate how power accumulation, the sharing rule, and possible behavioral factors, shape the way coalitions will be stable throughout time.

TABLE OF CONTENTS

1 COALITION FORMATION IN ECONOMICS: APPLICATIONS AND APPROACHES

While economic decisions are often made by groups rather than by individuals, it is only in the past 70 years or so where cooperative behavior has been systematically and formally examined. Starting from the seminal work by von Neumann and Morgenstern [102], the literature on cooperation and coalition formation in various contexts and applications has exploded exponentially, and unifying the fragmented nature of this literature proves to be a Herculean undertaking. The following classic economic applications illustrate some of the nuances of how to model strategic interaction in a group/coalition formation context.

Industrial Organization

The outcome of the strategic interaction of firms was first formally described by Cournot in 1838, whose analysis opened the floodgates of studies on cartels and the role of coalitions in industrial organization. The salient feature of these groups is that members of the coalition exert a positive externality over non-coalition members inasmuch as the increase in price due to the decrease in cartel output also benefits non-cartel members. An interesting question from this literature is: What is a stable coalition in the face of such externalities? Bloch [14] shows that the stability of cartels depends on how agents anticipate the actions of the coalition members after some agents leave. For instance, if remaining members of a coalition still stay together after a defection of some members (the δ model), then no cartel will be stable since there is always an incentive to defect since the payoff of being outside the cartel is higher than being a member of one. In contrast, if the coalition is dissolved after a defection (the γ model), then there will be no incentive to deviate and thus there will be some optimally sized cartel that can form.

Recent studies have studied various reasons why stable cartels exist. These justifications include synergies from horizontal mergers (Perry and Porter [77], Farrell and Shapiro [34]), capacity constraints (Nocke [73]), moral hazard (Espinosa and Macho-Stadler [31]), differentiated products (Deneckere and Davidson [26]), to name a few. By implicitly imposing arbitrary stability conditions, however, these studies leave a gap in the theory of cartel formation. For instance, the assumption of farsightedness may imply that a firm compares profits not only as a member of its current cartel but also from perspective of other cartels that might exist in the future.

Understanding how firms collude and the manner in which the cartel divides the surplus can also have important implications for competition policy.

Trade and International Economics

The popularity of free trade areas and customs union in international trade has raised concerns over the effect of these arrangements on the goal of creating a global free trade area. Several unsettled research questions in this topic include: "How can there be global free trade if there is an option to form free trade areas and customs union?" and "How can we divide gains and losses under global free trade?". A popular concept used in answering the first question is the concept of the core, including the condition that no coalition can block (or improve upon) the payoffs they get under global free trade. Riezman [86] develops a model where countries may prefer to form a free trade area or customs union and block the proposal of global free trade. That is, global free trade may not be inside the core. However, the reason why such proposals are blocked is that the model precludes the possibility of monetary transfers from global free trade winners to losers. Indeed, if there are efficiency gains from global free trade, then it may be possible to create a mechanism to divide these gains that will be amenable and respected by all parties. While there are papers that demonstrate that such transfer mechanisms can be implemented in the level of a custom union (e.g. Grinols [42], Kowalczyk and Sjostrom [57]), it is still an open question whether such transfer mechanisms can be extended globally.

To some extent, these theories rest on arbitrary assumptions. For instance, what sharing mechanism best satisfies both efficiency and some notion of "fairness"? Another gap is the role of power of the countries inasmuch as countries are heterogeneous in terms of income level, potential levels of exports and imports, and political influence and military might.

Public Economics

One of the earliest insights in modern economics is that the decentralized provision of public goods may be inefficient. Samuelson [90], formalizing Lindahl's approach of attaching personalized prices of the pure public good to each agent, represents the earliest attempt to reconcile efficiency of public goods provision with a quasi-market market mechanism. As

Samuelson acknowledged, this approach does not address certain game theoretic difficulties, including the problem of preference misrepresentation.

Tiebout [99] observed that rather than being "pure", most public goods are "local" in nature. That is, there is a class of public goods provided by local jurisdictions that are subject to congestion. He further hypothesized that these local jurisdictions, by offering various packages of (local) public good levels and tax payments, are able to incentivize agents to reveal their true preferences. That is, in a "large enough" economy with competition between jurisdictions and costless relocation, agents are incentivized to sort themselves into homogeneous jurisdictions (clubs) such that an equal sharing rule accurately reflects member willingness-to-pay for public goods (Conley and Smith [23]).

A celebrated result from cooperative game theory, due to Debreu and Scarf [25], is that the set of competitive equilibrium allocations is contained in the core.[1] Furthermore, under a sufficiently large replication of the economy, allocations that differ from competitive equilibria are blocked–that is, the core "shrinks" to the set of competitive (Walrasian) allocations as the economy enlarges. Since competitive allocations are Pareto efficient, and the core shrinks to competitive allocations, then by implication the core allocations are efficient. Wooders [103] shows that under certain assumptions (namely, that congestion does not depend on an agent's identity and that gains from scale in population can be realized by small coalitions), the shrunken core is equivalent to the set of Tiebout equilibrium states.

Environmental and Resource Economics

The environment exhibits a global commons problem: a country can free ride in efforts to improve environmental quality, that is, it benefits from other countries' efforts even if the country itself did not contribute. This presents a challenge to any cooperative agreement on an international scale. How can global environmental cooperation emerge when countries have the incentive to free ride? The popular approach to solve this problem is to model cooperation as the outcome of a non-cooperative game with two stages: in the first stage, a country noncooperatively chooses whether or not to sign an agreement (to join the coalition); in the second stage, a country decides its environmental policy given the outcome of the first stage and any agreed upon division of gains and losses. Carraro [21] argues that this approach leads to several pertinent findings: first, even absent any commitment to cooperation, countries can find an agreement that can be

[1] The core refers to the set of feasible allocations that cannot be improved upon (blocked) by a coalition of agents.

self-enforcing; second, the grand coalition where all countries cooperate may not emerge; third, linking issues such as trade to the environmental agreement may expand the coalition; and finally, membership rules (e.g. open membership, exclusive membership, unanimity) will be important in determining the shape of the coalitions that will emerge.

Another important application of group/coalition formation in resource economics relates to problems of resource extraction in common pool resources. A feature of these open-access resources (e.g., fishing grounds, water in a river) is that there is typically an incentive for agents to extract more than the optimal harvest level. While state takeover and imposition of private ownership have been traditional solutions to this problem, Ostrom [75] observed that there are other forms of governance without explicit private ownership, but where agents govern the resource through "common property".

Despite Ostrom's pioneering efforts demonstrating apparent effectiveness ("sustainability") of group management, the theory of resource governance by common property remains underdeveloped, particularly regarding endogenous coalition formation, including coalition size and stability, ability to punish deviators, and the extent to which information is spread around the group[2] (Dixit [28]). One component of common-property resource governance regards the sharing rules to be adopted, which hinge on questions of efficiency, perceived fairness, and implementability. For instance, Libecap's [60] example of drillers in Texas oil fields recognizes the need for "unitization"–an arrangement where several firms designate one firm (usually with the largest area) to operate and develop the field as a whole and share the returns based on a pre-negotiated formula based on each member's "units" (typically acres of land overlying the common oil pool). In some cases, private/asymmetric information may makes it difficult to arrive at an appropriate sharing scheme. Where agents are homogeneous, equal sharing becomes a plausible solution (Ostrom [75]), thereby simplifying contracting and administrative problems.

Political Economy

The application of coalition formation to political economy issues is pervasive: political institutions, formation of cabinets and legislatures, and examining the cause of public policy, to name a few.

[2] See Copeland and Taylor [24] for a promising approach, albeit without endogenous coalition formation and limited to steady state comparisons.

The phenomenon of the emergence of political parties has been studied in a coalition formation context. Besley and Coate [11] model an individual's decision to be a candidate as the outcome of the trade-off between the costs of fielding himself as a candidate and the benefits of obtaining the best policy that can be implemented by the candidate. Political parties emerge when a coalition of like-minded candidates are able to share the burden of candidacy or increase their probability of winning in an election.

In the context of formation of legislatures, the seminal paper by Baron and Ferejohn [7] develops a bargaining model where a random legislator is chosen and can propose a division of the surplus among a fixed number of identical districts. If a majority votes for approval, then the proposed division is implemented. If it is rejected a new proposer is chosen and the bargaining game is played anew. The implications of this model can be stark. Consider the example of three political parties with insufficient votes to constitute a majority by themselves. The prediction of the model is that the chosen proposer proposes a division that will just barely make the other party accept, which is just the other party's expected payoff if the bargaining game continued. Thus, it will be the case that the proposing party will want to form a coalition with the other political party that has the lowest expected payoff from continuation. Hence, when parties have high discount rates, the division will likely be unequal.

Another interesting application of coalition theory in political economy relates to the question, "What explains the public choice of agricultural policies?". For instance, Balisacan and Roumasset [6] propose a model of investment in political influence within opposing coalitions to explain the positive relationship between agricultural protection and per capita income. Coalitions form endogenously according to the benefits of increased political influence versus organizational costs as functions of investment by the opposing coalition. Equilibrium protection rates are thus explained as the outcome of the Nash-Cournot non-cooperative game between coalitions. The limitation of this promising theory is that marginal benefits of adding an additional member to a coalition are assumed to be known. Yet this would depend on being able to rank members from most to least productive members, which would depend in turn on sharing rules and assessments of stability.

1.1 Modern Approaches in Coalition Formation Theory

The preceding examples highlight the breadth of coalition formation applications. Each of the examples mentioned relates to a specific problem, whether it is to increase prices, divide gains or losses, reveal preferences, or influence policies. This naturally necessitates a proliferation of cooperation/coalition formation models according to the diverse contexts. Despite the diversity of coalition models, the above review suggests that there are some commonalities in the nature of the assumptions used to model these situations that motivate further game-theoretic exploration. To further improve and advance coalition formation theory, one should attempt to strengthen the commonalities of these diverse models and, at the same time, enrich these with reasonable extensions towards different notions of coalition stability, rules of the coalition formation game, and descriptions of interaction among agents. [3] Any advance in the theory, however, must navigate between two classical approaches in the coalition formation literature (Ray and Vohra [83]). These two approaches represent two main methodologies in development of the theory. The first approach, called *blocking approach*, involves the immunity of coalitional alliances against blocking by other potential coalitions. In the classic definition, a coalition blocks a feasible allocation if members of this coalition can find another feasible alternative allocation that improves the payoff for each of them. In this approach, the coalition is the fundamental unit of analysis and does not specify any explicit game based on moves by the individual. It is the process of blocking and counter-blocking by the coalitions that enables them to arrive at some notion of stability.

The second approach, named the *(noncooperative) bargaining approach*, involves agents making proposals to form a coalition and other responding agents to either accept or reject these proposals. Thus, the fundamental unit of analysis in this approach is the individual based on explicit decision-theoretic foundations.

1.1.1 Blocking Approach

Most models of what is to be known as "cooperative game theory" takes the coalition as the primitive decision making entity.[4] The basic question in this literature is "what can the coalition improve upon for its members?" (Wooders and Page [104]). The notion of a coalition "improving upon" a certain allocation is captured by the idea of *blocking*. When a coalition blocks a certain allocation by another coalition, this means that the original allocation achieved by that coalition is unstable and can therefore be ruled out. The most popular solution concept[5] in this literature is

[3] Wooders and Page [104], Gillies [40] as well as Ray and Vohra [83] provide an excellent overview on the evolution of economic theories on coalition formation.

[4] Binmore [13] differentiates noncooperative game theory from cooperative game theory not as the absence of conflict in the latter but rather that agents have access to a technology of resolving all the problems of commitment and trust. It is not also true that noncooperative game theory implies the absence of cooperation. Rather, cooperation is the noncooperative setting is the outcome of strategic interactions of agents.

[5] A "solution concept" is a formal rule that should be satisfied by any outcome to be viewed as stable. (Wooders and Page [104])

the concept of the *core* (Gillies [39]). The core consists of allocations that are never blocked. Put another way, an allocation is not in the core if some coalition can object to it by enforcing an alternative allocation that all its members prefer.

In many situations the core is empty.[6] In the context of our coalition formation game, there are instances where due to the nature of the way coalitions form, there will always exists a coalition that can improve upon (or block) the outcome of one coalition (see the example in Case 4 in Section 2.1.1). In some cases a restriction on the domain of games would be necessary for an equilibrium to exist.

There are several complications in the blocking approach that is pertinent to the Chapters in this study. First, there is the issue of blocking that is followed by further blocking in future moves. In this scenario, it is important to ask whether agents receive payoffs that directly emanate from their blocking or would they take into account the implications of their blocking as further blocks ensue in the future. Thus, the issue of of farsightedness (whether agents are forward-looking) becomes critical in prescribing a solution concept. If agents are farsighted, then they care only on the playoffs of being in the final coalition that forms rather than being part of any intermediate coalition that eventually dissolves.[7]

The second complication regards the notion of power. In the traditional cooperative game theory program, power emanates from the coalition's "inalienable attributes", reflected mostly on the ability of the coalition to appropriate value (Rowat and Kerber [88]).[8] In more recent studies, several authors (Jordan [50], Rowat and Kerber [88]) have described coalition power as emanating not only from "inalienable attributes" but also emanating from resource holdings (i.e., wealth). Jordan [50] proved that the stable sets in these *pillage games* can be represented as a (farsighted) core.

Most of the examples in the applications we have shown earlier present gaps in modelling these complications. In the theory of cartel formation, for instance, farsightedness may imply that membership in the current cartel must be immune to firm deviation from all possible cartels that may form in the future as opposed to the myopic membership where the firm assumes that the cartel he leaves will still continue to operate even without it. In the example of environmental coalition models, it may be possible that power may differ inasmuch as countries differ in terms

[6]In other situations the core is very large such that the core gives little guidance regarding the actual solution (Foley [36]; Sandler [91]).

[7]See for instance, Konishi and Ray [56], Chwe [22], Xue [105], Diamantoudi and Xue [27], Ray and Vohra [84].

[8]In transferable utility (TU) games, this value is represented by the "characteristic function" (see Moulin [66], Gillies [40]).

of income level, trade capacity, and political influence. Chapters 2 and 3 explore how to take these complications into account in the context of the blocking approach. Both chapters assume that agents are farsighted such that the equilibrium concept takes into account further blocking among coalitions. In addition, Chapter 2 incorporates bargaining power by representing power as an agent's share in the resource.

Another popular solution concept in the blocking approach is the Shapley value (Shapley [95]). As opposed to the Core that can be empty, Shapley [95] famously showed that a single-valued solution concept is guaranteed to exist for every cooperative game.[9] The Shapley value, assigns to each agent his expected marginal contribution to the different coalitions.[10] This solution concept has been used in a wide variety of applications, especially in the cost-sharing literature (e.g., Littlechild and Owen [61], Moulin [67]). In this line of literature, the main focus is how to divide the cost in a "fair" manner such that the grand coalition (the coalition composed of all agents in the society) can be induced to cooperate.

Rather than asking how the grand coalition can be induced to form, Chapter 3 asks a related question: given a fixed sharing rule, what are the coalitions that could possibly form? Since the sharing rule (which specifies the division of the resource) provides the incentive structure for agents to form coalitions, it is intuitive to expect that different sharing rules may lead to different coalitions that could possibly form.

1.1.2 Bargaining Approach

In the bargaining approach, the individual regains its status as the primitive decision making entity. In place of blocking by coalitions, individuals in this approach make proposals and responds to such proposals. This approach has its roots in the *Nash program* by using noncooperative game theory to test the predictions of agreements. The canonical *Nash bargaining solution* (Nash [69]) is a solution that satisfies several axioms including Pareto optimality. Hence, there always exists a solution to the two-person bargaining problem that will yield efficient outcomes. Nash pointed out that his bargaining solution can be the result of agents simultaneously committing to take-it-or-leave-it-demands (Binmore [13]). In his contribution to the Nash program, Rubinstein [89] showed that alternating offers until an agreement is reached (with sufficient discounting) lead to a unique subgame-perfect equilibrium.

[9]The Shapley value is a well-established solution concept when agents can transfer utility among themselves (the *transferable utility* assumption in cooperative game theory). In the case of n-person pure bargaining models (where the grand coalition forming or the total absence of cooperation are the only possible outcomes), which can be modeled with or without transferable utility, a widely popular solution concepth is the Nash bargaining solution (Nash [69]).

[10]See Moulin [66] for a discussion of the Shapley value.

Acemoglu, et al [1] used a dynamic model with this approach where agents propose a coalition with other agents either rejecting or accepting these offers. In this model, agents who are not part of the proposed coalition that eventually gets passed (accepted by all the members) will not be part of future coalition formation processes (that is, the agents outside the coalition are “killed”, in the sense that they drop out of the game and cannot organise nor join another coalition). Thus, these farsighted agents will only be interested in joining coalitions that will not ensue a deviation in the future, that is, only the ultimate or final coalition will be considered. Therefore strategies wherein agents who propose will only propose the final coalition and responders will only accept this final coalition (and nothing else) can support a subgame perfect equilibrium.

Chapter 4 uses this approach to investigate whether the predictions of the coalition formation models hold in the real world. Specifically, it uses the dynamic proposal-response model to test whether agents behave as theoretically predicted. Where the results deviate from theoretical predictions, we investigate whether certain behavioral factors come into play. This fills the gap in several political economy bargaining models presented earlier. First, we present a bargaining model with a different notion of stability–“self-enforcement”. Self-enforcement is appealing since it does not require members to commit to write binding agreements when they form coalitions. Since members are free to leave an existing forming coalition, member agents should not have sufficient power, nor sufficient incentives, to deviate from their current coalition. Second, we examine behavioral factors that come into play, for instance, bounded rationality (myopia), learning and the size of the resource that the coalitions will fight over. Adding these features to the traditional bargaining models will help enrich the literature and will take us a step closer to understanding fundamental coalition formation processes.

1.2 Conclusion

The models presented in this study illustrate how power accumulation, the sharing rule, and possible behavioral factors shape coalitional stability through time. The two approaches (blocking vs. bargaining) offer different methodologies regarding the modelling of how coalitions form. We use the strengths of these two approaches to investigate how different features introduced in the models affect the way coalitions form over time. Chapter 2 investigates the implications of agents in the winning coalition accumulating power over time according to their share of the resource. Chapter 3 disallows power accumulation but investigates the sensitivity of coalition formation to different ways of dividing the prize (sharing rules). Chapter 4 explores behavioral factors in an experimental setting that may explain deviations from the theoretical predictions of Chapters 2 and 3.

2 SELF-ENFORCING COALITIONS WITH POWER ACCUMULATION

2.1 Introduction

Recent developments in the coalition formation literature have focused on different notions of stability.[1] On one hand, this is motivated by popular results in political economy which show that stability in the ruling coalition leads to more favorable welfare outcomes in the long run. This is because stable coalitions have a higher "effective discount factor" that enables them to look at longer horizons for the consequences of their actions.[2] On the other hand, there have been recent attempts to test the robustness of different stability notions by adding dimensions relating to power, resource sharing and other institutional constraints. "Power"–which is measured in the literature in terms of wealth, military might or political influence–reflects the ability of a coalition to impose its will on the rest of the society. Another enrichment of recent coalition formation models involves how resources (prize or spoils) are divided within the winning coalition. Finally, variations in the institutional environment of the coalition formation process (in our case whether agents are "killed", that is, unable to participate in future coalition formation games, or are able to survive and still play in future games) will impinge on the possible coalitions that will form.

In this paper we present a dynamic model of coalition formation where farsighted agents are endowed with power and compete for a divisible resource by forming coalitions with other agents.[3] If a formed coalition has sufficient power to win, its members will divide the resource

[0]I thank Ruben Juarez for overall guidance in writing this chapter.

[1]For a recent survey of the growing literature on coalition formation, refer to Ray [82], Ray and Vohra [83], Bloch and Dutta [15].

[2]See, for example, McGuire and Olson [62] and Olson [74] where they compare public policy outcomes of a long-lived ruling coalition ("stationary bandit") to a short-lived one ("roving bandit"). In the context of a principal-agent political economy model, Barro [8], Ferejohn [35], Persson, Roland and Tabellini [78, 79] also find that politicians who are likely to remain in office for an extended period of time also support better policies. In contrast, Acemoglu, Golosov and Tsyvinski [3] disputes the finding that stable ruling groups lead to fewer economic distortions. This result happens when ruling coalitions can be rewarded or punished even after they have left power.

[3]The assumption of farsightedness assures that agents care about their payoff from being included in the limit coalition that forms rather than their immediate payoffs. See for instance, Konishi and Ray [56], Chwe [22], Xue [105], Diamantoudi and Xue [27], Ray and Vohra [84].

among themselves and their power will increase according to the member's share of the prize.[4] We focus on two sharing rules of the prize: sharing the prize in proportion to an agent's power in the winning coalition (*proportional sharing*) and sharing the prize equally among the winning agents (*equal sharing*). An institutional constraint emerging from the dynamic coalition formation model is the span of life of the agents. We focus on two polar cases. The first case is when agents who are not part of the winning coalition do not participate in the future (*agents are killed*). The second case is when all the agents remain in the game regardless of whether they are part of the winning coalition or not (*agents survive*).[5]

We employ an axiomatic approach to find the stable coalitions in the four cases above (combinations of proportional vs equal sharing and agents are killed vs agents survive). We focus on two main axioms, *self-enforcement* and *rationality*. In order to describe self-enforcement, notice that since our model is dynamic, it is possible that over time coalitions may disintegrate and new factions may form to overthrown the existing ruling coalition. We assume that agents are forward-looking and form coalitions that are stable from the moment of inception. That is, self-enforcement requires that no subcoalition of the winning coalition is powerful enough to encourage further deviations. Self-enforcement is a robustness property that ensures that the coalition that forms in round one never disintegrates afterwards. In addition, rationality requires that agents pick the coalition that gives them their highest limit payoff among self-enforcing

[4]The canonical example of power accumulation comes from non-democratic societies, where typically a ruling coalition can perpetuate itself in authority because it can use the state's resources to consolidate and accumulate power and wealth. Although data is sparse, there is some evidence that some world leaders amassed personal wealth either due to connections or investments made possible by (sometimes corrupt practices) using state resources. For instance, there is some speculation that Teodoro Obiang Nguema Mbasogo of Equitorial Guinea has amassed a fortune of over $600 million in his oil-rich country since ascending to power in 1979 (see http://www.huffingtonpost.com/2013/11/29/richest-world-leaders_n_4178514.html). Power accumulation is natural in many coalition formation settings ranging from labor unions to military alliances between countries, where the players generate benefits by being together and become more powerful as time passes.

[5]There are valid reasons to believe that the decision to kill agents in the society may be exogeneous, especially in certain political contests or military wars. For instance, Bueno de Mesquita, et al [18] asserts that "[w]hen the private benefits of office or coalition membership are large, people are more prepared to engage in horrendous acts of cruelty against others to ensure that their personal privileges are not lost". The size of these private benefits can be reflected by resources to be won in a coalition formation game. On the other hand, most of the biggest purges in the 20^{th} century are identified with specific personalities, e.g. Saddam Hussein, Adolf Hitler and Josef Stalin (Bueno de Mesquita, et al [18]). To the extent that these purges are sometimes personality-driven, the decision to kill opponents is exogenous. While the case where the decision to kill agents is endogenous is certainly an appealing study, it is beyond the scope of this paper and we leave it for future research. In this paper, we assume that the environment where agents are killed or are able to survive is exogenous.

coalitions. Rationality is related to the traditional axiom of coalitional stability, where no coalition will have the incentive to deviate.[6]

In the next subsection, we provide an illustrative example of how our coalition formation game works and how sharing of the prize and whether agents are killed or are able to survive shape the coalitions that form.

2.1.1 An illustrative example

Society is composed of 4 agents $\{1, 2, 3, 4\}$ with power profile $\pi = [1, 1.5, 8, 10]$. The game is played for an infinite number of periods and for each round the prize to be divided by the winning coalition is $I = 3$. An agent's share of the prize will be added to his power in the subsequent rounds of the game. A coalition's power is simply the sum of the agents' powers inside it. A winning coalition should have more than 50% of the total power in the society at any given round. Since agents are farsighted, they only care about their payoffs from being in the final (or limit) coalition.

Case 1: Non-winning agents are killed, proportional sharing.

First, no coalition of size 1 is winning. On the other hand, at the initial round (call this round 0) all winning coalitions of size 2 involves agent 4, i.e., $\{1, 4\}$, $\{2, 4\}$, $\{3, 4\}$. Suppose $\{3, 4\}$ forms in round 1. Power is added to agents 3 and 4 according to their share in the prize, that is, with proportional sharing agent 3 gets $\frac{8}{18}$ of the prize while agent 4 gets $\frac{10}{18}$. In round 1 after $\{3, 4\}$ forms, agents 1 and 2 are killed; agent 3's power is $8 + \frac{24}{18}$ while agent 4's power is $10 + \frac{30}{18}$. Left alone to themselves, agent 4 can now kill agent 3 since he has the higher power. Thus in round 2 and afterwards agent 4 will be the limit (singleton) coalition. Therefore, $\{3, 4\}$ is not self-enforcing because agent 3 will never agree to form this coalition since he will not be part of the limit coalition. This is true of all coalitions of size 2.

Furthermore, no coalition of size 3 is also self-enforcing. Since all size 3 coalitions involve either agent 3 or 4, at round 1 when we add power to the winning agents either agent 3 or 4 can kill the rest of the agents in the three person coalition. For instance, if $\{2, 3, 4\}$ forms and after sharing the prize, in round 1 agents 2's power is $1.5 + \frac{4.5}{19.5}$, agent 3's power is $8 + \frac{24}{19.5}$ and agent 4's

[6] This is related to immunity to group manipulations in models discussed by Bogomolnaia and Jackson [16], Ehlers [29], Juarez [53], Papai [76].

power is $10 + \frac{30}{19.5}$. Therefore agent 4 can kill both agents 2 and 3 since he has a higher power than the others combined. Following this argument, a three person coalition will never form.

Therefore, the only self-enforcing coalition is the grand coalition $\{1, 2, 3, 4\}$ since no subcoalition has the incentive to deviate following the logical process outlined above.

Case 2: Non-winning agents are killed, equal sharing.

With the prize shared equally among agents in the winning coalition, the coalition $\{2, 3, 4\}$ will be self-enforcing. To see this, note that at round 1 after power has been added agent 2's power is now 2.5, agent 3's power is 9 and agent 4's power is 11. Thus, in contrast to the previous case, agent 4 will not have sufficient power to kill agents 2 and 3. Furthermore, $\{2, 3, 4\}$ cannot deviate to a coalition of size 2 since the higher powered agent in the 2-person coalition can kill the lower-powered agent. Hence, coalition $\{2, 3, 4\}$ can form in round 1 and will be stable forever.

Case 3: Non-winning agents survive, equal sharing.

When non-winning agents survive and the prize is shared equally, coalition $\{3, 4\}$ can be self-enforcing. Unlike in the previous two cases where the higher-powered agent can kill the lower-powered agent in a 2-person coalition, agent 4 in this case will not have sufficient power to deviate from $\{3, 4\}$. To see this, note that agents 1 and 2 will still survive even if $\{3, 4\}$ forms, and thus in round 1 (after $\{3, 4\}$ forms and power has accumulated) agent 1's power (1) plus agent 2's power (1.5) plus agent 3's power $(8 + 1.5)$ is now greater than agent 4's power $10 + 1.5$. Thus, agent 4 cannot deviate to a singleton winning coalition. Coalition $\{3, 4\}$ will form in round 1 and will be stable forever.

Case 4: Non-winning agents survive, proportional sharing.

A curious feature in this case is that coalitions will have the incentive to "jump" from one coalition to another when agents survive. Take the example when $\{1, 2, 3\}$ forms. At round 1 after power has accumulated with proportional sharing, agent 1's power is $1 + \frac{3}{10.5}$, agent 2's power is $1.5 + \frac{4.5}{10.5}$, agent 3's power is $8 + \frac{24}{10.5}$ while agent 4's power remains at 10. At round 2, agents 1 and 2 can ditch agent 3 and align themselves with agent 4 since they can get a higher share of the prize by doing so (note that agent 3's power is larger than agent 4's power at round 1). At round 3, faced with the same incentive, agents 1 and 2 will again move to coalition $\{1, 2, 3\}$ since these agents can get a relatively higher share by deviating to this coalition. This phenomenon is true for any winning coalition other than the grand coalition. This example illustrates the difficulty of finding self-enforcing coalitions in this case. In a later section, we introduce a class of priority mechanisms that satisfy the axioms.

2.1.2 Overview of the results

Section 2.3 of the paper studies the case where agents are killed. Proposition 1 describes the unique transition correspondence that satisfies self-enforcement, rationality and *scale invariance*[7] under proportional sharing. On the other hand, under equal sharing the class of self-enforcing, rational and scale invariant transition correspondences may not exist. Proposition 2(i) provides the largest class of coalition formation games where a self-enforcing and rational transition correspondence exists for equal sharing. In this domain of games, Proposition 2(ii) shows the unique transition correspondence that meets self-enforcement, rationality and scale invariance. Roughly speaking, the transition correspondence characterized in Proposition 2(ii) picks the smallest winning coalition where the powers of the agents are relatively equal, and is of size $2^k - 1$ for some natural number k.

Section 2.4 of the paper examines the case where agents survive. Proposition 3 characterizes the unique transition correspondence under equal sharing that satisfies self-enforcement, rationality, scale invariance and *independence of zeros*.[8] This transition correspondence picks the smallest winning coalitions of size 2^k for some natural number k. Proposition 3 also shows that self-enforcement, rationality, scale invariance and independence of zeros are not compatible under proportional sharing. The class of self-enforcing, rational and scale invariant transition correspondences under proportional sharing is large (see Proposition 4).

The paper is organized as follows. Section 2.2 describes the model. Sections 2.3 and 2.4 study self-enforcing coalitions with power accumulation when agents are killed and survive, respectively. Section 2.5 concludes. All proofs are in Appendix A.

[7] A transition correspondence satisfies scale invariance if any scale in the power vector would not change the coalition chosen by the transition correspondence.

[8] A transition correspondence satisfies independence of zeros if agents with zero power do not affect the coalition chosen by the transition correspondence.

2.1.3 Related literature

There are several ways by which "power" is treated in contemporary coalition formation literature. Piccione and Razin [80] examine how power relations determine the ranking of agents in society. The identity of the coalitions (as characterized by the power of agents within that coalition) determines the social order and thus the structure of society. Jordan [50] characterizes the core and the stable set in a class of coalitional games called "pillage games". In this model, wealth is allocated among the finite agents in the game. A reallocation of wealth among the agents is only made possible by using force. A power function, which is monotonically increasing in membership and the members' wealth, regulates the ability of the agents to use force. The coalition is then able to appropriate the wealth of other less powerful coalitions. Although newer coalition formation models are dynamic in nature, most of them are limited to a static distribution of power among the agents. We fill this gap by exploring the possibility that agents are able to accumulate power over time as long as they are part of a winning coalition.[9][10]

The manner in which coalitions divide the prize among member agents has also received attention not only in the coalition formation literature but also in related political economy studies. Acemoglu, et al [1] [hereafter AES] examine a coalition formation process where agents split the prize in proportion to an agent's relative power in the coalition. On the other hand, Bartling and von Siemens [9] characterize the emergence of equal sharing as an optimal solution to an incentive problem in a partnership. In this paper, we focus on describing the stable coalitions that emerge under these two popular and commonly used sharing rules–proportional sharing and equal sharing. While we assume that the sharing rule in this paper is fixed and exogenously given, an interesting and natural extension not covered here would be to endogenize the sharing rule. There are several strategies to achieve this, for instance, in the context of rent-seeking tournaments, Nitzan [71], Lee [58, 59] and Baik [5] contend that the long-run group sharing rule can fall anywhere from being undefined to extremely equal depending on the context of the mechanism to determine these rules (Nitzan [72]).

Finally, other factors have been built in newer models to capture the institutional environment of the coalition formation process. AES, for instance, considers an extreme non-democratic institutional setting wherein non-winning agents are killed. On the other hand, Sekeris [92], also

[9] This was first articulated by Tullock [101], where he argues that since formal institutions are weak or absent regarding distribution and sharing of power, succession of leaders, and generating consensus, a ruling coalition ("junta") will degenerate into a dictatorship, that is, there will be power accumulation by one of the junta members until he becomes the sole ruler.

[10] Non-winning agents in our model "lose" power either by having lower relative power vis-a-vis agents who are part of the winning coalition or being "killed" or cut off from future coalition formation processes.

in a non-democratic setting, does not allow the possibility that the ruling coalition eliminate opponents. In this paper, we investigate these two polar cases.

Our work is closely related to AES wherein forward-looking agents are endowed with power and form coalitions with the goal of becoming the *ultimate ruling coalition.* In their paper, the winning coalition will split a given resource in proportion to its power and agents outside the winning coalition are killed. However, we extend AES on several fronts: First, we allow the possibility that non-winning agents survive throughout the whole coalition formation game.[11] Second, we allow the winning coalition to accumulate power over time. Third, while AES uses proportional sharing in their model, we also examine equal sharing where agents equally divide the prize among themselves.

2.2 The model

Consider the set $N = \{1, \ldots, n\}$ of initial agents who are endowed with powers $\pi = [\pi_1, \ldots, \pi_n]$, respectively. A coalition S is a subset of N, that is, $S \subseteq N$. The set of coalitions are all possible subsets of N, denoted by 2^N. A coalition formation game is a pair (S, π) where $S \subseteq N$ and $\pi \in \mathbb{R}^S_+$. The set of coalition formation games is denoted by $\mathbf{G}$. We assume that power is additive, that is, the power of coalition S is the sum of all powers of the agents inside the coalition, $\pi(S) = \sum_{i \in S} \pi_i$.[12] We denote by π_S the restriction of the vector $\pi \in \mathbb{R}^N_+$ over coalition S. Let $1_S \in \mathbb{R}^S$ be the vector where all coordinates are equal to 1.

Definition 1 *Given a game* (T, π)*, the* ***set of winning coalitions***[13] *is:*

$$W_{(T,\pi)} = \{S \subset T | \pi(S) > \pi(T \setminus S)\}$$

We fix a sharing rule that allocates the prize to the agents at every coalition formation game. Throughout the paper we devote special attention to two simple and commonly used sharing

[11] In another paper, Acemoglu, et al [2], show that the results in the AES case extend even to situations where agents are not killed. The stable ruling coalition in this case is the "minimally winning coalition" which is the coalition with the smallest power among all winning coalitions. In their dynamic game, this corresponds to the case when all agents are eligible to vote on the proposed coalition (i.e., the democratic case). In the case where the voting power is limited to the ruling coalition but agents are not killed, the mapping in AES is modified to include all alternative coalitions not just the subset of the ruling coalition.

[12] Juarez [52] considers a more general version where power can be any arbitrary monotonic function.

[13] This definition requires winning coalitions to have relative power larger than 50%. Our results below can be easily adapted to require winning coalitions to have relative power larger than α, where $\alpha \geq 50\%$. This is discussed in Section 2.5.1.

rules, equal sharing and proportional sharing (see Juarez [52]). That is, for the game (T, π), let $\xi(T, \pi) \in \mathbb{R}^T$ be the vector of shares under the sharing rule ξ. Under **equal sharing (ES)**, $\xi = ES$, the share of agents $i \in T$ equals $\xi_i(T, \pi) = \frac{1}{|T|}$. On the other hand, under **proportional sharing (PR)**, $\xi = PR$, the share of agents $i \in T$ equals $\xi_i(T, \pi) = \frac{\pi_i}{\pi(T)}$.

Given the sharing rule ξ, let $\xi^N(T, \pi) \in \mathbb{R}^N$ be the embedded vector $\xi(T, \pi)$ into $\mathbb{R}^N$. That is, $\xi_i^N(T, \pi) = \xi_i(T, \pi)$ if $i \in T$ and $\xi_i^N(T, \pi) = 0$ if $i \in N \setminus T$.

The major task of this paper is to examine the stable coalitions that emerge under the two sharing rules discussed above. As we will see below, the type of stable coalitions will greatly vary depending on the sharing rule and whether agents survive or are killed.

2.2.1 Dynamic Coalition Formation

The game is played in discrete rounds. Let t, where $t = 0, 1, \ldots$, denote the time of the game. We define a *transition correspondence* that maps from the set of coalition formation games to a particular set of winning coalitions.

Definition 2 *A* **transition correspondence** *is a continuous*[14] *correspondence* $\phi : \mathbf{G} \rightarrow 2^N$ *such that* $\forall\ (X, \pi) \in \mathbf{G}$*:* $\phi(X, \pi) \subset W_{(X,\pi)}$.

The transition correspondence ϕ selects all coalitions that could be winning at a given game (S, π). The evolution of the coalition formation games at every round depends on the transition correspondence along with the conditions on whether agents survive or are killed in future rounds. The main task of this paper is to axiomatize ϕ in the distinct four scenarios that we describe below.

We denoted by (S^t, π^t) the game at round t. We start with the game (S^0, π^0) and look at the evolution of the game throughout time. We denote this evolution by $(S^0, \pi^0), (S^1, \pi^1), (S^2, \pi^2), \ldots$.

In this paper, we focus at the evolution of the game in two polar cases. First, we consider the case where agents are killed if they are not part of the winning coalition, this is described by the case where $S^t \in \phi(S^{t-1}, \pi^{t-1})$ for all time t. That is, only a group of agents who were winners at time $t - 1$ could participate at time t. Second, we consider the case where agents survive regardless of

[14]A correspondence is continuous if for any sequence of power vectors $\pi^1, \pi^2, \cdots \rightarrow \pi^*$ where $S \in \phi(N, \pi^i)\ \forall i$ and S is winning in π^*, then $S \in \phi(N, \pi^*)$.

whether they are part of the coalition in the previous round. This is described by $S^t = N$ for all time t. That is, the agents participating at every time do not change.[15]

We now describe the manner in which power evolves throughout time. First, we fix the prize $I > 0$ that is divided at every round. The manner by which the power of the agents accumulates will be affected by both the sharing rule ξ and the transition correspondence ϕ. That is, we assume that the prize I is transformed to power in the same proportion. Formally, the power accumulation function for agent i at time $t > 0$ equals:

$$\pi_i^t = \begin{cases} \pi_i^{t-1} + \xi_i(S^{t-1}, \pi^{t-1})I & \text{if } i \in S^t \\ \pi_i^{t-1} & \text{if } i \in N \setminus S^t \text{ and agents survive} \end{cases}$$

Note that the power of the winning agent i increases by $\xi_i(S^{t-1}, \pi^{t-1})$ regardless of whether agents are killed or survive.[16] On the other hand, if non-winning agents survive, then their power does not increase. If agents are killed, they are removed from the game.

Finally, we assume that agents are infinitely forward-looking. That is, they care about their payoff at the limit coalition formation game. This payoff, denoted by $\xi(S^\infty, \pi^\infty)$, is defined as the limit of the sequence $\xi(S^0, \pi^0), \xi(S^1, \pi^1), \xi(S^2, \pi^2) \ldots$, when it exists. If this limit does not exist, $\xi(S^\infty, \pi^\infty) = 0$.[17]

For the two sharing rules discussed above this limit exists when agents are killed. To see this, notice that there exists a time $\bar{t}$ such that S^t does not change for all $t \geq \bar{t}$.[18] The convergence under PR happens because the relative power of the agents in (S^t, π^t) will not change for any $t > \bar{t}$. The convergence under ES occurs because the agents in S^t remains constant for any $t > \bar{t}$.

[15]First, observe that in the case where agents are killed and $\phi(S^{t-1}, \pi^{t-1})$ contains more than one element, we impose no restriction in which coalition from $\phi(S^{t-1}, \pi^{t-1})$ will equal S^t. This allows our results to be more robust, since the evolution of the game includes any potential path of coalitions such that $S^t \in \phi(S^{t-1}, \pi^{t-1})$ for all t.

[16]Implicit in our power accumulation function is the assumption that wealth and power are exactly convertible in the winning coalition, that is, an increase in wealth increases power in the same proportion. Jordan [50] makes a similar assumption in defining a coalition's power function that enables these coalitions to dominate others (see Section 1.3). We assume this particular function for tractability, and relegate to future research the extension to other monotonic functions of wealth and power.

[17]As we will see below, the axiom of self-enforcement guarantees that this limit always exists whether agents are killed or survive.

[18]Indeed, note that $S^0 \supseteq S^1 \supseteq S^2 \supseteq S^3 \supseteq \ldots$, thus the sequence $S^0, S^1, S^2, \ldots$ converges in finite time.

2.3 Agents are killed

2.3.1 Axioms

In this section we focus on the case where if a coalition S forms, agents outside S cannot participate in any future coalition formation process (that is, agents $i \in N \setminus S$ are "killed"). We are interested in finding transition correspondences that map to coalitions wherein agents within this coalition do not have the incentive nor the power to deviate in future rounds of the game. Agents who are chosen by the transition correspondence increase their power according to their share of their prize. We introduce an axiom where a chosen coalition should continue being chosen even after power accumulates.

Axiom 1 (Internal Self-Enforcement (ISE)) *A transition correspondence ϕ is* **internally self-enforcing** *(ISE) if for any game (X, π) and any coalition $S \in \phi(X, \pi)$, we have that $S \in \phi(S, \pi_S + I\xi(S, \pi_S))$.*

Note that $I\xi(S, \pi_S)$ is the accumulated power from the prize shared by the agents inside the coalition. Therefore, ISE requires that if coalition S is picked by the transition correspondence ϕ, the same coalition S will also be picked in future periods even if agents within S have accumulated power (and agents who are not part of S are killed). Given an ISE correspondence and a game (S, π), we say that a coalition S is an *ISE coalition* if it is generated by the transition correspondence.

Axiom 2 (Dynamic Internal Rationality (DIR)) *The transition correspondence ϕ meets* ***dynamic internal rationality*** *(DIR) if for any game (X, π), for any $T \in \phi(X, \pi)$ and for any $Z \subset X$ such that $Z \in W_{(X,\pi)}$ and $Z \in \phi(Z, \pi_Z + I\xi(Z, \pi_Z))$, we have that $Z \notin \phi(X, \pi) \Leftrightarrow \xi_i(T, \pi_T^\infty) > \xi_i(Z, \pi_Z^\infty)\ \forall i \in T \cap Z$.*

DIR requires that if there is another coalition Z that is winning and internally self-enforcing in the game, then coalition Z will not be chosen by the transition correspondence if and only if the agents in the intersection of a coalition T that is chosen by the transition correspondence and coalition Z receive a lower share of the prize in the limit when being at Z.

DIR ensures that the coalition picked by the transition correspondence will yield the highest payoff for all of its members as the number of rounds approaches infinity.[19] This is similar to

[19]DIR is weaker than the requirement that at every round the coalition chosen be the most preferred among ISE coalitions at that time as opposed to our axiom that only requires for the coalition to be preferred at the limit. The results of this section under equal or proportional sharing will not change whether we use this alternative definition of rationality.

other notions of coalitional stability previously discussed in the literature, where a coalition is chosen if it cannot be blocked by another coalition that is winning and self-enforcing.

Axiom 3 (Scale Invariance (SI)) *The transition correspondence ϕ is* ***scale invariant*** *(SI) in the vector of power if for any game (N, π), and any coalition $S \in \phi(N, \pi)$, we have that $S \in \phi(N, \gamma\pi)$ $\forall \gamma > 0$.*

Scale invariance requires that if the relative power of the agents does not change, then the coalition chosen should not change. This is a standard axiom in the literature.

2.3.2 Results

Proportional sharing

The class of transition correspondences that satisfy ISE, DIR and SI depends on the sharing rule. Proportional sharing will always induce an internally self-enforcing coalition because once an ISE coalition S^* forms at the initial stage, the relative power of agent $i \in S^*$ is unaffected by adding the share of the prize $\frac{\pi_i}{\pi(S^*)} \cdot I$.

Let the transition correspondence ϕ^* be defined as:

$$\phi^*(S, \pi) = \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\min} \; \pi(M) \tag{2.1}$$

where $Q(S, \pi) = \{T \subsetneq S \mid T \in W_{(S,\pi)}, T \in \phi^*(T, \pi_T)\}$

This transition correspondence defines for the game (S, π) a set $Q(S, \pi)$ of proper subcoalitions, which are both winning in S and self-enforcing. It picks the coalition(s) that has(have) the smallest power in $Q(S, \pi) \cup \{S\}$ (if there are multiple coalitions with the same power, then ϕ^* picks all of them). Every chosen coalition yields the highest limit payoff for the agents in the intersection of all the coalitions contained in $Q(S, \pi)$. If $Q(S, \pi)$ is empty then it picks coalition S itself. Thus, $\phi^*(S, \pi_S) \neq \emptyset$ for any game (S, π), hence ϕ^* is well defined.

Proposition 1 *Under proportional sharing, a transition correspondence ϕ satisfies ISE, DIR and SI if and only if $\phi = \phi^*$.*

The proof of Proposition 1 is in the Appendix.

The process by which ϕ^* chooses a coalition can be described inductively. We illustrate this process below.

Example 1 (ϕ^* and proportional sharing) *Consider the game*

$$(S, \pi) = (\{1, 2, 3, 4, 5\}, [16, 17, 18, 20, 29])$$

and prize $I = 10$. At the outset, note that with proportional sharing, the relative power within the winning coalitions will be the same as the case when powers have not increased. For instance, if winning coalition $T = \{1, 2, 3\}$ forms, then agent 1's relative power is still $\frac{16}{51}$ even when the power has accumulated by distributing the prize within this coalition.

The inductive process by which we choose a coalition with ϕ^* involves the following steps. First, start from any coalition of size 1. Note that all coalitions of this size will not be chosen by ϕ^* since none of them are winning. The same is true for all coalitions of size 2. On the other hand, all coalitions of size 3 are winning. One requirement of ϕ^* is that the game generated by the winning coalition should map into the same coalition when other non-winning agents are killed. This requirement is satisfied by any size 3 coalition. To see this, take for example the game $(T, \pi_T) = (\{1, 2, 3\}, [16, 17, 18])$ generated by the winning coalition T within the game (S, π). Coalition T is winning in S since $\pi(T) = 51 > \pi(S \setminus T) = 49$. Now within T, no coalition of size 1 would deviate since no agent has sufficient power to be winning within the game (T, π_T). Furthermore, no coalition of size 2 is internally self-enforcing. That is, a coalition of size 2 will not deviate from T since after the non-winning agent is killed, the agent with the higher power can kill the lower-powered agent in the next round. For instance if coalition $V = \{1, 2\}$ deviates from T, then in the next round after power has accumulated, agent 2 can kill agent 1 since he has the higher relative power ($\frac{17}{33}$). Thus, agent 1 (who is forward looking) will never agree to form the coalition V. In this case, any coalition of size 3 is internally self-enforcing. On the other hand, coalitions of size 4 and size 5, even though they are winning, will not be self-enforcing. This is because a coalition of size 3 will be winning in the games generated by coalitions of size 4 or 5, and those size 3 coalitions will be self-enforcing as we have shown earlier.

Finally, among all size 3 self-enforcing coalitions, the transition correspondence ϕ^* will choose only the coalition $\{1, 2, 3\}$. This is because of all size 3 coalitions that agents 1,2, or 3 are a part of, it is the coalition $\{1, 2, 3\}$ which maximizes their share of the prize since they have the highest relative shares in this coalition. Thus, $T = \{1, 2, 3\} \in \phi^*(S, \pi)$.

In contrast with the game above, for the game

$$(M, \pi) = (\{1, 2, 3, 4, 5, 6, 7\}, [11.5, 12.5, 13, 14, 15, 16, 18])$$

the coalition chosen by ϕ^* will be the grand coalition $\{1, 2, 3, 4, 5, 6, 7\}$. This is because no coalition of size 1,2 or 3 is winning within (M, π). In addition, a coalition of size 3 will deviate from any coalition of size 4, 5 or 6 and will be internally self-enforcing after the non-winning agents are killed. Thus, only the grand coalition will be internally self-enforcing within the game (M, π).

Equal sharing

In contrast with proportional sharing, the example below shows that with equal sharing, a transition correspondence that satisfies ISE, DIR and SI may not exist for all classes of games.

Example 2 (Equal sharing and agents are killed) *Consider the game*

$$(M, \pi) = (\{1, 2, 3, 4, 5, 6, 7, 8\}, [20, 15, 14, 13, 12, 11, 10, 5])$$

with $I = 10$ and equal sharing. Under this sharing rule, if a coalition S forms within (M, π) and continues to form forever, then the relative power of each agent at the limit approaches $\frac{1}{|S|}$. In this example, a coalition of size 3 cannot form since it does not have enough power to do so (the three highest powered agents only have power 49). A coalition of sizes 4, 5, 6 or 7 will not be internally self-enforcing since if we add the share of the prize to the agents, then a 3-person coalition can deviate and be self-enforcing, following the same logic as in Example 1. For instance, if $S = \{1, 2, 3, 4, 5, 6, 7\}$ forms, then after adding $\frac{10}{7}$ to each agent's power $T = \{1, 2, 3\}$ can deviate, since $\pi(T) = 49 + 4.3 > \pi(S \setminus T) = 46 + 5.7$. The grand coalition is not self-enforcing, since at the limit when relative powers are equalized, a 7-person coalition can deviate and be internally self-enforcing.

A way out of this dilemma is to impose some restrictions on the class of games allowed. We say that $\tilde{G}$ is a feasible domain of games if $(S, \pi) \in \tilde{G}$ implies that $(S, \lambda\pi + \lambda\frac{I}{|S|}1_S) \in \tilde{G}$ for all $\lambda > 0$ and $I > 0$.

A feasible domain of games is a class of games where a transition correspondence that meets DIR and SI is well defined.

Definition 3 (Strongly Balanced Game) *A coalition formation game (Y, π_Y) is* ***strongly balanced*** *if* *i.* $|Y| = 2^k - 1$ *for some* $k \in \mathbb{N}$*, and*

ii. *the coalition with the* 2^{k-1} *smallest agents is a winning coalition in the game* (Y, π_Y)*. That is, after renaming the agents in* (Y, π_Y)*, if* $\pi_1 \geq \cdots \geq \pi_{2^k-1}$ *then*

$$\pi_1 + \cdots + \pi_{2^{k-1}-1} < \pi_{2^{k-1}} + \cdots + \pi_{2^k-1}.$$

Condition i of Definition 3 restricts the cardinality of strongly balanced games to 1, 3, 7, 15, 31, etc. Condition ii of Definition 3 basically requires that agents in the society have powers that are not too "far off" from each other. This restriction ensures that there will be no subset of agents from the original forming coalition that will be powerful enough to ensue a subsequent deviation among themselves. This condition is weaker than size-monotonicity assumption in the literature, where coalitions of larger size have larger power.

Example 3 (Strongly balanced games) *Consider the game*

$$(Y, \pi_Y) = (\{1, 2, 3, 4, 5, 6, 7\}, [24, 16, 22, 18, 26, 20, 25]).$$

This game is strongly balanced. It clearly satisfies part i*. To see that it satisfies part* ii*, note that the coalition with* 2^{k-1} *(in this case, 4) lowest powers is winning within* Y*, that is,* $\{2, 3, 4, 6\}$ *has power* $\pi(\{2, 3, 4, 6\}) = 76$ *while the three remaining agents* $\{1, 5, 7\}$ *has power* $\pi(\{1, 5, 7\}) = 75$.

Now, consider another game

$$(V, \pi_V) = (\{1, 2, 3, 4, 5, 6, 7\}, [12, 16, 15, 22, 13, 17, 14]).$$

This does not satisfy Part ii *of the definition since the 4 agents with the least powers are not winning within the game, that is* $\pi(\{1, 3, 5, 7\}) = 54$ *while* $\pi(\{2, 4, 6\}) = 55$.

Definition 4 *The coalition formation game* (X, π) *is* ***balanced*** *if it contains a coalition formation game* (Y, π_Y) *such that*

i. Y *is winning in* (X, π)*, and*

ii. $(Y, \pi_Y + \frac{I}{|Y|} 1_Y)$ *is strongly balanced.*

A balanced game contains a winning coalition such that after adding power (under equal sharing) it will become strongly balanced.

If a coalition formation game (Y, π_Y) is strongly balanced, then it is balanced because the grand coalition Y is strongly balanced even after power has accumulated. The converse is clearly not true.

In general, for any coalition formation game (X, π), there exists a large enough I such that the game (X, π) is balanced, since the power of the agents equalize under equal sharing.

Example 4 (Balanced games) *To illustrate what a balanced game looks like, consider the game*

$$(X, \pi) = (\{1, 2, 3, 4\}, [1, 1.5, 8, 10]).$$

First, let us take the case where $I = 1.2$ and this prize is shared equally within the coalition. It is easy to see that no coalition of size 3 generates a strongly balanced game. For instance, suppose that $\{2, 3, 4\}$ forms. After winning and splitting the prize, the power profile at round 1 is now $\pi_{\{2,3,4\}} = [1.9, 8.4, 10.4]$. Note that the agents with the two lowest powers, agent 2 and agent 3, have a combined power of 10.3 which is less than agent 4's power. Thus, with our assumptions $\{2, 3, 4\}$ does not generate a strongly balanced game and this is true for any other coalition of size 3. Therefore, Part ii of the definition of a balanced game is not satisfied.

However if we increase the prize sufficiently, we can show that (X, π) is a balanced game. Consider the case where we double the prize to $I = 2.4$ and suppose that coalition $\{2, 3, 4\}$ forms. After splitting the prize equally, the new power profile for this coalition at round 1 is $\pi_{\{2,3,4\}} = [2.3, 8.8, 10.8]$. Agents 2 and 3—the two agents with the lowest powers—has a total power of 11.1 which is higher than agent 4's power. Thus, with the new prize, coalition $\{2, 3, 4\}$ generates a strongly balanced game. Therefore, the game (X, π) is balanced since it satisfies Part ii of the definition of a balanced game.

The next result finds the largest class of games where a transition correspondence that meets ISE, DIR and SI exists. It also finds the unique transition correspondence that meets these axioms.

Proposition 2 *Consider a feasible domain of generic*[20] *games $\tilde{G}$ and a transition correspondence $\tilde{\phi} : \tilde{G} \to 2^N$. Under equal sharing, if the transition correspondence $\tilde{\phi}$ satisfies ISE, DIR and SI, then:*

i. $\tilde{G}$ should only contain balanced games.

[20]This is defined by AES. We say the game (X, π) is generic if the power profile π has no ties in the power of any two coalitions; that is, $\pi(S) \neq \pi(T)$ for any $S, T \subset X$. Note that the class of non-generic games has a Lebesgue measure equal to zero, so this is a weak condition.

ii. $\tilde{\phi}(S,\pi) = \underset{M \in \tilde{Q}(S,\pi)}{\arg\min} |M|$

where $\tilde{Q}(S,\pi) = \{T|\ T \in W_{(S,\pi)} \text{ and } (T, \pi_T + \frac{I}{|T|} 1_T) \text{ is strongly balanced}\}$.

The proof of this result is in Appendix.

Note that the transition correspondence $\tilde{\phi}$ picks a coalition of the smallest size $2^k - 1$ for some $k \in \mathbb{N}$ that is strongly balanced. The intuition behind the proof of part ii is that as power accumulates the relative power of the winning agents equalizes. We show inductively that among vectors where power is relatively equalized, the only ISE coalitions are of size $2^k - 1$ for some $k \in \mathbb{N}$. Therefore, an ISE coalition must be of size $2^k - 1$ for some $k \in \mathbb{N}$, otherwise after enough rounds of power accumulation a winning subcoalition of size $2^k - 1$ for some $k \in \mathbb{N}$ will be able to deviate and be ISE.

2.4 Agents survive

2.4.1 Axioms

Let us begin this section with two examples to motivate why the two transition correspondences ϕ^* and $\tilde{\phi}$ introduced in the previous section do not work when agents survive.

Example 5 (Equal sharing and $\tilde{\phi}$) *Consider the game*

$$(S,\pi) = (\{1,2,3,4,5\}, [23,21,20,19,17])$$

with prize $I = 1$. We look at the case of equal sharing and transition correspondence $\tilde{\phi}$. To contrast with the previous section, first note that this game is balanced. Suppose we proceed with the transition correspondence $\tilde{\phi}$. Thus, the coalition $\{3,4,5\}$ is an element of $\tilde{\phi}$. However, at the 67^{th} round the power of agents 3 and 4 who originally had powers 20 and 19, respectively, now have powers 41.33 and 42.33. Thus, the combined powers of agents 3 and 4 is higher in that round than the rest of society composed of Agents 1,2 and 5 (since agents 1 and 2 still survive), that is, $\pi^{67}_{\{3,4\}} = 83.66 > \pi^{67}_{\{1,2,5\}} = 83.33$. Thus, $\{3,4\}$ can deviate from $\{3,4,5\}$ and be self-enforcing since neither 3 nor 4 have sufficient power to win, even if power has accumulated. Hence, the original coalition $\{3,4,5\}$ is not self-enforcing.

The coalition $\{2,3,4,5\}$, however, will be self-enforcing since no member of this coalition will want to deviate in any round. To see this, observe that any coalition of size 3 will not deviate from $\{2,3,4,5\}$ since in some future round a coalition of size 2 will have enough relative power to *deviate from the 3-agent coalition and be self-enforcing. To illustrate, suppose that* $\{3,4,5\}$ *deviates from* $\{2,3,4,5\}$ *at round 2. After sharing the prize equally among the three of them starting from round 2, at round 68 agents 3 and 4 can deviate since* $\pi^{68}_{\{3,4\}} = 84.17 > \pi^{68}_{\{1,2,5\}} = 83.83$ *and this new 2 person coalition will be self-enforcing (since neither 3 nor 4 can have higher individual power than the rest of society). Moreover, any size 2 coalition cannot deviate from* $\{2,3,4,5\}$ *since in no succeeding round will 2 agents have enough combined power to dominate the rest of the society.*

As we will show below, a self-enforcing coalition under equal-sharing must have size 2^k *instead of size* $2^k - 1$ *for some* $k \in \mathbb{N}$*.*

Example 6 (Proportional sharing and ϕ^***)** *For the same game as in example 5 above, we look at the case of proportional sharing and the transition correspondence* ϕ^**.*

With ϕ^**, coalition* $\{3,4,5\}$ *forms. At the third round after power accumulates, agent 3 (*$\pi^3_3 = 21.07$*) has now a higher power than agent 2 (*$\pi^3_2 = 21$*). Therefore agents 4 and 5 will get a higher proportion of the prize if they align with agent 2, and therefore* $\{3,4,5\}$ *will deviate to* $\{2,4,5\}$ *at the third round. The original coalition generated by* ϕ^* *is therefore not self-enforcing. This phenomenon of jumping from one coalition to another always occur for proportional sharing and* ϕ^* *when power accumulates. In order to avoid it, we introduce below a class of priority transition correspondences.*

We introduce the analogous version of ISE for the case where agents survive throughout rounds. Similar to ISE, once a coalition is chosen, it must continue chosen even after power accumulates and agents survive.

Axiom 4 (External Self-Enforcement (ESE)) *A transition correspondence* ϕ *is* **externally self-enforcing** *(ESE) if for any game* (N,π) *and* $S \in \phi(N,\pi)$*, we have that* $S \in \phi(N, \pi + I\xi^N(S,\pi))$*.*

When there is no confusion, we say that a coalition S is an *ESE coalition* if it is generated by the transition correspondence satisfying ESE. This modification shows that coalitions that are externally self-enforcing should map into the same coalition even though agents from $N \setminus S$ (agents outside S) can still form coalitions and threaten S.

We also introduce an analogous version of rationality in this section where agents survive. Dynamic external rationality (DER) requires that the transition correspondence will pick among all ESE coalitions the coalition that gives the highest payoff to the agents at the limit.[21]

Axiom 5 (Dynamic External Rationality (DER)) *A transition correspondence ϕ meets* ***dynamic external rationality*** *(DER) if for any $T \in \phi(N,\pi)$ and for any $Z \subset N$ such that $Z \in W_{(N,\pi)}$ and $Z \in \phi(N, \pi + I\xi^N(Z,\pi))$, we have that $Z \notin \phi(N,\pi) \Leftrightarrow \xi_i(N,\pi_T^\infty) > \xi_i(N,\pi_Z^\infty)\ \forall i \in T \cap Z$.*

There are many transition correspondences that satisfy ESE, DER and SI. The trivial transition correspondence $\phi(N,\pi) = N$ for any π is one of them. In order to avoid these trivial correspondences, we also introduce the axiom of *independence of zeros*, which says that the agents without power should not affect the winning coalition.

Axiom 6 *A transition correspondence ϕ is* ***independent of zeros (IZ)*** *whenever $\phi(N,[\pi_S, 0_{N\setminus S}]) = \phi(S,\pi_S)$.*

In particular, notice that IZ implies that if an agent has no power, he will not be chosen by the correspondence. This axiom is a consistency axiom related to the group of agents chosen. It is related to the consistency axiom in other settings such as resource allocation problems (Thomson [97, 98]). Section 2.4.2 introduces transition correspondences that do not satisfy IZ.

2.4.2 Results

We find that under equal sharing, there is always a transition correspondence that satisfies the axioms above. However, this is not true under proportional sharing.

To formalize the results, define the transition correspondence ϕ^{**} as follows:

$$\phi^{**}(N,\pi) = \arg\min_{M \in Q \cup \{N\}} |M| \tag{2.2}$$

where $Q = \{S \in 2^N \text{ such that } S \in W_{(N,\pi)} \text{ and } |S| = 2^m \text{ for some } m \in \mathbb{N}\}$.

Proposition 3 *i. Under equal sharing, the correspondence ϕ^{**} is the only transition correspondence that satisfies ESE, DER, SI and IZ.*

ii. Under proportional sharing, there is no transition correspondence that satisfies ESE, DER, SI, and IZ.

[21] Note this limit exists by ESE because the coalition that is chosen at time 0 is also chosen at any time in the future.

The proof of this result is in the Appendix. The intuition behind the proof of part i is that as power accumulates the relative power of the winning agents equalizes and becomes substantially larger than the losing agents. We show inductively that among vectors where power is relatively equalized, the only ESE coalitions are of size 2^k for some $k \in \mathbb{N}$. Therefore, for a general power vector, an ESE coalition must be of size 2^k for some $k \in \mathbb{N}$, otherwise after enough rounds of power accumulation a winning subcoalition of size 2^k for some $k \in \mathbb{N}$ will be able to deviate and be ESE.

The proof of part ii shows that for games with three or more agents the axioms are incompatible. The key argument relies on the fact that for coalition formation games with a dictator (an agent who has enough power to win by himself) the dictator must be chosen. Therefore, ESE implies that in games for three agents without a dictator a coalition of size 2 or less cannot be chosen, otherwise as power accumulates a game with a dictator will form. Finally, choosing the grand coalition in games for three agents without a dictator will contradict IZ.

In the context of the game in Example 5 above, ϕ^{**} will pick any coalition of size 4 since those are the smallest 2^k-sized coalition that are winning within the game.

Proportional sharing without IZ

In this subsection, we study transition correspondences that satisfy ESE, DER and SI but do not necessarily satisfy IZ.

Definition 5 *A **feasible sequence of coalitions** is a finite sequence of coalitions from N, denoted $\{S^0, S^1, \ldots, S^k\}$, such that*

i. $S^k = N$

ii. if $S^i \cap S^j \neq \emptyset$ *and* $i < j$ *then* $S^i \subsetneq S^j$

Denote the set of all feasible sequences of coalitions as $\mathcal{F}$.

Definition 6 *Given a feasible sequence of coalitions $\{S^0, S^1, \ldots, S^k\} \in \mathcal{F}$, we define a **sequential transition correspondence** as:*

$$\bar{\phi}(N, \pi) = \arg \min_{M \in Q \cup \{N\}} \pi(M)$$

where $Q(N, \pi) = \{S^i \mid S^i \in W_{(N,\pi)}, \text{ such that for all } S^j \subset S^i,\ S^j \notin W_{(S^i, \pi_{S^i})}\}$.

The sequential transition correspondence will pick the coalition with the least power that is winning and such that it does not contain a subset that is winning within that coalition. If that set is empty, then the grand coalition is picked.

Example 7 *Suppose there are seven agents* $\{1, 2, 3, 4, 5, 6, 7\}$ *a feasible sequence of coalitions given by*

$$\{4\}, \{3, 4, 5\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 5, 6, 7\}.$$

Consider the power profile $\pi = [12, 12.5, 13, 13.5, 14, 15, 16]$. *The winning coalition with the least power that appears in the sequence is coalition* $\{1, 2, 3, 4, 5\}$. *However, the sequential transition correspondence will not pick this coalition, since coalition* $\{3, 4, 5\}$ *is winning within* $\{1, 2, 3, 4, 5\}$. *The sequential transition correspondence then picks the grand coalition in this case.*

Now consider the power profile $\pi' = [10.5, 12, 14, 18, 19, 13, 13.5]$. *The winning coalition with the least power is* $\{3, 4, 5\}$ *and this will be picked by the sequential transition correspondence since* $\{4\}$ *is not winning within* $\{3, 4, 5\}$.

Proposition 4 *Under proportional sharing, any sequential transition correspondence satisfies ESE, DER and SI.*

The intuition behind the proof of this proposition is as follows. As power accumulates, the relative power of the winning agents remains the same while the relative power of the non-winning agents tends to zero. However, by construction of the correspondence, no winning subcoalition from the winning coalition is feasible. Hence, the winning coalition is chosen throughout time even after power accumulates.

2.5 Conclusion

This paper develops an axiomatic approach to a coalition formation model by focusing on two main axioms: self-enforcement and rationality. The variations in the model that takes into account whether non-winning agents can participate in future coalition formation processes, and different ways of sharing the prize, provide a rich characterization of possible transition correspondences that satisfies these axioms.

This is summarized in the table below.

In the case where agents are killed and the resource is distributed equally among agents in the winning coalition, a self-enforcing and rational transition correspondence exists only if we can

Scenario	Equal Sharing	Proportional Sharing
Agents Killed	$\tilde{\phi}$ for balanced games (See Proposition 2(ii))	ϕ^* (See Equation 3.1)
Agents Survive	ϕ^{**} (See Equation 2.2)	$\bar{\phi}$ for a feasible sequence of coalitions (See Definition 6)

find a subset of the grand coalition of size $2^k - 1$ that is strongly balanced after power has been added (Section 3.2, Proposition 2). Under the same case but with proportional sharing of the resource, a self-enforcing and rational transition correspondence always exists (Section 3.2, Proposition 1). When agents survive, under proportional sharing we have to restrict to feasible sequences of partitions in order to find a self-enforcing and rational transition correspondence (Section 4.2.1, Proposition 4). Under equal sharing we only have to pick the smallest coalition of size 2^k that is winning (Section 4.2, Proposition 3).

We believe that the attractiveness of our model's diverse results derives from highlighting the observation that the different institutions, or "rules of the game" (e.g., power, sharing rules, feasible coalitions) surrounding the coalition formation process play a large role in determining the types of coalitions that will be self-enforcing through time.

2.5.1 Extensions and open questions

The results in this paper can be easily extended to the case of supermajority, that is, when the set of winning coalitions is defined as $W_{(T,\pi)} = \{S \subset T | \pi(S) > \alpha\pi(T \setminus S)\}$ for $\alpha \in [0.5, 1)$. For instance, consider the case of ϕ^*. Under supermajority, ϕ^* picks the coalition with the highest share among the winning and self-enforcing coalitions.

The generalization of $\bar{\phi}$ is also straightforward. It picks the winning coalition with the least power in a feasible sequence without any subcoalition that is winning. For instance, in Example 7 with power profile π', if $\alpha = .70$ then the coalition $\{1, 2, 3, 4, 5\}$ will be the winning coalition with the least power in the feasible sequence. This will be picked by the transition correspondence $\bar{\phi}$ since the coalition $\{3, 4, 5\}$ is not winning within $\{1, 2, 3, 4, 5\}$ because it has less than 70% of the power.

The generalization of ϕ^{**} is more involved but still straightforward. The first step is to create for a given $\alpha \in [0.5, 1)$ a sequence of admissible sizes A^α that follows the next inductive process. The first element of the sequence is $A^{\alpha}_1 = 1$ since a coalition of size 1 is (externally) self-enforcing. For any $k > 1$, A^{α}_k is the smallest integer such that $A^{\alpha}_k > A^{\alpha}_{k-1}$ and $\frac{A^{\alpha}_{k-1}}{A^{\alpha}_k} \leq \alpha$.

For instance, for $\alpha = .5$ the set of admissible sizes is $\{1, 2, 4, 8, 16, 32, \ldots\}$. For $\alpha = .75$ the set of admissible sizes is $\{1, 2, 3, 4, 6, 8, \ldots\}$. For $\alpha = .6$ the set of admissible sizes is $\{1, 2, 4, 7, 12, 20, \ldots\}$.

The mapping ϕ^{**} is therefore modified as:

$$\phi^{**}(N, \pi) = \arg \min_{M \in Q \cup \{N\}} |M|$$

where $Q = \{S \in 2^N \text{ such that } S \in W_{(N,\pi)} \text{ and } |S| \in A^{\alpha}\}$.

There are multiple open questions for future research. A first extension of our study is the implementation of the stable coalitions found by this paper as the equilibrium of a non-cooperative game (see Acemoglu et al. [1]). Second, there is a need to develop a new theory describing how the different sharing rules emerge endogenously from the coalition formation process. Third, the case where the decision to kill agents is endogenous (as well as the possibility of growth in the number of agents in the society) should also be examined. Fourth, the extension to more general functions of power accumulation should be studied. Finally, another extension to study is the role of externalities (such as cultural characteristics or religion) in the coalition formation process (see Juarez [51, 52] for some advances).

3 THE EFFECT OF SHARING RULES ON THE FORMATION OF COALITIONS

3.1 Introduction

Searching for appropriate rules on how to divide an economic surplus (or cost) that will induce full cooperation among agents has long been the main focus of cooperative game theory (Moulin [66]). The structure of incentives provided by these sharing rules determines the extent to which agents can be induced to join a cooperative venture (Moulin [65]). The applications are wide-ranging, for instance, appropriating gains and costs from managing global public goods (Carraro [20]), cost and revenue sharing in irrigation networks (Jandoc, Juarez and Roumasset [49]), and market design (Roth [87]), to name a few.

It is less clear, however, how different sharing rules affect the formation of coalitions in a tournament setting. More recent coalition formation models where agents are endowed with power (e.g. Acemoglu, et al [1,2]), Jandoc and Juarez [47]) have focused primarily on a class of sharing rules that enables agents common between competing coalitions to unanimously prefer one coalition over the other. We call this condition *consistent ranking*[1] and two popular sharing rules, equal sharing and proportional sharing, satisfy this property. *Equal sharing* (or egalitarian sharing) is where the surplus is divided equally among the number of agents inside the coalition.[2]. *Proportional sharing* where the surplus is divided according to an agent's relative power inside the coalition. These sharing rules satisfies consistent ranking because the agents' share is always higher in smaller-sized coalitions (under equal sharing) and in lower-powered coalitions (under proportional sharing).

It is possible for sharing rules to create disagreements among agents' preferred coalitions. In this paper we use a particular sharing rule called *combination sharing*, that is, the convex combination of equal sharing and proportional sharing. This sharing rule creates a tension between coalition size and power, and some of the agents will prefer the coalition with smaller size while others prefer to be in the coalition with smaller power. This sharing rule has its roots in the *Sen share*

[0]Overall guidance by Ruben Juarez is deeply appreciated.

[1]Under consistent ranking, agents have the same ordinal ranking over coalitions in which they belong

[2]This is widely used not only on theoretical grounds but also for practical purposes such as inheritance bequest (Erixson and Ohlsson [30])

(Sen [94]) popular in the literature of team production and moral hazard. The Sen share is a convex combination of an equal share of the surplus among productive agents and a proportional share according to an agent's labor to the total labor supplied in the economy.

In this paper we examine how different sharing rules affect the manner in which coalitions form. We do this by presenting a dynamic coalition formation model where farsighted agents endowed with power compete for a divisible resource by forming coalitions with other agents.[3] If a coalition wins, its members will divide the resource among themselves according to a predetermined sharing rule. We focus on sharing rules that satisfy the consistent ranking property (equal and proportional sharing) and a sharing rule that does not satisfy it (combination sharing). We also examine the case of an institutional constraint that specifies the span of life of the agents. In particular, we focus on two polar cases where agents who are not part of the winning coalition do not participate in the future (*agents are killed*) and where all the agents remain in the game even though they are not part of the winning coalition (*agents survive*).[4]

We employ an axiomatic approach to find the stable coalitions in the different cases relating to the sharing rules and whether agents are killed or are able to survive. Our focus is on two main axioms, *self-enforcement* and *rationality*. Since our model is dynamic, we allow for the possibility of coalitions disintegrating and new factions forming to overthrown the existing ruling coalition. Since agents are forward-looking, they prefer to form coalitions that are stable from the moment of inception. Our axiom of self-enforcement requires that no subcoalition of the winning coalition is powerful enough to encourage further deviations. Self-enforcement is a robustness property that ensures that the coalition that forms in round one never disintegrates afterwards. In addition, rationality requires that agents pick the coalition that gives them their highest payoff among self-enforcing coalitions. Rationality is related to the traditional axiom of coalitional stability, where no coalition will have the incentive to deviate.[5]

In the next subsection, we provide an illustrative example of how our coalition formation game works and how sharing of the prize and whether agents are killed or are able to survive shape the coalitions that form.

---

[3] Farsightedness assures that agents care about their payoff from being included in the final coalition that forms rather than their payoffs from being in an intermediate coalition. See for instance, Konishi and Ray [56], Chwe [22], Xue [105], Diamantoudi and Xue [27], Ray and Vohra [84].

[4] In this paper, we assume that this institutional feature is exogenous. While the decision to kill may very well be endogenous, for instance, with the size of the resource (e.g. Bueno de Mesquita, et al [18]), it is beyond the scope of this paper and we leave it for future research.

[5] This is related to immunity to group manipulations in models discussed by Bogomolnaia and Jackson [16], Ehlers [29], Juarez [53], Papai [76].

3.1.1 An illustrative example

Suppose our society is composed of 6 agents $\{1, 2, 3, 4, 5, 6\}$ with power profile $\pi = [29, 26.5, 20, 2.499, 2.4, 2.39]$. The game is played for an infinite number of periods and for each round the prize to be divided by the winning coalition is $I = 3$. A coalition's power is simply the sum of the agents' powers inside it. A winning coalition should have more than 50% of the total power in the society at any given round. Since agents are farsighted, they only care about their payoffs from being in the final coalition. If an agent is part of a coalition S, his share of the prize under equal sharing will be $\frac{1}{|S|}$, under proportional sharing it is his power divided by the power of coalition S (that is, $\frac{\pi_i}{\pi(S)}$), under combination sharing it is $\lambda \frac{1}{|S|} + (1 - \lambda) \frac{\pi_i}{\pi(S)}$ for some $\lambda \in (0, 1)$.

Case 1: Non-winning agents are killed, sharing rule satisfies consistent ranking.

First, no coalition of size 1 is winning and therefore is incapable of forming. Although there are winning coalitions of size 2, we argue that this will not be self-enforcing. To see this, suppose coalition $\{1, 2\}$ forms. This is winning since their combined power (55.5) is higher than the rest of society composed of agents 3,4,5 and 6 (with combined power of 27.289). After forming, the non-winning agents 3,4,5, and 6 are killed and agents 1 and 2 continue on the next round. Left alone to themselves, agent 1 can now kill agent 2 since he has the higher power. Since agent 2 is farsighted, he will never agree to form $\{1, 2\}$. This will be true for any coalition of size 2.

Furthermore, note that there are several coalitions of size 3 that are winning but only $\{1, 2, 3\}$ is self-enforcing. This is because if $\{1, 2, 3\}$ forms and agents 4,5, and 6 are killed, then there can be no dictator among them and a deviation to a 2-person coalition is not feasible following the argument outlined above. On the other hand, take the case of the winning coalition $\{2, 3, 5\}$. This will not be self enforcing since when agents 1,4 and 6 are killed, agent 2 can be a dictator and can deviate from $\{2, 3, 5\}$.

Following the same arguments, it is straightforward to show that the coalitions $\{1, 2, 4, 5\}$, $\{1, 2, 5, 6\}$, $\{1, 2, 4, 6\}$ and $\{2, 3, 4, 5, 6\}$ are self-enforcing since they don't contain subcoalitions that are self-enforcing. In the same manner, the grand coalition $\{1, 2, 3, 4, 5, 6\}$ is not self-enforcing since it can deviate to either one of the self-enforcing coalitions mentioned.

With proportional sharing, the coalition $\{2, 3, 4, 5, 6\}$ will be preferred by this coalition's agents over any other self-enforcing coalition to which they could possibly belong. For instance, given the choice between $\{2, 3, 4, 5, 6\}$ and $\{1, 2, 3\}$ agents 2 and 3 (the agents common in these coalitions) will prefer the former since their share of the prize ($\frac{26.5}{53.789}$ and $\frac{20}{53.789}$ for agents 2 and 3, respectively) is higher than the latter coalition ($\frac{26.5}{75.5}$ and $\frac{20}{75.5}$ for agents 2 and 3, respectively). Hence, under proportional sharing, rationality implies that the coalition $\{2, 3, 4, 5, 6\}$ should form.

On the other hand, under equal sharing the coalition $\{1, 2, 3\}$ should form because the agents in this coalition will get a higher share $\frac{1}{3}$ than in any other self-enforcing coalition.

Case 2: Non-winning agents are killed, combination sharing.

We show in Example 8 later that when the combination sharing parameter $\lambda = 0.45$, there will be disagreement among agents on the preferred coalition. For instance, comparing coalitions $\{1, 2, 3\}$ and $\{2, 3, 4, 5, 6\}$, the share of agent 2 is higher in $\{2, 3, 4, 5, 6\}$ and the share of agent 3 is higher in $\{1, 2, 3\}$. Unlike in the previous case of equal or proportional sharing where these agents agree on the preferred coalition, in the case of combination sharing there may be no such agreement.

Case 3: Non-winning agents survive, consistent ranking.

When non-winning agents survive, the coalitions that are *minimally winning* are stable. Minimally winning coalitions are winning coalitions such that removing any one of the agents will make this coalition non-winning. In our particular example, the coalition $\{2, 3\}$ is minimally winning since removing either agent 2 or 3 will make this coalition non-winning. Another minimally winning coalition is $\{1, 2\}$. On the other hand, the coalition $\{2, 3, 4, 5, 6\}$ is winning but not minimally winning, since we can remove agents 4,5 and 6 and the coalition $\{2, 3\}$ will still be winning.

With proportional sharing, the coalition $\{2, 3\}$ will be preferred by the agents in this coalition among all possible minimally winning coalitions to which they can belong. Under equal sharing, all minimally winning coalitions of size 2 ($\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$) can form.

Case 4: Non-winning agents survive, combination sharing.

In this case, since all the minimally winning coalitions are of size 2, there will be no conflict between coalition size and power and the minimally winning coalition with the least power ($\{2, 3\}$) will be preferred by the agents in this coalition for any value of λ.

3.1.2 Overview of the results

Section 3.3 of the paper focuses on the case where agents are killed. Proposition 5 characterizes the unique transition correspondence[6] that is rational and self-enforcing under consistent ranking sharing rules. The proof of this result greatly simplifies the proof provided by Acemoglu, et al [1] in the case of proportional sharing. Moreover, the proof is based on the key Lemma 1, which highlights the importance of transition correspondences that do not necessarily satisfy rationality but meet a weaker requirement (*minimalistic,* introduced in Appendix A).[7] Under consistent ranking, the transition correspondence chooses the highest ranked coalition among self-enforcing and winning coalitions.

Under combination sharing, Proposition 6 and and Corollary 1 characterizes the unique transition correspondence that is self-enforcing and rational for a restricted class of coalition formation games. The restriction is either on the admissible power vectors (where we restrict to the condition of size-power monotonicity where larger coalitions have higher power) or a restriction on the values of λ either to be high enough to choose the smallest-sized self-enforcing coalition, low enough to choose the least-powered self-enforcing coalition, or just right for a "compromise" coalition to exist.

Section 3.4 of the paper studies the case where agents survive. Proposition 7 describes the unique transition correspondence that satisfies rationality under sharing rules that satisfy combination sharing. The transition correspondence in this case chooses the highest-ranked coalition among the minimally winning coalitions.

On the other hand, Proposition 8 and Corollary 3 provide the largest class of coalition formation games where a rational transition correspondence exists for combination sharing. Again, we need to make some restrictions on admissible games for these transition correspondence to satisfy rationality.

3.1.3 Related literature

The manner in which coalition members divide the prize has also received attention not only in the coalition formation literature but also in related fields such as team production as well as in political economy. Acemoglu, et al [1, 2] provides a coalition formation model where agents split

[6] A transition correspondence is a mapping that defines which coalitions form over time. The precise definition is given in Section 2.1.

[7] In Appendix B, Proposition 10 uncovers the entire class of sharing rules where self-enforcement and rationality are compatible. Proposition 10 also characterizes the unique transition correspondence that meets these two axioms and generalizes the results provided in Proposition 5.

the prize in proportion to an agent's relative power in the coalition. They show that there always exists a rule of choosing coalitions that satisfy self-enforcement and rationality and this rule can be supported as a subgame perfect Nash equilibrium. In a recent paper, Jandoc and Juarez [47] provides the conditions where rules satisfying self-enforcement and rationality extends to the case where winning agents accumulate power over time. In this paper, it is sometimes necessary to restrict the domain of games (for instance, for powers in the society to be relatively balanced) for these rules to exist.

Outside the coalition formation literature, the effect of sharing rules on incentivizing team production has been well investigated. For instance, a popular result due to Holmstrom [44] is that it is generally impossible to design a sharing rule that meets efficiency and budget-balance simultaneously. Subsequent extensions by Fabella [32, 33], shows that there exists a sharing arrangement that can support Pareto optimal outputs under certain conditions on the production function. Under constant returns to scale, proportional sharing of the surplus (based on effort) can achieve Pareto efficiency while under increasing returns to scale a generalized sharing mechanism where equal sharing and proportional sharing are special case can support Pareto optimal production. In the same manner, Sen [94] provides the conditions under which a convex combination of equal sharing and proportional sharing of the economic surplus leads to Pareto efficient production.

It has also been shown that certain sharing rules exacerbate the moral hazard problems in joint production while others enhance efficiency. For instance, Newhouse [70] finds that medical groups that practice equal sharing arrangements experience a high level of effort shirking and this level increases in partnership size. On the other hand, Reinhardt, Pauly, and Held [85] finds that some form of sharing based arrangements based on a proportion of some measure of effort enhanced productivity in medical partnerships.

While we assume a fixed and exogenous sharing rule, a natural extension (but beyond the scope of this paper) would be to endogenize the sharing rule. Some strategies employed in the context of rent-seeking tournaments include Nitzan [71], Lee [58, 59] and Baik [5] where they show the long-run group sharing rule can fall anywhere from being undefined to extremely equal depending on the context of the mechanism to determine these rules (Nitzan [72]).

The most related study to our paper is Acemoglu et al [1] wherein forward-looking agents are endowed with power and form coalitions with the goal of becoming the *ultimate ruling coalition.* In their paper, the winning coalition will split a given resource in proportion to its power and agents outside the winning coalition are killed. However, we extend AES on several fronts: First, we allow the possibility that non-winning agents survive throughout the whole coalition formation game.[8] Second, we allow for more robust sharing rules.

3.2 The model

Consider the set $N = \{1, \ldots, n\}$ of initial agents who are endowed with powers $\pi = [\pi_1, \ldots, \pi_n]$, respectively. A coalition S is a subset of N, that is, $S \subseteq N$. The set of coalitions are all possible subsets of N, denoted by 2^N. A coalition formation game is a pair (S, π) where $S \subseteq N$ and $\pi \in \mathbb{R}^S_+$. The set of coalition formation games is denoted by $\mathbf{G}$. We assume that power is additive, that is, the power of coalition S is the sum of all powers of the agents inside the coalition, $\pi(S) = \sum_{i \in S} \pi_i$.[9] We denote by π_S the restriction of the vector $\pi \in \mathbb{R}^N_+$ over coalition S.

Definition 7 *Given a game (T, π), the **set of winning coalitions**[10] is:*

$$W_{(T,\pi)} = \{S \subset T | \pi(S) > \pi(T \setminus S)\}$$

We assume a sharing rule that divides the prize of the agents. This sharing rule is fixed throughout the game.

Definition 8 *A **sharing rule** is a function $\xi : \mathbf{G} \to \mathbb{R}^N_+$ such that:*

i. If $k \notin S$, then $\xi_k(S, \pi) = 0$.

ii. $\sum_{i \in S} \xi_i(S, \pi) = 1$, and

iii. (Cross-Monotonicity) If $(S, \pi) \in \mathbf{G}$, $T \subset S$, $i \in T$ and $\pi_i > 0$, then $\xi_i(T, \pi_T) > \xi_i(S, \pi)$.

Cross-monotonicity of the sharing rule requires that the share of the prize of agent i in coalition S would be higher if he is part of any subcoalition of S that deviates compared to the share he will get if he stayed in coalition S.[11]

[8] In another paper, Acemoglu, et al [2], show that the results in the AES case extend even to situations where agents are not killed. The stable ruling coalition in this case is the "minimally winning coalition" which is the coalition with the smallest power among all winning coalitions. In their dynamic game, this corresponds to the case when all agents are eligible to vote on the proposed coalition (i.e., the democratic case). In the case where the voting power is limited to the ruling coalition but agents are not killed, the mapping in AES is modified to include all alternative coalitions not just the subset of the ruling coalition.

[9] Juarez [52] considers a more general version where power can be any arbitrary monotonic function.

[10] This definition requires winning coalitions to have relative power larger than 50%. Our results below can be easily adapted to require winning coalitions to have relative power larger than α, where $\alpha \geq 50\%$.

[11] This is contrary to the case where there are externalities, where agents might gain by associating with other agents of similar characteristics; see Juarez [52].

Throughout the paper we devote special attention to simple (and commonly used) sharing rules such as equal sharing and proportional sharing, or a convex combination of the two. That is, if $i \in S$, then the share of agent i when S is winning and the power profile is π equals:

$$\xi_i(S,\pi) = \begin{cases} \dfrac{1}{|S|} & \text{if equal sharing (ES)} \\ \dfrac{\pi_i}{\pi(S)} & \text{if proportional sharing (PR)} \\ \lambda \cdot \dfrac{1}{|S|} + (1-\lambda) \cdot \dfrac{\pi_i}{\pi(S)},\ \lambda \in (0,1) & \text{if combination sharing (CS)} \end{cases}$$

Note that these three basic sharing rules are cross-monotonic.

Suppose agents i and j belong to the intersection of coalitions S and T. A sharing rule satisfies consistent ranking if whenever agent i prefers S over T, then agent j also prefers S over T. In other words, between competing coalitions, a coalition S is picked if all agents in the intersection unanimously pick S over a competing coalition.

Definition 9 *(Consistent Ranking (CR)) The sharing rule ξ satisfies **consistent ranking (CR)** if for any two agents i and j, and coalitions S and T such that $i, j \in S \cap T$, if $\xi_i(S,\pi) > \xi_i(T,\pi)$, then $\xi_j(S,\pi) > \xi_j(T,\pi)$.*

If the sharing rule ξ satisfies consistent ranking, then there exists a ranking $R^{\xi} : \mathbf{G} \to \mathbb{R}$ for the society that coincides with individual rankings. That is, for any coalitions S and T such that $S \cap T \neq \emptyset$, we have that $R^{\xi}(S,\pi) > R^{\xi}(T,\pi) \Leftrightarrow \xi_i(S,\pi) > \xi_i(T,\pi)$ for any $i \in S \cap T$.

Equal sharing and proportional sharing satisfy consistent ranking. Under equal sharing, agents' share increases as they move to coalitions of smaller sizes; therefore, $R^{ES}(S,\pi) = \frac{1}{|S|}$ is an example of a consistent ranking for equal sharing. Similarly, under proportional sharing, agents' share increases as they move to coalitions of smaller power; therefore, $R^{PR}(S,\pi) = \frac{1}{\pi(S)}$ is an example of a consistent ranking for proportional sharing. On the other hand, combination sharing does not satisfy consistent ranking (see, Example 8).

The major task of this paper is to examine the stable coalitions that emerge under the different sharing rules discussed above.

3.2.1 Dynamic Coalition Formation

Let t, where $t = 0, 1, \ldots$, denote the discrete rounds of the game. A *transition correspondence* is a mapping from the set of coalition formation games to the set of winning coalitions.

Definition 10 *A* **transition correspondence** *is a continuous*[12] *correspondence* $\phi : \mathbf{G} \rightarrow 2^N$ *such that* $\forall\ (X, \pi) \in \mathbf{G}$*:* $\phi(X, \pi) \subset W_{(X,\pi)}$.

The transition correspondence ϕ selects all coalitions that could be winning at a given game (S, π). The evolution of the coalition formation games at every round depends on the transition correspondence along with the conditions on whether agents survive or are killed in future rounds.

We denoted by (S^t, π^t) the game at round t. We start with the game (S^0, π^0) and look at the evolution of the game throughout time. We denote this evolution by $(S^0, \pi^0), (S^1, \pi^1), (S^2, \pi^2), \ldots$.

In this paper, we focus at the evolution of the game in two polar cases. First, we consider the case where agents are killed if they are not part of the winning coalition, this is described by the case where $S^t \in \phi(S^{t-1}, \pi^{t-1})$ for all time t. That is, only a group of agents who were winners at time $t-1$ could participate at time t. Second, we consider the case where agents survive regardless of whether they are part of the coalition in the previous round. This is described by $S^t = N$ for all time t. That is, the agents participating at every time do not change.[13]

Finally, we assume that agents are infinitely forward-looking. That is, they care about their payoff at the limit coalition formation game. This payoff, denoted by $\xi(S^\infty, \pi^\infty)$, is defined as the limit of the sequence $\xi(S^0, \pi^0), \xi(S^1, \pi^1), \xi(S^2, \pi^2) \ldots$, when it exists. If this limit does not exist, $\xi(S^\infty, \pi^\infty) = 0$.[14]

For the two sharing rules discussed above this limit exists when agents are killed. To see this, notice that there exists a time $\bar{t}$ such that S^t does not change for all $t \geq \bar{t}$.[15] The convergence under PR happens because the relative power of the agents in (S^t, π^t) will not change for any $t > \bar{t}$. The convergence under ES occurs because the agents in S^t remains constant for any $t > \bar{t}$.

[12]A correspondence is continuous if for any sequence of power vectors $\pi^1, \pi^2, \cdots \rightarrow \pi^*$ where $S \in \phi(N, \pi^i)\ \forall i$ and S is winning in π^*, then $S \in \phi(N, \pi^*)$.

[13]First, observe that in the case where agents are killed and $\phi(S^{t-1}, \pi^{t-1})$ contains more than one element, we impose no restriction in which coalition from $\phi(S^{t-1}, \pi^{t-1})$ will equal S^t. This allows our results to be more robust, since the evolution of the game includes any potential path of coalitions such that $S^t \in \phi(S^{t-1}, \pi^{t-1})$ for all t.

[14]As we will see below, the axiom of self-enforcement guarantees that this limit always exists whether agents are killed or survive.

[15]Indeed, note that $S^0 \supseteq S^1 \supseteq S^2 \supseteq S^3 \supseteq \ldots$, thus the sequence $S^0, S^1, S^2, \ldots$ converges in finite time.

3.3 Agents are killed

3.3.1 Axioms

This section resembles the AES main features where if a coalition S forms, then agents outside S are killed in the sense that they cannot participate in any future coalition formation process. We want to find transition correspondences that are self-enforcing; that is, we are interested in finding transition correspondences that maps to coalitions that do not have the incentive or the power to deviate in future rounds of the game.

Axiom 7 (Self-enforcement (SE)) *The transition correspondence ϕ is **self-enforcing (SE)** if for any game $(X, \pi) \in \mathbf{G}$ and $S \in \phi(X, \pi)$, then $S \in \phi(S, \pi_S)$.*

When there is no confusion, given a transition correspondence ϕ and a game (S, π), we say that the coalition S is self-enforcing if $S \in \phi(S, \pi_S)$.

Self-enforcement requires that given any starting game (X, π), a coalition S is part of the transition correspondence from (X, π) only if there would be no further deviations into subcoalitions of S once S forms.

Since the sharing rule is cross-monotonic, we expect that in the presence of self-enforcing and winning coalitions that are strict subsets of the grand coalition, the grand coalition will not be chosen, since all of the agents gain by choosing its subset. This is reflected in the definition of a minimalistic transition correspondence.

Axiom 8 (Minimalistic (MIN)) *The transition correspondence ϕ is **Minimalistic (MIN)** if for the game $(S, \pi) \in \mathbf{G}$ such that there exists $T \subsetneq S$, where $T \in \phi(T, \pi_T)$ and $T \in W_{(S,\pi)}$, then $S \notin \phi(S, \pi)$.*

Next, we focus on comparing different transition correspondences based on the coalitions that they choose.

Axiom 9 (Superiority) *Consider two transition correspondences ϕ and $\hat{\phi}$. We say that ϕ is **superior** to $\hat{\phi}$ if for any game (X, π), $T \in \hat{\phi}(X, \pi)$ and $S \in \phi(X, \pi)$ such that $\xi_i(T, \pi_T) \geq \xi_i(S, \pi_S)$ for some $i \in T \cap S$ if and only if $T \in \phi(X, \pi)$.*

If a transition correspondence is superior to another, then it always picks outcomes that are preferred by common agents being chosen.

Axiom 10 (Rationality (RAT)) *The transition correspondence ϕ is **rational (RAT)** if for any $S \in 2^N$, for any $T \in \phi(S, \pi)$ and for any $Z \subset S$ such that $Z \in W_{(S,\pi)}$ and $Z \in \phi(Z, \pi_Z)$, we have that $Z \notin \phi(S, \pi) \Leftrightarrow \xi_i(T, \pi_T) > \xi_i(Z, \pi_Z)\ \forall i \in T \cap Z$.*

Rationality implies that agents prefer to form coalitions that give them a larger share of the resource. This is similar to other notions of coalitional stability previously discussed in the literature, where a coalition is chosen if it cannot be blocked by another coalition that is winning and self-enforcing.

Note that the cross-monotonicity of the sharing rule implies that if a rule satisfies RAT, then it also satisfies MIN.

AES discuss rationality for sharing rules that are proportional-like, where agents prefer to be in coalitions that give them a higher relative power. RAT extends the analysis of AES by considering sharing rules that satisfy or do not satisfy the consistent ranking property and by modifying the Rationality Axiom accordingly.

3.3.2 Result with Consistent Ranking

Let the transition correspondence ϕ^* be defined as:

$$\phi^*(S, \pi) = \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\max} R^{\xi}(M, \pi_M) \tag{3.1}$$

where $Q(S, \pi) = \{T \subsetneq S \mid T \in W_{(S,\pi)}, T \in \phi^*(T, \pi_T)\}$

This transition correspondence defines for the game (S, π) a set $Q(S, \pi)$ of proper subcoalitions, which are both winning in S and self-enforcing. It picks the coalition that yields the highest rank for the agents in the intersection of all the coalitions contained in $Q(S, \pi)$. If $Q(S, \pi)$ is empty then it picks coalition S itself. Thus, $\phi^*(S, \pi_S) \neq \emptyset$ and ϕ^* is well defined.

Proposition 5 *Consider a sharing rule that satisfies consistent ranking. Then, the following conditions are equivalent for the transition correspondence ϕ that is self-enforcing:*

i. *ϕ is superior to any other transition correspondence that is self-enforcing and minimalistic,*

ii. *ϕ is rational,*

iii. *$\phi = \phi^*$.*

AES's main result has a similar characterization to parts ii and iii under proportional sharing. This proposition shows that AES's result extends to a larger class of sharing rules that satisfy consistent ranking.

The proof of this result is provided in Appendix C. It greatly simplies the proof provided by AES, mainly by using a key observation that any two self-enforcing and minimalistic transition correspondences have the same sets of self-enforcing coalitions (see Lemma 1 in Appendix B).

In Appendix B we provide the necessary and sufficient conditions on the sharing rule that allows for the existence of a self-enforcing and rational transition correspondence. In particular, we show that the assumption of consistent ranking in Proposition 5 can be extended to an even larger class of sharing rules that satisfy a property called *generalized consistency*.

3.3.3 Results under Combination Sharing

An example of a sharing rule that does not satisfy consistent ranking is combination sharing.[16] This is illustrated in the next example:

Example 8 *Consider the game*

$$(N, \pi) = (\{1, 2, 3, 4, 5, 6\}, [29, 26.5, 20, 2.499, 2.4, 2.39]).$$

The self enforcing coalitions that are contained in (N, π) *are* $\{1, 2, 3\}$*,* $\{1, 2, 5, 6\}$*,* $\{1, 2, 4, 5\}$*,* $\{1, 2, 4, 6\}$*, and* $\{2, 3, 4, 5, 6\}$

To see this, note that the power of agents 3, 4, 5 and 6 together is less than the power of agent 1. Therefore, any self-enforcing coalition that contains agent 1 should also contain agent 2. In order for 1 not to be a dictator in a game that contains agents 1 and 2, the game must contain agents whose combined powers should exceed 2.5. Therefore, any self-enforcing coalition that contains agents 1 and 2 should include either agent 3 alone or any two agents from agents 4, 5 and 6.

On the other hand, note that if a self-enforcing coalition does not contain agent 1, then it is a subset of the coalition $\{2, 3, 4, 5, 6\}$*. The coalition* $\{2, 3, 4, 5, 6\}$ *is self-enforcing because any coalition of size 2, 3 or 4 has a dictator. Therefore,* $\{2, 3, 4, 5, 6\}$ *is the only self-enforcing coalition that does not contain agent 1.*

Let ξ *be the combination sharing rule with* $\lambda = .45$*. Notice that comparing coalitions* $\{2, 3, 4, 5, 6\}$ *and* $\{1, 2, 3\}$ *we have that*

[16] Combination sharing does not satisfy generalized consistency as defined in Appendix B.

$$\xi_2(\{2,3,4,5,6\},[26.5,20,2.499,2.4,2.39]) = 0.360966;$$
$$\xi_3(\{2,3,4,5,6\},[26.5,20,2.499,2.4,2.39]) = 0.294503$$

and

$$\xi_2(\{1,2,3\},[29,26.5,20]) = 0.343046;$$
$$\xi_3(\{1,2,3\},[29,26.5,20]) = 0.295695;$$

The share of agent 2 is higher in the coalition $\{2,3,4,5,6\}$ while the share of agent 3 is higher in the coalition $\{1,2,3\}$. Hence, combination sharing does not satisfy the property of consistent ranking.

There is a class of games, however, where combination sharing will yield consistent ranking for any value λ (and hence satisfies GC for the restriction to this games).

Definition 11 *(Size-Power Monotonicity) A game (N,π) is size-power monotonic (SPM) if for any $A,B \subset N$ such that $|B| > |A|$, we have that $\pi(B) > \pi(A)$. The set of SPM games is denoted by $\bar{G}$.*

What the SPM condition does is to take away the tension between coalition size and power (since power increases with size) and thus coalitions with smaller sizes (which, by definition, have lower power) will always give a higher share for any value of λ. Intuitively, since there is no disagreement with coalition size and power, then there will exist a ranking over these coalitions to which agents consistently agree. The result stated in Corollary 1 shows that if we want to find a transition correspondence that satisfies SE and RAT for all possible values of $\lambda \in (0,1)$, then we should restrict the domain of games to SPM games.

On the other hand, even if the game does not satisfy SPM (that is, there is a tension between coalition size and power), it is still possible to find conditions wherein a transition correspondence that satisfies SE and RAT will exist. To see this, recall from the definition of combination sharing that agents in the intersection of two coalitions S and T will agree on coalition S if and only if their share is higher in S than in T, that is:

$$\lambda\frac{1}{|S|} + (1-\lambda)\frac{\pi_i}{\pi(S)} > \lambda\frac{1}{|T|} + (1-\lambda)\frac{\pi_i}{\pi(T)} \;\forall\, i \in S \cap T$$

or, after rearranging,

$$\lambda\frac{|T|-|S|}{|T||S|}+(1-\lambda)\pi_i\frac{\pi(T)-\pi(S)}{\pi(T)\pi(S)}>0\,\forall\, i\in S\cap T \tag{3.2}$$

For instance, if the size of coalition S were smaller than T ($|S|<|T|$) but the coalition power were higher ($\pi(S)>\pi(T)$), the the only way that S will be preferred is when the first term on the left-hand side of Equation 3.2 is higher than the second term on the right hand side.

Since all the parameters (coalition sizes, coalition powers, and the agent's power) in Equation 3.2 are known, in principle we can find values of λ where we can make all agents in the intersection of two coalitions prefer one over the other. That is, we can find a λ high enough for agents to prefer the smaller-sized coalition (since higher λ puts more weight towards equal sharing) and a λ low enough for the agents to prefer lower-powered coalitions (where proportional sharing dominates). Thus, for any two coalitions $S,T\in(N,\pi)$ we can define:

$$\underline{\lambda}^{(S,T)}(N,\pi)=\frac{\underline{\pi}_i^{(S,T)}\left(\frac{\pi(S)-\pi(T)}{\pi(S)\pi(T)}\right)}{\underline{\pi}_i^{(S,T)}\left(\frac{\pi(S)-\pi(T)}{\pi(S)\pi(T)}\right)+\left(\frac{|T|-|S|}{|T||S|}\right)} \tag{3.3}$$

where $\underline{\pi}_i^{(S,T)}$ is the power of agent i in the intersection S and T with the lowest power.

and

$$\bar{\lambda}^{(S,T)}(N,\pi)=\frac{\bar{\pi}_i^{(S,T)}\left(\frac{\pi(S)-\pi(T)}{\pi(S)\pi(T)}\right)}{\bar{\pi}_i^{(S,T)}\left(\frac{\pi(S)-\pi(T)}{\pi(S)\pi(T)}\right)+\left(\frac{|T|-|S|}{|T||S|}\right)} \tag{3.4}$$

where $\bar{\pi}_i^{(S,T)}$ is the power of agent i in the intersection of S and T with the highest power.

In Equation 3.3, any $\lambda\leq\underline{\lambda}$ will convince agents in the intersection to choose S over T when S has the lower power by putting more weight on proportional sharing. Note that we only need to convince the agent with the lowest power $\underline{\pi}_i$ since he will have the least to gain under proportional sharing. If this lowest-powered agent's share is higher under this level of λ, then any other agent in the intersection of S and T with a higher power will also have a higher share in coalition S. In the same manner, Equation 3.4 provides the incentives to choose a smaller-sized coalition over a larger coalition (but with a lower power) by putting more weight on equal sharing. Note that only the agent in the intersection with the largest power must be convinced to choose the smaller-sized coalition since he has the most to lose in moving from a level of λ that puts more weight on proportional sharing than one that puts more weight on equal sharing.

As a corollary, λ values between $\underline{\lambda}^{(S,T)}(N,\pi)$ and $\bar{\lambda}^{(S,T)}(N,\pi)$ create disagreement between the agents in the intersection of S and T. Particularly, when $\lambda > \underline{\lambda}^{(S,T)}(N,\pi)$ the lowest powered agent prefers the coalition with the smaller size (and higher power) while for $\lambda < \bar{\lambda}^{(S,T)}(N,\pi)$ the highest powered agent prefers the lower powered coalition (with larger size).

In order to define the transition correspondence over these non-SPM games, we first note that when agents are killed, the set of self-enforcing coalitions coincide for all sharing rules, even those that does not satisfy consistent ranking. This is shown in Lemma 1 in Appendix A. What this means is that the set of self-enforcing coalitions does not depend on the sharing rule but only on the vector of powers in the game (N,π). Hence, we can construct the class of "stable games" generated by the self-enforcing coalitions in the game (N,π) (See Appendix A and Example 9).

From these stable games we can define the set

$$SEC(N,\pi) = \{T \subsetneq N \mid (T,\pi_T) \text{ is stable and } T \in W_{(N,\pi)}\}$$

as the set of coalitions that generate a stable game and are winning in the game (N,π).

Finally, we can define the set of *undominated coalitions*

$$UDC(N,\pi) = \{T \subseteq N \mid T \in SEC(N,\pi) \ \nexists S \in SEC(N,\pi), \xi_i(S,\pi_S) > \xi_i(T,\pi_T) \forall i \in S \cap T\}$$

The set $UDC(N,\pi)$ contains coalitions that can have potential disagreements between agents given a sharing rule ξ.

To illustrate these sets, let us refer to Example 8. In that game, all the self-enforcing coalitions are the singletons $\{1\}$, $\{2\}$,$\{3\}$,$\{4\}$,$\{5\}$,$\{6\}$ as well as the coalitions $\{1,2,3\}$, $\{1,2,5,6\}$, $\{1,2,4,5\}$, $\{1,2,4,6\}$, and $\{2,3,4,5,6\}$. The set of stable games which is generated by these coalitions are the following:

$$(\{1\},[29]),(\{2\},[26.5]),(\{3\},[20]),(\{4\},[2.499]),(\{5\},[2.4]),(\{6\},[2.39])$$

$$(\{1,2,3\},[29,26.5,20]),$$

$$(\{1,2,5,6\},[29,26.5,2.4,2.39]),$$

$$(\{1,2,4,5\},[29,26.5,2.499,2.4]),$$

$$(\{1,2,4,6\},[29,26.5,2.499,2.39]),$$

$$(\{2,3,4,5,6\},[26.5,20,2.499,2.4,2.39])$$

In the same example,

$$SEC(N,\pi) = \{\{1,2,3\},\{1,2,5,6\},\{1,2,4,5\},\{1,2,4,6\},\text{ and }\{2,3,4,5,6\}\}.$$

Here, the singletons are excluded since they are not winning in the game (N, π).

Finally, the set of undominated coalitions under combination sharing are

$$UDC(N,\pi) = \{\{1,2,3\},\{1,2,5,6\},\{2,3,4,5,6\}\}$$

since agents in the intersection of $\{1,2,5,6\}$ against either $\{1,2,4,5\}$ or $\{1,2,4,6\}$ prefer $\{1,2,5,6\}$ (since they have similar sizes but $\{1,2,5,6\}$ has the least power).

We are now in the position to calculate for the game $(N,\pi) \notin \bar{G}$ two threshold values $\underline{\Lambda}(N,\pi)$ and $\bar{\Lambda}(N,\pi)$ over the set $UDC(N,\pi)$. The threshold value $\underline{\Lambda}(N,\pi)$ ensures that for any $\lambda \leq \underline{\Lambda}(N,\pi)$ the lowest-powered coalition (with the largest size) in $UDC(N,\pi)$ will be unanimously preferred by all agents over all other coalitions in $UDC(N,\pi)$ to which they could possibly belong. On the other hand, for $\lambda \geq \bar{\Lambda}(N,\pi)$ the smallest-sized coalition in $UDC(N,\pi)$ will be unanimously preferred by all agents. These values are:

$$\underline{\Lambda}(N,\pi) = \min_{T} \underline{\lambda}^{(\underline{S},T)}(N,\pi) \tag{3.5}$$

where $\underline{S}, T \in UDC(N,\pi)$ and $|\underline{S}| > |T| \; \forall \, T \in UDC(N,\pi)$

and

$$\bar{\Lambda}(N,\pi) = \max_{T} \bar{\lambda}^{(\bar{S},T)}(N,\pi) \tag{3.6}$$

where $\bar{S}, T \in UDC(N,\pi)$ and $|\bar{S}| < |T| \; \forall \, T \in UDC(N,\pi)$

As an illustration, refer to the game in Example 8. Note that the game $(N,\pi) = (\{1,2,3,4,5,6\}, [29, 26.5, 20, 2.499, 2.4, 2.39])$ generated by these coalitions do not satisfy size-power monotonicity since smaller-sized coalitions have larger power. Recall in the example that the undominated winning and self-enforcing coalitions are $\{1,2,3\}$, $\{1,2,5,6\}$ and $\{2,3,4,5,6\}$. We calculate the following values based on Equations 3.3, 3.4, 3.5 and 3.6:

$\underline{\lambda}^{(\{1,2,3\},\{1,2,5,6\})}(N,\pi) = 0.515172$

$\underline{\lambda}^{(\{1,2,3\},\{2,3,4,5,6\})}(N,\pi) = 0.445036$

$\underline{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi) = 0.0874438$

$\bar{\lambda}^{(\{1,2,3\},\{1,2,5,6\})}(N,\pi) = 0.537643$

$\bar{\lambda}^{(\{1,2,3\},\{2,3,4,5,6\})}(N,\pi) = 0.515162$

$\bar{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi) = 0.515145$

$\underline{\Lambda}(N,\pi) = 0.0874438$

$\bar{\Lambda}(N,\pi) = 0.537643$

Figure 3.1 shows the same information graphically. The topmost line shows the interval between $\underline{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi)$ and $\bar{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi)$. Note that on the extreme left such that $\lambda \leq \underline{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi)$ the coalition $\{2,3,4,5,6\}$ will be preferred by the agents in the intersection of $\{2,3,4,5,6\}$ and $\{1,2,5,6\}$ (that is, agents 2, 5 and 6). On the other hand, if $\lambda \geq \bar{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi)$, the coalition $\{1,2,5,6\}$ is preferred. The other lines can be analogously described.

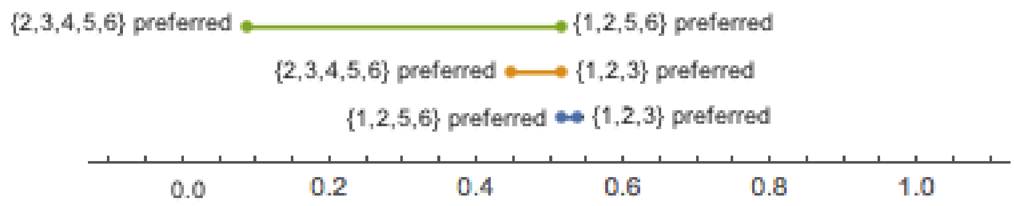

Figure 3.1: Threshold λ Values For The Game $(\{1, 2, 3, 4, 5, 6\}, [29, 26.5, 20, 2.499, 2.4, 2.39])$

Moreover, if $\lambda \leq \underline{\Lambda}(N, \pi)$, then we are assured that $\{2, 3, 4, 5, 6\}$ is preferred to either $\{1, 2, 5, 6\}$ or $\{1, 2, 3\}$. On the other hand, when $\lambda \geq \bar{\Lambda}(N, \pi)$, the coalition $\{1, 2, 3\}$ is preferred over $\{1, 2, 5, 6\}$ or $\{2, 3, 4, 5, 6\}$. Thus, we can always find a λ "low enough" to encourage agents to prefer the coalition with the lowest power and also a λ "high enough" to encourage agents to prefer the coalition with the smallest size.

In this particular example, however, we can also find an interval where a "compromise coalition" (with size between the smallest-sized and largest-sized cooalition) will be unanimously preferred by agents in the intersection. If λ is between 0.515145 and 0.515172, agents will prefer $\{1, 2, 5, 6\}$ over $\{2, 3, 4, 5, 6\}$ and $\{1, 2, 3\}$. Suppose $\lambda = 0.51516$, then:

Comparing $\{1, 2, 5, 6\}$ and $\{1, 2, 3\}$, we have

$$\xi_1(\{1, 2, 5, 6\}, [29, 26.5, 2.4, 2.39]) = 0.362002;$$

$$\xi_2(\{1, 2, 5, 6\}, [29, 26.5, 2.4, 2.39]) = 0.341898;$$

$$\xi_1(\{1, 2, 3\}, [29, 26.5, 20]) = 0.35795;$$

$$\xi_2(\{1, 2, 3\}, [29, 26.5, 20]) = 0.341896;$$

Comparing $\{1, 2, 5, 6\}$ and $\{2, 3, 4, 5, 6\}$, we have

$$\xi_2(\{1, 2, 5, 6\}, [29, 26.5, 2.4, 2.39]) = 0.341898;$$

$$\xi_5(\{1,2,5,6\},[29,26.5,2.4,2.39]) = 0.14809;$$

$$\xi_6(\{1,2,5,6\},[29,26.5,2.4,2.39]) = 0.14801;$$

$$\xi_2(\{2,3,4,5,6\},[26.5,20,2.499,2.4,2.39]) = 0.341896;$$

$$\xi_5(\{2,3,4,5,6\},[26.5,20,2.499,2.4,2.39]) = 0.124665;$$

$$\xi_6(\{2,3,4,5,6\},[26.5,20,2.499,2.4,2.39]) = 0.124575;$$

In this example every agent in the intersection prefers $\{1,2,5,6\}$. Thus, aside from agreement in the extreme ends of the spectrum of λ, there can also be agreements where agents will prefer the coalition that has neither the smallest size nor the smallest power. These compromise coalitions will exist whenever there exists a coalition Q where $|S| < |Q| < |T|$ and $S, Q, T \in UDC(N,\pi)$ such that:

$$\max_T \bar{\lambda}^{(T,Q)}(N,\pi) < \min_S \underline{\lambda}^{(S,Q)}(N,\pi) \tag{3.7}$$

Label these values as

$\bar{\Lambda}^Q(N,\pi) = \max_T \bar{\lambda}^{(T,Q)}(N,\pi)$ and

$\underline{\Lambda}^Q(N,\pi) = \min_S \underline{\lambda}^{(S,Q)}(N,\pi)$.

Of course there are games where we cannot find these compromise coalitions. For instance, suppose we decrease agent 4's power to 2.41 instead of 2.499 so that the game now becomes:

$$(N,\pi) = (\{1,2,3,4,5,6\},[29,26.5,20,2.41,2.4,2.39]).$$

The computed intervals are now the following:

$\underline{\lambda}^{(\{1,2,3\},\{1,2,5,6\})}(N,\pi) = 0.515172$

$\underline{\lambda}^{(\{1,2,3\},\{2,3,4,5,6\})}(N,\pi) = 0.446456$

$\underline{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi) = 0.0886686$

$\bar{\lambda}^{(\{1,2,3\},\{1,2,5,6\})}(N,\pi) = 0.537643$

$\bar{\lambda}^{(\{1,2,3\},\{2,3,4,5,6\})}(N,\pi) = 0.516597$

$\bar{\lambda}^{(\{1,2,5,6\},\{2,3,4,5,6\})}(N,\pi) = 0.518594$

$\underline{\Lambda}(N,\pi) = 0.0874438$

$\bar{\Lambda}(N,\pi) = 0.537643$

In this example compromise coalition $\{1,2,5,6\}$ does not exist because
$\bar{\Lambda}^{\{1,2,5,6\}}(N,\pi) = 0.518594 > \underline{\Lambda}^{Q}(N,\pi) = 0.515172$.

The next proposition establishes the largest class of non-SPM games for which a transition correspondence will satisfy SE and RAT. In these games we basically need the combination sharing parameter λ to be either low enough for agents in the agents to unanimously pick the coalition with the smallest power (but larger size), high enough for them to pick the coalition with the smallest size, or just right for a compromise coalition to be picked.

Proposition 6 *For the game (N,π), fix the sharing rule $\xi_i(N,\pi) = \lambda\frac{1}{|N|} + (1-\lambda)\frac{\pi_i}{\pi(N)}$. Consider the set of games $\breve{G}$ and define*

$$\breve{G}(N,\pi) = \{(S,\pi_S) \mid S \subseteq N \textit{ and } \pi_S \textit{ power vector } \pi \textit{ restricted to } S\}$$

.

The transition correspondence $\breve{\phi} : \breve{G} \to 2^N$ satisfies SE and RAT, if and only if:

i. For any game (S,π_S), $\breve{G}$ should only contain games where λ is an element of $[0, \underline{\Lambda}(S,\pi_S)]$ or $\left[\bar{\Lambda}(S,\pi_S), 1\right]$ or $\bigcup_{Q \in UDC(S,\pi_S)} \left[\bar{\Lambda}^{Q}(S,\pi_S), \underline{\Lambda}^{Q}(S,\pi_S)\right]$.

ii. if $\lambda \in \left[\bar{\Lambda}(S,\pi_S), 1\right]$,

$$\breve{\phi}(S,\pi_S) = \underset{V \in UDC(S,\pi_S)}{\arg\min} |V|$$

iii. if $\lambda \in [0, \underline{\Lambda}(S,\pi_S)]$,

$$\breve{\phi}(S, \pi_S) = \underset{V \in UDC(S,\pi_S)}{\arg\min} \; \pi(V)$$

iv. if $\bar{\Lambda}^Q(S, \pi_S) \leq \lambda \leq \underline{\Lambda}^Q(S, \pi_S)$

$$\breve{\phi}(S, \pi_S) = \left\{Q \mid Q \in UDC(S, \pi_S), \bar{\Lambda}^Q(S, \pi_S) < \underline{\Lambda}^Q(S, \pi_S)\right\}$$

The intuition behind Proposition 6 is that by restricting the combination sharing parameter λ to the values prescribed by part (i), we are assured that there is at least one coalition that will be preferred to other coalitions in the set $UDC(N, \pi)$. Parts (ii), (iii) and (iv) ensures that the chosen coalition will give the agents in the intersection of competing coalitions the highest possible share of the prize. This is achieved by choosing the coalition with the smallest size in the case where λ is high enough, choosing the coalition of the least power in the case where λ is low enough, and choosing all compromise coalitions whenever λ is just right for every agent in the intersection to agree on these coalitions.

The corollary to this proposition states that whenever we want a transition correspondence to satisfy SE and RAT for any value of λ, we must restrict the domain of games to SPM games. The transition correspondence will then pick the smallest-sized coalition in the set $UDC(N, \pi)$, which, by virtue of the restriction to SPM games, is also the coalition of the least power.

Corollary 1 *Suppose that the transition correspondence* $\breve{\phi} : \breve{G} \to 2^N$ *satisfies SE and RAT for any* $\lambda \in (0, 1)$*, then* $\breve{G} = \bar{G}$*, that is, the the set of games only admits SPM games. Moreover,* $\underline{\Lambda}(S, \pi_S) = 1$, $\bar{\Lambda}(S, \pi_S) = 0$ *and*

$$\breve{\phi}(S, \pi_S) = \underset{V \in UDC(S,\pi_S)}{\arg\min} |V| = \underset{V \in UDC(S,\pi_S)}{\arg\min} \; \pi(V)$$

3.4 Agents survive

In this section we tackle the case where non-winning agents survive the coalition formation game. Unlike the previous case, agents should be concerned not only with potential deviations within their coalitions but also possible deviations with agents outside the coalition. Hence, self-enforcement must ensure that the coalition picked by the transition correspondence should be immune to threats from all possible coalitions that could form from all the agents in the society.

We argue in this section that when agents survive, the set of winning coalitions completely characterizes the set of self-enforcing coalitions. In the previous section where agents are killed, self-enforcement requires that the coalition picked by the transition correspondence should not deviate into a subcoalition that will be self-enforcing. When agents survive, feasible deviations may include agents outside of the winning coalition. Thus, if the coalition deviates into another (winning) coalition, in the succeeding rounds nothing can stop the original deviating agents to form the initial winning coalition again. Hence, every deviating coalition from the initial winning coalition will never be assured that another deviation cannot occur. Therefore, the definition of the transition correspondence alone (where coalitions picked is from the set of winning coalitions, recalling Definition 10) is sufficient to guarantee that the coalitions picked will be self-enforcing.

For instance, consider the game $(N, \pi) = (\{1, 2, 3, 4, 5, 6\}, [29, 26.5, 20, 2.499, 2.4, 2.39]$. When agents are killed, the self-enforcing coalitions are $\{1, 2, 3\}$, $\{1, 2, 5, 6\}$, $\{1, 2, 4, 5\}$, $\{1, 2, 4, 6\}$ and $\{2, 3, 4, 5, 6\}$. When agents survive, we argue that the winning coalition $\{2, 3, 5\}$ is self-enforcing. To see this, suppose when $\{2, 3, 5\}$ forms, agents 2 and 3 want to deviate to coalition $\{2, 3, 6\}$ and this is feasible since $\{2, 3, 6\}$ is winning in (N, π). When coalition $\{2, 3, 6\}$ forms, agent 5 can enter into an agreement with agent 1 and 3 to form the coalition $\{1, 3, 5\}$ which can form because this coalition is winning. This incentive to counter-deviate will be present for all deviations from a winning coalition. Hence, any deviation from a winning coalition can engender another feasible counter-deviation to another winning coalition.

With this characterization of self-enforcement, we modify the definition of RAT appropriate to the case where non-winning agents survive. Analogous to our definition in the previous section, this version of rationality says that a winning (and hence, self-enforcing) coalition T will be chosen over another winning coalition Z whenever all the agents in the intersection receive a higher share of the prize in T.

Axiom 11 (Rationality (RAT2)) *A transition correspondence ϕ meets* ***dynamic external rationality*** *(DER) if for any $T \in \phi(N, \pi)$ and for any $Z \subset N$ such that $Z \in W_{(N,\pi)}$, we have that $Z \notin \phi(N, \pi) \Leftrightarrow \xi_i(N, \pi_T) > \xi_i(N, \pi_Z)\ \forall i \in T \cap Z$.*

Acemoglu, et al [2] has shown that under the assumption of proportional sharing, the transition correspondence that satisfies a version of RAT2 must choose from the set of "minimally winning coalitions". A minimally winning coalition is a winning coalition such that removing any of the agents within this coalition causes it to be non-winning. Formally, we define this set as

Definition 12 (Minimally winning coalition) *Given the game (T, π), the set of minimally winning coalitions is given by:*

$$MW_{(T,\pi)} = \{S \in W_{(T,\pi_T)} \mid S - \{j\} \notin W_{(T,\pi_T)} \; \forall \; j \in S\}$$

.

The intuition why minimally winning coalitions are important for a transition correspondence to satisfy RAT2 is straightforward. First, this ensures that there will be no deviations into a subset of this coalition, which means that the only possible deviations must involve agents outside this coalition. Second, minimally winning coalitions ensures that the shares of the agents inside it would be larger than any other (winning) coalition that contains them.

3.4.1 Consistent Ranking

This subsection presents the results with consistent ranking. The next proposition shows that the unique transition correspondence that satisfies RAT2 picks the highest ranked coalition among all winning coalitions. Furthermore, the corollary to this proposition characterizes the coalition that is picked by this transition correspondence. Under proportional sharing, the transition correspondence that satisfies RAT2 picks the "minimally winning coalition of minimal weight (MWCMW)" (the minimally winning coalition with the least power) while under equal sharing the transition correspondence picks the "minimally winning coalition of minimal size (MWCMS)" (the minimally winning coalition with the smallest size).

Proposition 7 *Consider a sharing rule that satisfies consistent ranking. Then a unique transition correspondence ϕ^{**} satisfies RAT2 if and only if:*

$$\phi^{**}(S, \pi) = \underset{M \in W_{(S,\pi)}}{\arg\max} \, R^{\xi}(M, \pi_M) \tag{3.8}$$

Corollary 2 *Consider the game $(T, \pi) \in \mathbb{G}$:*

i. Under equal sharing, if $S \in \phi^{**}(T, \pi)$, then S is the minimally winning coalition of minimal size, that is,

$$\{S | S \in MW_{(T,\pi)} \text{ and } |S| < |V| \; \forall \; V \in MW_{(T,\pi)}\}$$

.

ii. *Under proportional sharing, if $S \in \phi^{**}(T, \pi)$, then S is the minimally winning coalition with minimal weight, that is,*

$$\{S | S \in MW_{(T,\pi)} \textit{ and } \pi(S) < \pi(V) \; \forall \, V \in MW_{(T,\pi)}\}$$

.

Thus, there will always exists a well-defined transition correspondence that satisfies RAT2 when agents survive and the sharing rule satisfies consistent ranking. By choosing the MWCMS (under equal sharing) and the MWCMW (under proportional sharing), the agents in the chosen coalition are guaranteed their highest payoffs over all alternative coalitions to which they could belong.

3.4.2 Combination Sharing

As with the case where non-winning agents are killed, there may not exist a transition correspondence that satisfies RAT2 for the whole domain of games under combination sharing. If we want the transition correspondence to exist for all possible values of λ, then as with the previous section we should restrict the domain of games to SPM games.

Echoing the previous section, in the case where agents survive and games are not SPM, we can always find levels of λ such that a transition correspondence that satisfies RAT2 will exist. Instead of looking at the set UDC in the previous section, we now look at the set of undominated minimally winning coalitions (UDMW) over the game (N, π):

$$UDMW(N, \pi) = \{T \subseteq N \mid T \in MW_{(N,\pi)} \; \nexists S \in MW_{(N,\pi)}, \xi_i(S, \pi_S) > \xi_i(T, \pi_T) \forall i \in S \cap T\}$$

In Example 8, all the minimally winning coalitions are of size 2, that is $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$. With combination sharing, however, the set $UDMW(N, \pi)$ contains only $\{2, 3\}$ since the share of agent 2 is higher here than in $\{1, 2\}$ and agent 3 has a higher share here than in $\{1, 3\}$. Hence, coalitions $\{1, 2\}$ and $\{1, 3\}$ are dominated. But consider the game

$$(N, \pi) = \{(\{1, 2, 3, 4, 5, 6\}), [4, 4.5, 6, 9, 11, 12]\}$$

Here, there are many minimally winning coalitions, namely $\{4, 5, 6\}$, $\{1, 5, 6\}$, $\{2, 5, 6\}$, $\{3, 5, 6\}$, $\{3, 4, 5\}$, $\{1, 4, 5\}$, $\{4, 5, 6\}$, $\{2, 4, 5\}$, $\{3, 4, 6\}$, $\{1, 4, 6\}$, $\{2, 4, 6\}$, $\{1, 2, 3, 4\}$, $\{1, 2, 3, 5\}$ and $\{1, 2, 3, 6\}$. However, it is easy to verify that the set $UDMW(N, \pi) = \{\{1, 4, 5\}, \{1, 2, 3, 4\}\}$.

Hence, analogous to the case where agents are killed, we define the following thresholds for the case where agents survive:

$$\underline{\Lambda}(N, \pi) = \min_{T} \underline{\lambda}^{(\underline{S},T)}(N, \pi) \tag{3.9}$$

where $\underline{S}, T \in UDMW(N, \pi)$ and $|\underline{S}| > |T|$

and

$$\bar{\Lambda}(N, \pi) = \max_{T} \bar{\lambda}^{(\bar{S},T)}(N, \pi) \tag{3.10}$$

where $\bar{S}, T \in UDMW(N, \pi)$ and $|\bar{S}| < |T|$

Observe that $\underline{\lambda}^{(\underline{S},T)}(N, \pi)$ and $\bar{\lambda}^{(\bar{S},T)}(N, \pi)$ has the same definition as in Equations 3.3 and 3.4, respectively. The difference from the previous case where agents are killed is that in this case the computation of these λ values is made on the set $UDMW(N, \pi)$.

These values of λ in Equations 3.9 and 3.10 set the threshold for which the MWCMW and MWCMS are always chosen. That is, if $\lambda < \underline{\Lambda}(N, \pi)$ the minimally winning coalition of minimal weight will always be chosen and if $\lambda > \bar{\Lambda}(N, \pi)$ the minimally winning coalition of minimal size will always be chosen.

As with the previous section, there also may be λ values where a compromise coalitions is always chosen. In this case, the compromise coalition is a minimally winning coalition that is neither the MWCMS or MWCMW. These coalitions will exists whenever we can find a coalition Q where $|S| < |Q| < |T|$ and $S, Q, T \in UDMW(N, \pi)$ such that

$$\bar{\Lambda}^{Q}(N, \pi) = \max_{T} \bar{\lambda}^{(T,Q)}(N, \pi) < \min_{S} \underline{\lambda}^{(S,Q)}(N, \pi) = \underline{\Lambda}^{Q}(N, \pi) \tag{3.11}$$

The next proposition states the largest class of non-SPM games where a transition correspondence that satisfies RAT2 will exist:

Proposition 8 *For the game (S, π), let the cardinality of the set $|UDMW(S, \pi)| > 1$. Fix the sharing rule $\xi_i(S, \pi) = \lambda \frac{1}{|S|} + (1-\lambda)\frac{\pi_i}{\pi(S)}$ and consider the set of games $\tilde{G}$.*

The transition correspondence $\tilde{\phi} : \tilde{G} \to 2^N$ satisfies RAT2, if and only if:

i. For any game (S, π), $\tilde{G}$ should only contain games where λ is an element of $[0, \underline{\Lambda}(S, \pi)]$ or $\left[\bar{\Lambda}(S, \pi), 1\right]$ or $\bigcup\limits_{Q \in UDMW(S, \pi_S)} \left[\bar{\Lambda}^Q(S, \pi), \underline{\Lambda}^Q(S, \pi)\right]$.

ii. if $\lambda \in \left[\bar{\Lambda}(S, \pi), 1\right]$,

$$\tilde{\phi}(S, \pi) = \underset{V \in UDMW(S,\pi)}{\arg\min} |V|$$

iii. if $\lambda \in [0, \underline{\Lambda}(S, \pi)]$,

$$\tilde{\phi}(S, \pi) = \underset{V \in UDMW(S,\pi_S)}{\arg\min} \pi(V)$$

iv. if $\bar{\Lambda}^Q(S, \pi) \leq \lambda \leq \underline{\Lambda}^Q(S, \pi)$

$$\tilde{\phi}(S, \pi) = \left\{Q \mid Q \in UDMW(S, \pi_S), \bar{\Lambda}^Q(S, \pi) < \underline{\Lambda}^Q(S, \pi)\right\}$$

Akin to the case where agents are killed, if we want the transition correspondence to satisfy RAT2 for any value of λ, we need to restrict the domain of games. Corollary 3 shows this restriction.

Corollary 3 *Suppose that the transition correspondence $\tilde{\phi} : \tilde{G} \to 2^N$ satisfies RAT2 for any $\lambda \in (0, 1)$, then $\tilde{G} = \bar{G}$, that is, the the set of games only admits SPM games. Moreover, $\underline{\Lambda}(S, \pi) = 1$, $\bar{\Lambda}(S, \pi) = 0$ and*

$$\tilde{\phi}(S, \pi) = \underset{V \in W_{(S,\pi)}}{\arg\min} |V| = \underset{V \in W_{(S,\pi)}}{\arg\min} \pi(V)$$

3.5 Conclusion

This paper develops an axiomatic approach to a coalition formation model by focusing on two main axioms: self-enforcement and rationality. We investigate the effect of different sharing rules on the existence of transition correspondences that satisfy these axioms. We find that, indeed, the existence of these transition correspondences is very sensitive to the choice of sharing rules. The results are summarized in the table below.

Scenario	Consistent Ranking	Combination Sharing
Agents Killed	ϕ^* (See Equation 3.1)	$\breve{\phi}$ (See Proposition 6 and Corollary 1)
Agents Survive	ϕ^{**} (See Equation 3.8)	$\tilde{\phi}$ (See Proposition 8 and Corollary 3)

We find that when the sharing rule satisfies the property of consistent ranking (where agents have the same ordinal rank over coalitions) then we can always find a transition correspondence that satisfies these axioms regardless of whether non-winning agents are killed or are able to survive. Under combination sharing, however, these transition correspondences do not exist in general. We have to restrict the domain of games either to the case where coalition size and power move in the same direction or by allowing the sharing parameter λ to be high enough for agents to agree on the smallest-sized undominated coalition, low enough for agents to agree on the least-powered undominated coalition, or just enough for a compromise coalition to exist.

One lesson we have shown is that potential disagreements among agents on the choice of coalitions need not lead to a complete breakdown of cooperation. It is always possible to find some suitable compromises if only agents agree to set the correct method of sharing the prize. This is an optimistic result, and can very well be applied to any activity that requires cooperation. Hence, the world's biggest problems, for instance climate change and water scarcity, can be solved through collective action if only the cost and benefits are appropriately shared by everyone concerned.

4 AN EXPERIMENTAL STUDY OF SELF-ENFORCING COALITIONS

4.1 Introduction

The main economic rationale why agents strategically cooperate and form groups is that they can appropriate certain advantages unavailable to them if they acted on their own. In this framework, rational agents endogenously form coalitions depending on the incentive structure or the environment they face. Recent literature has been very active in modelling how agents form coalitions in the face of very diverse environments.[1]

One such recent advance is introducing heterogeneous *power* for the agents and examining how its configuration in the society affects the manner in which coalitions endogenously form. Power is the ability to influence the behavior of other agents in the society and could emanate from different sources, for instance, political, military or money. As an extreme example, dictatorships (a singleton coalition) may occur more frequently in environments when there is one agent who has a disproportionate share of power. For instance, Tullock [101] argues that juntas (a small group that wields military, economic or political power) always degenerates into a dictatorship since whenever formal institutions on distribution and sharing of power, succession of leaders, and generating consensus are weak or absent, there is a tendency for one person to dominate society. One key question is under what conditions Tullock's conjecture that a ruling coalition always degenerates into a dictatorship forms will hold. One plausible reason is that agents are myopic in choosing their strategies, that is, agents may not take into account the long-run consequences of the coalitions that they form.

An important paper by Acemoglu, Egorov, and Sonin [1] (hereafter AES) tackles this question heads-on and show, under a very specific bargaining context, how Tullock's conjecture will not hold when agents are farsighted. In their paper, agents with different powers make proposals and respond to such proposals to try to form coalitions with sufficient power to win throughout time. At each point in time, agents outside this winning coalition are "killed", that is, they are unable to

[0]My deepest appreciation to Ruben Juarez and Katya Sherstyuk for overall guidance in writing this Chapter. I also would like to thank Ryan Morisato and Bryson Yee for helping me run the experiments.

[1]An excellent introduction to the concepts in this literature is Ray [82].

participate in future time periods. The "ultimate ruling coalition" that emerges, which is the coalition that will be stable forever starting from a particular point in time, will win the prize and its members will divide this prize among themselves in proportion to their relative power inside the ultimate ruling coalition.[2] AES then show that the unique subgame perfect equilibrium of their bargaining process involves strategies that allow for an ultimate ruling coalition–not necessarily a dictator–to form after one transition, that is, there will be no intermediate coalitions forming.[3]

As a concrete example, imagine three agents in the society where agent 1 has power 20, agent 2 has power 35 and agent 3 has power 45. Notice that no single agent has sufficient power to win, and so he must propose to at least one other agent to form a "winning" coalition, that is, the coalition must have at least 50% of the power in the society (a majority). Suppose 35 was randomly selected to be a proposer and he proposed coalition $\{20, 35\}$ to agent 20. This coalition is winning since the sum of their powers (55) is more than the remaining agent 45 (that is, they have more than 50% of society's total power). Now 20 can either accept or reject this proposal. If he accepts, $\{20, 35\}$ forms and agent 45 is killed. In the next period another proposer is selected but this time only from agents 20 or 35. Agent 20 can only propose $\{20, 35\}$ since he is not winning by himself. Agent 35, however, can propose himself to be a dictator. If that happens, $\{35\}$ will be the ultimate ruling coalition from this period onwards. Since 20 is fully rational and farsighted, agent 20 will never agree to form $\{20, 35\}$ and thus rejects this proposal when faced with it. Following this argument, no 2-person coalition will ever form in this example. On the other hand, proposing the grand coalition is $\{20, 35, 45\}$ is feasible since this coalition will never deviate into a 2-person coalition following the same logical process outlined. Hence, this coalition is said to be "self-enforcing" or stable. In this simplified version of the AES's model, farsighted agents know that if they propose a 2-person coalition with a higher-powered agent, they will be killed in the future. Higher-powered agents, on the other hand, know they will be rejected by farsighted, lower-powered agents if a 2-person coalition is proposed. Hence, the unique subgame perfect equilibrium strategies requires that the proposer propose only the only self-enforcing coalition $\{20, 35, 45\}$ and nothing else, while the responders would accept the self-enforcing coalition $\{20, 35, 45\}$, a 2-person coalition if proposed by a lower-powered agent,

[2] AES's main stability concept, *self-enforcement*, requires that there will be no subcoalition of the winning coalition that will be powerful enough to encourage further deviations. Self-enforcement is a robustness property that ensures that the coalition that forms never disintegrates afterwards. Another appealing axiom they use is *rationality* that requires agents to pick the coalition that gives them their highest payoff among self-enforcing coalitions. Rationality is is related to immunity to group manipulations in models discussed by Bogomolnaia and Jackson [16], Ehlers [29], Juarez [53].

[3] Jandoc and Juarez [47] extends the analysis of Acemoglu et al [1] to the case where power accumulates, as well as to the case when the prize is shared equally among the agents.

and reject anything else. Playing these SPNE strategies would allow the grand coalition $\{20, 35, 45\}$ to form at once. In this particular example, we can clearly see that if agents are farsighted, then an ultimate ruling coalition that is not a dictator will form.

Hence we have two competing theories on how agents form coalitions over time. On one extreme is the Tullock position, where ruling coalitions always end up in a dictatorship because of the incentives and possibly because of the agents' bounded rationality. At the other extreme is the AES position where a non-dictatorship ruling coalition can form if agents are farsighted.

In this paper, we let actual subjects participate in a laboratory experiment to investigate which theory adheres more closely to agents' actual behavior. In this experiment we play a version of the AES coalition formation model corresponding to the setting we described in the example above.[4] Employing an experimental approach is useful in this setting for two reasons. First, adopting a highly simplified version of AES' coalition formation model enables us to examine the extent to which agents' behavior deviates from the predictions of this model. While the AES coalition formation process (see Section 4.3) provides clear cut predictions on what coalition will emerge starting from a society of agents with specific powers, it rests on the assumption that agents are farsighted and only care about their payoff from self-enforcing coalitions. If agents are myopic, then it is possible that dictatorships will be more likely to form since by trying to maximize immediate payoffs agents may fail to foresee that deviation into a non-self-enforcing coalition would lead to further deviations that may exclude them in the future. If the deviation is substantial, we can examine the agents' strategies in order to determine whether myopia has hindered the agents' from attaining the model's prediction. Second, we can determine whether agents are able to behave more rationally by increasing the incentives for them to do so or by gaining experience through time. The AES game is set up in such a way that the size of the resource does not have any bearing on the ultimate coalition that will form since only the agent's relative payoffs matter. In real-world settings, however, stake size may affect behavior because it could induce agents to behave more "cooperatively" since losing the pot entails a higher potential loss. Moreover, the AES model provides no scope for learning despite the dynamic nature of their game. In reality, however, boundedly rational agents may behave myopically at first but may learn to be strategic if given enough experience playing the game. In summary, we used a laboratory experiment to address the questions: Do agents form coalitions as the model predicts? If not, why does behavior deviate from prediction? Does myopia, learning and stake size affect the agents' decisions?

[4]The predictions of this particular model corresponds to the coalitions obtained in Section 3.3.2 of this dissertation.

Our experiment shows that the coalition formation model is not a perfect predictor of the actual coalitions that form. In our setting, the proportion of the grand coalitions that form almost never exceeds half of the total coalitions that form in every game. Over time, however, the proportion of grand coalitions forming tends to rise. Stake size slightly helps in increasing the proportion of grand coalitions forming. Examining the actual strategies of the agents, we find a fairly regular pattern for proposals and responses. First, the proposed coalitions are conditional on a particular agent's power. For instance, agents who have low power (20) tend to propose the grand coalition while agents assigned with higher power (35 or 45) tend to propose a 2-person coalition, especially to the lower-powered agent (20) where they can obtain a higher share of the prize. Second, most of the decision to accept or reject a proposed coalition is also contingent upon both an agent's drawn power and the coalition proposed. For example, almost everyone with power 45 accepts a proposed 2-person coalition, while only about half of agents with power 20 accepts a proposed 2-person coalition. Moreover, almost all agents with power 20 accept a proposed grand coalition.

Put in a larger perspective, our results suggest that many agents exhibit behavior deviating from pure farsightedness and full rationality but learn to behave more strategically with enough experience of playing the coalition formation game. This suggests that elements of both the Tullock and the AES theories may be at work in the real-world. Hence, examining these behavioral factors in a strategic context may prove useful to develop future coalition formation models that will help us understand better how agents form groups in reality.

To our knowledge this is the first and the only paper so far that investigates the implications of adding heterogenous power and investigating the stability of coalitions that form. Hence our unique contributions are the following. First, we fill in the gap on empirical studies in coalitions formation by using tools of experimental economics. Second, we establish a framework on how to conduct coalition formation experiments with power. Finally, we identify plausible reasons why actual behaviour may deviate from the theoretical predictions, specifically the importance of myopic strategies, learning and stake size. We then point out the need to refine theory to take into account these factors.

4.2 Literature

4.2.1 Experiments on coalition formation

The literature on coalition formation experiments is sparse. The earliest experiments were designed to test and compare solution concepts such as the core (Horowitz [45]) and the individually rational bargaining set (Rapport and Kahan [81]) using a package of software programs called *Coalitions* (Kahan and Helwig [55]).

Along with the rise in the popularity of the bargaining approach to coalition formation (e.g. Baron and Ferejohn [7]) came experiments to test the theoretical predictions of these models. For instance, several papers attempted to implement the procedures in the Baron and Ferejohn model to test its predictions (McKelvey [63], Frechette et al [37], Battaglini and Palfrey [10]).

We believe our paper is the first and the only one so far to investigate the implications of heterogeneous power to the stability of coalitions. In this paper, we implement the procedural rules from the bargaining game of the AES model, and thus examine the failure of some behavioral assumptions in the event that the outcome deviates from the theoretical prediction.[5]

4.2.2 Learning and experience

Fudenberg and Levine [38] assert that a dynamic and adaptive process of learning and evolution can be an alternative to the "common knowledge of rationality" assumption underpinning most game theory models (Aumann [4]). More forcefully, Ken Binmore [12] observes that "most theorist nowadays agree that people get to equilibrium...by an interactive process of trial-and-error learning". The interesting question in the nexus between theory and experiments is not whether the theory is debunked or not but whether the participants learn how to arrive at the equilibrium in repeated trials. More often than not, participants' behavior in each trial is dependent on the results of previous trials. According to Binmore, testing economic theories in a laboratory setting must satisfy three criteria:

- The problem must be framed to be "reasonably" simple;
- The incentives must be "adequate";
- The time allowed for trial-and-error adjustment is "sufficient"

Models of learning were developed that takes into account the possibility that agents adapt their beliefs to past experience instead of reasoning strategically. These models show that agents have a high degree of myopia. However, Hyndman, et al [46] shows through an experiment that economic agents actually rely on both adaptation and forward-looking behavior. Mengel [64] provides a learning model that ties the two features of adaptability and forward-looking behavior

[5] Tremewan and Vanberg [100] point out that if the outcome from these experiments deviate from theoretical predictions, the failure will emanate from behavioral rather than procedural assumptions.

together. The model is able to provide explanations for the possibility of cooperation in finitely repeated games.

Even though the environment of our coalition formation game is simple enough, we would expect that agents will undergo a trial-and-error process as they play our game. While there are several ways to model the dynamic and adaptive process of learning, we take a straightforward approach by simply enumerating the strategies employed over the progress of the games as well as the types of coalitions that will form in early and late games.

4.2.3 Stake size

Harrison and List [43] provide an excellent review of varying stakes in economic experiments. The question is whether stakes in economic experiments are non-trivial enough to induce agents to reveal their true behavior. To make the stakes appear more substantial, several experiments were conducted in developing countries where participants are poor enough to make the stakes more salient. For instance, Slonim and Roth [96] ran their experiments in the Slovak Republic, Kachelmeier and Shehata [54] in China, Cameron [19] in Indonesia, among others. The stakes in these experiments roughly equalled 62.5 hours of work in Slovak or three times the monthly expenditure of the average participant in Indonesia. The effect is somewhat mixed, in some cases high stakes, like the Slovak Republic study, induced a different behavior than that obtained with lower stakes while in the Indonesian study it did not.

Overall, it is still an open question how representative (small) experimental stakes translate into behavior that reflect situations in the real world. In this paper we increase the stake size in some games and examine the coalitions that will form as a result. In some games we increase the stake tenfold (from \$5 to \$50) while in some games we increase only from \$5 to \$20.

The rest of this paper is organized as follows: Section 4.3 describes the important elements of the theory of the coalition formation game we will conduct in the laboratory. Section 4.4 describes the design of the experiment. Section 4.5 discusses the result of the experiment and Section 4.6 concludes.

4.3 Theoretical considerations

Following AES, there is a set $N = \{1, \ldots, n\}$ of initial agents endowed with powers $\pi = [\pi_1, \ldots, \pi_n]$, respectively. A coalition S is a subset of N, that is, $S \subseteq N$. The set of coalitions are all possible subsets of N, denoted by 2^N. A coalition formation game is a pair (S, π) where $S \subseteq N$ and $\pi \in \mathbb{R}^S_+$. We assume that power is additive, that is, the power of coalition S is the sum of all powers of the agents inside the coalition, $\pi(S) = \sum_{i \in S} \pi_i$. The set of winning coalitions is given by $W_{(T,\pi)} = \{S \subset T | \pi(S) > \pi(T \setminus S)\}$.

We assume a sharing rule that divides the prize of the agents. This sharing rule is fixed throughout the game and we assume that if the agent is part of the final (limit) coalition S, then his share is proportional to his power in the coalition, that is, $\xi_i(S, \pi) = \frac{\pi_i}{\pi(S)}$ if $i \in S$. If the agent is not part of the limit coalition, his payoff is $\xi_i(S, \pi) = 0$.

We have the following extensive form of the game (N, π): Stage j of the game starts with coalition N_j and then the stage game proceeds as follows:

1. Nature picks proposer $a_{j,q} \in N_j$ for $q = 1$.
2. Agent $a_{j,q}$ proposes $P_{j,q} \in C_{N_j}, C_{N_j} \subseteq N_j$ such that $a_{j,q} \in P_{j,q}$
3. Agents in $P_{j,q}$ vote sequentially over the proposal. A random agent is selected to be the first voter $v_{j,q,1}$ who votes $\tilde{v}(v_{j,q,1}) \in \{yes, no\}$, then another voter is selected $v_{j,q,2} \neq v_{j,q,1}$ and the voting process continues. If those who voted yes have sufficient power such that $P_{j,q}$ is winning, that is, $P_{j,q} \in W_{(N_j, \pi_{N_j})}$ then proceed to step 4. If not, proceed to step 5.
4. If $P_{j,q} = N_j$, proceed to Step 6. If not, players from $N_j \setminus P_{j,q}$ are eliminated and the game reverts back to step 1 with $N_{j+1} = P_{j,q}$ (and j increases by 1).
5. If $q < |N_j|$, the next agenda setter $a_{j,q+1} \in N_j$ is randomly picked among those in N_j who have not yet proposed at this stage and the game proceeds to step 2 with q increased by 1. If $q = |N_j|$, the game proceeds to step 6.
6. N_j becomes the ruling coalition where each agent $i \in N_j$ receives payoff $\xi_i(N_j, \pi_{N_j})$.

The major result of AES can now be stated and the proofs can be found in Acemoglu, Egorov and Sonin [1].

Proposition 9 *For the bargaining procedure just described, there exists a pure strategy profile that constitutes a subgame perfect equilibrium and leads to an ultimate ruling coalition in at most one transition.*

The intuition is quite clear with the help of the example in Section 4.1. To reiterate, a farsighted, lower-powered agent knows that if they propose a 2-person coalition to a higher-powered agent they will be killed. Higher-powered agents know that if they propose a 2-person coalition to a farsighted, lower-powered agent they will be rejected. Hence, on the equilibrium path only the self-enforcing grand coalition $\{20, 35, 45\}$ will be the only one proposed and the same coalition will be the only one accepted. Off the equilibrium path, however, higher-powered agents will accept a 2-person coalition from a lower-powered agent. Playing these strategies would allow a self-enforcing coalition to be the ultimate ruling coalition in one transition, without forming any intermediate coalitions.

4.4 Experimental Design

We consider a very simple implementation of the procedures of the game described in the previous section. In this experiment our society is composed of three agents who are endowed with powers either 20, 35 or 45 as in the example in Section 4.1. Similar to the stages of the game in the previous section, we follow the steps below:

1. An agent is randomly chosen to be the proposer and he proposes to form a winning coalition.
2. Agents in this proposed coalition vote whether or not to join.
3. A coalition passes if the majority of the agents vote to join this coalition.
4. If a coalition passes, the game is repeated ONLY with the agents inside this passed coalition.
5. If the coalition fails, a new proposer is randomly chosen and the game begins again.
6. The game ends if a dictator is left or the same coalition gets to be passed after everyone has been given their chance to propose.

Under the theory, if all agents are farsighted no coalition of size two will be proposed nor accepted on the equilibrium path. Off the equilibrium path, however, a higher-powered agent will accept a 2-person coalition if proposed by a lower-powered agent. Hence, in this game the unique Subgame Perfect Nash Equilibrium (SPNE) strategy is to only propose the grand coalition, accept the grand coalition $\{20, 35, 45\}$ if proposed, accept a 2-person-coalition with a lower-powered agent if proposed, and reject anything else. The unique outcome under the SPNE is the grand coalition $\{20, 35, 45\}$ forming.

We developed our own internet-based software to implement the games. Figure 4.1 is a screen shot of the proposal stage. In this screen, available information includes the agent's power, other agents' power, the pot, the coalitions available for proposal (i.e. those with sufficient power to win) and the corresponding payoffs of each agent in any potentially proposed coalition (if the proposed coalition passes). The randomly chosen agent will then have to propose one coalition (with sufficient power) to the other agents in the game.

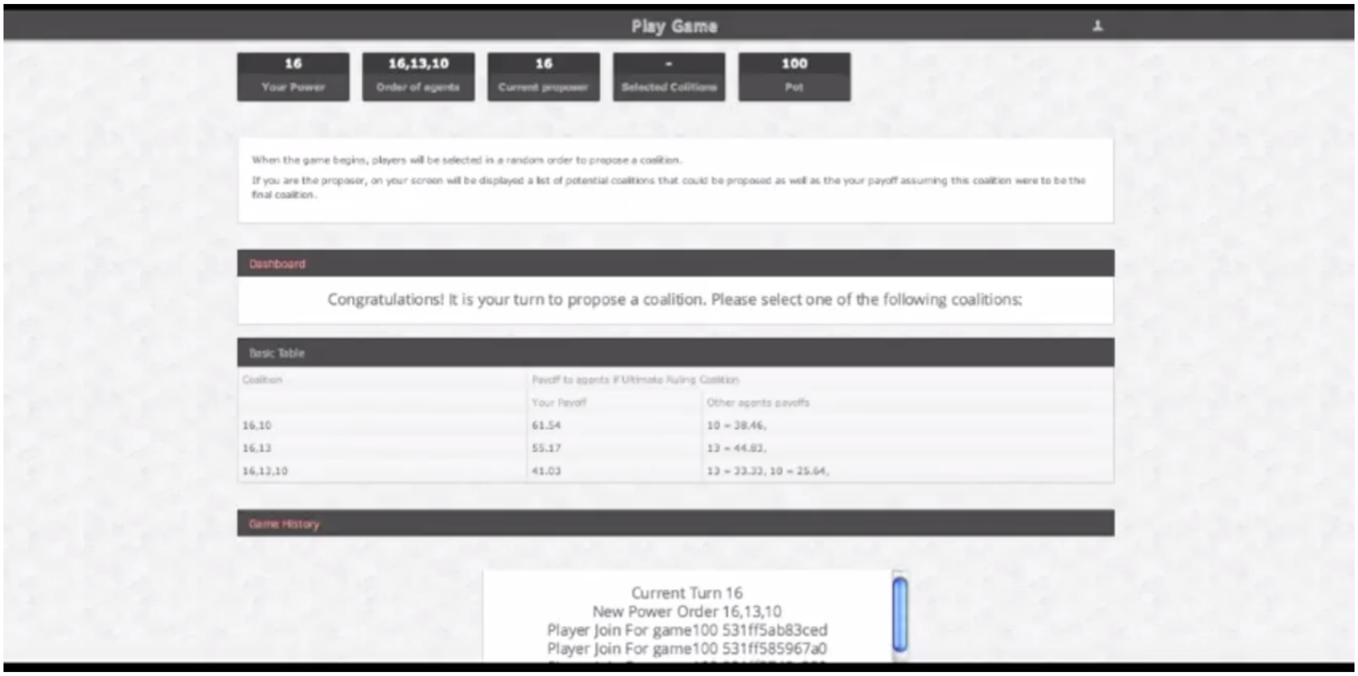

Figure 4.1: A screenshot of the proposal stage

Once a proposer has made his proposed coalitions, the other agents who are part of this proposed coalition will move to the response stage and will be shown a screen similar to Figure 4.2. In this screen, available information includes the agent's power, other agents' power, the pot, the selected proposed coalition, the voting screen, other winning coalitions that were not chosen for proposal, and the corresponding payoffs of each agent.[6] The agent then votes "yes" or "no" to the proposed coalition. The subsequent stages will follow the same procedures until a final coalition passes.

In this experiment the participants are undergraduate students from the University of Hawai'i and recruited via ORSEE (Greiner [41]). Subjects were allowed to participate in only one session in this experiment but they may have some prior experience with other non-related experiments. Subjects were shown a video of the instructions and were administered an online quiz to test how they understood the instructions.[7]

[6]Note that this screen is only viewed by the agents included in the proposed coalition. If they are not part of the proposed coalition, their vote is an automatic "No" and they will have to wait whether the proposed coalition passes or fails. If the proposed coalition passes, then they will have to wait until the game ends.

[7]The instructions can be viewed here: `https://www.youtube.com/watch?v=FKc0lPAQN_0`. Appendix D provides screenshots of the instructions and the quiz.

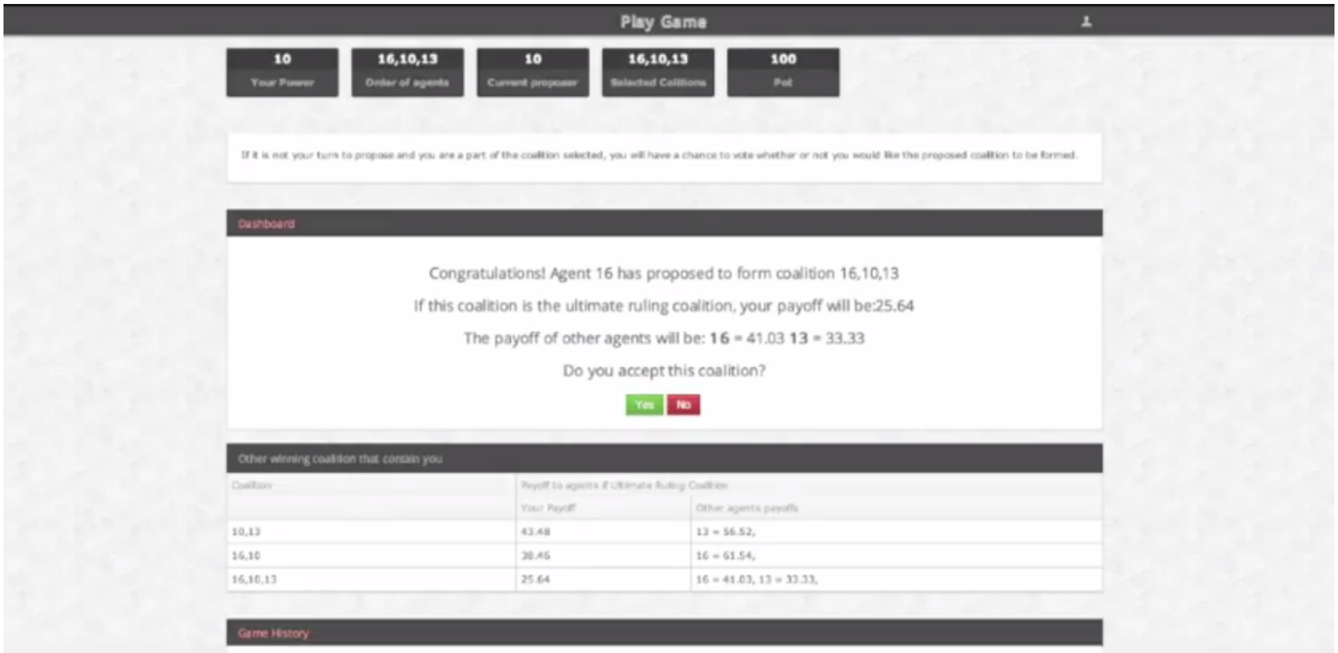

Figure 4.2: A screenshot of the response stage

To allow for learning in each session, the participants played up to 15 games with random rematching. We designate the first two games as "practice" games where the subjects will not be paid. After these practice games, the subjects will split a $5 pot in each of the "paid" games played. We pay the subjects the actual amount that they win for all of the paid games. In order to investigate the effect of increasing stakes, in some sessions we offered $20 or $50 in the last game. [8] In total we have five sessions with $50 final games, three sessions with $20 final games and three sessions with $5 final games.

Table 4.1 describes the summary of the experimental sessions. We held 11 sessions where in each session we conducted up to 15 games. Subjects are grouped into 3 for each game, and are randomly selected into one of these groups. The first two games of each session are designated as practice games where the participants receive no money. After these two practice games, we pay the subject's share of the $5 pot for each game. As just mentioned, there are several sessions where we increase the stake size to either $20 or $50. We pay out an average of (not including the $12 show-up fee) $ 20 for the low-stakes sessions, $25 for the $20-stake sessions, and $35 for the $50-stake sessions.

[8] We announced the stake prior to the start for each of these final games regardless whether it is a high stake (i.e., $20 or $50) or low stake (i.e., $5).

Table 4.1: Summary of experimental sessions

Session Date	Number of games	Number of subjects	Last game stake
Low stake sessions			
May 6, 2014	14	15	\$5
February 4, 2015	14	15	\$5
February 5, 2015	14	12	\$5
Medium stake sessions			
May 6, 2014	14	12	\$20
May 7, 2014	14	12	\$20
June 6, 2014	14	12	\$20
High stake sessions			
June 5, 2014	15	9	\$50
June 23, 2014	15	9	\$50
June 24, 2014	15	6	\$50
February 4, 2015	14	15	\$50
February 5, 2015	14	15	\$50

Note: Total number of sessions is 11. Total number of games is 157. The first two games per session are designated as practice games. There are 132 total participants.

4.5 Results

There is a total of 132 participants and 157 games in our experiment. Out of 157 games, 135 games are paid while 22 are practice games. Our analysis only covers the paid games in all sessions, that is, we do not include the first two practice games. In this section, we present our key results for the type of coalitions that form and Section 4.5.1 will present an analysis of the strategies employed by the agents.

Observation 1 *Very few grand coalitions form compared to the proportion of singletons (dictator) during the early games of the experiment. However, the proportion of grand coalitions forming significantly increases in later games.*

Considering all games, Figure 4.3 suggests that very few grand coalitions form which does not seem to accord with the model's prediction where the grand coalition $\{20, 35, 45\}$ should always form. Taking all sessions together, there is no game in which the proportion of grand coalitions forming exceed 50%. In contrast, during the first few games the proportion of dictators never goes below 50%. This indicates that the model is a weak predictor of the coalitions that will form. In the following we examine some factors that may induce behavior that will bring the results closer to the model's predictions.

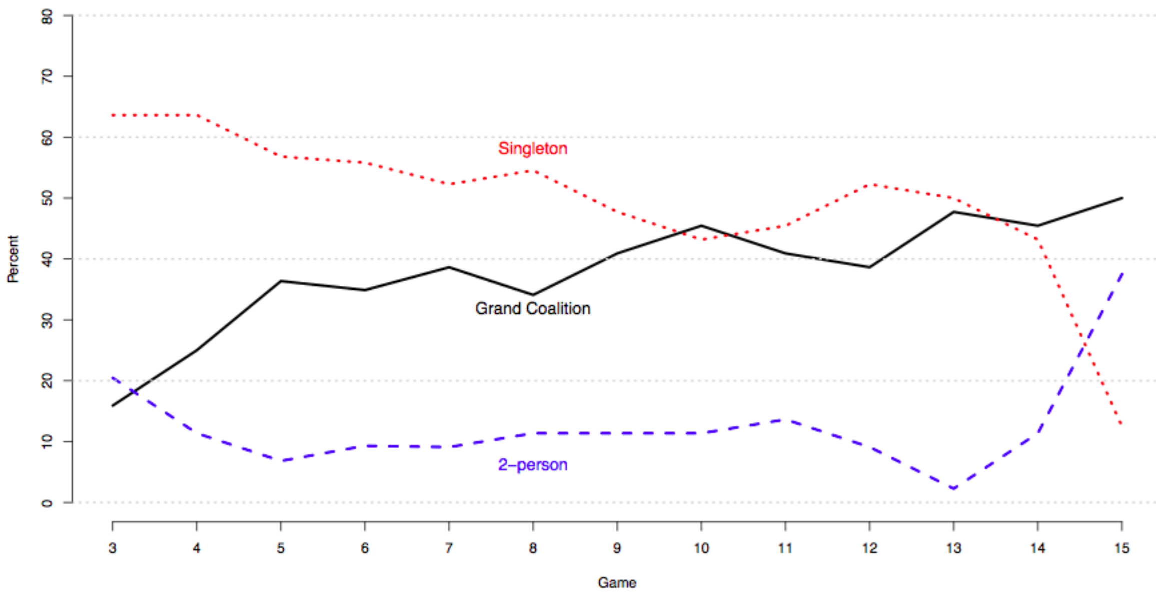

Figure 4.3: Percentage of final coalition through games, by type

In Figure 4.3 we notice that the proportion of grand coalitions steadily increases while that of singletons decreases. At the very end of the sessions, about half of formed coalitions are grand coalitions. This suggests that as agents learn how the game works, the more they tend to implement the grand coalition. The proportion of 2-person coalitions forming remain relatively steady throughout the games, except for the final game where the proportion increased to 40%.

In this paper we designated games 9 onwards as games in "late" time periods where we suppose that the subjects have gained some knowledge of the experiment. The first set of bars in Figure 4.4 shows the mean proportion of grand coalitions forming in early games (games 3-8) and late games (games 9 onwards) of the different sessions. The mean proportion of grand coalition in late games is around 40% while the mean proportion for early games is only 30%. This difference is statistically significant using sessions as the unit of observation (p-value= .013, Wilcoxon signed rank test).

The next set of results shows the effect of higher stakes.

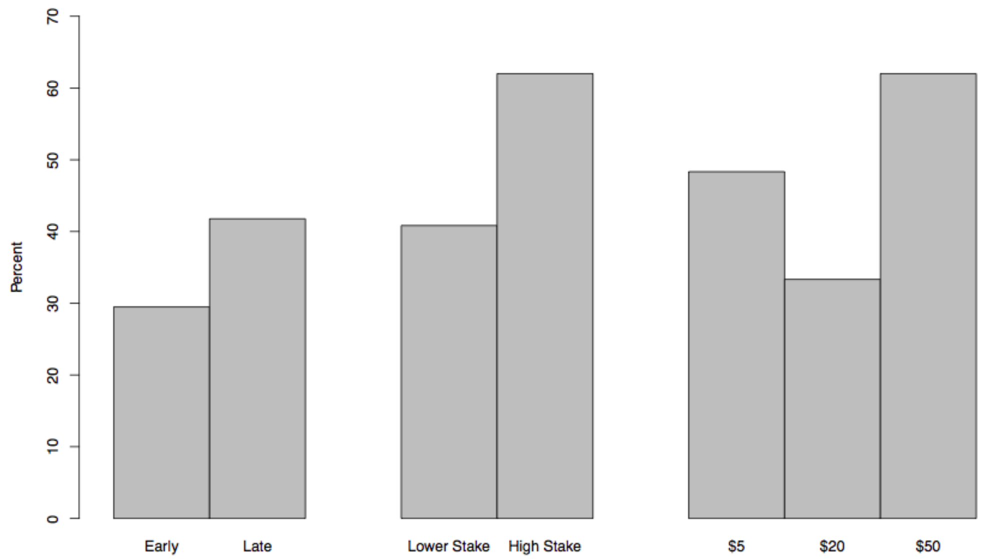

Figure 4.4: Percentage of grand coalitions forming, by time period and last game stake

Observation 2 *The difference in the proportion of grand coalitions in high stakes versus low and medium stakes is substantial and significant.*

The middle and rightmost set of bars in Figure 4.4 shows the percentage of grand coalition formation in the last game of the sessions by stake. In the rightmost set of bars, we see that when agents are offered a pot of $50, close to 70% of coalitions that formed were grand coalitions. However, when the pot was only $20, about 30% of the coalitions were grand, which is lower than the proportion of grand coalitions in the low stake games. In the middle set of bars in Figure 4.4, we lumped together the $20 and $5 into the lower-stakes treatment and calculated the proportion of grand coalitions to total formed coalitions in the last game of the sessions. Using the sessions as the unit of observations, we see that the difference between the proportion of grand coalitions forming in the low and high stakes is substantial (40% vs 62%, respectively) but not statistically significant (p-value$= 0.26$, Mann-Whitney test). We also tested the differences in mean proportions of grand coalitions forming for the last game vs the late games (except the last) for both the lower stakes and the higher stakes. The differences in the proportion of grand coalitions forming in the last game vs late games for lower stakes is statistically insignificant using sessions as the unit of observation (p-value> 0.10, Wilcoxon signed rank test). For the high stake games, this difference is statistically significant (p-value$= 0.10$, Wilcoxon signed rank test).

To provide an alternative (higher-powered) test to determine the effect of stakes on the formation of grand coalitions, we used a logit model (Table 4.2) with specifications that include late games as the sole control, stakes as the sole control, late games and stakes together, and a specification with session fixed effects. We find that the coefficients of the \$50-stake games are robustly positive and statistically significant, showing that increasing the stake size substantially will have a positive effect on the propensity of agents to form grand coalitions.

Table 4.2: Probability of grand coalition forming

	Dependent variable:			
	Grand coalition formed = 1			
	(1)	(2)	(3)	(4)
Late Games	0.832***		0.762***	0.840***
	(0.159)		(0.160)	(0.161)
Last game, \$5 stake		0.774	0.336	0.199
		(0.558)	(0.547)	(0.466)
Last game, \$20 stake		0.081	−0.357	−0.494
		(0.320)	(0.387)	(0.384)
Last game, \$50 stake		1.468***	1.030*	1.467***
		(0.564)	(0.536)	(0.465)
Constant	−1.099***	−0.774***	−1.099***	−0.797***
	(0.144)	(0.154)	(0.144)	(0.089)
Observations	1,872	1,872	1,872	1,872
Log Likelihood	−1,152.266	−1,171.191	−1,144.793	−1,071.061
Akaike Inf. Crit.	2,308.532	2,350.383	2,299.585	2,172.122
Session FE	No	No	No	Yes

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses are clustered at the session level (11 session-clusters).
Base category for stakes are the games before the last game of the sessions.

4.5.1 Analysis of strategies

With such low proportions of grand coalitions forming, it would be instructive to examine the strategies employed by the players. As a first cut, we examine whether subjects are playing myopic strategies and how they affect the proportion of grand coalitions forming. We regard myopic strategies as proposing a two-person coalition to an agent who has a higher power than him or accepting a two-person coalition with someone who has a higher power than him. In the context of our experiment, we tag a strategy as myopic, for instance, if the agent who has power 20 proposes $\{20, 35\}$ or $\{20, 45\}$ or if the agent with power 35 proposes $\{35, 45\}$. In addition, a strategy is deemed myopic if the agent with power 20 accepts to form either $\{20, 35\}$ or $\{20, 45\}$ or if agent with power 35 accepts to form $\{35, 45\}$. We consider alternative definitions of myopic agents:

- Agents that played myopic strategies in half their games (Myopia 1)
- Agents that played myopic strategies in $\frac{1}{3}$ of their games (Myopia 2)
- Agents that played myopic strategies in half of their latter games (Myopia 3)
- Agents that played myopic strategies at least once in their latter games (Myopia 4)

The presence of myopic agents drastically reduces the proportion of grand coalitions. Figure 4.5 shows this with our various definitions of myopic players. Without the presence of myopic players, the proportion of grand coalition can be as high as around 47% if we define myopic players as those who employed myopic strategies at least once in latter games (Myopia 4). Using sessions as unit of observation, except for Myopia 1 the differences in proportions are statistically significant (Wilcoxon signed rank test, p-value$= 0.19$ [Myopia 1], p-value$< .01$ [Myopia 2], p-value$< .01$ [Myopia 3], p-value$< .01$ [Myopia 4]).

In order to flesh out the actual strategies that our subjects employ, we now look at both the proposal and response stages *at the first round of every game*, that is, we look at the proposals and the responses of each agent every time a new game starts.

Observation 3 *The proposed coalition is dependent on an agent's power draw. The lower-powered agent (20) proposes the grand coalition more frequently whereas the higher-powered agents (35 and 45) propose a 2-person coalition with the agent with the lowest power (20) more frequently. This is consistent with agents playing their empirical best response which maximizes their expected payoff given the presence of myopic players.*

Table 4.3 shows the percentage of proposed coalitions by power. While the model predicts that the grand coalition $\{20, 35, 45\}$ forms regardless of the drawn power of the agent proposing, it seems that agents propose coalitions that benefits them in the short run. For instance, for agents that have drawn powers 35 and 45, a large proportion propose a coalition with the agent with the lowest power, that is, agent 20. About 73% of proposals for players who draw power 35 is the coalition

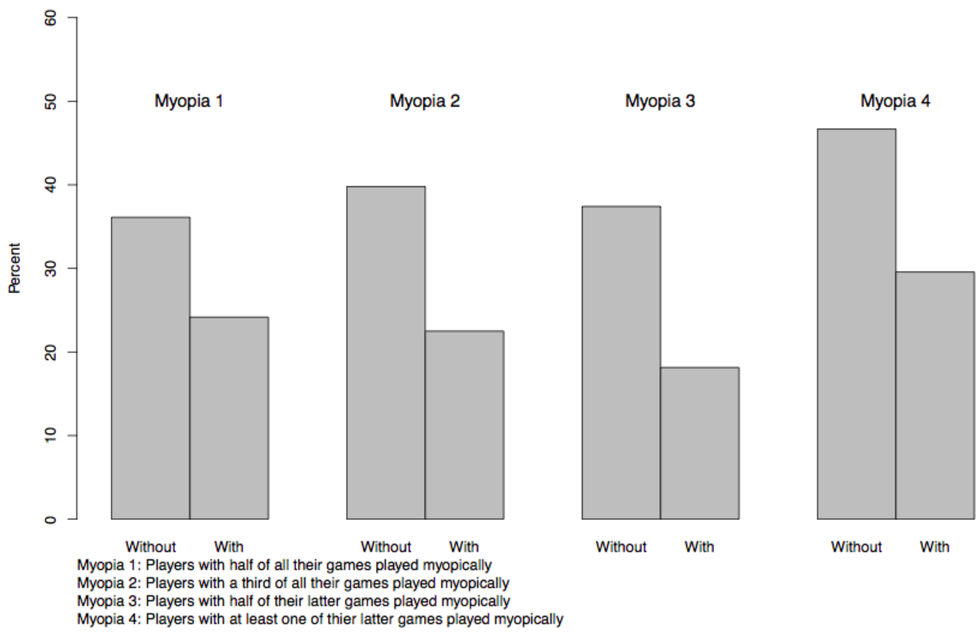

Figure 4.5: Percentage of grand coalitions forming with and without myopic players

$\{20, 35\}$ and 40% of proposals of agents who have drawn power 45 is $\{20, 45\}$. On the other hand, we see that if an agent draws power 20, he proposes the grand coalition $\{20, 35, 45\}$ 60% of the time. About 34% of the time the power-20 agent proposes a 2-person coalition with agent 35.

Table 4.3: Percentage of proposed coalition at the start of every game, by agent's power

	$\{20, 35, 45\}$	$\{20, 35\}$	$\{20, 45\}$	$\{35, 45\}$
Power = 20	60	34	6	-
Power = 35	20	73	-	7
Power = 45	33	-	40	27

Note: Rows sum to 100%

Table 4.4 shows the results of these proposals in detail. Line (1) of the Table shows the possible coalitions that can be proposed. For instance, for the agent who have drawn power 20 the possible coalitions he can propose is $\{20, 35, 45\}$, $\{20, 35\}$ or $\{20, 45\}$. Line (2) shows the percentage of the proposed coalition to the total number of proposals made by the agent drawing a particular power. Hence, coalition $\{20, 35, 45\}$ comprise 60% of the total proposals made by agent 20. Line (3) indicates whether the proposed coalition passed (i.e., a majority of agents voted "yes" for this proposed coalition) or failed (i.e., the proposer failed to muster enough "yes" votes), and line (4) shows the percentage of "pass" or "fail" for such coalition proposals.[9] After determining whether a proposed coalition passed or failed, Line (5) lists the possible final coalitions that can form. For instance, if coalition $\{20, 35\}$ passes for agent 20, the possible final coalitions can either be $\{20, 35\}$ (B) or the singleton $\{35\}$ (E). On the other hand, if $\{20, 35\}$ fails, then the possible final coalitions will be the whole gamut of coalitions that can form, that is, $\{20, 35, 45\}$, $\{20, 35\}$, $\{20, 45\}$, $\{35, 45\}$, $\{35\}$ or $\{45\}$ (See the footnote of the table for the labels of coalitions A-F). Conditional on the coalition passing or failing, Line (6) then shows the percentage of each of the possible final coalitions. For example, if $\{20, 35\}$ passes for agent 20, 79% of the time it degenerates into the dictator coalition $\{35\}$. Line (7) shows the final payoff if the specified coalition forms. So for agent 20, coalition $B = \{20, 35\}$ gives the agent $\frac{20}{20+35} \cdot 5 = \$1.82$. Finally, Line (8) shows the expected payoffs of each of the proposal strategies by weighting each final payoff with the appropriate proportions. Tables 4.3 and 4.4 present the same information for early and late games, respectively.

We can see in Table 4.4 that proposing to a lower-powered agent 20 may be consistent with the possibility of the strategies being an empirical best response given that agent 20 has a positive probability of behaving myopically (that is, accepting a 2-person coalition when he has the lower power). To investigate whether this strategy is indeed an empirical best response, we computed the expected payoffs of the agents given the empirical probabilities obtained in our dataset. In Table 4.4 we see that for agents who draw power 35 or 45, the strategy that maximizes expected payoff is to propose to the agent with the lowest power (i.e., agent 20). By proposing this coalition, agent 35 for instance will get \$2.84 in expected payoff, which is higher than the expected payoff obtained when he proposes something else. Similarly, agent 45 obtains his maximum expected payoff by proposing $\{20, 45\}$ given the possibility that the proposal will lead to him being a dictator. Playing this strategy yields a higher expected payoff during early games, when agents are still learning the experiment. In early games, proposing a 2-person coalition to the agent with the least power and becoming a dictator is even more lucrative, where the expected payoff is higher than the rest of the possible scenarios the proposer may face. For instance, a 35-powered agent will get \$3.11 in expected payoffs by proposing $\{20, 35\}$. As agents learn more about how to play the experiment in late games, the expected payoff of proposing to the lowest-powered agent diminishes in late games (Table 4.6). On the other hand, agent 20 maximizes expected payoff by proposing the grand coalition.

[9]Note that if the grand coalition is proposed, the coalition always passes since it is the default starting coalition when a majority of responders voted "no" for it.

Figure 4.6 shows the evolution of these proposals throughout the games. We can see that for agents with power 20 the proportion of the grand coalition being proposed is gradually rising. This may be an indication that learning is occurring, such that these agents comprehend that proposing the grand coalition will likely not exclude them from the final coalition. The trend of increasing proportion of grand coalition proposals is also true for agent with 35 and 45, although statistical test shows that the the increase is milder for agent 35 (see the logit model in Table 4.7). This is consistent with the results in Tables 4.5 and 4.6. There we can see that for agent 35, proposing the 2-person coalition with the lower-powered agent remains to be an empirical best response which provides him with the highest expected payoff among all his proposal strategies. On the other hand, for agent 45 we can see that in late games the gap in expected payoff from proposing the 2-person coalition with agent 20 and proposing the grand coalition becomes narrower. For agent 45, the expected payoffs of proposing $\{20, 45\}$ vs. $\{20, 35, 45\}$ is \$2.81 vs \$1.05 in early games and \$2.28 vs \$1.65 in late games. For agent 20, the expected payoff from proposing the grand coalition increases from \$0.59 in early games to \$0.66 in late games.

Proposer = 20																						
(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 35\}$								$\{20, 45\}$							
(2) Percent of Total Proposals	60						34								6							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						92		8						100		0					
(5) Final Coalition (see Note)	A	B	C	D	E	F	B	E	A	B	C	D	E	F	C	F	A	B	C	D	E	F
(6) Frequency (Percent)	57	4	0	1	25	13	21	79	40	20	0	20	0	20	8	92	0	0	0	0	0	0
(7) Final Payoff ($)	1	1.82	1.54	0	0	0	1.82	0	1	1.82	1.54	0	0	0	1.54	0	1	1.82	1.54	0	0	0
(8) Expected payoff ($)	0.64						0.42								0.12							
Proposer = 35																						
(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 35\}$								$\{35, 45\}$							
(2) Percent of Total Proposals	20						73								7							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						39		61						92		8					
(5) Final Coalition (see Note)	A	B	C	D	E	F	B	E	A	B	C	D	E	F	D	F	A	B	C	D	E	F
(6) Frequency (Percent)	55	0	3	0	17	25	30	70	70	4	0	3	8	15	0	100	0	0	0	0	0	100
(7) Final Payoff ($)	1.75	3.18	0	2.19	5	0	3.18	5	1.75	3.18	0	2.19	5	0	2.19	0	1.75	3.18	0	2.19	5	0
(8) Expected payoff ($)	1.80						2.84								0							
Proposer = 45																						
(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 45\}$								$\{35, 45\}$							
(2) Percent of Total Proposals	33						40								27							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						39		61						35		65					
(5) Final Coalition (see Note)	A	B	C	D	E	F	C	F	A	B	C	D	E	F	D	F	A	B	C	D	E	F
(6) Frequency (Percent)	47	0	0	0	46	7	25	75	45	14	0	0	34	7	12	88	0	0	0	0	0	0
(7) Final Payoff ($)	2.25	0	3.46	2.81	0	5	3.46	5	2.25	0	3.46	2.81	0	5	2.81	5	2.25	0	3.46	2.81	0	5
(8) Expected payoff ($)	1.40						2.62								2.49							

Note: Final Coalitions labels are $A = \{20, 35, 45\}$, $B = \{20, 35\}$, $C = \{20, 45\}$, $D = \{35, 45\}$, $E = \{35\}$, $F = \{45\}$. See the text for the explanation of each of the rows.

Table 4.4: Expected payoffs by proposal strategy for each power draw, all games

Proposer = 20																						
(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 35\}$								$\{20, 45\}$							
(2) Percent of Total Proposals	41						48								11							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						95		5						100		0					
(5) Final Coalition (see Note)	A	B	C	D	E	F	B	E	A	B	C	D	E	F	C	F	A	B	C	D	E	F
(6) Frequency (Percent)	54	3	0	0	17	26	23	77	50	50	0	0	0	0	11	89	0	0	0	0	0	0
(7) Final Payoff ($)	1	1.82	1.54	0	0	0	1.82	0	1	1.82	1.54	0	0	0	1.54	0	1	1.82	1.54	0	0	0
(8) Expected payoff ($)	0.59						0.46								0.17							

Proposer = 35																						
(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 35\}$								$\{35, 45\}$							
(2) Percent of Total Proposals	17						73								10							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						47		53						100		0					
(5) Final Coalition (see Note)	A	B	C	D	E	F	B	E	A	B	C	D	E	F	D	F	A	B	C	D	E	F
(6) Frequency (Percent)	58	0	3	0	21	21	22	78	67	10	0	0	6	17	0	100	0	0	0	0	0	0
(7) Final Payoff ($)	1.75	3.18	0	2.19	5	0	3.18	5	1.75	3.18	0	2.19	5	0	2.19	0	1.75	3.18	0	2.19	5	0
(8) Expected payoff ($)	2.07						3.11								0							

Proposer = 45																						
(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 45\}$								$\{35, 45\}$							
(2) Percent of Total Proposals	24						48								28							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						38		62						41		59					
(5) Final Coalition (see Note)	A	B	C	D	E	F	C	F	A	B	C	D	E	F	D	F	A	B	C	D	E	F
(6) Frequency (Percent)	38	0	0	0	58	4	6	94	52	17	0	0	24	7	18	82	50	6	0	0	44	0
(7) Final Payoff ($)	2.25	0	3.46	2.81	0	5	3.46	5	2.25	0	3.46	2.81	0	5	2.81	5	2.25	0	3.46	2.81	0	5
(8) Expected payoff ($)	1.05						2.81								2.54							

Note: Final Coalitions labels are $A = \{20, 35, 45\}$, $B = \{20, 35\}$, $C = \{20, 45\}$, $D = \{35, 45\}$, $E = \{35\}$, $F = \{45\}$. See the text for the explanation of each of the rows.

Table 4.5: Expected payoffs by proposal strategy for each power draw, early games

Proposer = 20

(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 35\}$								$\{20, 45\}$							
(2) Percent of Total Proposals	76						21								3							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						85		15						100		0					
(5) Final Coalition (see Note)	A	B	C	D	E	F	B	E	A	B	C	D	E	F	C	F	A	B	C	D	E	F
(6) Frequency (Percent)	59	4	0	1	29	7	18	82	33	0	0	33	0	34	0	100	0	0	0	0	0	0
(7) Final Payoff ($)	1	1.82	1.54	0	0	0	1.82	0	1	1.82	1.54	0	0	0	1.54	0	1	1.82	1.54	0	0	0
(8) Expected payoff ($)	0.66						0.32								0							

Proposer = 35

(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 35\}$								$\{35, 45\}$							
(2) Percent of Total Proposals	23						73								4							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						33		67						75		25					
(5) Final Coalition (see Note)	A	B	C	D	E	F	B	E	A	B	C	D	E	F	D	F	A	B	C	D	E	F
(6) Frequency (Percent)	55	0	5	0	13	27	39	61	72	0	0	4	9	15	0	100	0	0	0	0	0	100
(7) Final Payoff ($)	1.75	3.18	0	2.19	5	0	3.18	5	1.75	3.18	0	2.19	5	0	2.19	0	1.75	3.18	0	2.19	5	0
(8) Expected payoff ($)	1.63						2.61								0							

Proposer = 45

(1) Proposed Coalition	$\{20, 35, 45\}$						$\{20, 45\}$								$\{35, 45\}$							
(2) Percent of Total Proposals	43						31								26							
(3) Outcome of Response	Pass						Pass		Fail						Pass		Fail					
(4) Outcome by Percent	100						40		60						29		71					
(5) Final Coalition (see Note)	A	B	C	D	E	F	C	F	A	B	C	D	E	F	D	F	A	B	C	D	E	F
(6) Frequency (Percent)	54	0	0	0	37	9	60	40	33	7	0	0	53	7	0	100	34	13	0	0	40	13
(7) Final Payoff ($)	2.25	0	3.46	2.81	0	5	3.46	5	2.25	0	3.46	2.81	0	5	2.81	5	2.25	0	3.46	2.81	0	5
(8) Expected payoff ($)	1.65						2.28								2.44							

Note: Final Coalitions labels are $A = \{20, 35, 45\}$, $B = \{20, 35\}$, $C = \{20, 45\}$, $D = \{35, 45\}$, $E = \{35\}$, $F = \{45\}$. See the text for the explanation of each of the rows.

Table 4.6: Expected payoffs by proposal strategy for each power draw, late games

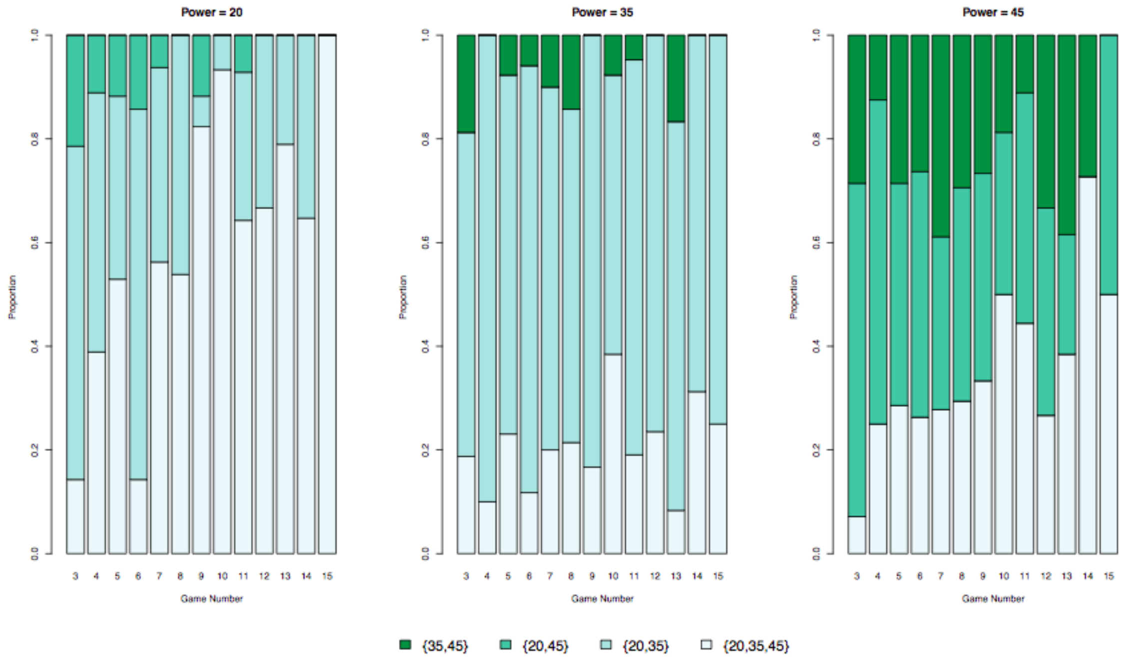

Figure 4.6: Percentage of proposed coalition at the start of every game, by game and agent's power

Observation 4 *Responses are dependent on an agent's drawn power. Higher-powered agents tend to accept 2-person coalitions with an agent with a lower power. A substantial proportion of responses accept the grand coalition when offered.*

Table 4.7: Percentage of responses at the start of every game, by agent's power and proposed coalition

	$\{20, 35, 45\}$		$\{20, 35\}$		$\{20, 45\}$		$\{35, 45\}$	
	Accept	Reject	Accept	Reject	Accept	Reject	Accept	Reject
Power = 20	31	1	17	26	10	15	-	-
Power = 35	51	10	20	2	-	-	6	11
Power = 45	81	5	-	-	7	0	7	0

Note: Rows sum to 100%

Table 4.7 shows some interesting patterns regarding the responses of the agents. First, higher power agents tend to overwhelmingly accept a 2-person coalition with an agent with a lower power. For instance, agent 35 tend to almost always accept $\{20, 35\}$ and agent 45 always accepts $\{20, 45\}$ and $\{35, 45\}$. This strategy may be consistent with both subgame perfect equilibrium strategy (off the equilibrium path) and also exhibiting other-regarding behavior (Bolton and Ockenfels [17]), in the sense that they care for the lower-powered agent to get a higher share than staying at the grand coalition. However, the latter may not be true since as Figure 4.3 shows, the proportion of 2-person final coalition forming seem to be relatively low. Hence, the 2-person coalition will more likely to deviate into a singleton rather than stay at the 2-person coalition. In Table 4.2, for instance, if agent 20 proposed $\{20, 35\}$ and it passes, 79% of the time it degenerates into a singleton coalition where 35 is the dictator.

Another interesting pattern is that agent who draw power 45 face a proposed grand coalition 86% of the time but the agent with power 35 only face the grand coalition around 61% of every first round proposals. This is indicative of a reluctance by the lower powered agents to include 45 in a 2-person coalition, since they know that it could lead to two things: first, they could get a relatively lower payoff compared to a 2-person coalition with a lower-powered agent and second, they know that agent 45 can kill them. This behavior may be consistent with both empirical expected payoff maximization and farsighted behavior.

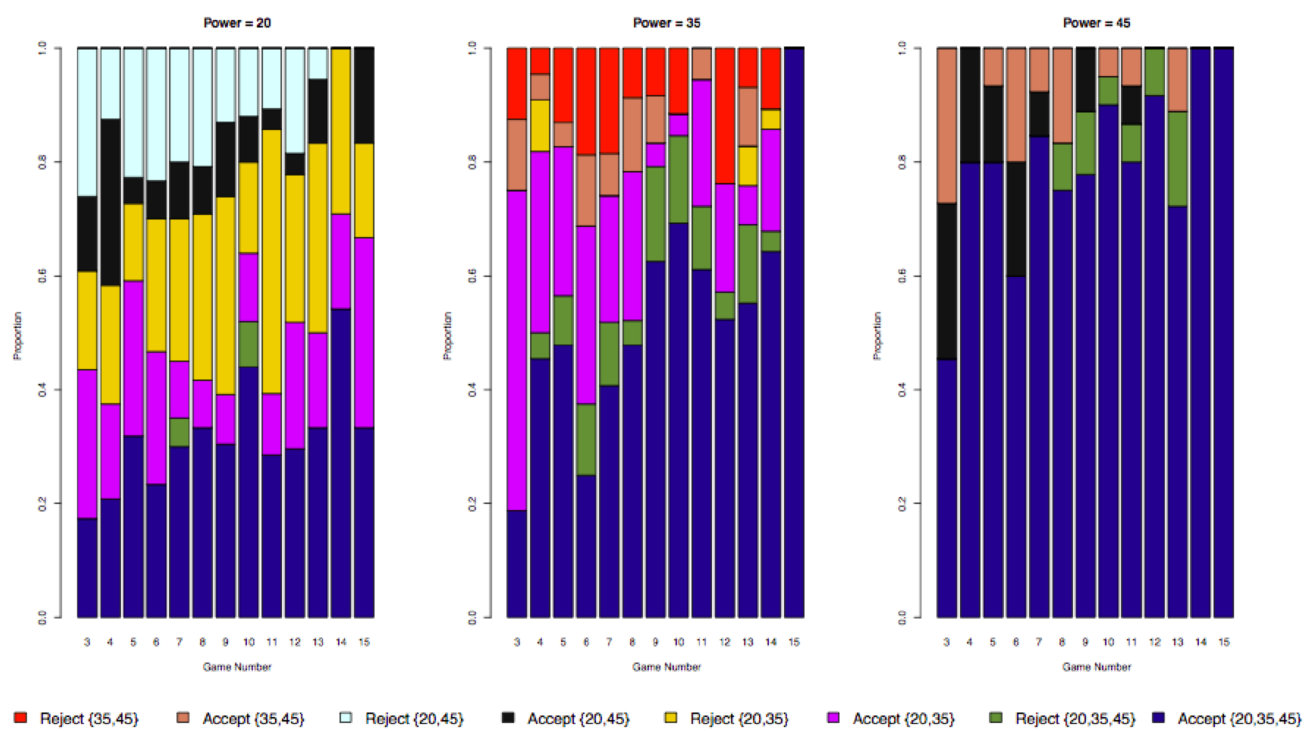

Figure 4.7: Percentage of responses at the start of every game, by game and agent's power

Figure 4.7 shows how these responses evolve throughout the games. For agent with drawn power 20, the proportion of rejecting 2-person coalitions increases throughout games and at the same time the proportion of accepting the grand coalition is also increasing (see Table 4.10). For both agents drawing power 35 and 45, the proportion of accepted grand coalitions increases over time. Agent 45 almost always face a proposed grand coalition, and most of them are accepted.

In order to assess the probability of proposing the grand coalition, we employ a logit analysis that takes into account an agents power as well as several possible scenarios of immediate past experience. These experience variables are (1) agents have no prior experience in proposing a coalition when currently chosen to be a proposer; (2) proposed a 2-person coalition in the last time the agent was chosen to be a proposer and being in the final coalition that formed during that time; (3) proposed a 2-person coalition in the last time the agent was chosen to be a proposer but being out of the final coalition that formed during that time; (4) proposed the grand coalition in the last time the agent was chosen to be a proposer and being in final coalition that formed during that time; and (5) proposed the grand coalition in the last time the agent was chosen to be a proposer but being out of the final coalition that formed during that time.

We ran five specifications: (1) with just the agents' power as controls; (2) with an interaction of power and late games; (3) with just prior experience as controls; (4) both power and experience as controls; and (5) power, experience, stakes, and late game interactions as controls.

Table 4.8 provides corroborating evidence to the analysis we have presented earlier. First, agents with relatively higher power (i.e., agents 35 and 45) have a lower probability of proposing the grand coalition, and this result is robust throughout all specifications. Another robust finding is that when agents propose the grand coalition when he was last a proposer, he will more likely propose the grand coalition if chosen as a proposer again. This is shown by the positive and statistically significant coefficients on the "Proposed grand" variable throughout all specifications. The interaction with late games shows up positive and significant, echoing our prior result that more grand coalitions are proposed during late games.

We likewise performed a similar analysis with response strategies. We first examined the probability of accepting the grand coalition conditional on being proposed a grand coalition. In these runs we also account for immediate past experience in responding to a proposed coalition. These possible scenarios are: (1) agents have no prior experience in responding when currently chosen to be a responder; (2) accepting a 2-person coalition in the last time that an agent was chosen to be a responder and being in the final coalition that formed during that time; (3) accepting a 2-person coalition in the last time that an agent was chosen to be a responder but being out of the final coalition that formed during that time; (4) rejecting a 2-person coalition in the last time that an agent was chosen to be a responder and being in the final coalition that formed during that time; (5) rejecting a 2-person coalition in the last time that an agent was chosen to be a responder but being out of the final coalition that formed during that time; (6) accepting the grand coalition in the last time that an agent was chosen to be a responder and being

Table 4.8: Probability of proposing the grand coalition

	Dependent variable:				
	Grand coalition proposed = 1				
	(1)	(2)	(3)	(4)	(5)
Power = 35	−1.743***	−1.194***		−1.961***	−1.259***
	(0.249)	(0.200)		(0.239)	(0.194)
Power = 45	−1.102***	−0.769***		−1.272***	−0.800***
	(0.281)	(0.236)		(0.256)	(0.219)
Late games		1.512***			1.973***
		(0.377)			(0.431)
Power = 35 x Late games		−1.160***			−1.482***
		(0.438)			(0.567)
Power = 45 x Late games		−0.659			−0.943**
		(0.487)			(0.382)
Proposed 2-person, in final			0.236	0.362	0.120
			(0.308)	(0.344)	(0.417)
Proposed 2-person, out of final			0.277	0.271	0.268
			(0.256)	(0.309)	(0.513)
Proposed grand, in final			1.517***	1.865***	1.667***
			(0.184)	(0.196)	(0.494)
Proposed grand, out of final			1.327***	1.483***	0.830
			(0.297)	(0.346)	(0.514)
$5 stake					0.044
					(0.427)
$20 stake					0.744
					(0.475)
$50 stake					1.215**
					(0.489)
Constant	0.392**	−0.357*	−1.019***	−0.134	−0.606**
	(0.179)	(0.198)	(0.222)	(0.277)	(0.290)
Experience/Late Game Interaction	No	No	No	No	Yes
Observations	535	535	535	535	535
Log Likelihood	−324.462	−308.903	−334.516	−299.883	−287.569
Akaike Inf. Crit.	654.923	629.806	679.032	613.766	609.137

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses are clustered at the session level (11 session-clusters).
Base category for stakes are the games before the last game of the sessions.

in the final coalition that formed during that time; (7) accepting the grand coalition in the last time that an agent was chosen to be a responder but being out of the final coalition that formed during that time; (8) rejecting the grand coalition in the last time that an agent was chosen to be a responder and being in the final coalition that formed during that time; and (9) rejecting the grand coalition in the last time that an agent was chosen to be a responder but being out of the final coalition that formed during that time.

The first specification in Table 4.9 show that while the coefficient is negative for agents 35 (significant) and 45 (not significant), the intercept is still sufficiently large such that there is still a positive probability in accepting a grand coalition. This is consistent with the findings in Table 4.7 where agents, when faced with a proposed grand coalition, tend to accept it. The experience variables do not seem to be robust, but some results are worth mentioning. First, the "Accept 2 person-out" may signify that agent may be willing to take risk with being in a 2-person coalition for the chance at a higher payoff. Second, the "Reject grand, in" decreases the probability of accepting the grand coalition, suggesting a form of hysteresis in the response strategies.

Finally, we examine the probability that the agents will accept a 2-person coalition conditional on being proposed a 2-person coalition. We likewise control for the same experience variable as in the case where we analyzed the probability of accepting the grand coalition. The difference is here we split the agent with power 35 into two scenarios, one when offered the coalition $\{20, 35\}$ and the other when offered the coalition $\{20, 45\}$ to account for the possible differences in strategies when faced with these different coalitions. In Table 4.10 we see that drawing a higher power robustly increases the probability that agents will accept a two person coalition. This effect is particularly strong for agent 45, since he can always propose to be a dictator once a two person coalition that includes him forms. The other robust variable shows that rejecting the grand coalition in the last time he was a responder makes an agent increase his probability that a 2-person coalition will be accepted. This also suggests some consistency and hysteresis in the behavior of these agents.

4.5.2 Typology of agents

In the previous subsection we looked at the type of strategies that were played in the experiment. Here we will attempt to categorize subjects into several types depending on the strategies they play. Note that the subjects may play different types of strategies throughout the games, and in this exercise we will attempt to assign subjects based on a scoring method called quadratic deviation measure (QDM) (see Nagel and Fang [68] as well as Selten [93] for an axiomatic

Table 4.9: Probability of accepting a grand coalition

	Dep. variable: Accepting the grand coalition=1				
	(1)	(2)	(3)	(4)	(5)
Power = 35	-1.686^{**}	-1.896^{*}		-1.609^{**}	-2.459^{**}
	(0.745)	(0.998)		(0.801)	(1.124)
Power = 45	-0.715	0.260		-0.585	0.179
	(0.768)	(0.161)		(0.863)	(0.280)
Late games		-0.297			15.787^{***}
		(0.784)			(1.386)
Power = 35 x Late games		0.331			1.164
		(0.838)			(1.073)
Power = 45 x Late games		-1.188			-0.653
		(0.896)			(0.899)
Accept 2-person, in final			-1.386	-1.259	-1.559
			(0.959)	(1.036)	(1.233)
Accept 2-person, out of final			-2.079^{**}	-1.906^{**}	-0.831
			(0.821)	(0.931)	(1.217)
Reject 2-person, in final			-1.144	-1.107	-1.827
			(0.912)	(0.934)	(1.221)
Reject 2-person, out of final			-1.553	-1.653	16.662^{***}
			(1.048)	(1.025)	(0.712)
Accept grand, in final			1.602	1.684	16.603^{***}
			(1.250)	(1.321)	(0.871)
Accept grand, out of final			-0.907	-0.763	16.574^{***}
			(0.658)	(0.705)	(1.134)
Reject grand, in final			-3.104^{***}	-3.080^{***}	-2.953
			(0.938)	(0.964)	(2.775)
Reject grand, out of final			-1.649	-1.480	17.440^{***}
			(1.377)	(1.323)	(2.022)
$5 stake					17.569^{***}
					(1.150)
$20 stake					18.185^{***}
					(1.222)
$50 stake					0.804
					(1.351)
Constant	3.423^{***}	3.611^{***}	3.258^{***}	4.220^{***}	4.584^{***}
	(0.675)	(0.945)	(0.666)	(0.846)	(1.179)
Experience/Late Game Interaction	No	No	No	No	Yes
Observations	406	406	406	406	406
Log Likelihood	-117.490	-116.122	-103.694	-98.239	-87.699
Akaike Inf. Crit.	240.979	244.243	225.387	218.477	225.398

Note: *p$<$0.1; **p$<$0.05; ***p$<$0.01. Standard errors in parentheses are clustered at the session level (11 session-clusters).
Base category for stakes are the games before the last game of the sessions.

Table 4.10: Probability of accepting a two person coalition

	Dep. variable: Accepting two-person coalition =1				
	(1)	(2)	(3)	(4)	(5)
Power = 35, $\{20, 35\}$	2.876***	3.297***		2.963***	3.341***
	(0.727)	(0.860)		(0.794)	(0.934)
Power = 35, $\{35, 45\}$	-0.141	-0.048		-0.178	-0.041
	(0.290)	(0.505)		(0.282)	(0.576)
Power = 45	18.026***	17.893***		18.044***	18.004***
	(0.463)	(0.479)		(0.473)	(0.467)
Late games		-0.288			0.451
		(0.360)			(1.255)
Power = 35, $\{20, 35\}$ x Late games		-0.948			-0.596
		(0.982)			(1.159)
Power = 35, $\{35, 45\}$ x Late games		-0.254			-0.470
		(0.928)			(0.956)
Power = 45 x Late games		0.288			-0.261
		(0.773)			(0.663)
Accept 2-person, in final			-0.285	-0.311	-0.319
			(0.311)	(0.328)	(0.467)
Accept 2-person, out of final			-0.144	0.165	-0.057
			(0.271)	(0.334)	(0.401)
Reject 2-person, in final			-1.534^{***}	-1.724^{***}	-1.185
			(0.379)	(0.455)	(0.745)
Reject 2-person, out of final			-0.818^{*}	-0.755^{*}	-0.838
			(0.457)	(0.446)	(1.018)
Accept grand, in final			-0.814^{***}	-0.660^{***}	-0.558
			(0.258)	(0.200)	(0.403)
Accept grand, out of final			-0.859^{**}	-0.534^{**}	-0.376
			(0.338)	(0.239)	(0.574)
Reject grand, in final			-0.076	-0.056	0.141
			(0.715)	(0.773)	(1.410)
Reject grand, out of final			13.930***	17.575***	17.173***
			(0.688)	(0.718)	(1.337)
\$5 stake					-1.488^{***}
					(0.529)
\$20 stake					-0.781
					(1.680)
\$50 stake					0.691
					(0.643)
Constant	-0.460^{**}	-0.327^{*}	0.636***	-0.009	-0.058
	(0.203)	(0.184)	(0.195)	(0.190)	(0.196)
Experience/Late Game Interaction	No	No	No	No	Yes
Observations	332	332	332	332	332
Log Likelihood	-181.300	-179.578	-218.975	-171.540	-170.199
Akaike Inf. Crit.	370.600	375.157	455.949	367.080	384.398

Note: *p$<$0.1; **p$<$0.05; ***p$<$0.01. Standard errors in parentheses are clustered at the session level (11 session-clusters).
Base category for stakes are the games before the last game of the sessions.

analysis of the method). What this measure does is to take the quadratic difference between the actual choice vector of a subject in a potential move and the choice vector predicted by the model. Let $c_{i,\alpha(m)}(t) = [c_{i,\alpha(m)}(1), \ldots, c_{i,\alpha(m)}(T)]$ be the strategy choices that an agents make in round t given a model m, whether he's a proposer or a responder. Note that the choice can be any coalition if he's a proposer, or whether to accept or reject the coalition if he's a responder. Specifically, each element in this choice vector takes on the value

$$c_{i,\alpha(m)}(t) = \begin{cases} 1 & \text{if strategy } \alpha \text{ prescribed by model m is chosen in round t} \\ 0 & \text{otherwise} \end{cases}$$

Here, we choose three types of models. First, the **equilibrium model** where agents play the subgame perfect equilibrium strategies. In our setting, this corresponds to subjects proposing $\{20, 35, 45\}$ and nothing else, and accepting $\{20, 35, 45\}$. Also part of the SPNE strategy is to accept a 2-person coalition whenever an agent has the higher power. Hence, the choice vector $c_{i,\alpha(eq)}(t)$ will get 1 in the t^{th} index if he is a proposer and he proposes $\{20, 35, 45\}$ or if he is a responder and he accepts $\{20, 35, 45\}$ or accepts a 2-person coalition if he has the higher power. Otherwise, the vector registers a zero in the t^{th} index. The second type of model is the **myopic model**. In this model, agents play myopic strategies where lower-powered agents propose a 2-person coalition with a higher-powered agent when it's his turn to propose, or accept a 2-person coalition with a higher-powered agent when it is time to respond. Here, the choice vector $c_{i,\alpha(myo)}(t)$ will get 1 in the t^{th} index if, for instance, a subject has power 20 and he proposes $\{20, 35\}$ or $\{20, 45\}$. If agent 20 is a responder and he accepts $\{20, 35\}$ or $\{20, 45\}$ then he also gets a 1 on the t^{th} index. The third type of model is where subjects play their **empirical best response**. Here, if an agent is a proposer the he will propose to a lower-powered agent or propose himself if a 2-person coalition has formed. If he is a responder, he will accept a 2-person coalition from a lower-powered agent or reject a 2-person coalition if he has a lower power. Here, the choice vector $c_{i,\alpha(br)}(t)$ will get 1 in the t^{th} index if, for instance, a subject has power 45 and he proposes $\{20, 45\}$ or $\{35, 45\}$. If agent 45 is a responder and he accepts $\{20, 45\}$ or $\{35, 45\}$ then he also gets a 1 on the t^{th} index. We have shown that this strategy is consistent to maximizing the expected payoffs given the empirical probabilities given in Tables 4.4 to 4.6.

Let $p_{i,\alpha(m)}(t) = [p_{i,\alpha(m)}(1), \ldots, p_{i,\alpha(m)}(T)]$ be the vector of predicted choices given a model. This vector will register a 1 on the t^{th} index if there is an opportunity to play a strategy prescribed by a model. For instance, in the best response model if you have power 35 or 45 and chosen to be a proposer in round 1, then you have the opportunity to play a best response strategy (i.e., propose to a lower-powered agent) and thus a 1 registered in the first index if you are a proposer. Hence,

$$p_{i,\alpha(m)}(t) = \begin{cases} 1 & \text{if strategy } \alpha \text{ can be played in round t} \\ 0 & \text{otherwise} \end{cases}$$

The quadratic deviation measure for a particular agent i in model m is thus given by:

$$QDM_{i,m} = \frac{\sum_t^T (c_{i,\alpha(m)}(t) - p_{i,\alpha(m)}(t))^2}{T} \tag{4.1}$$

Clearly, a $QDM_{i,m}$ closer to zero indicates a high proportion of strategies employed by the agents coinciding with those prescribed by model m. We then categorize the subjects into equilibrium players, myopic players, or empirical best response players by simply taking the minimum of QDMs for the three models for each subject i, that is,

$$type_i = \begin{cases} \text{Equilibrium} \quad \text{if } QDM_{i,eq} = \arg\min(QDM_{i,eq}, QDM_{i,myo}, QDM_{i,ebr}) \\ \text{Myopic} \quad \text{if } QDM_{i,myo} = \arg\min(QDM_{i,eq}, QDM_{i,myo}, QDM_{i,ebr}) \\ \text{Empirical Best Response} \quad \text{if } QDM_{i,br} = \arg\min(QDM_{i,eq}, QDM_{i,myo}, QDM_{i,ebr}) \end{cases}$$

Table 4.11 summarizes the subject types across sessions for all games, early games and late games. It is clear that there are very diverse types of agents in the experiment. About 44% of the agents can be classified as “equilibrium players”, 23% of the agents can be classified as “myopic” while 33% of the agents can be classified as mainly playing “equilibrium best response” strategies. There is also a substantial heterogeneity within sessions as some sessions are dominated by “equilibrium” players, such as the February 4, 2015, AM session and some sessions are dominated by myopic players, such as the June 23, 2014 session.

An interesting pattern is the decrease in number of myopic agents, as well as the increase in the number of “equilibrium” agents, when we move to late games. For instance, in early games 29% of the agents are classified as myopic but this percentage declines to 19% in late games. The proportion of “equilibrium” agents increases from 41% in early games to 55% in late games. This echoes the recurring finding that agents seem to learn the more they play our coalition formation game. This increase in the number of equilibrium subjects and the decrease of myopic subjects are probably the reasons why the proportion of grand coalitions forming in the late rounds is increasing.

4.6 Conclusion

The debate between AES and Tullock highlights how nuances in assumptions of human behavior affect how coalitions form and evolve. Dictatorships will be more common if agents behave myopically–that is, if agents do not take into account further possible ramifications of joining a particular group. If agents are farsighted, as in the AES case, the higher the chance that groups do not degenerate into a dictatorship. Our results in this paper suggest that agents' behavior is more complicated than what the dichotomy of the debate alleges. Indeed, we actually observe a lot of rational behavior, with some agents taking advantage of myopic agents especially in early games when agents do not have much experience playing. This behavior naturally leads to a proliferation of dictatorships. However, given sufficient time the agents learn how to behave more strategically and rationally, therefore avoiding being part of a coalition where they could be sidelined in the future.

The evidence in this paper has yielded some important insights. First, the agents' strategies are contingent on the power that they have drawn. In particular, many agents seem to be playing their empirical best response to the possibility that the other agents behave myopically. For instance, higher powered agents seem to propose the coalition with agent 20 in it. This is consistent with the behavior that agents maximize their expected payoff by taking advantage of the probability that the agent 20 may play myopically and accept the coalition which could then lead the higher-powered agent to form a dictatorship. Second, there is some hysteresis in the way agents are playing their strategies. For instance, agents who tend to accept a 2-person coalition or reject the grand coalition in the last time they were a responder increases the probability that they would accept a 2-person coalition when he is tasked to respond to such a coalition. This is an indication of the consistency of behavioral types. In addition, there is some evidence of an positive incentive effect of substantial stakes.

Sessions	Number of Subjects	All games			Early games			Late games		
		Equilibrium	Myopic	Empirical Best Response	Equilibrium	Myopic	Empirical Best Response	Equlibrium	Myopic	Empirical Best Response
May 6, 2014, AM	12	5	2	5	6	2	4	7	2	3
May 6, 2014, PM	15	9	2	4	8	2	5	7	4	4
May 7, 2014	12	7	3	2	5	4	3	8	2	2
June 5, 2014	9	1	4	4	3	4	2	4	3	2
June 6, 2014	12	3	4	5	1	4	5	5	2	5
June 23, 2014	9	0	7	2	3	5	1	0	6	3
June 24, 2014	6	2	1	3	1	2	3	4	1	1
February 4, 2015, AM	15	10	1	4	8	5	2	11	1	3
February 4, 2015, PM	15	9	1	5	8	4	3	11	1	3
February 5, 2015, AM	12	5	4	3	4	4	4	6	2	4
February 5, 2015, PM	15	7	2	6	8	2	5	10	1	4
Overall (Number)	132	58	31	43	55	38	39	73	25	34
Overall (Percent)	100	44	23	33	41	29	30	55	19	26
Min QDM		0	0.12	0	0	0	0	0	0	0
Max QDM		7.31	11.6	7.68	3.26	10.56	3.27	4.76	10.56	5.33
Mean QDM		1.39	3.17	1.40	0.83	1.51	0.65	0.74	1.88	0.89

Table 4.11: Subject type by session for all games, early games and late games

Given our experimental setup it is hard to disentangle strategies that exhibit some of the common behavioral phenomena such as other-regarding behavior. After all, proposing the grand coalition in this experiment is also consistent with having pro-social preferences. However, in the majority of cases there is evidence to suggest that this behavior can be ruled out. For instance, when faced with a 2-person coalition with a lower-powered agent, an overwhelming proportion of these agents will propose himself to be a dictator.[10]

All these results taken together suggest that elements of both Tullock and AES are important to understand the richness and complexity of how people form groups. From this experiment we see that people are rational, albeit through a trial-and-error learning process. Thus, the Tullock effect dominates until learning the game kicks in. The presence of a small number of myopic players is enough to drive rational agents to propose a 2-person coalition. This points out the importance of heterogeneity of behavioral types in the coalition formation process.

4.6.1 Extensions

There are several ways to extend our coalition formation experiment. First, the experiment can investigate how the sharing rules affect how coalitions are formed. In our particular coalition formation game, the theoretical prediction will not change if we change our sharing rule to equal sharing.[11] It is an open question whether the proportion of grand coalitions forming will be different under the two sharing rules.

Another interesting extension is the issue of endogenizing the sharing rule. Apart from the coalition that can form, agents can bargain over the split of the prize as well. This issue is still a loose end for both theory and experiment.

[10] The only case where it does not happen is in the late games when agent 45 proposes $\{20, 45\}$ and if passed, about 60% of the time this 2-person coalition still forms.

[11] Equal sharing or proportional sharing satisfy a property where agents have the same ordinal ranking over coalitions in which they belong. See Jandoc and Juarez [48] for a more detailed discussion of this property they call "consistent ranking".

5 SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

5.1 Summary

Forming groups is a pervasive human phenomenon. Individuals form groups because they derive personal benefits from so doing. Applications of the economics of cooperation–from oligopoly and public goods theories to international agreements to common property resource management to political economy–rest on specific and sometimes arbitrary assumptions about coalition formation. A crucial element in the way groups are formed is the relative influence of each member of the group. That is, an agent's power–whether derived from wealth, military might, or political influence–will be instrumental in determining the possible coalitions that may form.

The organising theme of this dissertation regards the question of stability: When agents have heterogenous power, can we find coalitions that are stable? The answer depends on how stability itself is conceived, the division of surplus, how non-winning agents are affected by the winning coalition, and whether or not power accumulates for the winning agents. This dissertation examined a specific notion of stability, the notion of "self-enforcement". Self-enforcement is reached whenever a set of agents who are part of a winning coalition does not have the incentive to deviate into another coalition. Moreover, since we assume that agents are rational, we expect them to join a coalition that will give them their highest possible payoff. Can we find rules on choosing coalitions that satisfy these desirable properties of self-enforcement and rationality?

Chapter 2 investigates such rules when power of the winning agents accumulates. Power accumulation introduces an interesting dynamic in the process of coalition formation: an agent wants to be in a winning coalition but must make sure that in the process of accumulating power no other set of agents will be strong enough to allow a subsequent deviation. In this chapter, power accumulates according to an agent's share in the prize, hence, the manner in which the prize is divided among winning agents will be crucial in determining which coalitions can form. In particular, we focus on two sharing rules: proportional sharing and equal sharing. Another factor that will determine which coalitions will form is how to treat non-winning agents: Will they survive or will they be eliminated? If they are eliminated then the winning agents will only be concerned by the balance of power inside the winning coalition. If they survive then the agents within the winning coalition also need to guard against the possibility that some of their coalition members can form another coalition with some of the non-winning agents in the future.

One of the findings of this chapter is that it is only in some of the cases that we can always find a rule that satisfies self-enforcement and rationality. For instance, under proportional sharing and where non-winning agents are "killed" we need to choose the self-enforcing coalition that has the least power. Under equal sharing and when agents survive we only have to pick the smallest coalition of size 2^k that is winning (Section 4.2, Proposition 3).

In other cases we need to restrict either the distribution of power in the society or the manner in which coalitions can form. For instance, in the case of equal sharing and where agents are killed, a self-enforcing and rational rule exists only if we can find a subset of the grand coalition of size $2^k - 1$ that is strongly balanced after power has been added (Section 3.2, Proposition 2). Under proportional sharing and when agents survive, we have to restrict to feasible sequences of partitions in order to find a self-enforcing and rational rule to select coalitions (Section 4.2.1, Proposition 4).

Chapter 3, on the other hand, delves a little deeper on the implications of sharing rules on the types of coalitions that will form when agents have heterogeneous power. We distinguish between two classes of sharing rules: First, there are sharing rules that satisfy the property of "consistent ranking", that is, given any competing coalitions, agents in the intersection of these coalitions will agree on which coalition gives them the highest payoff. Equal sharing and proportional sharing are examples of such sharing rules. Under equal sharing, agents will agree on coalitions of smaller sizes since they will divide the prize among fewer members. Under proportional sharing, common agents will prefer the coalition with the least power, since their relative power is highest within this coalition. The second class of sharing rules are those that do not satisfy consistent ranking, that is, those sharing rules that may create disagreement within common agents of any competing coalitions. An example of such sharing rule is what we call "combination sharing"–a convex combination of equal sharing and proportional sharing. Under this sharing rule some agents may prefer to be in the coalition with the smallest size while others may prefer to be in the coalition with the least power.

In this chapter we examine which rules on choosing coalitions satisfy self-enforcement and rationality under these different sharing rules when power does not accumulate. We find that there always exists a rule to choose coalitions regardless of whether agents are killed or are able to survive as long as the sharing rules satisfy consistent ranking. Under combination sharing, however, these rules of choosing coalitions do not exist in general. In order for these rules to simultaneously satisfy self-enforcement and rationality we need to restrict the domain of games either to the case where coalition size and power move in the same direction or by allowing the convex combination parameter to be high enough for agents to agree on the smallest-sized undominated coalition, low enough for agents to agree on the least-powered undominated coalition, or just enough for a compromise coalition to exist.

Chapter 4 implements a novel laboratory experiment to analyze actual behavior of agents when they form coalitions. There are several strategies that the agents may employ, some of which can deviate from the prescribed strategies of the particular coalition formation model corresponding to Section 3.3.2 of this dissertation. The results of this paper suggest that agents' behavior is more complicated in real life. We actually observe a lot of rational behavior, with some agents taking advantage of myopic agents especially in early games when agents do not have much experience playing the coalition formation game. In addition, the agents' strategies are contingent on the power that they have drawn and they play their empirical best response to the possibility that the other agents behave myopically. There is also some hysteresis in the way agents are playing their strategies. Over time, however, agents seem to learn how to behave more strategically and rationally. We also find a weak positive effect of increasing stakes, which suggests that incentives in the coalition formation game may still matter, but its effect needs to be studied more closely.

5.2 Further directions

There are several open questions for future research that thread through the chapters in this dissertation. Note that Chapters 2 and 3 utilize the blocking/axiomatic approach described in Chapter 1. One interesting extension for both these Chapters is to extend the analysis using the bargaining approach to be able to implement the self-enforcing coalitions as the equilibrium of a non-cooperative bargaining game. Doing this can open the analysis on several fronts such as developing a new theory describing how the different sharing rules emerge endogenously from the coalition formation process. This could be achieved, for example, by developing a model where apart from bargaining over coalitions the agents can also bargain over the sharing rules.

Another interesting development to the model is to extend the coalition formation process in these chapters to the case where the decision to sideline or "kill" agents is endogenous. Many real-world situations necessitate developing a theory that allows for this feature. For instance, transitions of power from one non-democratic regime to another usually mark a period of massive purges (e.g., Soviet Russia, Revolutionary China) and understanding the decision of eliminating opponents may explain why a particular set of leaders emerge.

In our models we disregard any role that externalities may play. In real life, though, factors such as culture, religion, and identity will affect how agents form coalitions. The models here should be extended to allow for the possibility of externalities, for instance, through modeling each agent's payoff function to incorporate utility and disutility of being in a group with people they approve and disapprove (see Juarez [51, 52]).

Finally, with regards to real-world behavior of agents, we must take into account cognitive biases, bounded rationality and issues of fairness. Thus, it is important to carefully examine the strategies that agents play in the real world. Will they be affected by changing the sharing rule, group size, or group homophily? This is a fruitful direction that will enrich the coalition formation model described in this dissertation.

5.3 Towards an Economics of Coalition Formation?

As the motivations to form groups are diverse, so are the perspectives and approaches to understand it. Sociologists assert that identity matters. Psychology stresses the importance of interpersonal attraction. Economists, on the other hand, look at the problem as a rational agent's optimization of the calculus of potential costs and benefits derived from joining a group.

The two methodological approaches in economics (blocking vs. bargaining) have been utilized in the different chapters in this dissertation to investigate the role of power, sharing rules, the span of life of the agents and behavioral factors affect the way coalitions form over time. Chapter 2 has investigated the role of power accumulation in the stability of coalitions. Chapter 3 emphasized the effect of sharing rules and its implication to stability. Chapter 4 examined possible factors that allow divergence of actual behavior from the theoretical predictions of the model.

As we have seen in Chapter 1 there has been an explosion in both the number of models and the applications related to coalition formation. In order to further improve our understanding of the different coalition formation processes underlying these applications, further research in the following area should be developed:

- the effect of externalities on agents' preferences
- the role of farsightedness in different coalition formation models
- the role of power in coalition formation models
- explicit treatment of information completeness, the implicit implications of binding or non-binding agreements, and the distribution and diffusion of costs and benefits

- careful treatment on the surplus division within coalitions
- considering the commitment structure of the model, which can allow for potential exit of agents in the coalition
- recognize the possibility of overlapping coalitions, in which agents can belong to more than coalition
- explicitly treatment of the framework of interactions between agents, which can be described as a network of these agents
- possible role of behavioral factors such as bounded rationality and cognition

The chapters in this dissertation can hopefully guide further modeling directions for the applications reviewed in Chapter 1. For instance, the issue of farsightedness and power can be incorporated to traditional oligopoly and cartel formation models. In addition, incorporating power and sharing rules to environmental and political coalitions would provide a richer characterization of equilibrium and stability of these coalitions. Finally, incorporating behavioral determinants in the different coalition formation models will help us understand more the way groups form in the real world. The remaining gaps just identified will make coalition formation a relevant and active research area for years to come.

Appendix A PROOFS OF CHAPTER 2

A.1 Proofs of Section 2.3

Proof of Proposition 1

We begin the proof by introducing two axioms which are variants of ISE and DIR.

Axiom 12 (Self-enforcement (SE)) *The transition correspondence ϕ is **self-enforcing (SE)** if for any game $(X, \pi) \in \mathbf{G}$ and $S \in \phi(X, \pi)$, then $S \in \phi(S, \pi_S)$.*

Axiom 13 (Rationality (RAT)) *The transition correspondence ϕ is **rational (RAT)** if for any $S \in 2^N$, for any $T \in \phi(S, \pi)$ and for any $Z \subset S$ such that $Z \in W_{(S,\pi)}$ and $Z \in \phi(Z, \pi_Z)$, we have that $Z \notin \phi(S, \pi) \Leftrightarrow \xi_i(T, \pi_T) > \xi_i(Z, \pi_Z)\ \forall i \in T \cap Z$.*

We will show in subsequent steps that ISE and SI imply the axiom SE above and DIR and SI imply RAT. We then show that ϕ^* is the unique transition correspondence that satisfy SE and RAT.

Proof. Step 1. Under proportional sharing, ISE and SI imply SE.

Proof. If $S \in \phi(X, \pi)$, then by ISE, $S \in \phi(S, \pi_S + I \cdot PR(S, \pi_S))$. Since the shares in $PR(S, \pi_S)$ are split in the ratio of π_S, then by scale invariance, $S \in \phi(S, \pi_S)$.

Step 2. Under proportional sharing, DIR and SI imply RAT.

Proof. Consider any $X \in 2^N$, and subset $T \in \phi(X, \pi)$ and any $Z \subset X$ such that $Z \in W_{(X,\pi)}$ and $Z \in \phi(Z, \pi_Z + I \cdot PR(Z, \pi_Z))$. Then, $Z \in \phi(Z, \pi_Z)$ by Step 1.

Therefore, we have that

$$Z \notin \phi(X, \pi) \Leftrightarrow \xi_i(T, \pi_T^\infty) > \xi_i(Z, \pi_Z^\infty)\ \forall i \in T \cap Z$$

$$\Leftrightarrow \xi_i(T, \pi_T) > \xi_i(Z, \pi_Z)\ \forall i \in T \cap Z$$

because the sharing rule is proportional. Hence, RAT is satisfied.

Step 3. There exists a unique transition correspondence that meets SE and RAT. That transition correspondence is ϕ^*.

The proof of this step is exactly similar to Theorem 1 of Acemoglu, et al [1] (AES) and will not be shown here. They prove this for the case where power does not accumulate, agents are killed and the sharing rule is proportional. Steps 1 and 2 in this proof basically transforms the axioms ISE and DIR to Acemoglu et al's axioms SE and RAT. Jandoc and Juarez [48] extends AES results to a more general class of sharing rules that satisfy the property of "consistent ranking" where agents have the same ordinal ranking over coalitions in which they belong. Equal and proportional sharing meet consistent ranking.

Step 4. We show that ϕ^* satisfies SI, ISE and DIR

- ϕ^* satisfies SI.

  Since for any $\gamma > 0$

$$\phi^*(S, \gamma\pi) = \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\min} \gamma\pi(M) = \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\min} \pi(M) = \phi^*(S, \pi)$$

- ϕ^* satisfies ISE.

  For the game (S, π) take a coalition $V \in \phi^*(S, \pi)$. By the definition of ϕ^*, either $V = S$ or $V \in Q(S, \pi)$. If $V = S$ then $S \in \phi^*(S, \pi)$. By SI, we have that $S \in \phi^*(S, \pi(1 + \frac{I}{\pi(S)})) = \phi^*(S, \pi + I \cdot PR(S, \pi_S))$.

  On the other hand, if $V \in Q(S, \pi)$, then $V \in \phi^*(V, \pi_V)$ by definition in Equation 3.1. By SI, we have that $V \in \phi^*(V, \pi_V(1 + \frac{I}{\pi(V)})) = \phi^*(V, \pi_V + I \cdot PR(V, \pi_V))$.

- ϕ^* satisfies DIR.

  Consider the game (S, π). Let $V \in \phi^*(S, \pi)$ and $Y \subset S$ such that $Y \in W_{(S,\pi)}$ and $Y \in \phi^*(Y, \pi_Y + I \cdot PR(Y, \pi_Y))$.

  First, consider the case where $Y \notin \phi^*(S, \pi)$. By definition,

$$V \in \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\min} \pi(M).$$

Therefore, $\pi(Y) > \pi(V)$. Therefore, for every agent $i \in Y \cap V$, $PR_i(V, \pi_V) > PR_i(Y, \pi_Y)$. Thus, $\xi_i(V, \pi_V^\infty) > \xi_i(Y, \pi_Y^\infty)$.

On the other hand, if $\xi_i(V, \pi_V^\infty) > \xi_i(Y, \pi_Y^\infty)$ for all $i \in V \cap Y$, then, $\pi(Y) > \pi(V)$. Therefore,

$$Y \notin \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\min} \pi(M).$$

Hence, $Y \notin \phi^*(S, \pi)$.

■

Proof of Proposition 2

Proof. We prove four intermediate steps before proving part i and ii in step 5.

For all the steps, we consider a transition correspondence ϕ that satisfies ISE, DIR and SI.

Step 1: If a coalition S is picked by any transition correspondence ϕ and continues to form, then over time the relative power of $i \in S$ approaches $\frac{1}{|S|}$.

Proof. Note that under equal sharing, the relative power of agent i in a coalition S that continues to form through the k^{th} stage is $\frac{\pi_i^0 + \frac{kI}{|S|}}{\sum_{i \in S} \pi_i^0 + kI}$. Evaluating this expression as $k \to \infty$ by using l'Hopital's rule yields:
$\lim_{k\to\infty} \frac{\pi_i^0 + \frac{kI}{|S|}}{\sum_{i \in S} \pi_i^0 + kI} = \lim_{k\to\infty} \frac{\frac{I}{|S|}}{I} = \lim_{k\to\infty} \frac{1}{|S|} = \frac{1}{|S|}$

Step 2: Any coalition that is chosen by a transition correspondence ϕ that satisfies ISE and DIR should be of size $2^m - 1$ for some $m \in \mathbb{N}$.

Proof. Consider any coalition S that is chosen by the transition correspondence ϕ, and suppose that $|S| = 2^m - 1 + r$ and $r \in [0, 2^m - 1]$. We will prove this step by induction on m.

Consider the base of induction $m = 1$. In this case,

$$|S| = \begin{cases} 1 \text{ if } r = 0 \\ 2 \text{ if } r > 0 \end{cases}$$

We know that $|S| = 1$ is an ISE coalition since a singleton maps into itself. On the other hand, if $|S| = 2$, then the agent i such that $\pi_i > \pi_j$ can always deviate from S and be self-enforcing (since he is a singleton coalition). Thus, S where $|S| = 2$ is not an ISE coalition.

Let our induction hypothesis be that the statement is true for $m = h$. That is,

$$\text{If } |S| = 2^h - 1 + r \text{ then } \begin{cases} S \text{ is an ISE coalition if } r = 0 \\ S \text{ is not an ISE coalition if } r > 0 \end{cases}$$

We now show that this relationship remains true for $m = h + 1$.

If $r = 0$, then:

- By Step 1 the relative power of $i \in S$ is $\frac{1}{2^{h+1}-1}$ as the rounds approach infinity. That is, $\lim_{k\to\infty} \frac{\pi^k}{\pi^k(S)} = \left[\frac{1}{2^{h+1}-1}, \frac{1}{2^{h+1}-1}, \ldots, \frac{1}{2^{h+1}-1}\right]$
- A coalition T that wishes to deviate from S must be at least $2^h - 1 + r$, where $2^h - 1 \leq 2^h - 1 + r < 2^{h+1} - 1$. Note that a $|T| = 2^h - 1$ will not be winning since $\pi(N \setminus T) > \pi(T)$
- In this case, by Step 1 we know that if T continues to form, then the relative power of $i \in T$ will approach $\frac{1}{2^h-1+r}$ in the limit.
- By the same reasoning, a coalition V, where $|V| = 2^h - 1$ can deviate from coalition T. This will be an ISE coalition by our induction hypothesis. Thus T is not an ISE coalition. Therefore, S where $|S| = 2^{h+1} - 1$ is an ISE coalition.

If $r > 0$, then:

- By Step 1 the relative power of $i \in S$ is $\frac{1}{2^{h+1}-1+r}$ as the rounds approach infinity. That is, $\lim_{k\to\infty} \frac{\pi^k}{\pi^k(S)} = \left[\frac{1}{2^{h+1}-1+r}, \frac{1}{2^{h+1}-1+r}, \ldots, \frac{1}{2^{h+1}-1+r}\right]$

- A coalition T where $|T| = 2^h - 1$ can deviate from S. From our induction hypothesis T will be an ISE coalition. Therefore S where $|S| = 2^h - 1 + r$ cannot be an ISE coalition if $r > 0$.

Step 3: Consider the transition correspondence ϕ that satisfies ISE and DIR and the coalition formation game (Y, π_Y) that is strongly balanced. Then, $\phi(Y, \pi_Y) = \{Y\}$.

Proof. By step 2, the coalitions chosen by ϕ must be of size $2^k - 1$. Since (Y, π_Y) is strongly balanced, the only winning coalition of this type is Y. Hence, $\phi(Y, \pi_Y) = \{Y\}$.

Step 4: If the coalition V is chosen by the transition correspondence that satisfies ISE and DIR at game (X, π), then $(V, \pi_V + \frac{I}{|V|} 1_V)$ is strongly balanced.
Proof.

By step 2, $|V| = 2^k - 1$ for some $k \in N$. Suppose that $(V, \pi_V + \frac{I}{|V|} 1_V)$ is not strongly balanced.

Partition V into the disjoint coalitions S, T and U, that is $V = S \cup T \cup U$, and such that coalition S contains the $2^{k-2} - 1$ largest elements in V, coalition T contains the 2^{k-2} largest elements in $V \setminus S$ and coalition U contains $V \setminus (S \cup T)$, which are the 2^{k-1} smallest elements in V.

Since $(V, \pi_V + \frac{I}{|V|} 1_V)$ is not strongly balanced then $\pi(S \cup T) > \pi(U)$.

By step 1, since power equalizes under ES, there exists a time t such that U becomes a winning coalition in V. That is,

$$\pi(S \cup T) + tI(2^{k-1} - 1) < \pi(U) + tI2^{k-1}.$$

Therefore,

$$\pi(S \cup T) - \pi(U) < tI \tag{A.1}$$

Also, note that by the choice of S, T and U,

$$\pi(S) - \pi(T) \leq \pi(S \cup T) - \pi(U) \tag{A.2}$$

To see this, assume that $\pi(S) - \pi(T) > \pi(S \cup T) - \pi(U)$, then $\pi(T) < \frac{\pi(U)}{2}$. Since $2|T| = |U|$, then $\frac{\pi(T)}{|T|} < \frac{\pi(U)}{|U|}$, which contradicts the choice of T and U, therefore inequality A.2 holds.

Combining inequalities A.1 and A.2:

$$\pi(S) - \pi(T) < tI$$

Therefore,

$$\pi(S) + tI(2^{k-2} - 1) < \pi(T) + tI2^{k-2} \tag{A.3}$$

The left and right hand side of inequality A.3 are the power of coalition S and T, respectively, after t rounds. Hence, coalition $S \cup T$ is strongly balanced at time t. Moreover, by the choice of t, coalition $S \cup T$ is winning in V at time $t-1$, hence it is balanced. Thus, $S \cup T$ can deviate from V at time $t-1$ and be internally self-enforcing by step 3. This contradicts DIR.

Step 5: Proofs of parts i and ii.

Part i follows directly from step 4, because if the transition correspondence selects Y at the coalition formation game (X, π), then Y is winning in (X, π) and (Y, π_Y) is strongly balanced. Hence, (X, π) is balanced.

To prove part ii, consider the balanced game (X, π) and suppose that $Y \in \phi(X, \pi)$. By step 2, $|Y| = 2^m - 1$ for some m. Suppose that $Z \in W_{(X,\pi)}$, (Z, π_Z) is strongly balanced and $Z \notin \phi(X, \pi)$. Since (Z, π_Z) is strongly balanced, then $Z \in \phi(Z, \pi_Z)$ by step 3. Hence, by ISE, $Z \in \phi(Z, \pi_Z + \frac{I}{|Z|}1_Z)$. Therefore, by DIR, $\xi_i(Y, \pi_Y^\infty) > \xi_i(Z, \pi_Z^\infty)$ for all $i \in Y \cap Z$. Hence, by step 1, $\xi_i(Y, \pi_Y^\infty) = \frac{1}{|Y|}$ and $\xi_i(Z, \pi_Z^\infty) = \frac{1}{|Z|}$. Thus, $|Y| < |Z|$.

■

A.2 Proofs of Section 2.4

Proof of Proposition 3

Proof of part i.

We fix the transition correspondence ϕ^{**} that satisfies ESE, DER, SI, IZ.

Step 1: If a coalition S is chosen by the transition correspondence ϕ^{**} that satisfies ESE, then over time the relative power of $i \in S$ approaches $\frac{1}{|S|}$.

Proof. This is similar to the proof of Step 1 in Proposition 2, therefore it is omitted.

Step 2: Consider a coalition formation game (N, π) such that $\phi^{**}(N, \pi) = \{S\}$. Then, there exists an open set $B \subset \mathbb{R}^N_+$ such that $\pi \in B$ and $S \in \phi^{**}(N, \tilde{\pi})$ for any $\tilde{\pi} \in B$.

Proof. Suppose that is not the case. Then, there exists a sequence of power vectors $\{\bar{\pi}^i\}_i$ where S is winning in the coalition formation $(N, \bar{\pi}^i)$ for every i, $\lim_{i\to\infty} \bar{\pi}^i = \pi$ and $S \notin \phi^{**}(N, \bar{\pi}^i)$ for all i.

Since $\phi^{**}(N, \bar{\pi}^i)$ is chosen from the finite set 2^N, then we can find a set that is chosen an infinite number of times in the sequence. That is, we can find $T \subset N$ and a subsequence $\{\tilde{\pi}^i\}_i \subset \{\bar{\pi}^i\}_i$ such that $T \in \phi^{**}(N, \tilde{\pi}^i)$ for all i. Hence, by continuity, $T \in \phi^{**}(N, \pi)$, which is a contradiction.

Step 3: Let $S^m = \{1, \ldots, m\}$ and consider the coalition formation game (N, π^m) such that $\pi^m_j = 1$ if $j \in S^m$, and $\pi^m_l = 0$ if $l \in N \setminus S^m$. We will show by induction on m that if $m = 2^k$ for some $k \in \mathbb{N}$ then $\phi^{**}(N, \pi^m) = \{S^m\}$; and if $m = 2^k + r$ for some $k \in \mathbb{N}$ and $0 < r \leq 2^k - 1$ then $\phi^{**}(N, \pi^m) = \{T | T \subset S^m \text{ and } |T| = 2^k\}$.

Proof. First, we start with the base of induction $m = 1$. By IZ, $\phi^{**}(N, \pi^1) = S^1$.

Second, suppose that the statement is true for $m < i$. We will show that it is also true for $m = i$.

Let $i = 2^k + s$ for $s \in [0, 2^k - 1]$ and $T \in \phi^{**}(N, \pi^i)$. By IZ, $T \subset S^i$. Since T is winning in (N, π^i), then $|T| > 2^{k-1}$. Thus, $|T| = 2^{k-1} + r$ for $0 < r \leq 2^{k-1} + s$. By step 1, scale invariance and continuity, $T \in \phi^{**}(N, \pi^T)$ where $\pi^T_i = 1$ if $i \in T$ and $\pi^T_i = 0$ if $i \in N \setminus T$. Thus, up to renaming the agents, $S^{2^{k-1}+r} \in \phi^{**}(N, \pi^{2^{k-1}+r})$. By the induction hypothesis, for $r \neq 2^{k-1}, 2^{k-1} + s$, we have that $S^{2^{k-1}+r} \notin \phi^{**}(N, \pi^{2^{k-1}+r})$. Thus, $r = 2^{k-1} + s$ or $r = 2^{k-1}$.

We analyze the next three cases depending on whether $s = 0$ or $s \neq 0$, and whether $r = 2^{k-1} + s$ or $r = 2^{k-1}$.

<u>Case 1. Suppose $s = 0$.</u>

Then, $r = 2^{k-1}$. Therefore $T = S^i$. Thus, $\phi^{**}(N, \pi^i) = \{S^i\}$.

Case 2.1. $s \in (0, 2^k - 1]$ and $r = 2^{k-1} + s$.

Then, $|T| = 2^{k-1} + 2^{k-1} + s = 2^k + s$. Thus, $S^{2^k+s} \in \phi^{**}(N, \pi^{2^k+s})$. Consider the vector $v^t \in \mathbb{R}_+^N$ such that

$$v^t = (tI)\pi^{2^k} + (\epsilon + \delta)\pi^{2^k+s} + (\delta)(\pi^{2^k+s} - \pi^{2^k}),$$

where $\delta < 2^k \epsilon < I$.

Note that for every $t \geq 0$, the size of the smallest winning coalition that is a power of 2 equals 2^k.

Given ϵ and δ, note that as t tends to infinity, the relative power equalizes. Thus, by step 2 there exists a large t^* such that $S^{2^k} \in \phi^{**}(N, v^{t^*})$.

Similarly, any winning coalition that is a power of 2 in the game (N, v^{t^*-1}) should be of size 2^k. Since $S^{2^k} \in \phi^{**}(N, v^{t^*-1} + I\pi^{2^k})$, and S^{2^k} is winning in (N, v^{t^*-1}), then by DER, $S^{2^k} \in \phi^{**}(N, v^{t^*-1})$.

Similarly, $S^{2^k} \in \phi^{**}(N, v^{t^*-2}), S^{2^k} \in \phi^{**}(N, v^{t^*-3}), \ldots$.

We can continue like that until $t = 0$. Therefore,

$$S^{2^k} \in \phi^{**}(N, (\epsilon + \delta)\pi^{2^k+s} + (\delta)(\pi^{2^k+s} - \pi^{2^k})).$$

By continuity, as $\delta \to 0$,

$$S^{2^k} \in \phi^{**}(N, \epsilon\pi^{2^k+s}).$$

By SI,

$$S^{2^k} \in \phi^{**}(N, \pi^{2^k+s}).$$

By assumption, $S^{2^k+s} \in \phi^{**}(N, \pi^{2^k+s})$, which contradicts DER, since the agents in S^{2^k} strongly prefer S^{2^k} to S^{2^k+s}.

Case 2.2. Supppose $s \in (0, 2^k - 1]$ and $r = 2^{k-1}$.

In this case, $S^{2^k} \in \phi^{**}(N, \pi^{2^k+s})$. Since all coalitions of size 2^k give exactly the same payoff and ϕ^{**} satisfies ESE, then $\phi^{**}(N, \pi^i) = \{T | T \subset S^i \text{ and } |T| = 2^k\}$.

Step 4: For any coalition formation game (N, π), if $S \in \phi^{**}(N, \pi)$, then $|S| = 2^k$ for $k \in \mathbb{N}$.

Proof.

Let $e^S \in \mathbb{R}^N$ be the vector such that $e^S_i = 1$ if $i \in S$ and $e^S_i = 0$ if $i \in N \setminus S$.

By ESE, $S \in \phi^{**}(N, \pi + tIe^S)$ for $t = 1, 2, \ldots$.

By SI, $S \in \phi^{**}(N, \frac{\pi}{tI} + e^S)$ for $t = 1, 2, \ldots$.

By continuity, letting t approach to infinity, $S \in \phi^{**}(N, e^S)$.

By step 3, $|S| = 2^k$ for $k \in \mathbb{N}$.

Step 5: For any coalition formation game (N, π)

$$\phi^{**}(N, \pi) = \arg \min_{M \in Q \cup \{N\}} |M|$$

where $Q(N, \pi) = \{S \in 2^N \text{ such that } S \in W_{(N,\pi)} \text{ and } |S| = 2^m$ for some $m \in \mathbb{N}\}$.

Proof.

By step 4, we know that $\phi^{**}(N, \pi) \subset Q(N, \pi)$.

Let S be a coalition of the smallest size in $Q(N, \pi)$. Since the sharing rule equalizes the shares of the agents, then S is also a coalition of the smallest size in $Q(N, \frac{\pi}{It} + e^S)$ for any $t \geq 1$.

By steps 2 and 3, there exists $t^* \in \mathbb{N}$ large enough such that $S \in \phi^{**}(N, \frac{\pi}{It^*} + e^S)$.

By SI, $S \in \phi^{**}(N, \pi + (It^*)e^S) = \phi^{**}(N, \pi + I(t^* - 1)e^S + \frac{I}{|S|}e^S)$.

Since S is a coalition of the smallest size in $Q(N, \pi + I(t^* - 1)e^S)$, then S will maximize the payoff of the agents among ESE coalitions. Thus, by DER, $S \in \phi^{**}(N, \pi + I(t^* - 1)e^S)$.

Continuing similarly for rounds $t^* - 1, t^* - 2, \ldots, 0$, we have that $S \in \phi^{**}(N, \pi)$.

Finally, by DER, all chosen coalitions should give the same share of the resource to the agents. Hence, by step 1 they are of the same size. Therefore, $\phi^{**}(N, \pi)$ contain only coalitions of the minimal size in $Q(N, \pi)$.

Proof of part ii.

Step 1. If agent i is a dictator in the game (N, π), that is $\pi_i > \pi(N \setminus i)$, then $\phi(N, \pi) = \{\{i\}\}$.

Proof. First notice that by IZ, $\phi(N, [\pi_i, 0_{N\setminus i}]) = \{\{i\}\}$, where $0_{N\setminus i}$ is the zero vector in $\mathbb{R}^{N\setminus i}$. Therefore, by continuity $i \in \phi(N, [\pi_i, \frac{1}{1+k^*\frac{I}{\pi_i}}\pi_{N\setminus i}])$ for k^* large enough. By SI, $i \in \phi(N, [\pi_i + (k^*)I, \pi_{N\setminus i}])$. By DER, $i \in \phi(N, [\pi_i + (k^* - 1)I, \pi_{N\setminus i}])$ because coalition $\{i\}$ is the smallest ESE coalition at the power profile $[\pi_i + (k^* - 1)I, \pi_{N\setminus i}]$. Continuing similarly $k^* - 2$ times, $i \in \phi(N, [\pi_i + (k^* - 2)I, \pi_{N\setminus i}])$, $i \in \phi(N, [\pi_i + (k^* - 3)I, \pi_{N\setminus i}])$, $\ldots$, $i \in \phi(N, [\pi_i, \pi_{N\setminus i}])$. Moreover, by DER, $\phi(N, \pi) = \{\{i\}\}$ because i's payoff is greater by getting the prize alone instead of sharing it with another agent.

Step 2. Consider the game for three agents $(\{1, 2, 3\}, [\pi_1, \pi_2, \pi_3])$ such that $\pi_1 > \pi_2 > \pi_3$ and $\pi_1 < \pi_2 + \pi_3$. Then, $\phi(\{1, 2, 3\}, [\pi_1, \pi_2, \pi_3]) = \{\{1, 2, 3\}\}$.

Proof. We prove this by contradiction. First, notice that any winning coalition contains at least two agents. Suppose that coalition S such that $|S| = 2$ is chosen, that is, $S \in \phi(\{1, 2, 3\}, [\pi_1, \pi_2, \pi_3])$. Then, by ESE, $S \in \phi(\{1, 2, 3\}, [(1 + k^*\frac{I}{\pi(S)})\pi_S, \pi_{-S}])$. For a large k^*, the game $(\{1, 2, 3\}, [(1 + k^*\frac{I}{\pi(S)})\pi_S, \pi_{-S}])$ has a dictator. Hence, by step 1, S such that $|S| = 2$ cannot be chosen. This is a contradiction.

Finally, consider the sequence of coalition formation games $(\{1, 2, 3\}, [1 + \frac{1}{k}, 1, \frac{2}{k}])$ for $k = 1, 2, 3, \ldots$. By step 2, $\{1, 2, 3\} \in \phi(\{1, 2, 3\}, [1 + \frac{1}{k}, 1, \frac{2}{k}])$. Also, $\{1, 2, 3\}$ is winning in $(1, 1, 0)$. Hence, $\{1, 2, 3\} \in \phi(\{1, 2, 3\}, [1, 1, 0])$. This contradicts IZ.

Proof of Proposition 4

Proof. SI is satisfied since the $\min$ function is scale invariant.

To prove that the sequential transition correspondence $\bar{\phi}$ satisfies ESE, let $S^j \in \bar{\phi}(N, \pi)$ and suppose $S^j \neq N$. Because we restrict on feasible sequences of coalitions, we know that the only possible deviations from S^j are only into feasible subsets. Since by definition of $\bar{\phi}$, there does not exist a coalition $S^m \subset S^j$ such that $S^m \in W_{(S^j, \pi_{S^j})}$ and since under proportional sharing, the relative power of every agent in S^j is unchanged and the relative power of the agents in $N \setminus S^j$ goes to zero, then coalition S^j will be chosen in further rounds. If $S^j = N$, with proportional sharing the relative powers are the same at every round and thus S^j has to be chosen again in every round.

To show that DER is satisfied, suppose $S^j \in \bar{\phi}(N, \pi)$ and there exist another coalition Z such that $Z \in W_{(N,\pi)}$ and $Z \in \bar{\phi}(N, \pi + I \cdot PR^N(Z, \pi))$ and $Z \notin \bar{\phi}(N, \pi)$. Since $Z \in \bar{\phi}(N, \pi + I \cdot PR^N(Z, \pi))$, then $Z = S^m$ for some m. By the definition of a feasible sequence of coalitions, $S^j \subset S^m$ or $S^m \subset S^j$. If $S^j \subset S^m$, then $PR(N, \pi^{\infty}_{S^m}) < PR(N, \pi^{\infty}_{S^j})\ \forall i \in S^j \cap S^m$. If $S^m \subset S^j$ and $S^m \notin \bar{\phi}(N, \pi)$, then there exists $S^h \subset S^m$ such that $S^h \in W_{(S^m, \pi_{S^m})}$. Hence, $S^m \notin \bar{\phi}(N, \pi + I \cdot PR^N(S^m, \pi))$ since S^h will be a winning coalition after enough rounds are played.

■

Appendix B RESULTS WITHOUT CONSISTENT RANKING

In this Appendix we discuss the necessary and sufficient conditions on the sharing rule that allows for the existence of a self-enforcing and rational transition correspondence. In particular, we show that the assumption of consistent ranking in Proposition 5 can be extended to a larger class of sharing rules that satisfies a property which we call "generalized consistency".

First, we show that the set of self-enforcing coalition coincide even for sharing rules that do not satisfy consistent ranking.

Lemma 1 *Consider the self-enforcing and minimalistic transition correspondences ϕ and $\tilde{\phi}$ for the sharing rules ξ and $\tilde{\xi}$, respectively. Then, for a given power vector π, the sets of coalitions that are self-enforcing coincide. That is,*

$$\{S|S \in \phi(S,\pi)\} = \{T|T \in \tilde{\phi}(T,\pi)\}.$$

Proof. Consider the sets $A^u = \{S|S \in \phi(S,\pi), |S| \leq u\}$ and $B^u = \{T|T \in \tilde{\phi}(T,\pi), |T| \leq u\}$.

We will prove by induction on the size of u that $A^u = B^u$.

This is clearly true if $u = 1$, because any singleton coalition is self-enforcing.

For the induction hypothesis, assume that $A^{u-1} = B^{u-1}$.

Consider $S \in A^u$. Then $S \in \phi(S,\pi)$. Therefore, since ϕ is minimalistic, there is no $Q \subsetneq S$ such that $Q \in W_{(S,\pi)}$ and $Q \in \phi(Q,\pi_Q)$.

Therefore, since $A^{u-1} = B^{u-1}$, there is no $Q \subsetneq S$ such that $Q \in W_{(S,\pi)}$ and $Q \in \tilde{\phi}(Q,\pi_Q)$.

Hence, $S \in \tilde{\phi}(S,\pi)$ and $S \in B^u$. Thus $A^u \subset B^u$.

We can similarly prove that $B^u \subset A^u$. ■

The game (S,π) is *stable* if the coalition S is self-enforcing at the vector of power π under any other self-enforcing and minimalistic transition correspondence (for instance, under PR or ES). By Lemma 1, this is well defined.

The class of stable games can be easily constructed. First, note that any coalition of size one is stable. Second, note that a game of size larger than 1 is stable if and only if no subgame that is winning is stable. Hence, the construction of the stable games is made inductively on the size of the coalition. Example 9 shows the class of coalitions that generate a stable game for a power profile with 8 agents.

Example 9 *Consider an 8-agent society with power profile*

$$\pi = [12, 11.5, 11, 10.5, 10, 9, 9.5, 8].$$

The table below gives the coalitions from the set of stable games for different coalition sizes:

Coalition size	Coalitions that generate a stable game for the power profile π
1	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}$
2	None, a coalition of size 1 can deviate
3	$\{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,2,6\}, \{1,2,7\}, \{1,2,8\}, \{2,3,4\}, \{2,3,5\},$ $\{2,3,6\}, \{2,3,7\}, \{2,3,8\}, \{3,4,5\}, \{3,4,6\}, \{3,4,7\}, \{3,4,8\}, \{4,5,6\},$ $\{4,5,7\}, \{4,5,8\}, \{5,6,7\}, \{5,6,8\}, \{6,7,8\}, \{1,3,4\}, \{1,3,5\}, \{1,3,6\},$ $\{1,3,7\}, \{1,3,8\}, \{1,4,5\}, \{1,4,6\}, \{1,4,7\}, \{1,4,8\}, \{1,5,6\}, \{1,5,7\},$ $\{1,5,8\}, \{1,6,7\}, \{1,6,8\}, \{1,7,8\}, \{2,4,5\}, \{2,4,6\}, \{2,4,7\}, \{2,4,8\},$ $\{2,5,6\}, \{2,5,7\}, \{2,5,8\}, \{2,6,7\}, \{2,6,8\}, \{2,7,8\}, \{3,5,6\}, \{3,5,7\},$ $\{3,5,8\}, \{3,6,7\}, \{3,6,8\}, \{3,7,8\}, \{4,6,7\}, \{4,6,8\}, \{4,7,8\}, \{5,7,8\}$
4	None, a coalition of size 3 can deviate
5	None, a coalition of size 3 can deviate
6	None, a coalition of size 3 can deviate
7	$\{1,2,3,4,5,6,7\}, \{2,3,4,5,6,7,8\}, \{1,3,4,5,6,7,8\}, \{1,2,4,5,6,7,8\}$ $\{1,2,3,5,6,7,8\}, \{1,2,3,4,6,7,8\}, \{1,2,3,4,5,7,8\}, (\{1,2,3,4,5,6,8\}$
8	None, a coalition of size 7 can deviate

All singleton coalitions generate a stable game since there are no other possible deviations. Coalitions of size 2 do not generate a stable game because the agent with the highest power can deviate and form a singleton coalition. Coalitions of size 3 generate a stable game because no single agent has enough power to deviate. Furthermore, coalitions of size 2 cannot deviate from a coalition of size 3 since they do not generate a stable game. Coalitions of size 4 do not generate a stable game since the three highest powered agents can deviate from this coalition and form a stable game. This is the same for coalitions of size 5 or 6. Coalitions of size 7 generate a stable game since a winning subcoalition needs to be at least of size 4, and we know that coalitions of sizes 4, 5, and 6 do not generate a stable game. The grand coalition does not generate a stable game because any coalition of size 7 can deviate and form a stable game.

Definition 13 • *Consider the game* (N, π) *and let*

$$SEC(N,\pi) = \{T \subsetneq N \mid (T, \pi_T) \text{ is stable and } T \in W_{(N,\pi)}\}$$

be the set of coalitions that generate a stable game and are winning in the game (N, π)*.*

- *Consider the game* (N, π) *and let*

$$UD(N,\pi) = \{(T, \pi_T) \mid T \in SEC(N,\pi) \; \nexists S \in SEC(N,\pi), \xi_i(S,\pi_S) > \xi_i(T,\pi_T) \forall i \in S \cap T\}$$

be the set of undominated stable games that are winning in (N, π)*.*

Definition 14 (Generalized consistency) *A sharing rule* ξ *satisfies generalized consistency (GC) if for any game* (N, π)*:*

a. *The set* $UD(N, \pi) \neq \emptyset$*.*

b. *For the game* $(S, \pi_S) \in UD(N, \pi)$ *and* $(V, \pi_V) \notin UD(N, \pi)$ *such that* $V \in SEC(N, \pi)$*, we have that* $\xi_i(S, \pi_S) > \xi_i(V, \pi_V)$ *for all* $i \in S \cap V$*.*

Generalized consistency requires that for every game, there exists a winning and stable subgame that is not dominated[1] by any other winning and stable subgame. Therefore, it cannot be that every game in $SEC(N, \pi)$ is dominated by another game in the same set. Moreover, it requires that every undominated game dominates every other game in $SEC(N, \pi)$ that is not undominated.

Note that if a sharing rule satisfies consistent ranking, then it satisfies GC. This is because the agents share the same ordinal ranking R^{ξ} over games. Therefore, $UD(N, \pi)$ coincides with the set of games that maximize R^{ξ}.

Example 10 *The class of sharing rules that satisfy generalized consistency contains a large class of important sharing rules not covered by consistent ranking.*

The definition of generalized consistency only imposes restrictions on the class of stable games. Thus, for instance, a sharing rule will meet GC if it satisfies consistent ranking within the class of stable games (and does not necessarily satisfy consistent ranking for games that are not stable). One example of such sharing rule might split the resource in proportion (or equally) within the class of stable games, and use combination sharing outside the class of stable games.[2]

[1] Coordinate by coordinate for the agents in the intersection.
[2] As long as cross-monotonicity is satisfied.

Alternatively, note that the sharing rule does not need to satisfy consistent ranking within the class of stable games. For instance, consider a convex combination of dictatorial sharing and proportional, where 90% of the resource is allocated to a single agent following a priority ordering and the remaining 10% of the resource is allocated to all the agents in proportion to their power. For instance, if the priority ordering is $1 \succ 2 \succ \cdots \succ n$*, then for all the stable games that contain agent 1, 90% of the resource is given to agent 1 and the remaining 10% is split between all the agents in proportion to their power. For all the stable games that contain agent 2 but do not contain agent 1, 90% of the resource is given to agent 2 and the remaining 10% is split between all the agents in proportion to their power, and so forth. This rule satisfies generalized consistency for the game* $(N, \pi) = (\{1, 2, 3, 4, 5\}, (18.5, 21, 20, 19, 18.6))$ *but does not satisfy consistent ranking. To see this, note that any game for three agents is stable. However, the only undominated game is the one with coalition* $\{1, 4, 5\}$ *because this coalition is prefered by agents 1,4 and 5 over any other coalition that contains three agents. Clearly consistent ranking is not satisfied, for instance, for the games with coalitions* $\{1, 2, 3\}$ *and* $\{2, 3, 4\}$*, agent 3 prefers the former whereas agent 2 prefers the latter.*

Note that the sharing rule does not need to be asymmetric. For instance, consider the sharing rule that allocates 90% of the resource to the agent with the largest power and 10% of the resource is split to all the agents in proportion to their power. Then, for the starting game $(N, \pi) = (\{1, 2, 3, 4, 5\}, (18.5, 21, 20, 19, 18.6))$*, only coalitions of size 3 are stable. Similarly as above,* $\{1, 4, 5\}$ *is the only undominated coalition. This rule does not satisfy consistent ranking.*[3]

The following result provides the complete class of sharing rules that allow the compatibility of SE and RAT. It also provides the unique transition correspondence that meets SE and RAT. Proposition 5(ii-iii) is a straighforward consequence of this result.

Proposition 10 *There exists a transition correspondence* ϕ *that satisfies SE and RAT under the sharing rule* ξ *if and only if* ξ *meets GC. Moreover, if* ϕ *satisfies SE and RAT then*

$$\phi(N, \pi) = \{T | (T, \pi_T) \in UD(N, \pi)\}.$$

Proof.

We prove this result in four steps.

[3] Note that the sharing rules provided in this example only work locally for this specific game. The extension of the above rules to the entire class of games requires more details, but they are omitted due to space constraints.

Step 1. Suppose ϕ and $\tilde{\phi}$ are transition correspondences that satisfy SE and RAT under the sharing function ξ. Then, $\phi = \tilde{\phi}$.

Proof. Suppose there is a game (N, π) such that $\phi(N, \pi) \neq \tilde{\phi}(N, \pi)$. Without loss of generality, let $T \in \tilde{\phi}(N, \pi) \setminus \phi(N, \pi)$ and let $S \in \phi(N, \pi)$. Note that S and T are self-enforcing under ϕ and $\tilde{\phi}$ by Lemma 1. Hence, by the rationality of ϕ, $\xi_i(S, \pi_S) > \xi_i(T, \pi_T)$ for all $i \in T \cap S$.

If $S \in \tilde{\phi}(N, \pi)$, then by the rationality of $\tilde{\phi}$, $T \notin \tilde{\phi}(N, \pi)$. This is a contradiction.

On the other hand, if $S \notin \tilde{\phi}(N, \pi)$, then by the rationality of $\tilde{\phi}$, we have that $\xi_i(S, \pi_S) < \xi_i(T, \pi_T)$ for all $i \in T \cap S$. This is a contradiction.

Therefore, $\tilde{\phi}(N, \pi) \setminus \phi(N, \pi) = \emptyset$. Hence, $\tilde{\phi}(N, \pi) \subset \phi(N, \pi)$. By a similar argument we can show that $\phi(N, \pi) \subset \tilde{\phi}(N, \pi)$.

Step 2. Let

$$DOM(N, \pi) = \{S \in SEC(N, \pi) | \exists T \in SEC(N, \pi) \text{ s.t. } \xi_i(T, \pi_T) > \xi_i(S, \pi_S) \forall i \in S \cap T\}$$

be the set of dominated coalitions in (N, π). Then, for any ϕ that satisfies RAT and SE, we have that

$$\phi(N, \pi) \cap DOM(N, \pi) = \emptyset.$$

Proof. We prove it by contradiction. Suppose that $S \in \phi(N, \pi) \cap DOM(N, \pi)$. Then, there exists $T \in SEC(N, \pi)$ such that $\xi_i(T, \pi_T) > \xi_i(S, \pi_S)$ for all $i \in S \cap T$. Since ϕ satisfies rationality, then it also satisfies minimalistic. Therefore, by Lemma 1, $T \in \phi(T, \pi_T)$.

If $T \in \phi(N, \pi)$, then $S \notin \phi(N, \pi)$ by rationality, which is a contradiction.

On the other hand, if $T \notin \phi(N, \pi)$, then $\xi_i(S, \pi_S) > \xi_i(T, \pi_T)$ for all $i \in S \cap T$. This is a contradiction.

Hence, $\phi(N, \pi) \cap DOM(N, \pi) = \emptyset$.

Step 3. We prove necessity.

Consider the transition correspondence ϕ that satisfies SE and RAT. First, note that RAT implies minimalistic; therefore, ϕ satisfies the conditions of Lemma 1. Let (N, π) be a coalition formation game and let $Q \in \phi(N, \pi)$. By SE, $Q \in \phi(Q, \pi_Q)$. Thus, by Lemma 1, the game (Q, π_Q) is stable. Therefore, $(Q, \pi_Q) \in SEC(N, \pi)$.

By step 2, $(Q, \pi_Q) \in UD(N, \pi)$. Therefore, $UD(N, \pi) \neq \emptyset$.

Let $(T, \pi_T) \in UD(N, \pi)$. Then, $T \in SEC(N, \pi)$. Thus, by Lemma 1, $T \in \phi(T, \pi_T)$.

If $T \notin \phi(N, \pi)$, then $\xi_i(Q, \pi_Q) > \xi_i(T, \pi_T)$ for all $i \in T \cap Q$. Thus, $(T, \pi_T) \notin UD(N, \pi)$. This is a contradiction. Therefore, $T \in \phi(N, \pi)$. Hence,

$$\{T|(T, \pi_T) \in UD(N, \pi)\} \subset \phi(N, \pi).$$

Let $S \in \phi(N, \pi)$. Then, S is self-enforcing. Thus, $S \in SEC(N, \pi)$. By step 2, $S \notin DOM(N, \pi)$. Hence, $(S, \pi_S) \in UD(N, \pi)$. Therefore,

$$\phi(N, \pi) \subset \{T|(T, \pi_T) \in UD(N, \pi)\}.$$

Hence,

$$\phi(N, \pi) = \{T|(T, \pi_T) \in UD(N, \pi)\}.$$

Step 4. We prove sufficiency by constructing a transition correspondence $\hat{\phi}$ that satisfies SE and RAT. Note that, by step 2, this will be unique.

Let $\hat{\phi}$ be the set of coalitions generated by an undominated game. That is,

$$\hat{\phi}(N, \pi) = SEC(N, \pi) \setminus DOM(N, \pi).$$

We show that $\hat{\phi}$ is self-enforcing. To see this, notice that if $S \in \hat{\phi}(N, \pi)$, then (S, π_S) is stable. Therefore, no game that is stable can be winning in (S, π_S). Thus, $SEC(S, \pi_S) = \{S\}$. Thus, $\hat{\phi}(S, \pi_S) = \{S\}$.

We show that $\hat{\phi}$ is rational. To see this, suppose that $Z \notin \hat{\phi}(N, \pi)$, where $Z \in \hat{\phi}(Z, \pi_Z)$ and $Z \in W_{(N,\pi)}$. Since $Z \in \hat{\phi}(Z, \pi_Z)$, then (Z, π_Z) is stable. Thus, $Z \in SEC(N, \pi)$. Since $Z \notin \hat{\phi}(N, \pi)$, then $Z \in DOM(N, \pi)$.

Let $\tilde{Q} \in \hat{\phi}(N, \pi)$. Then, $(Q, \pi_Q) \in UD(N, \pi)$. Therefore, by generalized consistency, $\xi_i(Q, \pi_Q) > \xi_i(Z, \pi_Z)$ for all $i \in Q \cap Z$. Therefore, rationality is satisfied.

■

Appendix C PROOFS OF CHAPTER 3

C.1 Proofs of Section 3.3.2

Proof of Proposition 5

Proof.

Note that, except for step 5, Proposition 5 and its proof is a trivial consequence of Proposition 10. Therefore, the proof will be removed from the final version of this paper. We leave this proof as an illustration of the main differences with the proof provided by AES in the special case of PR.

Step 1. ϕ^* is SE and minimalistic.

Proof. To show SE, take any $X \in \phi^*(S, \pi_S)$. There are two cases: either $X = S$ or $X \in Q$. If $X = S$, then $X \in \phi^*(S, \pi_S) = \phi^*(X, \pi_X)$. If $X \in Q$, then $X \in \phi^*(X, \pi_X)$ by definition of Q.

On the other hand, ϕ^* is minimalistic because ξ is cross-monotonic. That is, at the coalition formation game (S, π_S), the set S is chosen only if $Q(S, \pi_S) = \emptyset$.

Step 2. ϕ^* satisfies RAT.

Proof. Take $T \in \phi^*(S, \pi_S)$ and consider a coalition Z such that $Z \in W_{(S,\pi)}$ such that $Z \in \phi^*(Z, \pi_Z)$.

($\Rightarrow$) First assume that $Z \notin \phi^*(S, \pi_S)$. Since $T \in \phi^*(S, \pi_S)$ we have that

$$T \in \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\max} R^\xi(M, \pi_M)$$

Notice that Z is winning and self-enforcing within S, therefore $Z \in Q(S, \pi) \cup \{S\}$. Moreover, since $Z \notin \phi^*(S, \pi_S)$, then $Z \notin \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\max} R^\xi(M, \pi_M)$. Hence, $R^\xi(T, \pi_T) > R^\xi(Z, \pi_Z)$.

($\Leftarrow$) Now, assume that $R^\xi(T, \pi_T) > R^\xi(Z, \pi_Z)$. Then, $Z \notin \underset{M \in Q(S,\pi) \cup \{S\}}{\arg\max} R^\xi(M, \pi_M)$. Hence, $Z \notin \phi^*(S, \pi_S)$

Step 3. Consider any cross-monotonic sharing rule and transition correspondences ϕ and $\tilde{\phi}$ that are self-enforcing and minimalistic. Then, the sets of coalitions that are self-enforcing coincide. That is,

$$\{S|S \in \phi(S,\pi)\} = \{T|T \in \tilde{\phi}(T,\pi)\}.$$

Proof. The proof of this result is a straighforward consequence of Lemma 1.

Step 4. There exists a unique transition correspondence that meets SE and RAT.

Proof. Consider a transition correspondence ϕ that is SE and RAT. Then, ϕ is minimalistic because the sharing rule is cross-monotonic. We will show that $\phi = \phi^*$.

Since ϕ and ϕ^* are SE and RAT, then by step 3,

$$\{T|T \in \phi(T,\pi_T)\} = \{T|T \in \phi^*(T,\pi_T)\} \tag{C.1}$$

Suppose $X \in \phi(X,\pi)$. Then, by equation C.1, $X \in \phi^*(X,\pi)$. Thus, $\phi(X,\pi) \subset \phi^*(X,\pi)$. Similarly, $\phi^*(X,\pi) \subset \phi(X,\pi)$. Hence, $\phi(X,\pi) = \phi^*(X,\pi)$.

On the other hand, suppose $S \in \phi(X,\pi)$, where $S \neq X$. Then, by RAT, $\xi_i(S,\pi_S) \geq \xi_i(V,\pi_V)$ for any $V \in \{T|T \in \phi(T,\pi_T), T \in W_{(X,\pi)}\}$ and $i \in V \cap T$. Therefore, by consistent ranking, $R^{\xi}(S,\pi_S) \geq R^{\xi}(V,\pi_V)$ for any $V \in \{T|T \in \phi(T,\pi_T), T \in W_{(X,\pi)}\}$. Thus, $R^{\xi}(S,\pi_S) \geq R^{\xi}(V,\pi_V)$ for any $V \in \{T|T \in \phi^*(T,\pi_T), T \in W_{(X,\pi)}\}$. Hence, $S \in \phi^*(X,\pi)$ and $\phi(X,\pi) \subset \phi^*(X,\pi)$. Similarly, $\phi^*(X,\pi) \subset \phi(X,\pi)$. Hence, $\phi(X,\pi) = \phi^*(X,\pi)$.

Step 5. ϕ^* is superior to any transition correspondence that is SE and minimalistic.

Proof. We prove this step by contradiction. Suppose ϕ^* is not superior to the SE and minimalistic transition correspondence $\hat{\phi}$. Then, there exists a game (N,π) such that $S,T \subset N$ where $S \in \phi^*(N,\pi)$ and $T \in \hat{\phi}(N,\pi)$ such that $\xi_i(T,\pi_T) \geq \xi_i(S,\pi_S)$ for some $i \in T \cap S$, and $T \notin \phi^*(N,\pi)$.

By step 3, since T is self-enforcing for $\hat{\phi}$, then it is also self-enforcing for ϕ^*. Therefore, $T \in Q(N,\pi) = \{L \mid L \in W_{(N,\pi)}, L \in \phi(L,\pi_L)\}$.

Since $T \notin \phi^*(N,\pi)$, then $T \notin \underset{M \in Q(N,\pi)\cup\{N\}}{\arg\max} R^{\xi}(M,\pi_M)$. Since $S \in \phi^*(N,\pi)$, then $S \in \underset{M \in Q(N,\pi)\cup\{N\}}{\arg\max} R^{\xi}(M,\pi_M)$. Since S and T are winning within N (by the definition of a transition correspondence), we have that $T \cap S \neq \emptyset$. Therefore, $R^{\xi}(S,\pi_S) > R^{\xi}(T,\pi_T)$, which implies that $\xi_i(S,\pi_S) > \xi_i(T,\pi_T)$ for $i \in T \cap S$. This is a contradiction.

■

C.2 Proofs of Section 3.3.3

Proof of Proposition 6

Proof. To prove part (i), we prove some ancillary steps. Suppose $(N,\pi) \notin \bar{G}$. Note that for any pair of coalitions $S,T \in UDC(N,\pi)$ where $|S| < |T|$, we have that $\underline{\lambda}^{(S,T)}(N,\pi) < \bar{\lambda}^{(S,T)}(N,\pi)$. This is because $\frac{\left(\frac{|T|-|S|}{|T||S|}\right)}{\underline{\pi}_i^{(S,T)}\left(\frac{\pi(S)-\pi(T)}{\pi(S)\pi(T)}\right)} > \frac{\left(\frac{|T|-|S|}{|T||S|}\right)}{\bar{\pi}_i^{(S,T)}\left(\frac{\pi(S)-\pi(T)}{\pi(S)\pi(T)}\right)} \Leftrightarrow \underline{\lambda}^{(S,T)}(N,\pi) < \bar{\lambda}^{(S,T)}(N,\pi)$.

Case 1: $\bar{\Lambda}^Q(S,\pi_S) > \underline{\Lambda}^Q(S,\pi_S)$ for all $Q \in UDC(S,\pi_S)$.

This illustrates the case when there is no range for λ where we can find a compromise coalition for the game (S,π_S). Suppose $\bar{\Lambda}(S,\pi_S) > \lambda > \underline{\Lambda}(S,\pi_S)$ and $V \in \phi(S,\pi_S)$ such that $|M| > |V| > |T|$ where M is the coalition of the largest size and T is the coalition of the smallest size in the set $UDC(S,\pi_S)$.

Suppose $\bar{\Lambda}(S,\pi_S) \geq \bar{\lambda}^{(M,V)}(S,\pi_S) > \lambda > \underline{\Lambda}(S,\pi_S) = \underline{\lambda}^{(M,V)}(S,\pi_S)$. Then, for the lowest-powered agent in the intersection of V and M (call him $\underline{i}$), we have that $\xi_{\underline{i}}(V,\pi_V) > \xi_{\underline{i}}(M,\pi_M)$ and there exist some agent $j \in V \cap M$, $j \neq \underline{i}$ such that $\xi_j(V,\pi_V) < \xi_j(M,\pi_M)$. Hence, by RAT we have that $V \notin \phi(S,\pi_S)$, a contradiction.

Case 2.1: $\lambda \notin \bigcup_{Q \in UDC(S,\pi_S)} \left[\bar{\Lambda}^Q(S,\pi_S), \underline{\Lambda}^Q(S,\pi_S)\right]$ and $\lambda > \underline{\Lambda}(S,\pi_S)$.

This examines the case where λ lies out of the interval of any compromise coalition but is larger than $\underline{\Lambda}(S,\pi_S)$. Suppose $M \in \phi(S,\pi_S)$ where $|M| > |Q|$.

If $\lambda < \bar{\Lambda}^Q(S,\pi_S)$, then for the highest-powered agent in the intersection of M and Q (call him $\bar{i}$), we have that $\xi_{\bar{i}}(M,\pi_M) > \xi_{\bar{i}}(Q,\pi_Q)$ and there exist some agent $j \in M \cap Q$, $j \neq \bar{i}$ such that $\xi_j(M,\pi_M) < \xi_j(Q,\pi_Q)$. Hence, by RAT we have that $M \notin \phi(S,\pi_S)$, a contradiction.

Case 2.2: $\lambda \notin \bigcup_{Q \in UDC(S,\pi_S)} \left[\bar{\Lambda}^Q(S,\pi_S), \underline{\Lambda}^Q(S,\pi_S)\right]$ and $\lambda < \bar{\Lambda}(S,\pi_S)$.

This illustrates the case where λ lies out of the interval of any compromise coalition but is lower than $\bar{\Lambda}(S,\pi_S)$ Suppose $M \in \phi(S,\pi_S)$ where $|M| < |Q|$.

If $\lambda > \underline{\Lambda}^Q(S,\pi_S)$, then for the lowest-powered agent in the intersection of M and Q (call him $\underline{i}$), we have that $\xi_{\underline{i}}(M,\pi_M) > \xi_{\underline{i}}(Q,\pi_Q)$ and there exist some agent $j \in M \cap Q$, $j \neq \underline{i}$ such that $\xi_j(M,\pi_M) < \xi_j(Q,\pi_Q)$. Hence, by RAT we have that $M \notin \phi(S,\pi_S)$, a contradiction.

Note that this set of games is well defined. That is, for any game $(S,\pi_S) \in \breve{G}(N,\pi)$ we can define $\underline{\Lambda}(S,\pi_S)$, $\bar{\Lambda}(S,\pi_S)$, $\underline{\Lambda}^Q(S,\pi_S)$ and $\bar{\Lambda}^Q(S,\pi_S)$ for $Q \in UDC(S,\pi_S)$. The intersection of the intervals $[0,\underline{\Lambda}(S,\pi_S)] \cup \left[\bar{\Lambda}(S,\pi_S),1\right] \cup [\bar{\Lambda}^Q(S,\pi_S),\underline{\Lambda}^Q(S,\pi_S)]$ for all $(S,\pi_S) \in \breve{G}(N,\pi)$ is nonempty.

Proof of part [ii]

Suppose $\lambda \in \left[\bar{\Lambda}(S,\pi_S),1\right]$ and that $T \in \breve{\phi}(S,\pi_S)$. Assume that $T \neq \arg\min_{V \in UDC(S,\pi_S)} |V|$. Then for the coalition $M \in UDC(S,\pi_S)$ where $M = \arg\min_{V \in UDC(S,\pi_S)} |V|$ we have that for all the agents $i \in M \cap T$ we have that $\xi_i(M,\pi_M) > \xi_i(T,\pi_T)$. Hence, by RAT, $T \notin \breve{\phi}(S,\pi_S)$, a contradiction.

Proof of part [iii]

Suppose $\lambda \in [0,\underline{\Lambda}(S,\pi_S)]$ and that $T \in \breve{\phi}(S,\pi_S)$. Assume that $T \neq \arg\min_{V \in UDC(S,\pi_S)} \pi(V)$. Then for the coalition $M \in UDC(S,\pi_S)$ where $M = \arg\min_{V \in UDC(S,\pi_S)} \pi(V)$ we have that for all the agents $i \in M \cap T$ we have that $\xi_i(M,\pi_M) > \xi_i(T,\pi_T)$. Hence, by RAT, $T \notin \breve{\phi}(S,\pi_S)$, a contradiction.

Proof of part [iv]

Suppose $\bar{\Lambda}^Q(S,\pi_S) \leq \lambda \leq \underline{\Lambda}^Q(S,\pi_S)$ and that $T \in \breve{\phi}(S,\pi_S)$. Assume that $T \notin \left\{Q \mid Q \in UDC(S,\pi_S), \bar{\Lambda}^Q(S,\pi_S) < \underline{\Lambda}^Q(S,\pi_S)\right\}$. Then for the coalition $M \in UDC(S,\pi_S)$ where $M \in \left\{Q \mid Q \in UDC(S,\pi_S), \bar{\Lambda}^Q(S,\pi_S) < \underline{\Lambda}^Q(S,\pi_S)\right\}$ we have that for all the agents $i \in M \cap T$ we have that $\xi_i(M,\pi_M) > \xi_i(T,\pi_T)$. Hence, by RAT $T \notin \breve{\phi}(S,\pi_S)$, a contradiction.

■

Proof of Corollary 1

Proof. Suppose for the game (S, π_S), we have coalitions $M, V \in UDC(S, \pi_S)$. Coalition M is preferred to coalition V for any value of $\lambda \in (0,1)$ if and only if

$$\xi_i(M, \pi_M) = \lambda \frac{1}{|M|} + (1-\lambda)\frac{\pi_i}{\pi(M)} > \lambda \frac{1}{|V|} + (1-\lambda)\frac{\pi_i}{\pi(V)} = \xi_i(V, \pi_V)$$

This only happens when $|M| < |V|$ **and** $\pi(M) < \pi(V)$.

Furthermore, if $T \in \breve{\phi}(S, \pi_S)$ such that $T = \underset{V \in UD(S,\pi_S)}{\arg\min}\ \pi(V) = \underset{V \in UDC(S,\pi_S)}{\arg\min}\ |V|$ then it is guaranteed that $\xi_i(T, \pi_T) > \xi_i(M, \pi_M)$ for any other $M \in UDC(S, \pi_S)$. ■

C.3 Proofs of Section 3.4.1

Proof of Proposition 7

Proof.

We show that ϕ^{**} satisfies RAT2. Let $T \in \phi^{**}(N, \pi)$ and let $Z \in W_{(N,\pi)}$ but $Z \notin \phi^{**}(N, \pi)$. By the definition of the transition correspondence ϕ^{**} we have that $R^{\xi}(Z, \pi_Z) < R^{\xi}(T, \pi_T)$. By the consistent ranking property of the sharing rule, we have that $\xi_i(Z, \pi_Z) < \xi_i(T, \pi_T)$.

To show uniqueness, we first show that $\phi^{**} \subseteq \underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M)$. Suppose $T \in \phi^{**}(S, \pi)$ but $T \neq \underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M)$. Then, for the coalition $V = \underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M)$, we have that for all $i \in T \cap V$, $\xi_i(V, \pi_V) > \xi_i(T, \pi_T)$. Hence, by RAT2 $T \notin \phi^{**}(S, \pi)$, a contradiction.

We then show that $\underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M) \subseteq \phi^{**}$. If $\underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M)$ is unique, because ϕ^{**} is nonempty (because the set of winning coalitions is nonempty) then we have that
$\phi^{**} = \underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M)$.

Suppose $\underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M)$ is not unique, that is, suppose $T, \tilde{T} \in \underset{M \in W_{(S,\pi)}}{\arg\max} R^{\xi}(M, \pi_M)$.
Suppose that $T \in \phi^{**}$ but $\tilde{T} \notin \phi^{**}$. Since $\tilde{T} \notin \phi^{**}$, there exists an agent $i \in T \cap \tilde{T}$ such that $\xi_i(\tilde{T}, \pi_{\tilde{T}}) < \xi_i(T, \pi_T)$. Hence, $R^{\xi}(T, \pi_T) > R^{\xi}(\tilde{T}, \pi_{\tilde{T}})$, a contradiction.

Hence, the transition correspondence is unique.

■

Proof of Corollary 2

Proof. This follows from the definition of the transition correspondence. We have that for equal sharing if $S \in \phi^{**}$ then $\{S|S \in W_{(T,\pi)}$ and $|S| < |V| \; \forall \; V \in W_{(T,\pi)}\}$ which is equivalent to the definition of MWCMS. The case for proportional sharing can be similarly proven.

■

C.4 Proofs of Section 3.4.2

Proof of Proposition 8

Proof. The proof of Part [i] is very similar to the proof of the steps in Proposition 6. As in the proof of that proposition, we examine the case where a compromise coalition does not exist and the case where a compromise coalition exists but λ lies outside the feasible intervals.

Case 1: $\bar{\Lambda}^Q(S,\pi) > \underline{\Lambda}^Q(S,\pi)$ for all $Q \in UDMW(S,\pi)$.

This illustrates the case when there is no range for λ where we can find a compromise coalition for the game (S,π). Suppose $\bar{\Lambda}(S,\pi) > \lambda > \underline{\Lambda}(S,\pi)$ and $V \in \phi(S,\pi)$ such that $|M| > |V| > |T|$ where M is the MWCMW and T is the MWCMS game (S,π).

Suppose $\bar{\Lambda}(S,\pi) \geq \bar{\lambda}^{(M,V)}(S,\pi) > \lambda > \underline{\Lambda}(S,\pi) = \underline{\lambda}^{(M,V)}(S,\pi)$. Then, for the lowest-powered agent in the intersection of V and M (call him $\underline{i}$), we have that $\xi_{\underline{i}}(V,\pi_V) > \xi_{\underline{i}}(M,\pi_M)$ and there exist some agent $j \in V \cap M$, $j \neq \underline{i}$ such that $\xi_j(V,\pi_V) < \xi_j(M,\pi_M)$. Hence, by RAT2 we have that $V \notin \phi(S,\pi)$, a contradiction.

Case 2.1: $\lambda \notin \bigcup_{Q \in UDMW(S,\pi)} \left[\bar{\Lambda}^Q(S,\pi), \underline{\Lambda}^Q(S,\pi)\right]$ and $\lambda > \underline{\Lambda}(S,\pi)$.

This illustrates the case where λ lies out of the interval of any compromise coalition but is larger than $\underline{\Lambda}(S,\pi)$. Suppose $M \in \phi(S,\pi)$ where $|M| > |Q|$.

If $\lambda < \bar{\Lambda}^Q(S,\pi)$, then for the highest-powered agent in the intersection of M and Q (call him $\bar{i}$), we have that $\xi_{\bar{i}}(M,\pi_M) > \xi_{\bar{i}}(Q,\pi_Q)$ and there exist some agent $j \in M \cap Q$, $j \neq \bar{i}$ such that $\xi_j(M,\pi_M) < \xi_j(Q,\pi_Q)$. Hence, by RAT2 we have that $M \notin \phi(S,\pi)$, a contradiction.

Case 2.2: $\lambda \notin \bigcup_{Q \in UDMW(S,\pi)} \left[\bar{\Lambda}^Q(S,\pi), \underline{\Lambda}^Q(S,\pi)\right]$ and $\lambda < \bar{\Lambda}(S,\pi)$.

This illustrates the case where λ lies out of the interval of any compromise coalition but is lower than $\bar{\Lambda}(S,\pi)$. Suppose $M \in \phi(S,\pi)$ where $|M| < |Q|$.

If $\lambda > \underline{\Lambda}^Q(S,\pi)$, then for the lowest-powered agent in the intersection of M and Q (call him $\underline{i}$), we have that $\xi_{\underline{i}}(M,\pi_M) > \xi_{\underline{i}}(Q,\pi_Q)$ and there exist some agent $j \in M \cap Q$, $j \neq \underline{i}$ such that $\xi_j(M,\pi_M) < \xi_j(Q,\pi_Q)$. Hence, by RAT2 we have that $M \notin \phi(S,\pi)$, a contradiction.

Proof of part [ii]

Suppose $\lambda \in \left[\bar{\Lambda}(S,\pi), 1\right]$ and that $T \in \tilde{\phi}(S,\pi_S)$. Assume that $T \neq \underset{V \in MW_{(S,\pi)}}{\arg\min} |V|$. Then for the coalition $M \in MW_{(S,\pi)}$ where $M = \underset{V \in UDMW(S,\pi)}{\arg\min} |V|$ we have that for all the agents $i \in M \cap T$ we have that $\xi_i(M,\pi_M) > \xi_i(T,\pi_T)$. Hence, by RAT2, $T \notin \tilde{\phi}(S,\pi)$, a contradiction.

Proof of part [iii]

Suppose $\lambda \in [0, \underline{\Lambda}(S,\pi)]$ and that $T \in \tilde{\phi}(S,\pi)$. Assume that $T \neq \underset{V \in UDMW(S,\pi)}{\arg\min} \pi(V)$. Then for the coalition $M \in MW_{(S,\pi)}$ where $M = \underset{V \in UDMW(S,\pi)}{\arg\min} \pi(V)$ we have that for all the agents $i \in M \cap T$ we have that $\xi_i(M,\pi_M) > \xi_i(T,\pi_T)$. Hence, by RAT2, $T \notin \tilde{\phi}(S,\pi)$, a contradiction.

Proof of part [iv]

Suppose $\bar{\Lambda}^Q(S,\pi) \leq \lambda \leq \underline{\Lambda}^Q(S,\pi)$ and that $T \in \tilde{\phi}(S,\pi)$. Assume that $T \notin \left\{Q \mid Q \in UDMW(S,\pi), \bar{\Lambda}^Q(S,\pi) < \underline{\Lambda}^Q(S,\pi)\right\}$. Then for the coalition $M \in UDMW(S,\pi)$ where $M \in \left\{Q \mid Q \in UDMW(S,\pi), \bar{\Lambda}^Q(S,\pi) < \underline{\Lambda}^Q(S,\pi)\right\}$ we have that for all the agents $i \in M \cap T$ we have that $\xi_i(M,\pi_M) > \xi_i(T,\pi_T)$. Hence, by RAT2 $T \notin \tilde{\phi}(S,\pi)$, a contradiction.

■

Proof of Corollary 3

Proof. Suppose for the game (S, π), we have coalitions $M, V \in W_{(S,\pi)}$. Coalition M is preferred to coalition V for any value of $\lambda \in (0, 1)$ if and only if

$$\xi_i(M, \pi_M) = \lambda \frac{1}{|M|} + (1 - \lambda) \frac{\pi_i}{\pi(M)} > \lambda \frac{1}{|V|} + (1 - \lambda) \frac{\pi_i}{\pi(V)} = \xi_i(V, \pi_V)$$

This only happens when $|M| < |V|$ **and** $\pi(M) < \pi(V)$.

Furthermore, if $T \in \tilde{\phi}\,(S, \pi)$ such that $T = \underset{V \in W_{(S,\pi_S)}}{\arg\min}\ \pi(V) = \underset{V \in W_{(S,\pi_S)}}{\arg\min}\ |V|$ then it is guaranteed that $\xi_i(T, \pi_T) > \xi_i(M, \pi_M)$ for any other $M \in W_{(S,\pi_S)}$. ■

Appendix D SCREENSHOT OF EXPERIMENT INSTRUCTIONS AND QUIZ

The instructions for the experiment can be viewed here: `https://www.youtube.com/watch?v=FKc0lPAQN_0`. The following are screenshots of the instructions.

The video begins with the introduction frame.

Figure D.1: Screenshot of the instruction video

The instruction then proceeds to describe the proposal stage.

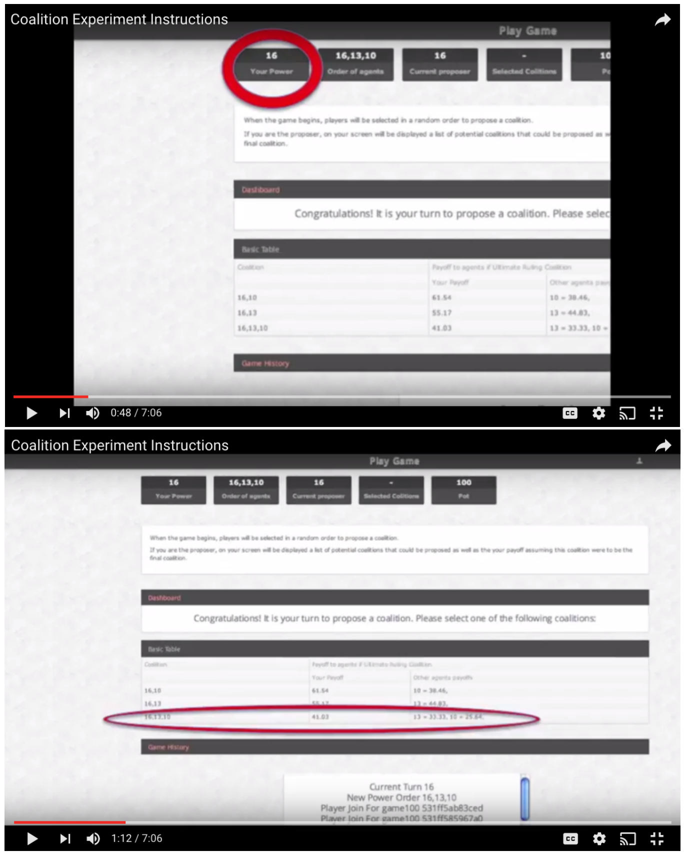

Figure D.2: Screenshot of the instruction video: proposal stage

The response stage instructions are shown next.

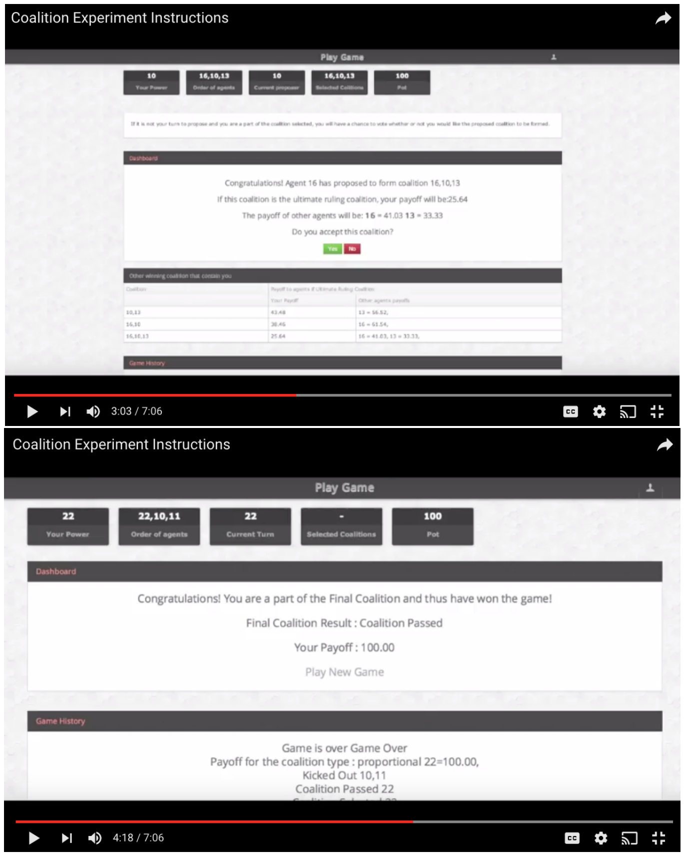

Figure D.3: Screenshot of the instruction video: response stage

After showing the instructions for the game, subjects were given an online quiz to test their knowledge of the general instructions. There are five questions in the quiz corresponding to the different different parameters of the game (e.g. subjects own power), the payoffs, the order of proposals, and the different stages of the game. The subject will not be able to proceed to take part in the experiment if he has not finished the quiz. The following are the screenshots of the questions:

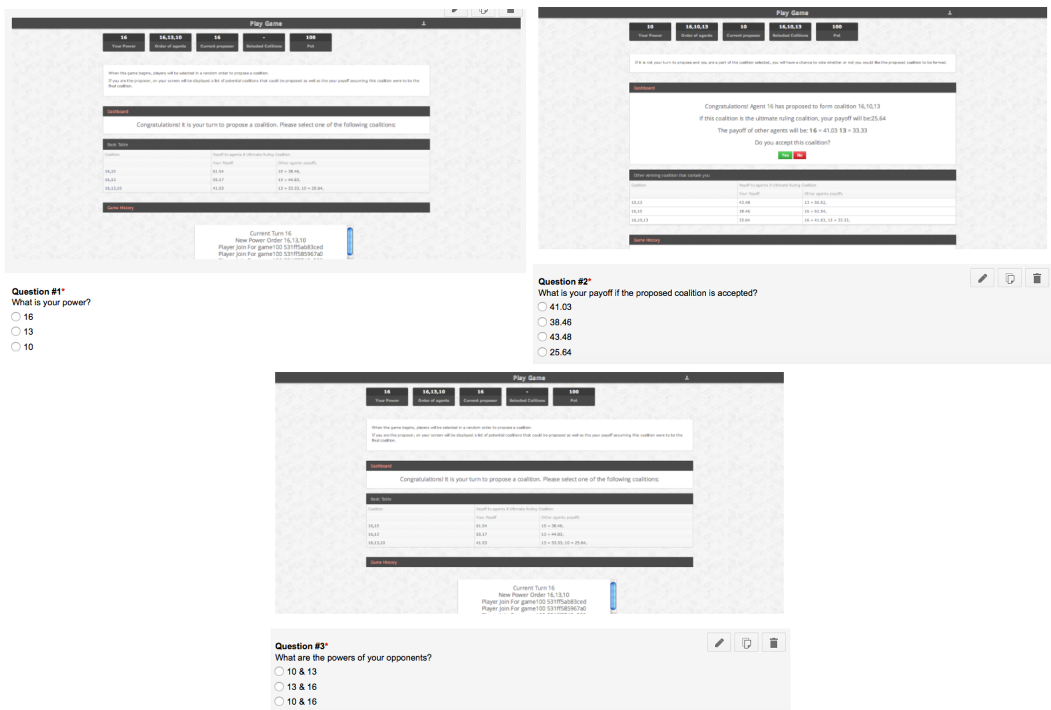

Figure D.4: Screenshot of the quiz: questions1-3

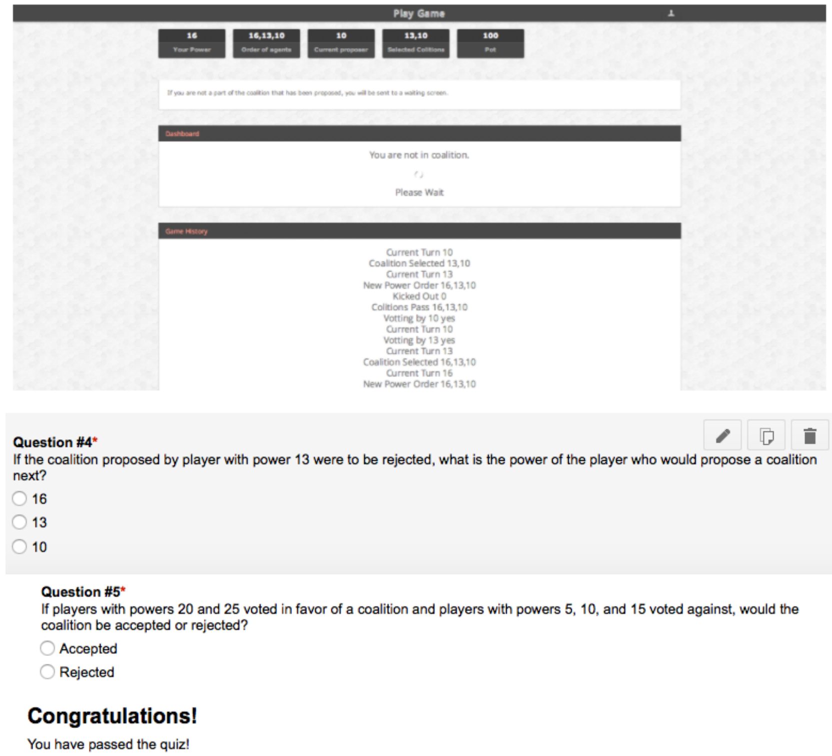

Figure D.5: Screenshot of the quiz: questions 4 and 5

REFERENCES

[1] D. Acemoglu, G. Egorov, and K. Sonin. Coalition Formation in Non-Democracies. *Review of Economic Studies*, 75(4):987–1009, 2008.

[2] D. Acemoglu, G. Egorov, and K. Sonin. Do Juntas Lead to Personal Rule? *American Economic Review: Papers and Proceedings*, 99(2):298–303, 2009.

[3] D. Acemoglu, M. Golosov, and A. Tsyvinski. Power Fluctuations and Political Economy. *Journal of Economic Theory*, 146(3):1009–1041, 2011.

[4] R. Aumann. Agreeing to Disagree. *The Annals of Statistics*, pages 1236–1239, 1976.

[5] K.H. Baik. Winner-Help-Loser Group Formation in Rent-Seeking Contests. *Economics and Politics*, 6(2):147–162, 1994.

[6] A. Balisacan and J. Roumasset. Public Choice of Economic Policy: the Growth of Agricultural Protection. *Weltwirschaftliches Archiv*, 123(2):232–248, 1987.

[7] D. Baron and J. Ferejohn. Bargaining in Legislatures. *The American Political Science Review*, 83(4):1181–1206, 1989.

[8] R. Barro. The Control of Politicians: An Economic Model. *Public Choice*, 14(1):19–42, 1973.

[9] B. Bartling and F.A. von Siemens. Equal Sharing Rules in Partnerships. *Journal of Institutional and Theoretical Economics*, 166(2):299–320, 2010.

[10] M. Battaglini and T. Palfrey. The Dynamics of Distributive Politics. *Economic Theory*, 49(3):739–777, 2012.

[11] T. Besley and S. Coate. An Economic Model of Representative Democracy. *The Quarterly Journal of Economics*, 112(1):85–106, 1997.

[12] K. Binmore. Why Experiment in Economics? *Economic Journal*, 109(453):F16–F24, 1999.

[13] K. Binmore. *Game Theory: A Very Short Introduction*. Oxford University Press, Oxford, 2007.

[14] F. Bloch. Group and Network Formation in Industrial Organization: A Survey. In G. Demange and M. Wooders, editors, *Group Formation in Economics: Networks, Clubs and Coalitions*. Cambridge University Press, Cambridge, MA, 2005.

[15] F. Bloch and B. Dutta. Formation of Networks and Coalitions. *Mimeo University of Warwick*, 2010.

[16] A. Bogomolnaia and M. Jackson. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

[17] G. Bolton and A. Ockenfels. ERC: A Theory of Equity, Reciprocity and Competition. *American Economic Review*, 90(1):166–193, 2000.

[18] B. Bueno de Mesquita, A. Smith, R. Siverson, and J. Morrow. *The Logic of Political Survival*. The MIT Press, Cambridge MA, 2003.

[19] L. Cameron. Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia. *Economic Inquiry*, 37(1):47–59, 1999.

[20] C. Carraro. Environmental Conflict, Bargaining and Cooperation. In J. Van den Bergh, editor, *The Handbook of Environmental and Resource Economics*. E. Elgar, Cheltenham, 1999.

[21] C. Carraro. Institution Design for Managing Global Commons: Lessons from Coalition Theory. In G. Demange and M. Wooders, editors, *Group Formation in Economics: Networks, Clubs and Coalitions*. Cambridge University Press, Cambridge, MA, 2005.

[22] M. Chwe. Farsighted Coalitional Stability. *Journal of Economic Theory*, 63(2):299–325, 1994.

[23] J. Conley and S. Smith. Finite Decentralization in a Tiebout Economy with Crowding Types. *Social Choice and Welfare*, 20(1):49–75, 2003.

[24] B. Copeland and M. Taylor. Trade, Tragedy and the Commons. *American Economic Review*, 99(3):725–749, 2009.

[25] G. Debreu and H. Scarf. A Limit Theorem on the Core of an Economy. *International Economic Review*, 4(3):235–246, 1963.

[26] R. Deneckere and C. Davidson. Incentives to Form Coalitions with Bertrand Competition. *RAND Journal of Economics*, 16(4):473–486, 1985.

[27] E. Diamantoudi and L. Xue. Farsighted Stability in Hedonic Games. *Social Choice and Welfare*, 21(1):39–61, 2003.

[28] A. Dixit. *Lawlessness and Economics: Alternative Modes of Governance*. Princeton University Press, Princeton, NJ, 2004.

[29] L. Ehlers. Coalitional Strategy-proof House Allocation. *Journal of Economic Theory*, 105(2):298–317, 2002.

[30] O. Erixson and H. Ohlsson. Estate Division: Equal Sharing as Choice, Social Norm, and Legal Requirement. *IFN Working Paper No. 1006*, 2014.

[31] M. Espinosa and I. Macho-Stadler. Endogenous Formaton of Competing Partnerships with Moral Hazard. *Games and Economic Behavior*, 44(1):183–194, 2003.

[32] R. Fabella. Natural Team Sharing and Team Productivity. *Economics Letters*, 27(2):105–110, 1988.

[33] R. Fabella. Generalized Sharing, Membership Size and Pareto Efficiency in Teams. *Theory and Decision*, 48(1):47–60, 2000.

[34] J. Farrell and C. Shapiro. Horizontal Mergers: An Equilibrium Analysis. *American Economic Review*, 80(1):107–126, 1990.

[35] J. Ferejohn. Incumbent Performance and Electoral Control. *Public Choice*, 50(1):5–25, 1986.

[36] D. Foley. Resource Allocation and the Public Sector. *Yale Economic Essays*, 7(1):43–98, 1967.

[37] G. Frechette, J. Kagel, and M. Morelli. Nominal Bargaining Power, Selection Protocol, and Dicounting in Legislative Bargaining. *Journal of Public Economics*, 89(8):1497–1517, 2005.

[38] D. Fudenberg and D. Levine. *The Theory of Learning in Games*. The MIT Press, Cambridge MA, 1998.

[39] D. Gillies. Solutions to General Non-Zero-Sum Games. In A.Tucker and D. Luce, editors, *Contributions to the Theory of Games IV*. Princeton University Press, Princeton, 1959.

[40] R. Gillies. *The Cooperative Game Theory of Networks and Hierarchies*. Springer-Verlag, Berlin, 2010.

[41] B. Greiner. Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125, 2015.

[42] E. Grinols. An Extension of the Kemp-Wan Theorem on the Formation of Customs Unions. *Journal of International Economics*, 11(2):259–266, 1981.

[43] G. Harrison and J. List. Field Experiments. *Journal of Economic Literature*, 42(4):1009–1055, 2004.

[44] B. Holmstrom. Moral Hazard in Teams. *Bell Journal of Economics*, 13(2):324–340, 1982.

[45] A. Horowitz. A Test of the Core, Bargaining Set, Kernel, and Shapley Models in n-person Quota Games with One Weak Player. *Theory and Decision*, 8(1):49–65, 1977.

[46] K. Hyndman, E. Ozbay, A. Schotter, and W. Ehrblatt. Convergence: An Experimental Study of Teaching and Learning in Repeated Games. *Journal of the European Economic Association*, 10(3):573–604, 2012.

[47] K. Jandoc and R. Juarez. Self-enforcing Coalitions with Power Accumulation. *International Journal of Game Theory, forthcoming*.

[48] K. Jandoc and R. Juarez. The Effect of Sharing Rules on the Formation of Coalitions. *Mimeo University of Hawai'i*, 2015.

[49] K. Jandoc, R. Juarez, and J. Roumasset. Towards an Economics of Irrigation Networks. In K. Burnett, K. Howitt, J. Roumasset, and C.A. Wada, editors, *Handbook of Water Economics and Institutions*. Routledge, New York, 2015.

[50] J.S. Jordan. Pillage and Property. *Journal of Economic Theory*, 131(1):26–44, 2006.

[51] R. Juarez. Three Equilibria in Tournament Games with Externalities. *Mimeo University of Hawaii at Manoa*, 2012a.

[52] R. Juarez. The No-Threat Equilibrium in Tournament Games with Externalities. *Mimeo University of Hawaii at Manoa*, 2012b.

[53] R Juarez. Group Strategyproof Cost Sharing: The Role of Indifferences. *Games and Economic Behavior*, 82:218–239, 2013.

[54] S. Kachelmeier and M. Shehata. Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the Peoples Republic of China. *American Economic Review*, 82(5):1120–1141, 1992.

[55] J.and R. Helwig Kahan. Coalitions: A System of Programs for Computer-Controlled Bargaining Games. *General Systems*, 16(3):31–41, 1971.

[56] H. Konishi and D. Ray. Coalition Formation as a Dynamic Process. *Journal of Economic Theory*, 110(1):1–41, 2003.

[57] C. Kowalczyk and T. Sjostrom. Bringing GATT into the Core. *Economica*, 61(243):301–317, 1994.

[58] S. Lee. Inter-Group Competition For A Pure Private Rent. *The Quarterly Review of Economics and Finance*, 33(3):251–266, 1993.

[59] S. Lee. Endogenous Sharing Rules in Collective-Group Rent Seeking. *Public Choice*, 85(1):31–44, 1995.

[60] G. Libecap. *Contracting for Property Rights*. Cambridge University Press, Cambridge, MA, 1989.

[61] S. Littlechild and G. Owen. A Simple Expression for the Shapley Value in a Special Case. *Management Science*, 20(3):370–372, 1973.

[62] M. McGuire and M. Olson. The Economics of Autocracy and Majority Rule: The Invisible Hand and the Use of Force. *Journal of Economic Literature*, 34(1):72–96, 1996.

[63] R. McKelvey. An Experimental Test of a Stochastic Game Model of Committee Bargaining. In T. Palfrey, editor, *Laboratory Research in Political Economy*. University of Michigan Press, Ann Arbor, 1991.

[64] F. Mengel. Learning by (Limited) Forward Looking Players. *Journal of Economic Behavior and Organization*, 108:59–77, 2014.

[65] H. Moulin. Equal or Proportional Division of a Surplus and Other Methods. *International Journal of Game Theory*, 16(3):161–186, 1987.

[66] H. Moulin. *Axioms of Coperative Decision Making*. Cambridge University Press, Cambridge MA, 1988.

[67] H. Moulin. An Application of the Shapley Value to Fair Division with Money. *Econometrica*, 60(6):1331–1349, 1992.

[68] R. Nagel and T. Fang. Experimental Results on the Centipede Game in Normal Form: an Investigation on Learning. *Journal of Mathematical Psychology*, 42(2):356–384, 1998.

[69] J. Nash. The Bargaining Problem. *Econometrica*, 18(2):155–162, 1950.

[70] J. Newhouse. The Economics of Group Practice. *Journal of Human Resources*, 8(1):37–56, 1973.

[71] S. Nitzan. Collective Rent Dissipation. *Economic Journal*, 101(409):1522–1534, 1991.

[72] S. Nitzan. Modelling Rent Seeking Contests. *European Journal of Political Economy*, 10(1):41–60, 1994.

[73] V. Nocke. Cartel Stability Under Capacity Constraints: The Traditional View Restored. *STICERD Discussion Paper EI/23, London School of Economics*, 1999.

[74] M. Olson. Dictatorship, Democracy, and Development. *The American Political Science Review*, 87(3):567–576, 1993.

[75] E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge, MA, 1990.

[76] S. Papai. Unique Stability in Simple Coalition Formation Games. *Games and Economic Behavior*, 48(2):337–354, 2004.

[77] M. Perry and R. Porter. Oligopoly and the Incentive for Horizontal Merger. *American Economic Review*, 75(1):219–227, 1985.

[78] G. Roland Persson, T. and G. Tabellini. Separation of Powers and Political Accountability. *The Quarterly Journal of Economics*, 112(4):1163–1202, 1997.

[79] G. Roland Persson, T. and G. Tabellini. Comparative Politics and Public Finance. *Journal of Political Economy*, 108(6):1121–1161, 2000.

[80] M. Piccione and R. Razin. Coalition Formation under Power Relations. *Theoretical Economics*, 4(1):1–15, 2009.

[81] A. Rapoport and J. Kahan. When Three is Not Always Two Against One: Coalitions in Experimental Three-person Cooperative Games. *Journal of Experimental Social Psychology*, 12(3):253–273, 1976.

[82] D. Ray. *A Game-Theoretic Perspective on Coalition Formation*. Oxford University Press, New York, 2007.

[83] D. Ray and R. Vohra. Coalition Formation. In P.Young and S. Zamir, editors, *Handbook of Game Theory, Volume 4*. North-Holland, Amsterdam, 2013.

[84] D. Ray and R. Vohra. The Farsighted Stable Set. *Econometrica*, 83(3):977–1011, 2015.

[85] U. Reinhardt, M. Pauly, and P. Held. Financial Incentives and the Economic Performance of Group Practitioners. In P. Held and U. Reinhardt, editors, *Analysis of Economic Performance in Medical Group Practices*. Princeton, New Jersey, 1979.

[86] R. Riezman. Customs Union and the Core. *Journal of International Economics*, 19(3):355–365, 1985.

[87] A. Roth. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy*, 92(6):991–1016, 1984.

[88] C. Rowat and M. Kerber. Sufficient Conditions for Unique Stable Sets in Three Agent Pillage Games. *Mathematical Social Sciences*, 69(3):69–80, 2014.

[89] A. Rubinstein. Perfect Equilibrium in a Bargaining Model. *Econometrica*, 50(1):97–109, 1982.

[90] P. Samuelson. The Pure Theory of Public Expenditure. *Review of Economics and Statistics*, 36(4):387–389, 1954.

[91] T. Sandler. *Collective Action: Theory and Applications*. University of Michigan Press, Ann Arbor, 1992.

[92] P. Sekeris. Endogenous Elites: Power Structure and Patron-Client Relationships. *Economics of Governance*, 12(3):237–258, 2011.

[93] R. Selten. Axiomatic Characterization of the Quadratic Scoring Rule. *Experimental Economics*, 1(1):43–62, 1998.

[94] A. Sen. Labour Allocation in a Cooperative Enterprise. *Review of Economic Studies*, 33(4):361–371, 1966.

[95] L. Shapley. A Vvalue for n-Person Games. In A.Tucker and D. Luce, editors, *Contributions to the Theory of Games II*. Princeton University Press, Princeton, 1953.

[96] R. Slonim and A. Roth. Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic. *Econometrica*, 66(3):569–596, 1998.

[97] W. Thomson. Axiomatic and Game-theoretic Analysis of Bankruptcy and Taxation Problems: a Survey. *Mathematical Social Science*, 45(3):249–297, 2003.

[98] W. Thomson. On the Existence of Consistent Rules to Adjudicate Conflicting Claims: a Constructive Geometric Approach. *Review of Economic Design*, 11(3):225–251, 2007.

[99] C. Tiebout. A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64(5):416–424, 1956.

[100] J. Tremewan and C. Vanberg. The Dynamics of Coalition Formation: A Multilateral Bargaining Experiment With Free Timing of Moves. *Univeristy of Heidelberg Department of Economics Discussion Paper Series No. 582*, 2015.

[101] G. Tullock. *Autocracy*. Kluwer Academic Publishers, Dorderecht, 1987.

[102] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 1944.

[103] M. Wooders. Equilibria, the Core and Jurisdiction Structures in Economies with a Local Public Good. *Journal of Economic Theory*, 18(2):328–348, 1978.

[104] M. Wooders and F. Page. Coalitions. In S.Durlauf and L. Blume, editors, *The New Palgrave Dictionary of Economics (Second Edition)*. Palgrave Macmillan, London, 2008.

[105] L. Xue. Coalitional Stability Under Perfect Foresight. *Economic Theory*, 11(3):603–627, 1998.