

# Power indices when players can commit to reject coalitions\*

László Á. Kóczy<sup>†</sup>

## Abstract

Power indices have been used to evaluate the allocation of power in a wide range of voting situations. While they use the language of game theory known measures of a priori voting power are hardly more than statistical expectations assuming the random behaviour of the players. We introduce a model where players can reject certain partnerships in cooperation. For normalised indices strategic rejection may increase power. Our notion of a strategic power index is well defined if power is measured by an index that takes only minimal winning coalitions into account.

**Keywords and phrases:** quarrelling, rejected coalitions, a priori voting power, power indices, minimal winning coalitions, rational players.

**JEL codes:** C71, D71.

---

\*An earlier version of this paper was circulated under the title “Strategic power indices: Quarrelling in coalitions”. The author is grateful for discussions with Manfred Holler and Stefan Napel; he also thanks funding by OTKA (Hungarian Scientific Research Fund) for the project NF-72610 and the support of the Hungarian Academy of Sciences under its Momentum programme (LD-004/2010).

<sup>†</sup>Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budaörsi 45., H-1112 Budapest and Keleti Faculty of Economics, Óbuda University, Tavaszmező 15-17. H-1084 Budapest. Email: [koczy@krtk.mta.hu](mailto:koczy@krtk.mta.hu)

# 1 Introduction

A power index measures the statistical probability of a voter being instrumental in a voting situation. Power indices have been used to evaluate the distribution of power in currency unions (Berger and Mueller, 2007), the International Monetary Fund (Leech, 2002), the Bretton Woods institutions (Leech and Leech, 2005), the European Union (Leech, 2002; Kóczy, 2012) or the US Supreme Court (Kaniowski and Leech, 2009). Several such indices have been defined and studied by game theory since Shapley and Shubik (1954) defined the Shapley-Shubik index by applying the Shapley value to simple games, despite the fact that there is nothing “game theoretical,” strategic about them.

In cooperative games strategies are implicit, but players do maximise their payoffs. The fairness represented by *values* assumes unconditional cooperation. There is nothing wrong with fairness in problems of cost sharing or risk allocation, but when we turn to voting situations, this approach seems less fit. The current power indices seem to measure voting *luck*.

So what *is* power? Power is the “ability to act or produce an effect.” (Definition by the Merriam-Webster online dictionary.) Strategic decisions that increase a voter’s ability to make decisions are a manifestation of power. What are the strategic decisions a voter can take? Kilgour (1974, 1977) introduced the idea of *quarrelling* to voting games and axiomatised the Shapley value for games with a quarrelling subset of players. Quarrelling players refuse to cooperate, two such players refuse to join the same coalition so a coalition may contain at most one quarrelling player. As the paradox of quarrelling members (Brams, 1975) illustrates, two players may mutually benefit from their quarrelling. Highly artificial these may seem (Laruelle and Valenciano, 2005), the paradoxes of voting power arise in real voting situations (van Deemen and Rusinowska, 2003).

In this paper we allow for similar strategic decisions, but as opposed to quarrelling, coalitions can be rejected unilaterally.

Should all winning coalitions form? The structure of communication (Myerson, 1977, 1980; Faigle and Kern, 1992) or physical or ideological position (Bilbao, Jiménez, and López, 1998) may put restrictions on the set of feasible coalitions. Would all coalitions form in the absence of such restrictions? When a player is offered to join a winning coalition, accepting will clearly cause no harm, although a voter may turn down membership in a coali-

tion where she gets no, or too little credit for membership expecting more elsewhere. A priori, however, she may strictly benefit from a strategic commitment to *refuse* to join this coalition.

Our model formalises this possibility: we augment voting games with a previous stage, where players can choose the coalitions they want to join in the voting game. This is a simple noncooperative game, where “the acquisition of power is the payoff” (Shapley, 1962, p. 59). We show that all known normalised indices are affected by such strategic behaviour.

The use of the word *strategic* in the context of power indices is not entirely new. There is a stream of literature who consider tactical voting and define power indices for such situations (Steunenberg, Schmidtchen, and Koboldt, 1999; Napel and Widgrén, 2010).

The structure of the paper is as follows. We start with a brief introduction to voting games and an overview of the known indices. We briefly explain the paradox of quarrelling members, introduce a framework for strategic indices and prove that our definition of strategic power index is well defined for a wide class power indices.

## 2 Power indices

A voting situation or voting game is a pair  $(N, \mathcal{W})$ , where  $N$  is the set of voters and  $\mathcal{W} \subseteq 2^N$  denotes the set of *winning coalitions*. A game is simple if

1.  $\emptyset \notin \mathcal{W}$ ,
2.  $N \in \mathcal{W}$
3. if  $C \subset D$  and  $C \in \mathcal{W}$ , then  $D \in \mathcal{W}$
4. If  $C \in \mathcal{W}$  and  $D \in \mathcal{W}$ , then  $C \cap D \neq \emptyset$ .

Condition 1 states that the empty set is not winning, while the coalition of all players is winning. Condition 3 is a monotonicity property: if a coalition is winning additional members cannot make it losing. Condition 4 requires the game to be *proper*, that is, no two subsets of the voters can make decisions simultaneously.

## 2.1 Convex voting games

We consider *convex* voting games without the assumption that the grand coalition is winning.

1.  $\emptyset \notin \mathcal{W}$ ,
2. if  $C \subset D \subset E$  and  $C, E \in \mathcal{W}$ , then  $D \in \mathcal{W}$
3. If  $C \in \mathcal{W}$  and  $D \in \mathcal{W}$ , then  $C \cap D \neq \emptyset$ .

Condition 2 is a *convexity* condition on the poset formed by the winning coalitions. The Condition 2 replaces Condition 3, but the two coincide if  $N \in \mathcal{W}$ . Let  $\Gamma$  denote the collection of proper convex voting games satisfying the above properties. Note that our model permits the trivial game  $(N, \emptyset)$ .

A player  $i$  is *critical* in the winning coalition  $C \in \mathcal{W}$  if  $C \setminus \{i\} \notin \mathcal{W}$ . Let  $k^C(\mathcal{W})$  denote the number of critical players in coalition  $C$  when the set of winning coalitions is  $\mathcal{W}$ . When this does not lead to confusion, we drop the reference to  $\mathcal{W}$ .

Let  $\mathcal{M} \subseteq \mathcal{W}$  denote the set of *minimal winning coalitions*: the set of coalitions without proper winning subsets. This implies that if  $C \in \mathcal{M}$  and  $i \in C$ , then  $C \setminus \{i\} \notin \mathcal{W}$ . *Surplus coalitions* are winning, but non-minimal. All members of a minimal winning coalition are critical.

contributes to a decision provided that a decision has been reached.

Given a game  $\Gamma$  a *power measure*  $\kappa : \Gamma \rightarrow \mathbb{R}_+^N$  assigns to each player  $i$  a non-negative real number  $\kappa_i$ , its *power*; a power measure  $\kappa$  satisfying  $\sum_{i \in N} \kappa_i = 1$  is called a *power index*. The 0-game  $(N, \emptyset)$  is an exception as here each player has a power 0.

## 2.2 A characterisation

In the following we introduce a class of power indices  $\kappa$ , that can be expressed as follows:

$$\kappa_i = \sum_{C \in 2^N \setminus \emptyset} a^C \mu_i^C, \text{ where } \sum_{C \in 2^N \setminus \emptyset} a^C = 1, \quad (2.1)$$

that is, the power of a player can be expressed as a weighted average of her *power or contribution*  $\mu_i^C$  in the (winning) coalitions  $C \in 2^N \setminus \emptyset$  weighted by the (relative) impact  $a^C \geq 0$  of a coalition  $C$ . Note that  $a^C$  accounts for both the importance and the success of a coalition.

This class is sufficiently large to include all known indices. Moreover we can only talk about contributions by members, so  $\mu_i^C = 0$  when  $i \notin C$  and these indices recognise the contribution of *critical* players only, but for these equally.

$$\mu_i^C = \begin{cases} \frac{1}{k^C} & \text{if } i \text{ is critical} \\ \frac{1}{|C|} & \text{if no } i \in C \text{ is critical,} \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

This also means that the differences are limited to the weights  $a_C$ .

Let  $\mathcal{K} = \{C \in \mathcal{W} | k^C > 0\}$  denote the set of winning coalitions with at least one critical player. Clearly  $\mathcal{M} \subseteq \mathcal{K} \subseteq \mathcal{W}$ . Then we have

$$a_\phi^C = \frac{(|C| - 1)! k^C (n - |C|)!}{n!}. \quad (2.3)$$

for the *Shapley-Shubik index* (Shapley and Shubik, 1954),

$$a_\beta^C = \frac{k^C}{\sum_{C \in \mathcal{W}} k^C}. \quad (2.4)$$

for the *Banzhaf index* (Penrose, 1946; Banzhaf, 1965),

$$a_\gamma^C = \begin{cases} \frac{1}{|\mathcal{K}|} & \text{if } C \in \mathcal{K} \\ 0 & \text{otherwise.} \end{cases} \quad (2.5)$$

for the *Johnston index*  $\gamma$  (Johnston, 1978),

$$a_\rho^C = \begin{cases} \frac{1}{|\mathcal{M}|} & \text{if } C \in \mathcal{M} \\ 0 & \text{otherwise.} \end{cases} \quad (2.6)$$

for the *Deegan-Packel index*  $\rho$  (Deegan and Packel, 1978), and

$$a_h^C = \begin{cases} \frac{k^C}{\sum_{C \in \mathcal{M}} k^C} & \text{if } C \in \mathcal{M} \\ 0 & \text{otherwise.} \end{cases} \quad (2.7)$$

for the Holler-Packel or *Public Good Index*  $h$  (Holler and Packel, 1983).

### 3 Strategic power indices

There may be various power indices, but they all work with an exogenously given set  $\mathcal{W}$  of winning coalitions and it is implicitly assumed that all these coalitions will form. This seems indeed natural – why would players give up part of their power? If, for instance, two players start to “quarrel” and refuse to cooperate making any coalition they both belong to losing, their power should decrease. Not necessarily. In the so-called “Paradox of Quarrelling Members” (Kilgour, 1974; Brams, 1975) two players mutually benefit from refusing to cooperate with each other.

Paradoxical or not is a matter of interpretation, but players can certainly *acquire* a larger share of power by approving/rejecting coalitions. In this paper we extend voting games to allow for strategic considerations where player  $i$  can specify an arbitrary subset of  $2^{N \setminus \{i\}}$  they reject. We do not claim that rejected coalitions will not form, but that these should not be taken into account when calculating voting power Braham and Holler (See also 2005, on the understanding of *power*). We define strategic power indices by applying power indices to this reduced set of winning coalitions.

#### 3.1 Examples

As a motivation we present a number of games based on *weighted voting*. Here  $N$  is a collection of  $n$  interest groups, or *parties* having  $w_1, w_2, \dots, w_n$  individual representatives ( $w_i \in \mathbb{N}_+$ ). Let  $w = \sum_{i=1}^n w_i$ . We assume that a quota of  $w \geq q > w/2$  is *required* to pass a bill. A coalition  $C$  of parties is winning if and only if  $\sum_{i \in C} w_i \geq q$ . Since  $w \geq q$  and  $w_i \geq 0$  weighted voting games are simple and proper.

**Example 1.** The game  $G_1$  consists of four players represented by their weights (subscripts distinguish players with identical weights from each other):  $3_1, 3_2, 2_1, 2_2$  and voting has a quota of 6. The set winning coalitions is  $\mathcal{W} = \{\underline{3_1 3_2}, \underline{3_1 3_2 2_1}, \underline{3_1 3_2 2_2}, \underline{3_1 2_1 2_2}, \underline{3_2 2_1 2_2}, 3_1 3_2 2_1 2_2\}$  (with critical players underlined). The Banzhaf index is  $\beta = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}\}$ .

Notice that in coalition  $3_1 3_2 2_1$  player  $2_1$  is not critical, while the two larger players are. If  $2_1$  can prevent the formation of this coalition, the latter are critical in fewer coalitions, so in relative terms (in a power *index*)  $2_1$  gains.

Given  $\mathcal{W}' = \{\underline{3_1 3_2}, \underline{3_1 3_2 2_2}, \underline{3_1 2_1 2_2}, \underline{3_2 2_1 2_2}, 3_1 3_2 2_1 2_2\}$  the recalculated Banzhaf index is  $\beta' = \{\frac{3}{10}, \frac{3}{10}, \frac{1}{5}, \frac{1}{5}\}$ . Player  $2_1$ 's rejection increased its relative

power. It is therefore *not* in player  $2_1$ 's interest to join every winning coalition it is invited to. This finding is not really surprising. In coalition  $3_13_22_1$  player  $2_1$  assisted players  $3_1$  and  $3_2$  in forming a winning coalition, but without getting any credit for it.

Similarly, we find that it is not in player  $2_2$ 's interest to join  $3_13_22_2$  and we get  $\mathcal{W}'' = \mathcal{M} = \{\underline{3_13_2}, \underline{3_12_12_2}, \underline{3_22_12_2}, 3_13_22_12_2\}$  and  $\beta^* = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$ .

where they are not critical. By this they do not lose power, on the other hand, as the absolute powers of other, in the coalition critical players decrease, relative power increases.

Minimal winning coalitions may also be subject to rejections:

**Example 2.**  $G_2$  is a 9-player game with players  $5_1, 5_2, 5_3, 1_1, 1_2, 1_3, 1_4, 1_5, 1_6$  and a quota of 11. Here  $\mathcal{M} = \{5_15_25_3, 5_i5_j1_k, 5_l1_11_21_31_41_51_6\}$ , where  $k \in \{1, 2, 3, 4, 5, 6\}$  and  $i, j \in \{1, 2, 3\}$  with  $i \neq j$ . Let  $\mathcal{W} = \mathcal{M}$ . Then the Banzhaf index is given by  $\beta = \{\frac{7}{39}, \frac{7}{39}, \frac{7}{39}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}\}$ .

Now consider  $\mathcal{W}' = \{5_15_25_3, 5_i5_j1_k, 5_l1_11_21_31_41_51_6\}$ , where  $k \in \{1, 2, 3, 4, 5, 6\}$ ,  $i, j \in \{1, 2, 3\}$  and  $l \in \{2, 3\}$ . Then  $\beta' = \{\frac{13}{71}, \frac{14}{71}, \frac{14}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}\}$ . The set  $\mathcal{W}'$  does not contain the minimal winning coalition  $5_11_11_21_31_41_51_6$ , yet the critical player  $5_1$  is better off as  $\frac{13}{71} > \frac{7}{39}$ .

Similarly we find that players  $5_2$  and  $5_3$  will respectively reject the coalitions  $5_21_11_21_31_41_51_6$  and  $5_31_11_21_31_41_51_6$  resulting in  $\mathcal{W}'' = \{5_i5_j1_k\}$ , where  $k \in \{1, 2, 3, 4, 5, 6\}$  and  $i, j \in \{1, 2, 3\}$ . Then  $\beta'' = \{\frac{2}{9}, \frac{2}{9}, \frac{2}{9}, \frac{1}{18}, \frac{1}{18}, \frac{1}{18}, \frac{1}{18}, \frac{1}{18}, \frac{1}{18}\}$ .

For completeness note that it is not in the interest of any of the players to further reduce the set of winning coalitions, so  $\beta^* = \beta''$ .

In this example  $5_1$  is a large player with a large share of power, so he does not benefit from participating in coalitions, where he plays a relatively minor role as one of the many members. Such large players will only benefit from joining coalitions with few members. Some players *can* increase their power beyond what is given by the standard power indices simply by refusing to participate in certain coalitions. Such coalitions that never form should not be included in the calculations to determine voting power.

### 3.2 Model

We generalise the idea of quarrelling to coalitions: a coalition  $Q$  is *rejected* if *any* of its members rejects  $Q$ . Player  $i$ 's strategy  $s_i$  therefore corresponds to

quarrelling in certain coalitions that  $i$  belongs to, thus  $s_i \subseteq \{C \mid i \in C\}$  and its strategy space  $S_i \subset 2^{\{C \ni i\}}$ .

A rejected coalition is therefore infeasible, so we might as well limit the term winning to coalitions that are feasible and have the ability to pass decisions. Note that rejection extends to supersets of coalitions, so  $C \in s_i$  and  $C \subseteq D$  imply that  $D$  is, in effect, rejected, too. Therefore, given  $s = \{s_i\}_{i \in N}$  the *strategy profile*, they are collected by

$$\mathcal{W}(s) = \{W \in \mathcal{W} \mid W \not\supseteq R \forall R \in s_i, \forall i \in N\} = \mathcal{W} \setminus \bigcup_{i \in N} s_i. \quad (3.1)$$

Observe that  $(N, \mathcal{W}(s))$  is a voting game, thus each strategy profile  $s$  determines a voting game. In this game Conditions 1 and 3 clearly hold, since no new winning coalitions have been added. On the other hand, as the addition of new members to a rejected coalition does not make it winning, convexity, that is: Condition 2 holds, too.

**Definition 1.** A *strategic voting game* is a quadruple  $(N, S, \mathcal{W}, \kappa)$  consisting of a set of players  $N$ , a strategy space  $S$ , a collection of initial winning coalitions  $\mathcal{W}$  and a power index  $\kappa$ .

As the objective of this game is to maximise power, the utility function is nothing, but  $\kappa : S \rightarrow \mathbb{R}_+^N, s \mapsto \kappa(N, \mathcal{W}(s))$ . Strategies are sets of coalitions, the strategy space can be derived from the player set, therefore the triple  $(N, \mathcal{W}, \kappa)$  fully defines the game.

The game consists of two stages: a first, noncooperative game of rejecting coalitions and a second, implicit, cooperative game of power allocation. We seek allocations under equilibrium rejection, and the resulting equilibrium sets of winning coalitions  $\mathcal{W}^*$ .

### 3.3 Equilibria with no more rejection

Note the asymmetry in rejecting coalitions: a single player can reject a coalition, but all members must stop rejecting it to make the coalition acceptable again. Since our interest is not so much in the strategies selected by the players, but the *equilibrium sets of winning coalitions*  $\mathcal{W}^*$ , it is rather natural to seek equilibria in a game where only more rejection is permitted.

**Proposition 1.** *The equilibrium sets of winning coalitions  $\mathcal{W}^*$  under this game coincide with those of the original game.*



*Proof.* Firstly note that the restricted game allows for less deviations. Clearly, the profitability of a particular deviation is the same in the two cases. Therefore, if for  $\mathcal{W}$  and for all  $s$  such that  $\mathcal{W}(s) = \mathcal{W}$  rejecting more coalitions is profitable for one of the players, this will be true also in the original game. Therefore the set of equilibrium sets of winning coalitions in the restricted game contains the set of equilibrium sets of winning coalitions in the unrestricted case.

On the other hand for all equilibrium sets of winning coalitions  $\mathcal{W}^*$  from the restricted game the strategy vector  $s$  where  $s_i = \{C | C \ni i, C \notin \mathcal{W}^*\}$  is also an equilibrium in the unrestricted game.  $\square$

In the following we focus on the “restricted” game, where only more rejection is possible. Once a coalition has been rejected there is no renegotiation despite incentives to make peace ex-post, and therefore only deviations with more rejection are possible. A strategy profile is a Nash equilibrium if  $s$  is a best response to itself, that is,  $\kappa(s) \geq \kappa(s'_i, s_{-i})$  for all  $i$  and  $s'_i \in S_i$  such that  $s'_i \supseteq s_i$ . The latter condition expresses the commitment to rejection, they can only reject more and not less coalitions.

Given a power index  $\kappa$  the *strategic  $\kappa$  power index* is then a vector of equilibrium payoffs, that is  $\kappa(s^*) = \kappa(N, \mathcal{W}(s^*))$ , where  $s^*$  is a Nash equilibrium: for all  $i \in N$  and all  $s_i \subseteq s_i^*$ ,  $s_i \in S_i$  we have  $\kappa_i(s^*) \geq \kappa_i(s_i, s_{-i}^*)$ . Such a strategic power index always exists ( $\mathcal{W}(s^*) = \emptyset$  is an equilibrium), but is generally not unique.

In the sequel we provide a unique refinement for indices that can be written as Equation 2.1.

### 3.4 Only minimal winning coalitions

Rejecting a coalition  $B$  affects a player in two ways. On the one hand for all  $C \supseteq B$  the coalition’s weight (recall the definition in Section 2) becomes  $(a^C)' = 0$  and hence the player loses  $\sum_{C \supseteq B} a^C \mu_i^C$ , on the other hand, due to the normalisation, the weight of other coalitions increases, and hence the credit it gets from other coalitions is scaled up by

$$\frac{\sum_{C \in 2^N \setminus \emptyset} a^C}{\sum_{C \in 2^N \setminus \emptyset} a^C - \sum_{C \supseteq B} a^C}. \quad (3.2)$$

Null players are unaffected and are therefore ignored in our analysis.

**Proposition 2.** *Surplus coalitions containing critical players are rejected.*

*Proof.* Consider a coalition  $B$  containing a surplus player  $i$ . If  $i$  is not critical in  $B$ , it is also not critical in  $C \supset B$  (as, by monotonicity if  $B \setminus \{i\}$  is winning, so is  $C \setminus \{i\} \supset B \setminus \{i\}$ ) and therefore  $a^C \mu_i^C = 0$  for all  $C \supseteq B$ . In sum, neither  $B$  nor  $C \supset B$  yields any profit for  $i$ .

On the other hand  $a^B > 0$  (and possibly  $a^C > 0$  for some  $C \supset B$ ), so when rejecting  $B$  the power of player  $i$  is scaled up according to Expression 3.2 making the rejection profitable.  $\square$

**Corollary 1.** *For power indices we have  $\mathcal{M} \supseteq \mathcal{W}^*$ .*

However, even minimal winning coalitions can be rejected (see Example 2).

In the following we focus on power indices for which  $a^C > 0$  only if  $C \in \mathcal{M}$ . Holler and Packel (1983, p. 24.) argue that “since a non-critical member ... has no incentive to vote ... only these coalitions should be considered for measuring a priori voting power.” Thus a player cannot count on the formation of coalitions that are not due to his or her power. A similar prediction is made by *aspiration* solution concepts (Bennett, 1983, p. 15.).

### 3.5 Elementary rejections

**Definition 2.** Given a strategy profile  $s$ , let the deviation  $s'_i$  is *elementary* if  $|s'_i| - |s_i| = 1$ , that is, if  $s'_i$  rejects a single additional coalition.

**Proposition 3.** *Given a strategy profile  $s$ , let  $s'_i$  be  $i$ 's best response to  $s_{-i}$ . Then  $s'_i$  can be reproduced by a sequence of elementary deviations.*

*Proof.* Proof by construction. Consider the best response  $s'_i$  and let  $s_i \setminus s'_i = \{C_1, \dots, C_k\}$  where, without loss of generality,  $\mu_i^{C_1} \geq \dots \geq \mu_i^{C_k}$ .

We show that  $\kappa_i(s') \geq \mu_1$ . Consider the deviation  $s''_i = s'_i \setminus C_1$ . By assumption

$$\kappa_i(s'_i, s_{-i}) \geq \kappa_i(s''_i, s_{-i}) \quad (3.3)$$

$$\frac{\sum_{C \neq \emptyset} a^C}{\sum_{C \neq \emptyset} a^C - \sum_{s''_i} a^C} \left( \sum_{C \notin s'_i} a^C \mu_i^C + a^{C_1} \mu_i^{C_1} \right) \quad (3.4)$$

$$\kappa_i(s'_i, s_{-i}) \geq \frac{\left( \sum_{C \neq s'_i} a^C \right) \kappa_i(s'_i, s_{-i}) + a^{C_1} \mu_i^{C_1}}{\sum_{C \neq s'_i} a^C + a^{C_1}} \quad (3.5)$$

The right hand side is a weighted average of  $\kappa_i(s'_i, s_{-i})$  and  $\mu_i^{C_1}$ , and hence  $\kappa_i(s'_i, s_{-i}) \geq \mu_i^{C_1}$ .  $\square$

In the following we will only consider elementary deviations.

### 3.6 Friendly equilibrium selection

The strategy profile, where  $\mathcal{W}^* = \emptyset$  is clearly a Nash-equilibrium, while this is neither the only one nor the one we want; out of the many Nash equilibria we select a focal equilibrium.

We now move on to define our refinement.

The literature of power indices has been built on the assumption that all winning coalitions form. We agree that it is a reasonable starting point to assume that, unless for good reasons, players will be friendly and not reject coalitions. Therefore also in our model we take this as the status quo; when strategic considerations do not play a major role, the equilibrium remains  $s = \emptyset$ , and a coalition is only rejected if this increases a player's power. This last observation translates to the fact that for any acceptable equilibrium, there is a sequence of elementary deviations where each of the rejections are introduced all the way back to the status quo. Put it differently: any other equilibrium builds on the assumption that at least one coalition has been rejected irrationally.

The *friendly set*  $F$ , defined below collects the acceptable strategy profiles.

$$s \in F \text{ if } \begin{cases} s_i = \emptyset \ \forall i \in N, \text{ or} \\ \exists i \in N, \exists (s'_i, s_{-i}) \in F, \text{ such that } \kappa_i(s) > \kappa_i(s'_i, s_{-i}). \end{cases}$$

We select *friendly equilibria*  $s^* \in F$  that are Nash equilibria and are maximal for inclusion. The *equilibrium set of winning coalitions* is  $\mathcal{W}^* = \mathcal{W}(s^*)$  and the *strategic  $\kappa$  power index* is defined as

$$\kappa^* = \kappa(s^*) = \kappa(N, \mathcal{W}^*).$$

**Example 1.** continued. Figure 1 shows the rejection game of Example 1. Each node of this lattice corresponds to a set of winning coalitions corresponding to a strategy profile. At the top all winning coalitions are accepted, at the bottom all are rejected. Dashed lines indicate where would the game continue if a particular coalition would be rejected by one of the players,

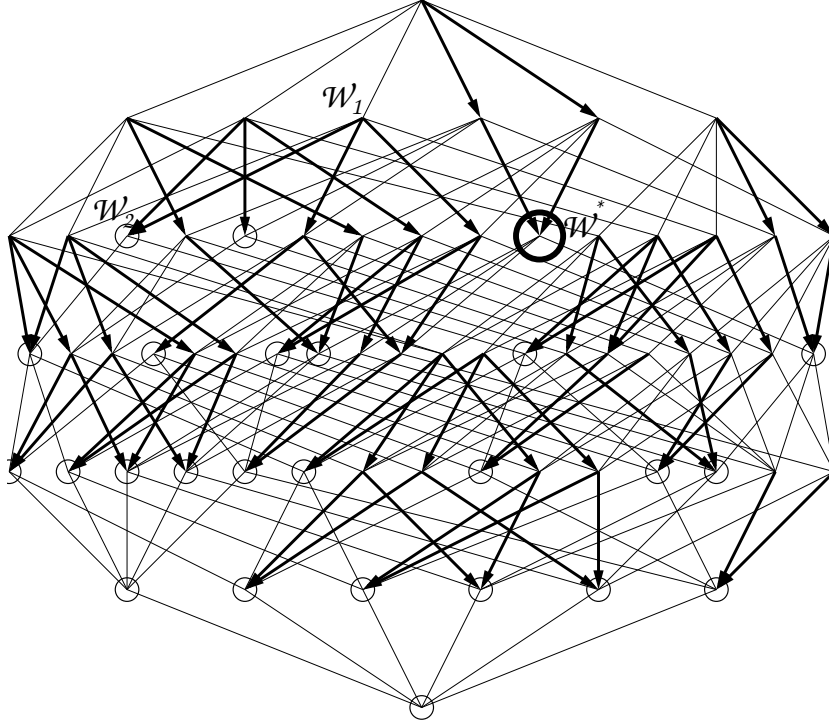


Figure 1: The rejection game for Example 1: At the top no, at the bottom all coalitions are rejected. Arrows are deviations, Nash equilibria are circled.

arrows show the relations where such a rejection is profitable. Nodes where no profitable deviations are possible are circled and are the Nash-equilibria of the game.

Node  $\mathcal{W}_1$ , for instance, has already coalition  $3_12_12_2$  rejected, here we have  $\beta(N, \mathcal{W}_1) = (\frac{3}{9}, \frac{4}{9}, \frac{1}{9}, \frac{1}{9})$ . Clearly no player in the rejected coalition benefits to get these payoffs from the original  $\beta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6})$ . For this node there are three profitable deviations: rejecting coalition  $3_22_12_2$  by  $3_2$  giving  $\beta(N, \mathcal{W}_2) = (\frac{1}{2}, \frac{1}{2}, 0, 0)$ , or rejecting  $3_13_22_1$  by  $2_1$  or  $3_13_22_2$  by  $2_2$  to get  $(\frac{2}{7}, \frac{3}{7}, \frac{1}{7}, \frac{1}{7})$  in both cases.

Note that no further profitable rejections exist at  $\mathcal{W}_2$ , that is, after  $3_12_12_2$  and  $3_22_12_2$  have been rejected, so  $\mathcal{W}_2$  corresponds to a Nash equilibrium of the game.

On the other hand this equilibrium assumes that the coalition  $3_12_12_2$  has also been rejected and for no particular reason. Without arbitrary rejections we would get to  $\mathcal{W}^*$ , which is also a Nash-equilibrium, but this is a *friendly*

Nash equilibrium. As a friendly equilibrium it is unique and we will see that this is a general result.

Now observe that for minimal winning coalitions  $C \neq D$  we have neither  $C \subset D$  nor  $D \subset C$ , therefore by rejecting  $C$  a player will not reject  $D$  and vice versa, a player has the possibility to reject each minimal winning coalition separately. In sum, our model can be reduced to players picking which coalitions they do not want to form. This result makes it particularly easy to work with coalitions rather than strategies. Then an equilibrium is simply  $\mathcal{W}^*$  instead of  $\mathcal{W}(s^*)$  and let  $\mathcal{F} = \{\mathcal{W}(s) | s \in F\}$ .

Player  $i$  profitably rejects coalition  $B$  iff

$$\kappa_i(N, \mathcal{W} \setminus \{B\}) > \kappa_i(N, \mathcal{W}) \quad (3.6)$$

$$\frac{\sum_{C \in \mathcal{W}} a^C}{\sum_{C \in \mathcal{W}} a^C - a^B} \left( \sum_{C \in \mathcal{W}} a^C \mu_i^C - a^B \mu_i^B \right) > \sum_{C \in \mathcal{W}} a^C \mu_i^C \quad (3.7)$$

After some rearrangements we get

$$\frac{\sum_{C \in \mathcal{W}} a^C \mu_i^C}{\sum_{C \in \mathcal{W}} a^C} = \kappa_i(N, \mathcal{W}) > \mu_i^B, \quad (3.8)$$

which gives the following result.

**Lemma 1.** *A rejection by player  $i$  is profitable if and only if player  $i$  has less power in the rejected coalition than on average, as given by the strategic power index.*

Proposition 2 can also be seen as a corollary of this lemma.

Lemma 1 also suggests a relation to the theory of *aspirations* (Bennett, 1983), although this relation turns out to be superficial. In the theory of aspirations it is not some coalition's payoff that is bargained over: it is the players who make their claims, and unless their claims are satisfied, certain coalitions will or will not form. Here this claim is expressed by their power index, the "credit they receive in general" and players demand the same credit in coalitions. Unfortunately the link between the two concepts does not go much beyond that. While a power index satisfies  $\sum_{i \in N} \kappa_i^* = 1$  a vector of aspirations will almost always be larger. For instance, Bennett (1983, p. 15.) provides an example with 5 players with weights 2, 2, 1, 1, and 1, and a quota of 5. Here the unique partnered, balanced, equal gains aspiration is  $(0.4, 0.4, 0.2, 0.2, 0.2)$ , while the public good index is  $h = (\frac{4}{17}, \frac{4}{17}, \frac{3}{17}, \frac{3}{17}, \frac{3}{17})$ .

Now we move on to our main result.

**Theorem 1.** *The friendly equilibrium set of winning coalitions is uniquely defined and is given by*

$$\mathcal{W}^* = \bigcap_{s \in F} \mathcal{W}(s). \quad (3.9)$$

In order to prove this theorem we need some additional results.

**Proposition 4.** *Let  $C_i, C_j \in \mathcal{W}$  be coalitions that both contain both  $i$  and  $j$  and such that  $i$  and  $j$  can profitably reject  $C_i$  and  $C_j$  respectively. Then either  $i$  can profitably reject  $C_j$  or  $j$  can profitably reject  $C_i$ .*

*Proof.* Assume that the proposition is false. This means the following. Player  $j$  rejects  $C_j$ , hence  $\mu_j^{C_j} < \kappa_j(\mathcal{W})$ . By our assumption  $j$  does not reject  $C_i$ , hence  $\mu_j^{C_i} \geq \kappa_j(\mathcal{W})$ . Therefore  $\mu_j^{C_j} < \mu_j^{C_i}$ . Similarly  $i$  rejects  $C_i$ , hence  $\mu_i^{C_i} < \kappa_i(\mathcal{W})$ . By our assumption  $i$  does not reject  $C_j$ , hence  $\mu_i^{C_j} \geq \kappa_i(\mathcal{W})$ . Therefore  $\mu_i^{C_i} < \mu_i^{C_j}$ . In sum  $\mu_j^{C_j} < \mu_j^{C_i}$  and  $\mu_i^{C_i} < \mu_i^{C_j}$ . Since  $C_i$  and  $C_j$  are minimal winning coalitions  $\mu_i^{C_i} = \mu_j^{C_i} = \frac{1}{|C_i|}$  and  $\mu_j^{C_j} = \mu_i^{C_j} = \frac{1}{|C_j|}$ . Thus  $\frac{1}{|C_j|} < \frac{1}{|C_i|}$  and  $\frac{1}{|C_i|} < \frac{1}{|C_j|}$ . Contradiction.  $\square$

**Proposition 5.** *For all  $\mathcal{W}_i, \mathcal{W}_j \in \mathcal{F}$  we have  $\mathcal{W}_i \cap \mathcal{W}_j \in \mathcal{F}$ .*

*Proof.* The proof is by induction on the differences between  $\mathcal{W}_i$  and  $\mathcal{W}_j$ .

First we deal with the elementary step. Assume  $\mathcal{W}_i = \{A, C_1, C_2, \dots, C_m\}$ ,  $\mathcal{W}_j = \{B, C_1, C_2, \dots, C_m\}$ , that is, the two sets only differ in 1 element each. This ensures that their intersection is non-trivial.  $\mathcal{W}_i$  and  $\mathcal{W}_j$  are descendants of a common ancestor  $\mathcal{W}_0 = \{A, B, C_1, C_2, \dots, C_m\}$ , but after rejecting  $B$  and  $A$ , respectively by some players  $i$  and  $j$ . The proposition merely states that either rejecting  $A$  is profitable from  $\mathcal{W}_i$  or rejecting  $B$  is profitable from  $\mathcal{W}_j$ .

$\mathcal{W}_i$  is the result of rejecting  $B$  by  $i$ . If  $j \notin B$  then  $\kappa_j(\mathcal{W}_0) \leq \kappa_j(\mathcal{W}_i)$ . We know that  $j$  rejects  $A$  at  $\mathcal{W}_0$  and hence  $\kappa_j(\mathcal{W}_0) > \mu_j^A$ . Hence  $\kappa_j(\mathcal{W}_i) > \mu_j^A$ , which implies that  $j$  also rejects  $A$  at  $\mathcal{W}_i$ . Thus  $\mathcal{W}_{ij} = \{C_1, C_2, \dots, C_m\} \in \mathcal{F}$ .

The symmetric case gives the corresponding result for  $i$  and  $B$  at  $\mathcal{W}_j$ .

Finally we must consider the case where none of the previous two cases applied, that is where  $j \in B$  and  $i \in A$ . As only a member can reject a coalition, we also have  $j \in A$  and  $i \in B$ . Therefore we can apply Proposition 4 to show that  $i$  rejects at  $\mathcal{W}_j$  or  $j$  at  $\mathcal{W}_i$ , which, as before, gives the result.

We have discussed all possible cases, which completes the first part of the proof. Now we move on to the general case. Assume that we have shown the result for all pairs of sets with differences up to  $k - 1$ .

Now consider  $\mathcal{W}_i = \{A_1, A_2, \dots, A_k, C_1, C_2, \dots, C_m\}$  as well as  $\mathcal{W}_j = \{B_1, B_2, \dots, B_l, C_1, C_2, \dots, C_m\}$ , where  $A_1, A_2, \dots, A_k$  and  $B_1, B_2, \dots, B_l$  represent the rejections that did *not* take place and  $l \leq k$ . (Possibly  $A_p = B_q$  for some  $p$  and  $q$ .) The question is whether this difference can be eliminated.

By definition if  $\mathcal{W}_i \in \mathcal{F}$  there exists a sequence of rejections starting from  $\mathcal{W}_0$  that lead to  $\mathcal{W}_i$  and a similar sequence exists to  $\mathcal{W}_j$ . Let  $\mathcal{W}_i^0$  and  $\mathcal{W}_j^0$  be the first elements that are not common, without loss of generality, as results of rejecting  $B_1$  and  $A_1$  respectively. By the elementary step  $\mathcal{W}_j^1 = \mathcal{W}_i^0 \cap \mathcal{W}_j^0$  belongs to  $\mathcal{F}$ . (Our notation is slightly misleading given that  $\mathcal{W}_j^1$  is not necessarily on the path to  $\mathcal{W}_j$ , but this should not lead to confusion.) Now take the next set  $\mathcal{W}_2$  along the path to  $\mathcal{W}_i$ ,  $\mathcal{W}_i^1$ . By the same argument  $\mathcal{W}_i^1 \cap \mathcal{W}_j^1$  also belongs to  $\mathcal{F}$ . Repeating this argument we travel parallel to the path and in the penultimate step we get  $\mathcal{W}_j^p \in \mathcal{F}$ . For the last time by the same argument  $\mathcal{W}_i \cap \mathcal{W}_j^p = \{A_2, \dots, A_k, C_1, C_2, \dots, C_m\}$  also belongs to  $\mathcal{F}$ . If  $l < k$ , the inductive assumption completes the proof.

In case  $l = k$  it is necessary to apply the same argument once more, but on the other side: to show that  $\{B_2, \dots, B_l, C_1, C_2, \dots, C_m\} \in \mathcal{F}$ .  $\square$

*Proof of Theorem.* By Proposition 5 pairwise intersections of elements of  $\mathcal{F}$  also belong to  $\mathcal{F}$ . As the number of winning coalitions is finite the result on pairwise intersections implies that  $\mathcal{W}^*$  as defined in Equation 3.9 belongs to  $\mathcal{F}$ . Clearly  $\mathcal{W}^* \subseteq \mathcal{W}^{\mathcal{F}}$  for all  $\mathcal{W}^{\mathcal{F}} \in \mathcal{F}$ . Therefore  $\mathcal{W}^*$  is the smallest friendly set and it is trivially an equilibrium.  $\square$

**Corollary 2.** *The strategic power index  $\kappa^*$  is well-defined.*

## 4 Conclusion

The power index approach sees power as the probability of being critical to a winning coalition. As such, it ignores the strategic aspects of forming winning coalitions. In our model the implicit cooperative game is preceded by a noncooperative stage where players decide which (winning) coalitions may form: only those with unanimous agreement. It turns out that there is a well-defined refinement of the Nash-equilibria of this game allowing us to define strategic power indices.

While the calculation of strategic power indices for larger problems is computationally difficult, strategic considerations may influence power distribution in such well-studied weighted voting situations as the voting in the

Council of the European Union (formerly Council of Ministers). According to the Lisbon Treaty the support of 55% of the countries (when acting on a proposal from the Commission or from the High Representative) and 65% of the population is required to form a coalition. We have calculated the Public Good Index values using the software IOP-Indices of Power 2.05 (Bräuning and König, 2005) and the population data currently used in the Council (Council of the European Union, 2014). We found that Germany, the largest EU member state has 5.56% or a little over  $\frac{1}{18}$  of the voting power, while it is a member of minimal winning coalitions with up to 25 members. Our model suggests that it will prefer to cooperate with smaller coalitions, ideally with less than 18 members. Besides Germany, also the other large members not included in the large coalitions (that consist of small countries, mostly) also benefit from such a move. Technical limitations inhibit us from conducting a full analysis, but it is expected that large member states will only participate in minimal winning coalitions that are also minimal in size, that is, with 16 members resulting in a higher power concentration than estimated using non-strategic power indices.

There is an interesting alternative to the underlying problem. To see this consider an initial set  $\mathcal{W}$  of winning coalitions and a rejected coalition  $C$ . By Lemma 1  $\mu_i^C > \kappa_i(N, \mathcal{W})$  for the rejecting player(s). Since  $\sum_{i \in C} \mu_i^C = 1$  and  $\sum_{i \in C} \kappa_i(N, \mathcal{W}) < 1$  the total power of the members of the coalition decreases after the rejection. While rejecting  $C$  benefits some members, it harms others. It would be more efficient to renegotiate the  $\mu_i^C$  values in the coalition. Currently power indices do not allow for that, but it would be interesting to see if there exist rejection-proof allocations of credit within coalitions. In practice, such as in the Council example above, it is very natural that less powerful voters make smaller claims than larger members of a minimal winning coalition.

Two other choices we have made was to assume that rejecting coalition  $C$  also rejects  $D \supset C$ , and to work with power indices defined over minimal winning coalitions only. Rejecting a single coalitions would not preserve null players who could gain power for “mediation” (turning a rejected coalition into a (non-rejected) winning one by their entry – of course this coalition would be rejected soon, too) and would allow non-minimal winning coalitions that are not surplus coalitions as they would only consist of critical players. While our original model considered a variant of this alternative, in order to avoid such odd phenomena one has to separate the notions of winning and



feasibility.

At last we note that the uniqueness proof of the friendly equilibrium relies on the assumption that only minimal winning coalitions are considered. In the aforementioned model we could find counterexamples, but not here. Even a systematic search for them was in vain, so now we are inclined to believe the result extends to all power indices. The proof of this claim remains, however, open.

## References

- Banzhaf, J. F. (1965), “Weighted voting doesn’t work: A mathematical analysis,” *Rutgers Law Review* 19: 317–343.
- Bennett, E. (1983), “The aspiration approach to predicting coalition formation and payoff distribution in sidepayment games,” *International Journal of Game Theory* 12: 1–28.
- Berger, H., and T. Mueller (2007), “How should large and small countries be represented in a currency union?,” *Public Choice* 132: 471–484.
- Bilbao, J. M., A. Jiménez, and J. López (1998), “The Banzhaf power index on convex geometries,” *Mathematical Social Sciences* 36: 157–174.
- Braham, M., and M. J. Holler (2005), “The Impossibility of a Preference-Based Power Index,” *Journal of Theoretical Politics* 17: 137–157.
- Brams, S. J. (1975), *Game theory and politics*. Dover, 2nd edn.
- Bräuningner, T., and T. König (2005), “Indices of Power IOP 2.0,” Accessible at <http://www.tbraeuningner.de/download/>.
- Council of the European Union (2014), “Council Decision of 29 September 2014 Amending the Council’s Rules of Procedure,” *Official Journal of the European Union* 57(L289): 18.
- Deegan, J., and E. W. Packel (1978), “A New Index of Power for Simple n-Person Games,” *International Journal of Game Theory* 7: 113–123.
- Faigle, U., and W. Kern (1992), “The Shapley value for cooperative games under precedence constraints,” *International Journal of Game Theory* 21: 249–266.

- Holler, M. J., and E. W. Packel (1983), “Power, luck and the right index,” *Journal of Economics* 43: 21–29.
- Johnston, R. J. (1978), “On the measurement of power: some reactions to Laver,” *Environment and Planning A* 10: 907–914.
- Kaniovski, S., and D. Leech (2009), “A behavioral power index,” *Public Choice* 141: 17–29.
- Kilgour, D. M. (1974), “A Shapley value for cooperative games with quarrelling,” in *Game theory as a theory of conflict resolution*, ed. by A. Rapoport, vol. 26, pp. 193–206. D. Reidel Publishing Co.
- (1977), “Axioms for Shapley values in games with quarrelling,” *Theory and Decision* 8: 193–207.
- Kóczy, L. A. (2012), “Beyond Lisbon: Demographic trends and voting power in the European Union Council of Ministers,” *Mathematical Social Sciences* 63: 158–152.
- Laruelle, A., and F. Valenciano (2005), “A critical reappraisal of some voting power paradoxes,” *Public Choice* 125: 17–41.
- Leech, D. (2002), “Designing the Voting System for the Council of the European Union,” *Public Choice* 113: 437–464.
- Leech, D., and R. Leech (2005), “Voting power in the Bretton Woods institutions,” *Homo Oeconomicus* 22: 605–627.
- Myerson, R. B. (1977), “Graphs and cooperation in games,” *Mathematics of Operations Research* 2: 225–229.
- (1980), “Conference structures and fair allocation rules,” *International Journal of Game Theory* 9: 169–182.
- Napel, S., and M. Widgrén (2010), “Strategic versus non-strategic voting power in the EU Council of Ministers: the consultation procedure,” *Social Choice and Welfare* 37: 511–541.
- Penrose, L. S. (1946), “The elementary statistics of majority voting,” *Journal of the Royal Statistical Society* 109: 53–57.

- Shapley, L. S. (1962), "Simple games: an outline of the descriptive theory," *Behavioral Science* 7: 59–66.
- Shapley, L. S., and M. Shubik (1954), "A method for evaluating the distribution of power in a committee system," *American Political Science Review* 48: 787–792.
- Steunenberg, B., D. Schmidtchen, and C. Koboldt (1999), "Strategic Power in the European Union: Evaluating the Distribution of Power in Policy Games," *Journal of Theoretical Politics* 11: 339–366.
- van Deemen, A., and A. Rusinowska (2003), "Paradoxes of Voting Power in Dutch Politics," *Public Choice* 115: 109–137.