Strategic power indices: Quarrelling in coalitions*

László Á. Kóczy†

Abstract

While they use the language of game theory known measures of a priory voting power are hardly more than statistical expectations assuming voters behave randomly. Focusing on normalised indices we show that rational players would behave differently from the indices predictions and propose a model that captures such strategic behaviour.

Keywords and phrases: Banzhaf index, Shapley-Shubik index, a priori voting power, rational players.

*The author is grateful for discussions with Manfred Holler, Stefan Napel, Paul Harrenstein and for comments by Max Albert; he also thanks the funding of the Netherlands Organisation for Scientific Research as well as the hospitality of the Economic Research Institute of the Hungarian Academy of Sciences in the summer 2005.

†Keleti Faculty of Economics, Budapest Tech, Tavaszmezõ 15-17. H-1084 Budapest and Department of Economics, Maastricht University, P.O.Box 616, NL-6200MD Maastricht. Email: koczy.laszlo@kgk.bmf.hu

1 Introduction

Since Shapley and Shubik (1954) adopted the Shapley value to measure a priori voting power game theory has contributed an enormous literature to this topic: established theoretical underpinnings for the existing or rediscovered indices, introduced new ones, but the plethora of power indices hints that there is no single best. What is best depends on the institutional details concerning the voting that cannot be captured by the voting game the index is applied to (Laruelle, 1999). Whether one is more interested in comparing powers of different players in the same game or the powers of the same player in different games is one crucial choice. Since we are more interested in the first we focus on normalised indices.

Game theory embraced power indices despite the fact that none of the power indices are really "game theoretical." Voting situations are games where "the acquisition of power is the payoff" (Shapley, 1962, p. 59.), but 'acquisition' is an overstatement as players have no strategies: it seems voting indices are hardly more than statistical measures of the voters' random behaviour. We like to believe that this is not a realistic model of most voting situations; we assume that voters are rational who can and want to influence (that is: maximise) their power. (Albert, 2003, makes a similar point).

Motivated by the paradox of quarrelling members (Brams, 1975/2003) we extend voting games to *strategic voting games* where players can choose which coalitions are they willing to join. We show that all known normalised indices are affected by such strategic behaviour.

Our paper is not the first to disallow certain (winning) coalitions in values or power indices. Aumann and Drèze (1975) assume that property rights may make it impossible to form every coalition. Owen (1977, 1982) assume that coalitions are formed exactly in order to increase power. Myerson (1977, 1980) presents a model where players communicate via *conferences* and not all conferences may occur (Faigle and Kern, 1992). The application of such restrictions to power indices are more recent (Bilbao et al., 1998). Lastly, Steunenberg et al. (1999) introduce a notion of strategic power based on the theory of political institutions.

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 some properties.

2 Power indices

A voting situation is a pair $(N, \mathcal{W})$, where N is the set of voters and $\mathcal{W}$ denotes the set of winning coalitions. We consider games where

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

Condition 3 requires the game to be *proper*, Condition 2 is a *convexity* condition on the poset formed by the winning coalitions. It is often assumed that $N \in \mathcal{W}$ and then Condition 2 is expressed in terms of N (in the place of E), in which case we have *simple games*.

Let Γ denote the collection of proper convex voting games satisfying the above properties.

Let $\mathcal{M}$ denote the set of *minimal winning coalitions*: the set of coalitions without proper winning subsets. Formally: if $S \in \mathcal{M}$ and $i \in S$, then $S \setminus \{i\} \notin \mathcal{W}$. Clearly $\mathcal{M} \subseteq \mathcal{W}$. *Surplus coalitions* are winning, but non-minimal.

Given a game Γ a *power measure* $\kappa : \Gamma \longrightarrow \mathbb{R}_+^N$ assigns to each player i a non-negative real number κ_i, its *power*; if $\sum_{i \in N} \kappa_i = 1$ then it is also a *power index*.

In the following we explain some of the well-known indices.

The *Shapley-Shubik index* (Shapley and Shubik, 1954) applies the Shapley value (Shapley, 1953) to simple games: Voters arrive in a random order; if and when a coalition turns winning the full credit is given to the last arriving, the *pivotal* player. A player's power is given as the proportion of orderings where it is pivotal, formally (for simple games) $\phi_i = \frac{\#\text{ times } i \text{ is pivotal}}{n!}$.

While in simple games any order will yield a unique pivotal player, when only Condition 2 is satisfied, there may be none. (To see this, start the order with a destructive player, whose membership turns any coalition losing.) In such cases, to obtain an index a further normalisation is required.

The *Banzhaf measure* (Penrose, 1946; Banzhaf, 1965) is the probability that a party is *critical* for a coalition, that it can turn winning coalitions into losing ones. Formally $\psi_i = \frac{\#\text{ times } i \text{ is critical}}{2^{n-1}}$; when normalised to 1, we get the *Banzhaf index* β (Coleman, 1971).

Numerous variants of the (normalised) Banzhaf index exist. In the *Johnston index* γ (Johnston, 1978) the credit a critical player gets is inversely proportional to the number of critical players in the coalition. In effect, coalitions of different sizes have the same contribution to the distribution of power. Deegan and Packel (1978) argue that only those coalitions form where the benefits are least divided (Riker, 1962): the *Deegan-Packel index* ρ only considers minimal winning coalitions. Finally the Holler-Packel or *Public Good Index* h (Holler and Packel, 1983) modifies the Deegan-Packel index: here the benefit of forming a winning coalition is given to each and every player in the coalition. With the normalisation in simple games the index is nothing but a normalised Banzhaf index, where only minimal coalitions are taken into account.

Although there is some disagreement on what should a power index be like, the ones in use are very much alike. They give credit precisely to the critical (or swing) players, and give them all the same disregarding their size. The sole difference lies in weighting winning coalitions differently. We consider a general power indices along these lines.

For coalition $C \in 2^N \setminus \varnothing$ let a^C denote its weight such that $\sum_{C\in 2^N\setminus\varnothing} a^C = \sum_{C\in\mathcal{W}} a^C = 1$ and let k^C denote the number of critical players in C. The power index $\kappa(N, \mathcal{W})$ can be rewritten as

$$\kappa_i = \sum_{C\in 2^N\setminus\varnothing} a^C \mu_i^C, \quad \text{where} \tag{2.1}$$

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

is the credit player i gets for being in the coalition (therefore $\sum_{i\in C} \mu_i^C = 1$). Observe that $a^C \neq 0$ iff C contains critical players.
For instance for the Banzhaf index $a^C = \frac{k^C}{\sum_{C\in\mathcal{W}} k^C}$.

Players with no power are *null*. As the set of winning coalitions does not expand in this model, their situation does not improve. As they merely create multiplicities in our model, they will simply be ignored. To be precise, whenever we say surplus player, we exclude null players, and by surplus coalition we mean a coalition that contains such surplus players.

3 Strategic voting

All existing indices assume an exogenously given set of winning coalitions and that players join winning coalitions at all times. 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. The "Paradox of Quarrelling Members" (Kilgour, 1974; Brams, 1975/2003) arises when two players mutually benefit from refusing to cooperate with each other.

Paradoxical or not is a matter of interpretation, but players can certainly *acquire* (relative) power by approving/rejecting coalitions. In this paper we extend voting games to allow for such strategic considerations and define strategic power indices.

3.1 Examples

As a motivation we present a number of games based on *weighted voting games*. Here N is a collection of n interest groups, or *parties* having $w_1, w_2, \ldots, w_n$ 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 iff $\sum_{i \in C} w_i \geq q$. Since $w > q$ 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[1]: $3_1, 3_2, 2_1, 2_2$ and voting has a quota of 6. The set winning coalitions is $\mathcal{W} = \left\{\underline{3_1 3_2}, \underline{3_1 3_2} 2_1, \underline{3_1 3_2} 2_2, \underline{3_1 2_1 2_2}, \underline{3_2 2_1 2_2}, 3_1 3_2 2_1 2_2\right\}$ (with critical players underlined). The vector of Banzhaf indices is $\beta = \left\{\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}\right\}$.

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 less coalitions, so in relative terms (thus: in a power *index*) 2_1 gains.

Given $\mathcal{W}' = \left\{\underline{3_1 3_2}, \underline{3_1 3_2} 2_2, \underline{3_1 2_1 2_2}, \underline{3_2 2_1 2_2}, 3_1 3_2 2_1 2_2\right\}$ the recalculated Banzhaf index is $\beta' = \left\{\frac{3}{10}, \frac{3}{10}, \frac{1}{5}, \frac{1}{5}\right\}$. 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 $3_1 3_2 2_1$ player

[1]Subscripts are used to distinguish players with identical weights from each other.

2_1 assisted players 3_1 and 3_2 in forming a winning coalition, but without getting any credit for it.

Minimal winning coalitions may also be subject to blocks:

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 $\mathcal{M} = \{5_15_25_3, 5_i5_j1_k, 5_i1_11_21_31_41_51_6\}$, 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 $\beta = \left\{\frac{7}{39}, \frac{7}{39}, \frac{7}{39}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}\right\}$.
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 $\beta' = \left\{\frac{13}{71}, \frac{14}{71}, \frac{14}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}, \frac{5}{71}\right\}$. 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 $\frac{13}{71} > \frac{7}{39}$.

While the aforementioned indices claim to measure power, it seems, players have actually little power to influence their power: hence they are no more than probabilistic values. The paradox of quarrelling members as well as the above examples illustrate that players *can* increase their power by refusing to participate in certain coalitions. If a player credibly refuses to participate in a coalition neither him nor his colleagues should get credit for being critical to a coalition that never forms.

3.2 Model

The idea of quarrelling is generalised to coalitions: a coalition Q quarrels if *any* of its members quarrels. 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 $S_i \subset 2^{\{C \mid i \in C\}}$. Note that as $C \in s_i$ and $C \subseteq D$ imply $D \in s_i$ not all combinations of quarrelled coalitions are possible.

Only winning coalitions without quarrelling remain winning. Given $s =$

$\{s_i\}_{i \in N}$ the *strategy profile*, they are collected by

$$\mathcal{W}(s) = \{w \in \mathcal{W} |\ w \notin s_i,\ \forall i \in N\} = \left\{ w \in \mathcal{W} \setminus \bigcup_{i \in N} s_i \right\}. \tag{3.1}$$

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 quarrelling 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 κ.

The acquisition of power is the payoff, so the utility function is simply $\kappa : S \longrightarrow \mathbb{R}_+^N, s \longmapsto \kappa(N, \mathcal{W}(s))$. Strategies are in fact 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 quarrelling and a second, implicit, cooperative game of power allocation. Quarrelling is for good despite incentives to make peace ex-post, which implies that only asymmetric deviations are possible, introducing quarrelling to additional coalitions, but not allowing players to reconcile.

Definition 2. A *strategic 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 $\kappa_i(s^*) \geq \kappa_i(s_i, s_{-i}^*)$.

A strategic power index always exists (s^* where $\mathcal{W}(s^*) = \varnothing$ is an equilibrium) but is generally not unique. In the sequel we provide a unique refinement for certain indices.

4 Results

4.1 Only minimal winning coalitions

Blocking 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

$$\frac{\sum_{C \in 2^N \setminus \varnothing} a^C}{\sum_{C \in 2^N \setminus \varnothing} a^C - \sum_{C \supseteq B} a^C}. \tag{4.1}$$

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

Proposition 3. *Surplus coalitions containing critical players are blocked.*

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 blocking B the power of player i is scaled up according to Expression 4.1 making the block profitable. □

Corollary 4. *For power* indices *we have* $\mathcal{M} \supseteq \mathcal{W}(s^*)$.

Not all minimal coalitions are quarrel-free (see Example 2).

In the following we allow $a^C > 0$ 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. Interestingly, a similar prediction is made by *aspiration* solution concepts (Bennett, 1983, p. 15.).

4.2 Elementary blocks

Definition 5. Given a strategy profile s the deviation s_i' is *elementary* if $|s_i'| - |s_1| = 1$, that is, if s_i' extends quarrelling to a single new coalition.

Proposition 6. *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 $s_i \setminus s_i^* = \left\{C^1, \dots, C^k\right\}$ where, without loss of generality, $\mu_i^{C^1} \leq \cdots \leq \mu_i^{C^k}$. Consider the sequence of elementary deviations $s_i^h = s_i^{h-1} \cup \left\{C^h\right\}$, $h = 1, \dots k$, $s_i^0 = s_i$ and $s_i^k = s_i^*$. Now suppose that from (s_i^{h-1}, s_{-i}), the elementary deviation s_i^h is not profitable:

$$\kappa_i(s_i^h, s_{-i}) \leq \kappa_i(s_i^{h-1}, s_{-i}) \tag{4.2}$$

$$\kappa_i(s_i^h, s_{-i}) \leq \frac{\left(\sum_{C \notin \mathcal{W}(s_i^h, s_{-i})} a^C\right) \kappa_i(s_i^h, s_{-i}) + a^{C^h} \mu_i^{C^h}}{\sum_{C \notin \mathcal{W}(s_i^h, s_{-i})} a^C + a^{C^h}}. \tag{4.3}$$

The right hand side is a weighted average of $\kappa_i(s_i^h, s_{-i})$ and $\mu_i^{C^h}$ hence $\kappa_i(s_i^h, s_{-i}) \leq \kappa_i(s_i^{h-1}, s_{-i}) \leq \mu_i^{C^h}$. Since $\mu_i^{C^h} \leq \mu_i^{C^{h+1}} \leq \cdots \leq \mu^{C^k}$ by a similar argument reversed $\kappa_i(s_i^*, s_{-i}) = \kappa_i(s_i^k, s_{-i}) \leq \kappa_i(s_i^{h-1}, s_{-i})$, hence s_i^* is not a best response. Contradiction, therefore each elementary deviation is profitable.

We have considered a particular sequence of elementary deviations. Clearly if C^h cannot be blocked profitably, the property extends to all C^g, $g \geq h$ as $\mu^{C^h} \leq \mu^{C^g}$. Would an alternative sequence of blocking $C^1 \dots C^{h-1}$ work? By the profitability of previous deviations player i's power has strictly monotone increasing until $\kappa_i(s_i^{h-1}, s_{-i})$. Assuming the same coalitions can be blocked profitably in different orders to get to (s_i^{h-1}, s_{-i}) (if not, the sequence can clearly be dropped), at an intermediate stage i's power is less, making the same blocking of C^h even more difficult. □

A similar argument (that we skip here) proves a more general result.

Proposition 7. *Given a strategy profile s and consider a deviation s_i'. Then s_i' can be reproduced by a sequence of elementary deviations if and only if for all s_i'', such that $s_i' \supset s_i'' \supset s_i$, we have $\kappa_i(s_i', s_{-i}) \geq \kappa_i(s_i'', s_{-i}) \geq \kappa_i(s)$.*

The message is clear: a player may quarrel too much, while blocking some "bad" coalitions, a few good ones may get blocked, too. While this may still result in an overall profitable deviation, a player should be more conservative in choosing its strategies. The above results show that there is a simple rule of thumb: use elementary deviations; we even have a recipe: first eliminate the worst coalitions. In the following, by deviation, we mean elementary deviations.

Now observe that for minimal winning coalitions $C \neq D$ we have neither $C \subset D$ nor $D \subset C$, therefore by blocking C a player will not block D and vice versa, a player has the possibility to block 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.

4.3 Friendly equilibrium selection

While s^*, where $\mathcal{W}(s^*) = \varnothing$ is a Nash-equilibrium this is neither the only one nor the one we want (for one, power indices are undefined here); out of the many Nash equilibria we make a selection.

Our approach is conservative: A whole literature has been built on the idea that all coalitions should form; we accept this status quo. If this is not a Nash-equilibrium, a player deviates and the status quo is replaced by another, and so on. We consider strategy profiles that can arise as results of such sequences of elementary deviations from the classical setting. These are collected by the *friendly set* F :

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

Note that while we speak in terms of elementary deviations, these can be aggregated into single deviations. We do, however forbid deviations that fail to meet the conditions of Proposition 7, that is, deviations that go too far. This, too, could be regarded as an aspect of our conservative approach.

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* κ *power index* is defined as

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

In the following we prove the uniqueness of this equilibrium for the class of power indices that take only minimal coalitions into account.

Lemma 8. *A block by player* i *is profitable if and only if the blocked coalition gives less credit to player* i *than the average credit it gets, that is, than its power index.*

Proof. Given a strategy profile s player i profitably blocks coalition B iff

$$\kappa_i(s_i \setminus \{B\}, s_{-i}) \quad > \quad \kappa_i(s) \tag{4.4}$$

$$\frac{\sum_{C \neq \varnothing} a^C}{\sum_{C \in \mathcal{W}(s)} a^C - a^B} \left(\sum_{C \in \mathcal{W}(s)} a^C \mu_i^C - a^B \mu_i^B \right) \quad > \quad \sum_{C \in \mathcal{W}(s)} a^C \mu_i^C \tag{4.5}$$

After some rearrangements we get

$$\frac{\sum_{C \in \mathcal{W}(s)} a^C \mu_i^C}{\sum_{C \in \mathcal{W}(s)} a^C} = \kappa_i(s) > \mu_i^B, \tag{4.6}$$

□

Proposition 3 can also be seen as a corollary of this lemma.

Lemma 8 hints 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: players 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 "typical" credit they receive and players make the same claim in all 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. Bennett (1983, p. 15.) provides the following example:

Example 3. A game 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 9. *Let $(N, S, \mathcal{W}, \kappa)$ be a strategic voting game, such that in $\kappa_i = \sum_{C \in 2^N \setminus \varnothing} a^C \mu_i^C$ we have $a^C = 0$ for all $C \notin \mathcal{M}$. The friendly equilibrium set of winning coalitions is uniquely defined and is given by*

$$\mathcal{W}^* = \bigcap_{s \in F} \mathcal{W}(s). \tag{4.7}$$

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

Proposition 10. *Let $C_i, C_j \in \mathcal{W}$ be coalitions such that $\{i, j\} \subseteq C_i \cap C_j$ and i and j want to block C_i and C_j respectively. Then either i wants to block C_j or j wants to block C_i.*

Proof. Assume that the proposition is false: Player j blocks C_j, hence $\mu_j^{C_j} < \kappa_j(\mathcal{W})$ but i does not block, hence $\mu_i^{C_j} \geq \kappa_i(\mathcal{W})$. Therefore $\mu_j^{C_j} < \mu_i^{C_j}$. Similarly i blocks C_i, hence $\mu_i^{C_i} < \kappa_i(\mathcal{W})$. By our assumption j does not block, hence $\mu_j^{C_i} \geq \kappa_j(\mathcal{W})$. In sum $\mu_i^{C_i} < \mu_i^{C_j}$ and $\mu_j^{C_i} < \mu_j^{C_j}$. Since C_i

and C_j are minimal coalitions $\mu_i^{C_i} = \mu_j^{C_i} = \frac{1}{|C_i|}$ and $\mu_j^{C_j} = \mu_i^{C_j} = \frac{1}{|C_j|}$. Contradiction $\square$

Proposition 11. *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 blocking B and A, respectively by some players i and j. The proposition merely states that either blocking A is profitable from $\mathcal{W}_i$ or blocking B is profitable from $\mathcal{W}_j$.

$\mathcal{W}_i$ is the result of blocking B by i. If $j \notin B$ then $\kappa_j(\mathcal{W}_0) \leq \kappa_j(\mathcal{W}_i)$. We know that j blocks 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 blocks 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 block a coalition, we also have $j \in A$ and $i \in B$. Therefore we can apply Proposition 10 to show that i blocks 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 blocks 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 blocks 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 blocking 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 $\mathcal{F}$.[2] 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 $\mathcal{W}_i \cap \mathcal{W}_j^p = \{A_2, \ldots, A_k, C_1, C_2, \ldots C_m\}$ also belongs to $\mathcal{F}$. If $l < k$, our inductive assumption can be used to complete the proof.

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

Proof of Theorem. By Proposition 11 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 4.7 belongs to $\mathcal{F}$. Clearly $\mathcal{W}^* \subseteq w$ for all $w \in \mathcal{F}$. Therefore $\mathcal{W}^*$ is the smallest friendly set of winning coalitions and is trivially an equilibrium. □

Corollary 12. *The strategic power index κ^* is well-defined.*

5 Conclusion

We have developed a model that measures power taking the rational, utility maximising behaviour of players into account. We have also shown that none of the well-known power indices account for this behaviour. It appears that these supposedly game theoretic concepts are not more than statistical measures of random behaviour.

There are at least two possibilities to resolve this conflict. The one we chose is to modify existing power indices so that no credit is given for coalitions that do not form. The advantage of this solution is that it is directly motivated by the problem and gives a perfect answer to it without affecting the concepts a great deal.

[2]Our notation is slightly misleading as $\mathcal{W}_j^1$ is not necessarily on the path to $\mathcal{W}_j$, but this should not lead to confusion.

While this is the option we choose here there is an interesting alternative. Observe that blocking a winning coalition may be advantageous to some players, but it will hurt others in the coalition. The only players whose power will surely increase are those *outside* the coalition. This indicates that overall members of the coalition lose by not forming the coalition. Hence forming the coalition increases the power of the members and therefore there exists distributions of this power that benefit all members. Giving room for renegotiation would lead us to cooperative, probably set-like solutions and would make us lose the advantages of a single-point solution concept.

Two other choices we have made are to assume that blocking coalition C also blocks $D \supset C$ and to work with power indices defined over minimal winning coalitions only. Blocking single coalitions would not preserve null players who could gain power for "mediation" (turning a blocked coalition into a winning one by their entry – of course this coalition would be blocked 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 a feasible coalition.

Finally, the uniqueness of the friendly equilibrium for power indices also looking at surplus coalitions remains an open problem. With the aforementioned model counterexamples can be presented here a systematic search for them was in vain, now we believe the result to hold, but the present proof does not directly extend to those indices.

References

Albert M (2003) The voting power approach. Measurement without theory. European Union Politics 4(3):351–366

Aumann RJ, Drèze JH (1975) Cooperative games with coalition structures. International Journal of Game Theory 3:217–237

Banzhaf, III. JF (1965) Weighted voting doesn't work: A mathematical analysis. Rutgers Law Review 19(2):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):1–28

Bilbao JM, Jiménez A, López JJ (1998) The banzhaf power index on convex geometries. Mathematical Social Sciences 36:157–173

Brams SJ (1975/2003) Game theory and politics, 2nd ed. Dover, Mineola, N.Y.

Coleman JS (1971) Control of collectives and the power of a collectivity to act. In: Lieberman B (ed.) Social choice, Gordon and Breach, New York pp. 192–225

Deegan J, Packel EW (1978) A new index of power for simple n-person games. International Journal of Game Theory 7(2):113–123

Faigle U, Kern W (1992) The shapley value for cooperative games under precendence constraints. International Journal of Game Theory 21(3):249–266

Holler MJ, Packel EW (1983) Power, luck and the right index. Journal of Economics (Zeitschrift für Nationalökonomie) 43(1):21–29

Johnston RJ (1978) On the measurement of power: Some reactions to Laver. Environment and Planning A 10(8):907–914

Kilgour DM (1974) A shapley value for cooperative games with quarrelling. In: Rapoport A (ed.) Game theory as a theory of conflict resolution, Theory & Decision Library D. Reidel Publishing Co., Dordrecht pp. 193–206

Laruelle A (1999) On the choice of a power index. Discussion Paper WP-AD 99-10, Instituto Valenciano de Investigaciones Económicas, Valencia

Myerson RB (1977) Graphs and cooperation in games. Mathematics of Operations Research 2:225–229

Myerson RB (1980) Conference structures and fair allocation rules. International Journal of Game Theory 9(3):169–182

Owen G (1977) Values of games with a priori unions. In: Henn R, Moeschlin O (eds.) Mathematical economics and game theory: Essays in honor of Oskar Morgenstern, Lecture Notes in Economics and Mathematical Systems 141. Springer, Berlin pp. 76–88

Owen G (1982) Modification of the Banzhaf-Coleman index for games with a priori unions. In: Holler MJ (ed.) Power, voting and voting power, Physica-Verlag, Würzburg–Wien pp. 232–238

Penrose LS (1946) The elementary statistics of majority voting. Journal of the Royal Statistical Society 109(1):53–57

Riker WH (1962) The Theory of Political Coalitions Yale University Press, New Haven

Shapley LS (1953) A value for n-person games. In: Kuhn HW, Tucker AW (eds.) Contributions to the theory of games II, Annals of Mathematics Studies 28. Princeton University Press, Princeton pp. 307–317

Shapley LS (1962) Simple games: An outline of the descriptive theory. Behavioral Science 7:59–66

Shapley LS, Shubik M (1954) A method for evaluating the distribution of power in a committee system. American Political Science Review 48(3):787–792

Steunenberg B, Schmidtchen D, Koboldt C (1999) Strategic power in the European Union. Evaluating the distribution of power in policy games. Journal of Theoretical Politics 11(3):339–366