The Possibility of a Preference-Based Power Index*

Stefan Napel
University of Hamburg
Department of Economics
Von-Melle-Park 5
20146 Hamburg
Germany
Fax +49-40-42838-3957
napel@econ.uni-hamburg.de

Mika Widgrén
Turku School of Economics,
ETLA, CEPR and CESifo
Rehtorinpellonkatu 3
20500 Turku
Finland
Fax +358-2-4814302
mika.widgren@tukkk.fi

January 20, 2005

Abstract

This paper replies to the claim that preference-based power indices are impossible and that preferences should be ignored when assessing actors' influence in different interactions (Braham and Holler, 2005). The paper argues that preferences are an important determinant of potential and actual outcomes of social interaction and thereby a valuable ingredient of power analysis.

Keywords: strategic power, power indices, preferences

*This research has been supported by the Yrjö Jahnsson Foundation. We thank Hannu Nurmi for helpful comments.

1 Introduction

Power in political institutions is a *social* concept, referring to several human decision-makers with potentially conflicting interests. Max Weber's (1924/1948, p. 180) definition is still most authoritative:

> "In general, we understand by 'power' the chance of a man or of a number of men to realize their own will in a communal action even against the resistance of others who are participating in the action."

Therefore, it is somewhat surprising that one of the two most prominent measures of decision power, the Penrose-Banzhaf index, formally coincides with a measure used by engineers to evaluate the reliability importance of *electrical* or *mechanical parts* in multi-component systems, the Birnbaum index (cf. Birnbaum, 1969, or Barlow and Proschan, 1975). Similarly, it seems a quite peculiar coincidence that the probabilistic assumptions underlying the Penrose-Banzhaf index and the equally prominent Shapley-Shubik index are essentially the Maxwell-Boltzmann statistics and the Bose-Einstein statistics used in physics to describe the behavior of particles such as *photons, electrons*, or *nuclei* (see e. g. Feller, 1968, p. 40f).

The interaction of (sub-)atomic particles or electrical circuits differs in multiple ways from that of *human agents*, who – in many contexts – form expectations and think strategically before consciously taking their decisions. This has recently motivated several attempts to introduce measures of power which explicitly take the difference between human players and electrons into account (see Steunenberg, Schmidtchen, and Koboldt, 1999, Garrett and Tsebelis, 1999, and Napel and Widgrén, 2004). Unfortunately, these measures have met with great skepticism amongst advocates of traditional power indices.

In a very eloquent critique, Matthew Braham and Manfred J. Holler (2005, henceforth B&H) have claimed that indices based on utility-maximizing and strategizing agents are not only unneeded, but undesirable and conceptually meaningless. They argue that "the notion of strategic power ... cannot be taken to be saying anything intelligible about power in the fundamental sense of a 'generic ability'" (p. 154). Their central tenet – provocatively mislabelled "Core Theorem of the Measurement of Power" (p. 146) – is a dogmatic exclusion of *any* reference to a player i's preferences in the evaluation of i's power to affect outcomes under a given decision procedure.

The benefits of preference-based analysis of decision procedures seem to be self-evident in studies not related to power indices. For example, debate in the literature on voting procedures (investigating criteria like outcome-based distance between voting rules, Condorcet efficiency, etc.) is not on whether or not one should refer to preferences, but which preference distributions are best considered (see e. g. Nurmi, 1988 and 1992, in contrast to Regenwetter, Adams, and Grofman, 2002). However, in order to make intelligible statements about decision power one must, according to B&H, rely on preference-free indices like the ones mentioned above.[1] We strongly disagree and feel it necessary to respond

[1] For a comprehensive overview see Felsenthal and Machover (1998).

to B&H's article, which can well serve as an elaborate summary of the reactions also by other advocates of conventional indices. Our comment consists of four main points, that correspond to the following sections of this paper.

2 Analysis of Rules for Strategic Interaction Requires Analysis of Strategic Interaction

Not *all* political decisions are taken in a strategic way. For example, there is little scope for sophisticated behavior and mutual outguessing in a binary referendum involving millions of voters. So, it would be futile to argue that *every* measure of political power should be founded on a model of strategic interaction.[2] *Some* – or rather quite many – political decisions, however, involve significant strategic interaction. If the number of players is sufficiently small, the stakes are sufficiently high, and a reasonable amount of information about others is available, rational players will use whatever resources they have (voting weights, rights to set the agenda, to make the final proposal, to abstain from an election, etc.) in a *strategic* way. That is, unlike electromechanical parts in a vacuum-cleaner or photons passing through a slit, they respond consciously to anticipated behavior of others taking into account that others try to do the same.

Therefore, if rules for decision-making which make the involved actors engage in strategic interaction are to be evaluated, the Penrose-Banzhaf or Shapley-Shubik indices are not particularly useful tools. They *may* yield useful first approximations of how likely a given player is to affect an outcome in special cases. In general, accurate analysis of decision-making procedures and institutions in which players behave strategically for whatever preferences they happen to be endowed with must take the latter fact into account. Since strategic interaction rests on known or suspected preferences of all players, a given player's power cannot reside "in, and only in, the strategies available to her given by the game form" (B&H, p. 139), but also depends on how alternative outcomes are evaluated by the players.

This is not saying that we postulate any particular monotonic relation between "the power index value assigned to a player" and "how that player values the outcome", as B&H (p. 143) rather inaccurately describe "the basic intuition of a strategic power index." What preferences are needed for is to screen outcomes that are possible and can potentially be affected or forced by a given player from those that are not possible given all players' strategic efforts to realize their own will – not more, but also not less.

In line with the traditional literature, B&H (p. 139) give the following very narrow definition of a power index:

> "A *power index* assigns to each player of an n-person *simple game* – a game in which each coalition that might form is either all powerful (*winning*) or

[2]That is to say, we do not question the general legitimacy of power measures that are non-strategic. In fact, we have been using them ourselves where we deem their use appropriate (see e. g. Baldwin and Widgrén, 2004).

> completely ineffectual (*losing*) – a non-negative real number which purportedly indicates a player's ability to determine the outcome of the game. This ability is a player's power in a game given the rules of the game."

Power indices that treat a given collective decision problem as an aptly-called *simple game* can produce useful information if, and *only if*, the characteristics of the problem are indeed adequately captured by this Spartan model of winning and losing coalitions, which traditional indices take to result from independent coin tosses corresponding to – from an *a priori* perspective – fully random *yeah* or *nay*-votes. If coalitions and, more relevantly, the political outcomes associated with them are the result of strategic pursuit of objectives – which from an *a priori* perspective may well be taken to be random – in a decision framework that allows for more sophisticated behavior than just a simultaneous vote, then measures of the ability to affect outcomes (i. e., power) must be more sophisticated, too.

3 Bush's Response to Gaddafi's Preferences Determines Gaddafi's Set of Outcome Alternatives

Preferences and social interaction sometimes make for complicated dependencies between agents' decisions – be they simple *yeah* or *nay*-votes or complex bargaining strategies. These dependencies are what rational choice models and non-cooperative game theory try to predict, and what classical power indices do not capture. Which outcomes a given player i can actually select between – and thus be non-redundant for – depends crucially on the action or strategy choices of *other players* $j \neq i$ (except in degenerate cases of 'social interaction' such as Robinson Crusoe deciding between *Treasure Island* and *Parcival* on a lonely desert island, which is B&H's not particularly fortunate illustrative example).

Let a social decision problem be described by a *game form*, i. e. a complete account of the strategies available to each of n players, in what order and based on which information these players take their actions, and how strategy profiles map to outcomes. Denote by X the entire *set of outcomes* that can be realized by some strategy profile. We will refer to the subset of those outcomes which player i might have the ability to select between as *player i's set of outcome alternatives*.[3] An outcome $x \in X$ becomes part of this set along with one or more other outcomes y if for some circumstances – in particular a combination of strategy choices by all *other* players – player i induces outcome x for one of her strategies and outcome y for another. We will write $X_{i|R}$ to denote the subset of i's outcome alternatives under context-specific conditions or restrictions described by R. Generally, R would refer to the behavior of other players as expressed by a set of possible strategy combinations – those which survive iterated deletion of dominated strategies or which are rationalizable – or even a unique combination of their strategies.

Since a player's power or ability to affect the outcome of interaction is determined by her conditional set of outcome alternatives, $X_{i|R}$, player j's behavior directly affects i's

[3] We do not use the simpler term *set of alternatives* to avoid confusion with the set of feasible strategies.

power. There are two determinants of j's behavior and hence, for given game form, also of i's power. The first are *j's preferences* over the various elements in his respective set of outcome alternatives. Player j could, for example, regard one of his strategies as very undesirable and therefore all outcomes requiring this strategy to be chosen by j cannot be elements of $X_{i|R}$. This might implicitly be acknowledged by B&H (cf. p. 138, fn. 3). The second determinant is *j's anticipation of i's behavior* (unless j has a strategy dominating all others). This is in turn determined by *i's preferences* and i's expectations about j's behavior, etc.[4]

As an example, consider a man and a woman dating for the first time in a fancy restaurant. Whether the set of outcome alternatives of the man includes the option "continue the evening by having an additional glass of wine or even more in my or her apartment" or only the element "say good-bye and perhaps meet again next week" depends, first, on whether she – despite her definite original plans to only have dinner – has a preference for having an additional drink but nothing more, for having an additional drink or even more, for having an additional drink if this implies more, or for just saying good-bye. Second, it depends on whether she anticipates him moving or (not) resisting her move towards 'even more' once they have arrived at the apartment. These expectations – and thus ultimately his set of outcome alternatives – are partly determined by his observed or supposed *preferences*. Neither her nor his preferences can therefore be omitted if the man's ability to affect the outcome of this stylized interaction is evaluated. A similar, more political story could be told of the changing powers of Middle East dictators such as Muammar Gaddafi of Libya or, formerly, Saddam Hussein of Iraq: assumed or known preferences over alternative uses of imported engineering parts and chemicals and for (de-)stabilizing the region have a tremendous impact on the purchasing power of their oil fortunes as well as their political set of outcome alternatives and power. The link between own preferences and power via the strategic response to these preferences by others is in our view convincingly illustrated by Mr Gaddafi's recent change of mind, and the US administration's response to it.

We agree with B&H that "power is about events or potential outcomes themselves, not the utility attached to these outcomes" (p. 155). The utilities attached to outcomes, however, are the primary source of dependency between the decisions of different players. It is important to judge which *are* the possible events or potential outcomes. Because classical power indices rest on the notion that player's decisions are *independent*,[5] they cannot identify those potential outcomes that are ruled out by interdependency (or, alternatively, rendered particularly prevalent).

The independence assumption in the analysis of political decision-making has an interesting analogue in the economic analysis of imperfectly competitive markets. Models of *monopolistic competition* (Chamberlin, 1933) consider firms which compete with slightly differentiated goods but *de facto* behave as completely independent monopolists serving different markets. Such models can, amongst other things, be used to analyze the equi-

[4]Resolution of this circularity is the chief topic of game theory.

[5]Note that even the so-called *homogeneity assumption* yielding the Shapley-Shubik index and, similarly, *partial homogeneity* in Straffin's framework (1977) imply uncorrelated *yeah* or *nay* decisions (despite correlation in the respective probability of a *yeah* or *nay*).

librium number of firms and market power in an industry for which strategic aspects of market entry (e. g. limit pricing or possible post-entry price wars) do not matter. If they matter, models using non-cooperative game theory with equilibrium concepts designed to capture strategic interdependence have proved far more successful. Both approaches have their merits and complement each other.

4 Power Measurement Demands More than Power Ascription

B&H argue that power "[c]onceived of as a 'potential', 'capacity', or 'ability' ... [is] a *dispositional* concept akin to terms such as 'soluble', 'brittle', 'flammable', etc." (p. 150f). Therefore, according to B&H, a player either has the disposition of being powerful (in the sense of being able to affect the outcome) or does not have it "independent of ... [its] manifestation or exercise." Both this black-or-white view (fitting the losing-or-winning-coalition setting of simple games) and the supposed irrelevance of actual manifestations of the disposition are questionable.

'Powerful' can indeed be understood as a dispositional concept like 'soluble' or, e. g., 'carcinogenic'. This does, however, neither preclude nor render superfluous quantitative analysis of the degree of the disposition. Some varieties of salt (even of table salt) are clearly more soluble than others in a specifiable sense. Both the xenon isotope (Xe-133) used by hospitals to investigate patients' pulmonary ventilation or regional brain perfusion and plutonium have the dispositional property of being carcinogenic. But while one poses a minimal risk, largely outweighed by the potential benefits of an accurate diagnosis, the other – emitting mostly α rather than β and γ-rays and having a half-period of decay of over 24,000 years – is a very powerful killer.

Evaluation of a player i's power requires more than just determining *whether or not* a player has the considered ability for some 'state of the world', referring to the prevalence (or absence) of the conditions necessary for player i being able to force a particular outcome or of "i [being] non-redundant for the outcome" (cf. B&H, p. 145f). While a power *ascription* can stop with dichotomous statements, power *measurement* is about *how much* power i has. This requires, first, a quantitative statement about how much power a player has in a given scenario which, we agree with B&H, can be preference-free even in contexts where the *given* scenario (set $X_{i|R}$) is the result of a particular form of strategic interaction. Second, it is necessary to aggregate measurements (rather than just "ascriptions of non-redundancy", p. 146) in a reasonable way. If the considered interaction is of a strategic character, this 'reasonable way' must attach positive weight (greater weight) to scenarios that are possibly realized (more likely realized) such that power can be exerted and the disposition may be manifested, and zero weight to scenarios which are impossible for equilibrium behavior.

5 There is No Clear Boundary between Physical and Preference-Based Constraints

B&H make quite arbitrary distinctions between restrictions of a player's set of outcome alternatives that are imposed by physical constraints and those that are imposed by different types of subjective mental constraints usually modelled by preferences. Physical constraints that make certain potential elements of a player's set of outcome alternatives unavailable for a given state of the world are to be reflected by excluding these elements from this set. B&H propose to proceed in the same way, i. e. to redefine the game form, in case of the mental constraint of a *phobia*, which may make a player "unable to want to do or experience something" (p. 149). Thus, the fact that, e. g., a claustrophobic person will never take the elevator if stairs are available should be reflected in the deletion of the option 'arrive by elevator' from X and $X_{i|R}$ whenever 'use the stairs' is an element. However, when the same aversion against lift rides results from, say, the person's strong preference for dieting and physical exercise – rendering use of the elevator simply a *dominated strategy* – the game form should, according to B&H, stay as it is and the fact that the option 'arrive by elevator' in $X_{i|R}$ will never be selected should be ignored. Since strong aversion continuously blends to phobia, such a discontinuity in modelling has little foundation.

The conceptual problem of how to correctly discriminate between various types of physical and mental constraints on agents' choices and power is an old one. David Hume's statement (1739/1951, Book II, Part I, Sect. 10):

> "I do not think to have fallen into my enemy's power, when I see him pass me with a sword by his side, while I am unprovided of any weapon. I know that the fear of the civil magistrate is as strong a restraint as any of iron, and that I am in as perfect safety as if he were chain'd or imprison'd."

suggests that this eminent philosopher did not deem it worthwhile to make a conceptual distinction at all. Many others have begged to differ. Note that even if the enemy in Hume's example were chained or imprisoned, i. e. the enemy were *physically* restricted in his ability to stage an attack, a successful escape and subsequent attack are still possibilities – with a probability anywhere in the range from zero to high.

In our view, one cannot draw any irrefutable boundary line between those potential elements of an agent's completely unrestricted set of outcome alternatives which are actually impossible, which are, in principle, possible but have zero probability, which are possible but have negligible probability, and which are possible with non-negligible probability. Therefore, B&H's claim that preferences must be excluded from an evaluation of an agent's ability to affect outcomes while phobias can be taken into account is unconvincing. Again, they refer to a world either black or white, while reality has many shades of grey. To us, looking at all potential choice problems of a player and weighting those that are objectively impossible with zero and others with weights derived from plausible preference assumptions and rational behavior is a reasonable way to deal with these shades. Dogmatic rules for the assignment of either weight 0 or 1 are bound to fail.

As illustration of the unresolvable difficulty in distinguishing between purely exogenous physical constraints and preferences precluding certain choices under rational behavior, consider two equally skillful NHL hockey players who score the same average number of goals per season. From a macro perspective both are equally powerful members of the team, where power refers to the ability to score a pivotal goal and win the match. However, player 1 is often tired on Saturday since he meets his buddies in a bar the night before, and player 2 is frequently unfocussed during Wednesday's games since he has taken care of his undisciplined kids in the afternoon. Knowing this, their docile coach gives the latter more ice time on Saturdays and the former on Wednesdays, resulting in doubly asymmetric abilities to affect the outcome at the micro level. The respective per-game sets of outcome alternatives (related to number and precision of shots and passes, speed of sprints, etc.) on the one hand differ at an exogenous objective level – physical strength, mental strength, ice time –, and so B&H might acknowledge different ability to affect the outcome. On the other hand, these objective differences have their reason in player 1's subjective preference for a Friday night out, player 2's preference against arguments with his spouse, and the coach's strategic response to both. Leaving preferences and strategic interaction out of the picture may yield a reasonable approximation for the players' power at a macro level, but implies great inaccuracy in more refined analysis.

More generally, it can be argued that many of the objective physical or legal constraints[6] we find ourselves in are, ultimately, the consequence of earlier preference-based choices. The ability of George W. Bush to veto bills, for example, might be taken as independent of his preference for actually doing so, but certainly has a lot to do with his preference for running for office and US voters' (and Supreme Court judges') preferences in the 2000 and 2004 elections.

6 Concluding Remarks

Blanket claims about a supposed impossibility of taking preference into power measurement – called a "Core Theorem" by B&H – are not particularly constructive. Some of B&H's and our positions are probably reconcilable. Their key motivation seems to be opposition against a direct link between the utility attached to an outcome by a given player and the deduction of power from this subjective valuation of the outcome. Since the index proposed by Steunenberg et al. (1999) indeed interprets average utility levels as 'strategic power' (in a spatial voting framework where players try to minimize distance between the actual outcome and their most preferred one), B&H, as well as others, may have misperceived this problematic link to be the chief characteristic of strategic measures of power.

In the framework discussed in detail by Napel and Widgrén (2004), no such connection

[6]It is not clear if B&H would argue in favor of restricting moves to what is legal when analyzing, say, the EU Council's power to impose fines on countries in violation of the so-called Stability and Growth Pact, or whether respect for the law is merely a preference that according to them has no place in power analysis.

exists. Ex ante power is taken to be the expected impact that a player's possibly but not necessarily preference-driven and strategic behavior has on the outcome, i. e. the *sensitivity* of a reference outcome with respect to the considered player's behavior on which the constraints of utility maximization may or may not be imposed. The reference outcome ex post refers to a particular strategy profile that realizes a particular element from the set of all potential outcomes, associated with a terminal node of a game form; ex ante, it is considered a random variable. This is also what traditional power indices do, albeit they restrict sensitivity analysis to the 0-1 world of simple games and assume stochastically independent decisions by the players. In the mentioned framework, preferences may enter to determine, first, which potential outcomes serve as relevant reference points and, second, which weight they should be given in the light of dependencies when moving from possible ex post realizations of power to an ex ante measure.

In a recent application of our framework for strategic power analysis to the European Union's *co-decision procedure* (Napel and Widgrén, forthcoming), we have attached positive weight only to outcomes supported by a subgame perfect Nash equilibrium. This formalizes that all players j have the chance to reappraise their own strategy as player i contemplates a different choice after a change of i's preferences (induced e. g. by a lobbyist). This is neither necessary nor would we argue that it is the best option in every context. As one alternative, one may use the equilibrium $s^* = (s_1^*, \ldots, s_i^*, \ldots, s_n^*)$ for a particular realization of all players' preferences in order to pin down the behavior of all players j except that of player i (all players j, however, have to *believe* that i picks her equilibrium strategy) and then consider i's set of outcome alternatives, $X_{i|R}$, without attaching any special importance to s_i^*. This amounts to what we have labelled global *action trembles* in Napel and Widgrén (2004) or, equivalently, to global *unanticipated preference trembles*. One may then, for example, consider the maximal impact by *any* strategy choice of i or weight different choices $s_i' \neq s_i^*$ differently, e. g. based on 'closeness' to expected choice s_i^* either in terms of a metric on the set of i's strategies or by referring to the preferences that would make s_i' rather than s_i^* optimal given $s_{-i}^* = (s_1^*, \ldots, s_{i-1}^*, s_{i+1}^*, \ldots, s_n^*)$.

Braham and Holler (2005) have ruled out *any* reference to player i's preferences when i's power is to be measured. This precludes consideration of strategic interaction at all levels, even in determining the choices i really has in the light of at least mildly sophisticated behavior by other players. At the very least one should refer to i's preferences whenever they affect player j's behavior and, thus, i's set of outcome alternatives. If one then proceeds to measure i's power by studying the ex post choice set without further reference to preferences, some elements in i's resulting choice set would have to be selected in a surprise move in order not to be prevented by player j. In many contexts such non-equilibrium stupefaction is problematic. In many others – e. g. general elections for given party platforms, with virtually no effect of i's preferences on j's behavior – we agree with B&H that it is not.

References

Baldwin, Richard and Mika Widgrén (2004) 'Winners and Losers Under Various Dual Majority Rules for the EU Council of Ministers', in Matti Wiberg (ed.) *Reasoned Choices – Essays in Honor of Academy Professor Hannu Nurmi on the Occasion of His 60th Birthday*, The Finnish Political Science Asssociation.

Barlow, Richard E., and Frank Proschan (1975) *Statistical Theory of Reliability and Life Testing: Probability Models*, New York, NY: Holt, Rinehart and Winston.

Birnbaum, Z.W. (1969) 'On the importance of different components in a multi-component system', in: Paruchuri R. Krishnaiah (ed.), *Multivariate Analysis*, Vol. 2: 581–92, New York: Academic Press.

Braham, Matthew, and Manfred J. Holler (2005) 'The Impossibility of a Preference-Based Power Index', *Journal of Theoretical Politics* 17: 137–57.

Chamberlin, Edward H. (1933) *The Theory of Monopolistic Competition*, Cambridge, MA: Harvard University Press.

Feller, William (1968) *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., New York, NY: John Wiley & Sons.

Felsenthal, Dan, and Moshé Machover (1998) *The Measurement of Voting Power: Theory and Practice, Problems and Paradoxes,* Cheltenham: Edward Elgar.

Garrett, Geoffrey, and George Tsebelis (1999) 'Why Resist the Temptation to Apply Power Indices to the EU', *Journal of Theoretical Politics* 11: 291–308.

Hume, David (1739/1951) *A Treatise of Human Nature*, Oxford: Oxford University Press.

Napel, Stefan, and Mika Widgrén (forthcoming) 'The Inter-Institutional Distribution of Power in EU Codecision', *Social Choice and Welfare*.

Napel, Stefan, and Mika Widgrén (2004) 'Power Measurement as Sensitivity Analysis – A Unified Approach', *Journal of Theoretical Politics* 16: 517–38.

Nurmi, Hannu (1998) 'Discrepancies in the Outcomes Resulting from Different Voting Systems', *Theory and Decision* 25: 193–208.

Nurmi, Hannu (1992) 'An Assessment of Voting System Simulations', *Public Choice* 73: 459–487.

Regenwetter, Michel, James Adams, and Bernard Grofman (2002) 'On the (Sample) Condorcet Efficiency of Majority Rules: An Alternative View of Majority Cycles and Social Homogeneity', *Theory and Decision* 53: 153–186.

Steunenberg, Bernard, Dieter Schmidtchen, and Christian Koboldt (1999) 'Strategic Power in the European Union: Evaluating the Distribution of Power in Policy Games.' *Journal of Theoretical Politics* 11: 339–66.

Straffin, Philip D. (1977) 'Homogeneity, Independence and Power Indices.' *Public Choice* 30: 107–18.

Weber, Max (1924/1948) 'Class, Status and Party', in: Hans H. Gerth and C. Wright Mills, *Essays from Max Weber*, London: Routledge and Kegan Paul.