COMMON KNOWLEDGE IS POWER

ALEXIS V.BELIANIN

International College of Economics and Finance,
Higher School of Economics,
and IMEMO RAS, Moscow, Russia

icef-research@hse.ru

Abstract. The paper studies power in social affairs, with special implications of the role of power perceptions in determining the outcome of repeated social interactions. In political science, power is traditionally viewed as either an asymmetric *group* relation, or as an intrinsic *nonsectional* property of the social order. The present paper offers a framework which enables a 'dual' characteristic of power relations either as 1) the limiting outcome of the process of mutual deductions of the players' payoffs, or as 2) believed perception of the limiting state of an arbitrary deductions process which is *almost common knowledge* in the sense of Monderer and Samet (1989). This duality facilitates a unified approach to a wide range of social interactions, including coexistence of formal law and informal arrangements, prevalence of particular norms in different societies, and social coordination on some of multiple competing equilibria. Some illustrations of the applications of this approach to the analysis of social systems are provided.

1. Introduction: main ideas

This paper is about power in social affairs.

Power is everywhere. Two firms which bargain over a business contract attempt to exercise power to impose on each other the terms that would be most favourable for themselves. Spouses in a family quarrel over whose duty is it to do housekeeping. Political parties bargain over the right to govern the country, trying to impress the electorate by the authority of their slogans, and then (often) abuse their electorate's rights while holding the office.

Given this overwhelming importance of power in social sciences, it is somewhat surprising that the nature of this concept remains at large unexplored in the economics literature. 'Threat points' in bargaining or modern contract theory (Maskin and Moore, 1999) may be treated as partial exceptions, but the existence of such points is postulated axiomatically rather than derived from the first principles of the underlying social processes. By contrast, political scientists have several related concepts of power. However, unlike economists and natural scientists, they widely acknowledge this notion to be *essentially contestable*, that is, inevitably value-based, thus neither logically deducible from any first principles nor testable empirically (Gallie, 1955). Taken literally, this statement would mean that *the* definition of power does not exist: any that is set forth would inevitably be

---

Date: May 16, 2006.

The author is grateful to the audience of ICEF and VIth International Conference of HSE for valuable comments. The usual caveat applies.

value-based. At the same time, there exist several conceptual traditions in which political scientists characterize what is power (Ledyaev, 1998).

One of these, called *sectional* or *group*, comes up to Hobbes, and views power as asymmetric relation between individuals in which one of the actors has authority to impose his preferences on those of the other. In particular, *"G has power on X over K if G takes part in making decisions of X over K"* (Laswell and Kaplan, 1950, p. 76). In a classical book by Dahl, power is defined as *"the ability of one agent (A) to force another agent (B) to do something she (B) would not have done otherwise"* (Dahl, 1969, p. 80). Dahl views power as an instance of control over *behaviour*, which is based on the intrinsic ability of one actor to impose sanctions on the the other in case of disobedience of this latter. Lukes (Lukes, 1974) extends these characteristics from direct imposition of particular lines of *behavioural* to the ability to influence other people's *preferences and judgments*. A key feature of this tradition is that power is essentially viewed as a *zero-sum game*: whichever the power-taking side gains, the other one loses.

Another, *nonsectional* tradition views power not as a nonzero-sum game, but rather as a coordination game. One of the most influential proponents of this view, Talcott Parsons understands power as a *"generalized ability of the elements of a system to ensure its functioning, which is legitimized by the need to reach collective goals"* (Parsons, 1986, p. 103). In this way, power is not a disposition or faculty of specific actors, but a systemic property of social relations, a tool for social transformation, a mechanism of social coordination (Foucault, 1986).

The present paper takes seriously these two characteristics, and shows their inner connection with each other as well as with many issues which arise in economics in connection with the allocation of power, bargaining and political economy. All these situations are naturally modelled as noncooperative games with independent players who, having choices among several courses of actions, would like to *persuade* or *impose upon* the opponent a particular course of actions. A natural translation of the sectional view of power in the language of game theory is simply this: the more powerful a player is (or believes herself to be) the larger can be the (actual or believed) ability to affect payoffs in the game (her own as well as those of the opponent). In section 3 we build a stylized example of such pressure, concluding the section with a discussion of the limitations of this approach. These limitations are tackled in in a more general framework build on the interactive epistemology approach (Aumann and Brandenburger, 1995) developed in section 4. The main result of this section consists of two theorems that characterize power in dual ways: as a limit of a dynamic process of *deductions*, and in terms of common knowledge of the outcomes of such process. This method, *inter alia*, suggests possible solution to the conceptual problems of the sectional approach, and tights it up to the nonsectional one. Yet the exact implications of these findings await further research, with applications to institutional imperfections, political manipulations and multilateral bargaining in section. Some directions of this research are collected in section 5.

2. Related literature

The problems discussed herein were of interest to many researchers in economics, although none of them seem to have endorsed the same line of approach. Thus, Mailath, Morris and Postlewaite in an unpublished paper (Mailath e.a., 2001) observe that laws *by themselves* do not change payoffs of the players, from which they derive that coordination on either legal or non-legal actions stems from *authority* of one of the parties. Their notion of *authority*, however, is defined by the ability of one party to force the equilibria preferred by this party by means of *cheap talk*, which essentially leaves unanswered the main question: where this authority comes from? The framework of these authors fails to provide an answer to that question. Mailath e.a. concede that laws may affect behaviour "because they affect the utilities individuals derive from particular strategy choices" (Mailath e.a., 2001, p. 5); — however, they attribute this influence to individual "psychology", which they explicitly put aside of their framework.

Geanakoplos and coauthors (Geanakoplos e.a., 1989) aim at capturing the role of players' beliefs by defining the utility of the players in a strategic form game as a function of the product of strategies' profile and players' beliefs. This extension defines *psychological game* whose equilibria are defined in terms of strategies and beliefs. A fixed point of the psychological best reply correspondences, however, can be guaranteed only if individual beliefs conform to *reality*, i.e. actual beliefs held by all players. To close this model, Geanakoplos e.a. have to assume coherence of the various levels of beliefs an common knowledge of coherence in the sense of (Brandenburger and Dekel, 1993). This setup treats *reality* as exogenously given, and effectively rules out the main question: where these beliefs come from?

Our approach to the analysis of these problems is "classical" in the sense that it draws on the standard economic notions of common knowledge (Aumann, 1976) and hierarchy of types (Mertens and Zamir, 1985), (Tan and Werlang, 1988), (Brandenburger and Dekel, 1993). By combining these concepts with the notion of an hierarchical belief system (Aumann and Brandenburger, 1995), we obtain a dual characteristic of *power* in terms of both common knowledge and dynamic exploration of the environment in question.

3. A first approach: deduction games

3.1. **Example.** As a stylized story, consider two players who bargain over the terms of a contract which is in their mutual interests. This situation can be represented as a standard coordination game with two strategies (interpreted as possible terms of the contract) and symmetric payoffs (see Table 1, referred as to Game 1). In Game 1, player 1 (he) prefers strategy 1 (payoff 2) to strategy 2 (payoff 1), which is preferred to the disagreement payoff of 0; preferences of player 2 (she) are symmetric, with strategy D2 preferred to strategy D1.

1 ╲ 2	D1	D2
C1	2,1	0,0
C2	0,0	1,2

Table 1

Game 1 has two equilibria in pure strategies, which we call *contracts* 1 and 2, corresponding to coordination on first or second strategies, respectively[1]. Players' preferences about the two contracts, however, are opposite, which creates incentives to persuade the opponent to choose the equilibrium they prefer. In the spirit of political science literature (see, e.g., the work (Dahl, 1969) cited in the previous section), we interpret this potential as the *ability to affect payoffs of the opponents from the strategies they do not want*

[1]A mixed strategy equilibrium with probability 1/3 on the preferred strategy of both players shall not be considered, as it invokes different description of players' rationality.

them to play. At first sight, this specification might appear unduly restrictive, but little reflection suggests it is not. Indeed, exercise of *power* most often means enforcement of specific action by making *the other* actions less attractive for the opponents. In case of contracting, player 1 may say to his opponent: if you want to have contract 2, then you will have to do additional job for me, whose extra disutility might force you to think of accepting contract 1 which I prefer. Other instances of similar persuasion include such practices as: promise to send thieves to jail (which is presumably less attractive than a short benefit from theft), legal prosecution of fraudulent businesses (including compensation and penalty fees which make frauds unattractive), pressure on foreign governments in international politics (threatening sanctions in case of "disobedience") etc. The bottom line of these and other similar cases is that a threat to decrease opponent's payoffs from an undesirable course of action is used as the major tool of persuasion: if you don't do what I request you to do, you will suffer. Whether (and when) this threat shall to be implemented in reality is of course a key issue, to which we shall return in a while.

To formalize the above argument, assume that player 1 can threaten player 2 to deduce her payoff by a fraction η in case she plays D2. To make sure this threat works, given the payoff in Game 1 it should be the case that $\eta > 0.5$, as this would make the strategy D2 strictly dominated for player 2. However, exercising power is costly for player 1 as well — we denote the cost of this threat through ϵ^1, which has to be paid lump-sum. Hence we obtain Game 2 (see Table 2):

$1 \diagdown 2$	D1	D2
C1	$2-\epsilon^1, 1$	$-\epsilon^1, 0$
C2	$-\epsilon^1, 0$	$1-\epsilon^1, 2(1-\eta)$

TABLE 2

In Game 2, deductions of $2\eta > 1$ from the payoffs of player 2 make (C1,D1) more desirable as the contract than (C2,D2). What should player 2 do then? First, she may treat this threat as empty, and not believe the ability of the opponent to do so. This possibility reveals fundamental uncertainty about the actual limits of bargaining power of the opponent, which player 2 might want to verify over the long run. Another option is to make similar deduction from player 1's payoffs. Such strategy is clearly relevant; yet in a sense it reduces to the third case, namely defensive measures undertaken by player 2 against the threat of player 1. In case of a contract, this will be a counteroffer: I can do extra job for you, but if you pay me extra money. Other examples may include: taking more precaution before committing a theft, hiring extra lawyers (or mafia) to protect one's business, or making extra stocks of goods and equipment in case of sanctions. We assume such defence blocks a fraction $\zeta < 1$ of the potential damage invoked by player 1's threat[2], so that the total payoff of player 2 would be not $2 - 2\eta$ but $2 - 2\eta(1 - \zeta)$. This defense comes at a cost $\xi^1 < 1$ to be borne in either case[3], resulting in Game 3 (see Table 3):

This endeavor is clearly worthwhile if $\eta(1 - \zeta) - \xi^1 < 1$, which restores desirability of contract 2 for player 2, and effectively returns the situation to the original coordination game, but this time with *lower* payoff to both players: if contract 2 is implemented in

[2]Parameters η and ζ in this context can be interpreted as *bargaining powers* of the players.

[3]Series ϵ^t and ξ^t can be interpreted as the *cost of conflict*.

$1 \diagdown 2$	D1	D2
C1	$2-\epsilon^1, 1-\xi^1$	$-\epsilon^1, -\xi^1$
C2	$-\epsilon^1, -\xi^1$	$1-\epsilon^1, 2(1-\eta(1-\zeta))-\xi^1$

Table 3

Game 3, it will involve deadweight losses to both parties in comparison to Game 1. Facing such defensive actions, player 1 who initiated this bargaining, will probably raise the stake, and we assume this happens in a form of defence against the defence of player 2, pursued with the same strength η as the original attack. As a result, payoff of player 2 in the (C2,D2) profile becomes

$$2(1-\eta(1-\zeta\eta))-\xi^1$$

which increment comes at a cost ϵ^2 for player 1. This latter will respond by decreasing the strength of 2's attack by ζ at a cost ξ^2, leaving to player 2:

$$2(1-\eta(1-\zeta\eta)(1-\zeta))-\xi^1-\xi^2$$

Continuing in the same way, payoffs of player 2 at the (C2,D2) profile at each stage of this bargaining procedure *short of the cost series* ξ^t can be represented by the signed product series of the form

$$\begin{cases} 2-2\eta+2\eta^2\zeta-\ldots(\mathbb{I})^t\cdot 2\eta^t\zeta^{t-1} & \text{for the } t^{th} \text{ order attack;} \\ 2-2\eta+2\eta^2\zeta-\ldots(\mathbb{I})^t\cdot 2\eta^t\zeta^{t-1}+(\mathbb{I})^{t-1}\cdot 2\eta^t\zeta^t & \text{for the } t^{th} \text{ order defence.} \end{cases}$$

where

$$\begin{cases} \mathbb{I}= \;\;1 & \text{for } t \text{ odd;} \\ \mathbb{I}= -1 & \text{for } t \text{ even.} \end{cases}$$

A *sequence* of games introduced above is an example of what we call a **threat game** and denote $\eth$. Stage games of $\eth$ are called **deduction games**, they are denoted Γ^t. The outcome of any threat game is clearly determined by three factors: 1) value of the series of payoffs in the deduction games, 2) cost of conflict series $\{\xi^t\}$ and $\{\epsilon^t\}$ and 3) position of the limit point of the series of deduction games (if any) about 1, the equilibrium payoff under contract 1. Namely, the series of subsequent attacks and defences will stop whenever expected gains of player 2 from forcing the contract 2 exceed the aggregate cost of conflict for her, or whenever the outcome of the conflict short of conflict cost would result in payoff less than 1.

3.2. **Discussion.** A sequence of deduction games just described represents a very specific instance of possible bargaining process between two players. Our aim is to make much more general claims about manifestations of power in social interactions, as expressed in the deliberate changes of the players' payoffs. The tools for this analysis will be developed in the next section, when it will become clear that the example of deduction games is illuminating in many respects.

First, deduction games can be very general, including negative deductions (i.e. augmentation of some payoffs of the opponents), changes of payoffs in more than one profile, launching counter-attacks on the opponent's payoffs etc. Thus, the deduction games merely offer a *general framework to describe the ability of players to impose one's will on the choices of the opponents*. To put differently, deduction games offer a set of possible arguments (be these explicit or implicit) in the bargaining process, without actual play of all (or any subset) of these games before colluding at a particular strategies' profile in the main game. In view of this, instead of further analysis of just one of many possible processes of that kind, we shall aim at a general characteristic of the class of threat games. Our example and the above discussion suggest that the set of all stage games for the threat games should belong to the class of *generic* games, that is, open and dense in the Euclidean space of dimension $I \times \prod_{i=1}^{I} |A_i|$, where I is the number of players and $|A_i|$ is the cardinality of the strategies' space of player $i \in I$ (Fudenberg and Tirole, 1991, p.479). If not, the players will have "no room" in their payoff spaces for deductions, attacks and defences, making these latter void.

Another important aspect of the deduction games is the extent to which the opponent is ready to go along the sequence of stage games. This possibility reflects an inherent *uncertainty* of the players about each other's bargaining capacities, which uncertainty is captured by the set of all deduction games (not necessarily the one described in the previous subsection) as formed by all possible sequences of payoff deductions. Call the set of all such games X — this notation will be used in the next section. Bayesian approach to interactive decisions (Tan and Werlang, 1988) implies that every player i should have probabilistic beliefs over X. The space of all probability distributions over the first-order belief space is denoted $\Delta(X)$, with generic element β_i^1. Since beliefs of any player are unobservable for the others, each player should have belief about 1) the game to be played from the set X, and 2) beliefs of the other players over this set. Cartesian product $X \times \Delta(X)$ constitutes the second-order uncertainty space, over which all players should have second-order beliefs. The second-order beliefs are unobservable either, so the sequence of beliefs continues *ad infinitum*. An infinite sequence of beliefs $(\beta^1, \beta^2 \dots)$ is the type of each player, denoted θ_i. Two consistency conditions have to be imposed on these beliefs, and all higher order beliefs. First, the marginal of β_i^k on X^{k-2} should equal β_i^{k-1}. That is, any distribution of order k refines distribution of order $k-1$ while keeps intact the beliefs over all all lower-level spaces[4]. Second, the marginal distribution over X^0 of any level k must agree with that of level $k-1$. These notations are standard in the literature, and will be used to develop a more general framework for the notion of power presented in the next section. This construction simultaneously generalizes the example of deduction game and provides an alternative and more intuitive characteristic to the notion of power.

4. Analytical framework

Fix a strategic *game form* $\langle I, \{A_i\}\rangle$, where $I = \{1, \dots i \dots n\}$ is a finite set of players, and A_i is the set of actions (pure strategies) available to player i (with generic pure strategy a_i and strategies' profile $a \in A \equiv \prod_i A_i$; in the previous section, $A_1 = \{C1, D1\}$, $A_2 = \{C2, D2\}$). Analogous definitions apply to the profiles of mixed strategies $\sigma = (\sigma_1 \dots \sigma_n)$, where each σ_i is the probability distribution on the set A_i. Payoff of each player $u_i(a, x)$ (or $u_i(\sigma, x)$) is a function of the strategies' profile $a = \prod_i a_i$ (resp., $\sigma = \prod_i \sigma_i$), and of random vector $x = \{x_i\}_{i \in I}$. In the light of the previous discussion, we interpret this vector as possible realizations of the deductions, so that $x_i = (\eta_i, \zeta_i)$, $\forall i$. The set of all possible values of x, denoted X, determines the set of all generic games which reflect players' beliefs about the objective (exogenous, nonstrategic) characteristics of their opponents i.e. their

[4]We use indices k to refer to *levels*, and indices t — to *time* variables.

powers. The key question is of course how these powers can be revealed to all players. To do so we need a broader concept of interactive belief system introduced by Robert Aumann and used in (Aumann and Brandenburger, 1995); we adapt their notation to our case.

An *interactive belief system* for the game form Γ is defined as the Cartesian product of a finite number n of possible *states spaces* S_i, with generic state s_i – one for each player. This space is denoted $S = \prod_i S_i$, with typical element $s = \prod_i s_i$. Each state is associated with a unique collection of

- probability distribution $p(S_{-i}, s_i)$ of player i in state s_i on the set $S_{-i} = S_1 \times \ldots S_{i-1} \times S_{i+1} \times \ldots S_n$ of types of the other players (i's *theory*);
- an *action* $a_i(s)$ of player i in state s;
- *payoff function* $u_i(s) : A \to \mathbb{R}$.

Any subset E of S is called *event*; we denote through $\mathscr{E}$ the Borel σ-algebra of events over S. The definition of $p(S_{-i}, s_i)$ admits canonical extension to S as follows: for any $E \in \mathscr{E}$ let $p(E, s_i)$ be the probability that i at S_i assigns to the set $\{s_{-i} \in S_{-i} : (s_i, s_{-i}) \in E\}$. Functions defined on the space $(S, \mathscr{E})$ are denoted by bold letter ($\mathbf{a}, \mathbf{u}$ etc.); in our settings all such functions will be measurable, and thus can be interpreted like random variables. We use square brackets $[\cdot]$ to refer to particular events: thus, $[u]$ where $u \equiv (u_1 \ldots u_n)$ is the event $\{s \in S : \mathbf{a}(s) = u\}$ that the game actually played has payoffs u (with any possible beliefs and actions profiles); and $[a] \equiv \{s \in S : \mathbf{a}(s) = a\}$ is the event that the profile a is played in state s. Naturally, if $E \equiv s_i$, then $p(E, s_i) = 1$ — that is, player i in the state s_i assigns unit probability to being in that state (Brandenburger and Dekel, 1987). Beliefs of player i in s_i about the strategies' profile of other players, $a_{-i} \in A_{-i}$ is called i's *conjecture* at s_i and denoted $p([a_{-i}], s_i)$.

Thus defined, the interactive belief space can be thought as an extension of the standard definition of types space for general games of incomplete information (Harsanyi, 1967-68), developed in (Mertens and Zamir, 1985), (Brandenburger and Dekel, 1993) and other papers. The set X of possible games in game form Γ is the zero-level uncertainty X^0 over which each player i has beliefs in the space $\Delta(X^0)$. The spaces of player i's belief of level $k \geq 1$ are defined recursively as $X^k = X^{k-1} \times \Delta(X^{k-1})$. Each infinite hierarchy $\theta_i \in \times_{k=0}^{\infty} \Delta(X^k)$ is called *type* of player i; the set of all types of player i is denoted Θ_i; finally, the product $\Omega \equiv X^0 \times (\Theta_i)^n$ is the state space. The construction presented in the previous paragraph essentially "augments" this states space by the set of possible strategies A, i.e. $S \equiv \Omega \times A$ (Dekel and Gul, 1997).

Finally, let us introduce basic knowledge structures over the set S, where knowledge is associated with probability 1 (Brandenburger and Dekel, 1987); see a discussion in (Aumann and Heifetz, 2002), (Dekel and Gul, 1997), (Fagin e.a., 1995). Let $\mathcal{P}_i$ be the (countable) partition of S into measurable sets P_i with positive measure, defined for all i. Denote by $\mathscr{F}_i$ the σ-field generated by $\mathcal{P}_i$. Elements of partition $P_i(s)$ are defined as the set of states which player i cannot distinguish from S. Player i is said to *know* event E at state s if $P_i(s) \subseteq E$, or alternatively, if $Prob_{s_i}(E) = 1$. *Knowledge operator* of player i, $\mathcal{K}_i : 2^S \to 2^S$ associates with every $E \subset S$ the set of states in which the player knows this E, i.e. $\mathcal{K}_i E = \{s : P_i(s) \subseteq E\}$. An event is called *mutual knowledge*, denoted $\mathcal{K}^1 E$ if it belongs to $\mathcal{K}^1 E = \mathcal{K}_1 E \cap \ldots \cap \mathcal{K}_n E$, *self-evident* if $E \subseteq \mathcal{K}_i E \; \forall i$, and *common knowledge* ($\mathcal{CK}E$) if it belongs to $\mathcal{K}^1 E \cap \mathcal{K}^1 \mathcal{K}^1 E \cap \ldots$. Alternative definition of common knowledge of an event E at s is that there exists an event E^1 such that $s \in E^1 \subseteq E$ and $E^1 \in \mathscr{F}_i \; \forall i$ (Monderer and Samet, 1989). There exists a canonical representation of this knowledge structure (called also Aumann structures — (Fagin e.a., 1995)) through the state space (semantic) description of uncertainty (Brandenburger and Dekel, 1993).

Similarly to knowledge we may define beliefs. An event E is said to be *r-believed* at $s \in S$ if $p(E, s_i|P_i(s)) \geq r$ for all i and some r close to 1, formally $\mathcal{B}_i^r(E) = \{s : p(E, s_i|P_i(s)) \geq r\}$. In words, player i r-believes in E if in all states of the same element P_i of her information partition at s_i her probabilistic belief in E is at least r. Event E is called *almost common knowledge* (Monderer and Samet, 1989) if all players r-believe in E, r-believe that all players r-believe in E etc.

We use these specifications to extend the results obtained by (Mertens and Zamir, 1985) and (Brandenburger and Dekel, 1993) to the interactive belief systems. First, we state several technical assumptions.

Assumption 1. *The space X is Polish (complete separable metric).*

Assumption 2. *The space X is compact.*

Assumption 3. *All players have rational beliefs in the sense that if $p([a_{-i}], s_i) = 0$ then*

$$\begin{aligned} u_i((a_i', a_{-i}'), x)(s) &< u_i((a_i'', a_{-i}'), x)(s), \ \forall [a_{-i}'] \neq [a_{-i}], \ a_i'' \neq a_i' \\ &\Rightarrow \sigma_i(a_i') = 0. \end{aligned}$$

A few comments are in place. Both assumptions 1 and 2 are standard in the literature on hierarchies of beliefs, if not on the interactive belief systems. To qualify, recall that the set X of possible deduction games (as determined by the mutually uncertain bargaining power of the players) should allow for some profitable deductions from all payoffs — i.e. all such games should be *generic*. As a subspace of the Euclidean space of dimension $I \times \prod_{i=1}^{I} |A_i|$, this space clearly satisfies Assumption 1. Assumption 2 is also not difficult to justify provided the set X contains all its limit points, which holds because the payoffs in all games in X are contained in the simplex spanned by all players' payoffs, and is therefore bounded. Note that if X is Polish, so is the space $\Delta(X)$ of all probability distributions over X^0, and recursively, all spaces of higher order $\times_{k=0}^{\infty} \Delta(X^k)$ (e.g. (Dellacherie and Meyer, 1978, p.73). The same is true of Assumption 2: the space of all probability distributions over a compact set is compact, and so is any product of compact sets by the Tikhonov's theorem.

Assumption 3 says that if players think of some set of states in which the opponents want to play the profile $[a_{-i}]$ as of impossible, they will never want to play the strategy a_i' that is strictly dominated in the complement to that set. To put it at work, we need one more definition:

Definition 1. *A sequence of games $\Gamma^t \in X$ is called* ***monotone*** *if $\mathcal{P}_i^{t-1} \subset \mathcal{P}_i^t, \ \forall i$*

In words, a sequence of games is monotone (and accordingly, the σ-fields $\mathscr{F}_i^t$ generated by the partitions $\mathcal{P}_i^t$ are *filtrations*) if it refines the players' information partitions. Stated as such, monotonicity is essentially a statement of information efficiency of the games from $\eth$.

These properties are used in the proof of the following result which bears on an explicit construction of *some* threat game $\eth = \{\Gamma^t\}, \ t = 1, 2 \ldots$. The intuition is as follows: inasmuch as players are ignorant about each other's strategies and payoffs, they want to have more specific information about these in the course of repeated interactions. This is obtained through sequences of the following two-stage games. At stage 1 of each period

t, the players choose independently from each other the values of deductions x_i^t from each other's payoffs in such a way that

$$\hat{x}_i^t \in \arg\max_{x_i^t} \mathsf{E}_{p_i^t([a_{-i}],s_i)} u([\sigma_{BR_i}^t, a_{-i}^t], x_i^t, s_i | P_i^t(s)) \tag{1}$$

In words, players choose the value of deduction which maximizes their payoff in all states given their best response strategy $\sigma_{BR_i}^t$ at t and their conjecture $p_i^t([a_{-i}], s_i)$ about the opponents' actions at t, conditional upon the element of period t's information partition which the state s_i belongs to. At stage 2, players actually choose an action which maximizes their expected payoff given the information they have about the state of the game, i.e.

$$\hat{\sigma}_i^t = \arg\max_{\sigma_i^t} \mathsf{E}_{p_i^t([a_{-i}],s_i)} u((\sigma_i^t, a_{-i}^t), s_i | P_i^t(s)) \tag{2}$$

We call the Γ^t games played according to the above rules the **BR games**.

Theorem 1. *Any sequence of BR games which is monotone has a limit, and the state at which this limit is obtained is common knowledge.*

Proof. Consider a sequence of strategies' profiles $(\hat{x}_i^t, \hat{\sigma}_i^t)$ of player i along some $\eth$. Because all players have rational beliefs in the sense of 3, and each sequence is monotonic, individual partitions $\mathcal{P}_i^t$, and the σ-algebra $\mathscr{F}_i^t$ are refined at each stage t. Define the sequence of measures $p_i^{t+1}([a_{-i}], s) = \int_{\mathcal{P}_i^t} p_i^t((\cdot)|\mathscr{F}^t)ds$; by the Radon-Nykodim theorem this sequence is well-defined up to a set of measure zero. Taking consecutive expectations with respect to this measure in (1), observe that the set of possible values of $\hat{x}_i^t$ also decreases with time — hence decreasing is the set of games Γ^t that the player considers possible at each t. Since all players are identical, let $\mathscr{F}^t \equiv \cap_{i\in I}\mathscr{F}_i^t$ be the intersection of all players' partitions, and let $\mathcal{P}_i \equiv \bigcap_{t\to\infty} \mathcal{P}_i^t$ be the element of the limit partition that player i considers possible. Taking the intersection of all $\mathscr{F} \equiv \bigcap_{t\to\infty} \mathscr{F}^t$, observe that by (2), the strategies of all players in best reply correspondence must converge to the same limit point in the strategy space. As a sequence of closed set, this limit is a point state s^* which is self-evident ($s^* \subseteq \mathcal{K}_i s^*$) for all players, and coincides with itself. Hence s^* is common knowledge. □

The above theorem establishes the connection between the limiting distribution of the deductions process and common knowledge of the outcome: the former specifies the actual state "from above", the latter — "from below", as the finest common coarsening, according to the classical definition (Aumann, 1976). While useful in principle, it might be of limited value in practical applications for two reasons. First, the connection is explicitly based on the limit of an infinite sequence, hence the stated equivalence is difficult to realize in practice. Second, and more substantially, the present derivation was based on a specific learning mechanism. Third, common knowledge as such is rather restrictive — in particular, most results obtained from it, including the famous "agreement to disagree" and "no speculation", explicitly assume that players know each other's information partitions (Aumann and Heifetz, 2002). Finally, even public announcements can be rather "highly certain" to be commonly known than strictly commonly known — e.g., because of absent-mindedness of some players (Monderer and Samet, 1989).

A remarkable feature of the following theorem is that it establishes very moderate conditions under which one could obtain essentially the same result using the concept of *almost common knowledge* (Monderer and Samet, 1989). To proceed, we need a few more definitions.

Definition 2. *A profile $\sigma = (\sigma_1, \dots, \sigma_n)$ is called an interim ε-equilibrium if for all i and all $s \in S$ the strategy $\sigma_i(s)$ is within ε of the maximizer of i's expected payoffs in all $s' \in P_i(s)$.*

Formally, this condition is

$$\mathsf{E}\left[u_i(\sigma_i(s'), \sigma_{-i}(s')|P_i(s)\right] \geq \mathsf{E}\left[u_i(a_i, \sigma_{-i}(s')|P_i(s)\right] - \varepsilon \tag{3}$$

where E denotes the expectation operator with respect to measure p. This concept, while being weaker than the Nash equilibrium, allows one to reach an "agreement" under broader set of beliefs and hence also in shorter time (although this last claim can be made only in an informal sense). This notion was used by (Monderer and Samet, 1989) to prove the following theorem (see also (Fudenberg and Tirole, 1991)):

Theorem 2. *Let $r \in (0.5, 1]$ and let*

$$q = p\left([\exists x \in X : u_i(a,x) \textit{are common } r\textit{-beliefs at } s], s_i\right)$$

for all players. Then for any collection σ^* of equilibrium profiles in the games which are commonly known to be played, there exists a profile σ for the games from the set X which is an interim ε-equilibrium such that $p\left([\sigma(s) = \sigma^*(s)\right) > 1 - \varepsilon$.

Proof. See (Fudenberg and Tirole, 1991, p.566). □

Using this theorem, we now state the following result.

Theorem 3. ***Any*** *interim ε-equilibrium of the game of incomplete information with uncertainty space X can be obtained as the limit point of* ***some*** *sequence of games from X.*

Proof. Let σ_i^* be the equilibrium strategy of any player i in *any given* game from X, and let the beliefs of that player be given by a distribution clustered around the state s in which she wants to play this equilibrium. Consider a neighborhood $\varepsilon > 0$ of that s. The smaller is that neighborhood, the larger is the probability of the states which warrant playing the strategy $\sigma_i = \sigma_i^*$ even given that the actual state remains *uncertain.* Since X is complete and compact, there will exist a closed set $k \equiv [\sigma_i(s) = \sigma_i^*(s)] \subset S_i$ such that $p[k] < 1 - \varepsilon$. By symmetry this is true for all i, and hence also for the product of these profiles:

$$p[(\sigma_1(s) = \sigma_1^*(s), \dots \sigma_n(s) = \sigma_n^*(s))] \Leftrightarrow p[K] < 1 - \varepsilon.$$

Hence we have obtained the set of states which depends on ε, is compact and has which are strictly larger for any measure than $1 - \varepsilon$. Probability measures which satisfy these requirements are called *tight.* Tightness of the measures is the conclusion of Theorem 2; in turn, by the Prohorov theorem (Billingsley, 1968), if the sequence of probability measures is tight, the space of probability distributions will be relatively weakly compact (or sequentially compact), so that from any sequence of measures one could extract converging subsequence. Taking this subsequence, we conclude that it will have an ε-interim equilibrium as its limit. □

This theorem generalizes the intuition developed in the previous section) that *any monotone* sequence of games $\{\Gamma_t\}$ (not necessarily based on the specific mechanism of deductions) has as its limit an equilibrium in some game of the family X whose payoffs are almost common knowledge. In other words, there are two *identical* ways to characterize power: 1) as the limit of a "muddling" bargaining process with gradual tâtonnement of the specific equilibrium, or 2) as almost common knowledge of some state $s \in S$. What this theorem does *not* tell us, however, is which of the equilibria in the family X is going to be attained, and under what conditions.

Also, one technical issue should be mentioned. Throughout the paper we have been working with an infinite space S, in line with the traditional construction of the universal types space. This approach is somewhat problematic if one wants a complete characteristic of all equilibria — in particular, the set of equilibria need not be continuous in the topology of weak convergence (Fudenberg and Tirole, 1991, p.570). For our present purposes, this fact is not crucial, as our primary goal was to establish the equivalence between "common knowledge" and "deductions" characteristics of power. Exact specification of the set of equilibria might however be desirable in applications, where it is likely to depend on the specificity of particular cases.

5. Qualifications and extensions

The above theorems establish the main results of this paper — namely, the duality and equivalence of the two interpretations of *power*. The interest in this result is not purely theoretical, but rather bears on several substantive issues, and suggests several applications.

Why such 'benign' practices as honesty or mutual trust are quite widespread in some economies, while being rather uncommon in others? For instance, why people in some countries, like Finland, virtually do not cheat with their taxes, while people in most transitional and even many mature market economies (such as the US[5]) would gladly evade despite substantial and credible threat of prosecution? It appears that *social habit* is the key answer; yet how does this habit arise? Our approach suggests there are two ways which led to essentially the same result. One is *prosecution of bad practice*: to prevent tax evasion, the government should intervene with inspections and send offenders to jail. An *equivalent* method is to make sure everyone believes that paying taxes is an appropriate line of behaviour in the sense that everyone is almost sure to pay them. This belief plays exactly the same role as the entire army of tax inspectors backed by courts, fines and prisons.

This fundamental duality illustrates another important point: almost common belief can be *the only* mean to impose some norms whenever direct force cannot be used. Some social norms, such as trust or honesty, which are known to be favourable for economic prosperity, are typically missing in many transforming societies, and in far too many countries there seem to be little progress with their gradual emergence. Why instead of the *good* institutions the society can end up with *bad* ones, which are inferior from social viewpoint[6], even if their inferiority is commonly known to the individual agents? Why society fails to elaborate these efficient norms, and why the authorities, acting on behalf

[5]There was a famous experiment in the state of Minnesota, where tax authorities instead of inspections used moral suasion: they bombarded corporate taxpayers by the brochures, newsletters etc. explaining the social duty of taxpayers and importance of the programs they fund. The result was complete absence of the effect of such messages on tax morale (Blumenthal e.a., 2001).

[6]In today's Russia, for instance, it became commonplace to contrast social interactions conducted in the legal way - by *law* ("po zakonu") and by *arrangement* (custom, rule — "po ponyatiyam").

of it, do not succeed in approval and legalization of those norms which support socially desirable activities and oppress the undesirable ones?

Our approach suggests two lines of explanation. First, even though some, if not most people would benefit from switching from bad to good equilibrium, they are not certain enough this move is going to be followed by the majority of their partners in the social game. the outcome is of the "stag-hare" type, namely failure to coordinate on the good outcome. Persistence of bad practices is indeed backed by the social consensus that these are too widespread to be neglected. To appeal again to a example of transitional economy, consider a popular joke about traffic inspectors, who are among the most renowned corruptioneers in Russia and other FSU countries. A traffic inspector stops a car drivers for whatever small offence, and salutes him "Sergeant Ivanov. Three kids", alluding thereby not only on the need to pay a bribe for permission to go, but also on the size of the requested bribe. By saying so, the inspector in fact appeals to a set of practices common and shared by the driver. Further examples include corporate greenmail or hostile takeovers backed by the corrupted courts or (to depart with transition examples) the well-known American example of don Corleone from *Il Padrino* who used to make offers that could not be refused.

A second, maybe more important lesson is that more powerful explanation is that it may be against the interests of some of the players to reach a particular outcome, even though it may be beneficial for the others. In that sense, common knowledge becomes particular method to impose one's power: it only suffices to persuade everybody that a given course of event is what everyone does. This method is actually quite common to many politicians who make use of public suasion through media control in order to gain public's belief that only some particular courses of events are the *right* ones. This warrants the question of potential abuse of power and "hidden" imposition of one's will on the opponents; yet it also raises a parallel question: is it possible that several players coordinate on an equilibrium which becomes their common knowledge despite there is some other player who do not want this outcome to happen? A detailed exploration of this and the related issues warrants closer look at the dynamics of multi-players games with an explicit consideration of the dynamics of individual beliefs.

References

Aumann R.J. (1976) Agreeing to disagree. *Annals of statistics*, v.4, p.1236-1239.

Aumann R.J., with Heifetz A. (2002) Incomplete information. In: R.J.Aumann and S.Hart, Eds. *Handbook of game theory with economic applications*, v.3, ch.43.

Aumann R.J. and Brandenburger A. (1995) Epistemic conditions for Nash equiliblrium. *Econometrica*, v.63, no.5, p.1161-1180.

Billingsley P. (1968) Convergence of probability measures. N.Y.: Wiley.

Blumenthal M., Christian C., Slemrod J. (2001) Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota, *National Tax Journal*, v.54, p.125138.

Brandenburger A. and Dekel E. (1987) Common knowlege with probability 1. *Journal of Mathematical Economics*, v.16, p.237-245.

Brandenburger A. and Dekel E. (1993) Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, v.59, p.189-198.

Dahl R. (1969) The concept of power. In: R.Bell, D.V.Edwards, R.H.Wagner, Eds. *Political power: a reader in theory and research.* N.Y.: Free press.

Dekel E. and Gul F. (1997) Rationality and knowledge in game theory. In: D.Kreps and K.Wallis, Eds. *Advances in economics and econometrics: the VIIth World Congress.* Vol.1. Cambridge: CUP.

Dellacherie C. and Meyer P.-A. (1978) *Probability and potential.* Paris: Hermann.

Engelking R. (1986) General topology. M.: Mir (Russian).

Fagin R., Halpern J.Y., Moses M., Vardi M.Y. (1995) *Reasoning about knowledge.* Cambridge: MIT Press.

Foucault M. (1986) Disciplinary power and subjction. In: S.Lukes (ed.). *Power.* Oxford: Blackwell.

Fudenberg D, and Tirole J. (1991) *Game theory.* MIT Press.

Gallie W.B. (1955) Essentially contestable concepts. *Procdedings of the Aristotelean Society*, v.56, p.67-198.

Geanakoplos J., Pearce D., Stacchetti E. (1989) Psychological games and sequential rationality *Games and Economic Behavior*, v.1, p.60-79.

Harsanyi J. (1967-68). Games with incomplete information played by Bayesian players. *Management science*, v.14, Pts.I-III, pp.159-182, 320-334, 486-502.

Harsanyi J. and Selten R. (1988) *General theory of equilibrium selection in games.* MIT Press.

Lasswell H.D. and Kaplan A.K. (1950) *Power and society.* New Haven: Yale UP.

Ledyaev V.G. (1998) *Power: a conceptual analysis.* N.Y.: Nova Science.

Lukes S. (1974) *Power: a radical view.* L.: Macmillan.

Mailath G., Morris S., Postlewaite A. (2001) Laws and authority. Unpublished mimeo, University of Pennsylvania.

Maskin E. and Moore J. (1999) Implementation and renegotiation. *Review of Economic Studies*, v.61, p.39-56

Mertens J.-F. and Zamir S. (1985) Foundations of Bayesian Analysis for Games with Incomplete Information. *International Journal of Game Theory*, v.14, p.1-29.

Monderer D. and Samet D. (1989) Approximating common knowledge with common beliefs. *Games and Economic Behavior*, v.1, p.170-190.

Parsons T. (1986) Power and the Social system. In: S.Lukes (ed.). *Power.* Oxford: Blackwell.

Rubinstein A. (1989) The electronic mail game: strategic behaviour under almost common knowledge. *American Economic Review*, v.79, p.385-391.

Tan T.C.-C. and Werlang S.R. da Costa (1988) The Bayesian foundations of solution concepts of games. *Journal of Economic Theory*, v.45, p.370-391.