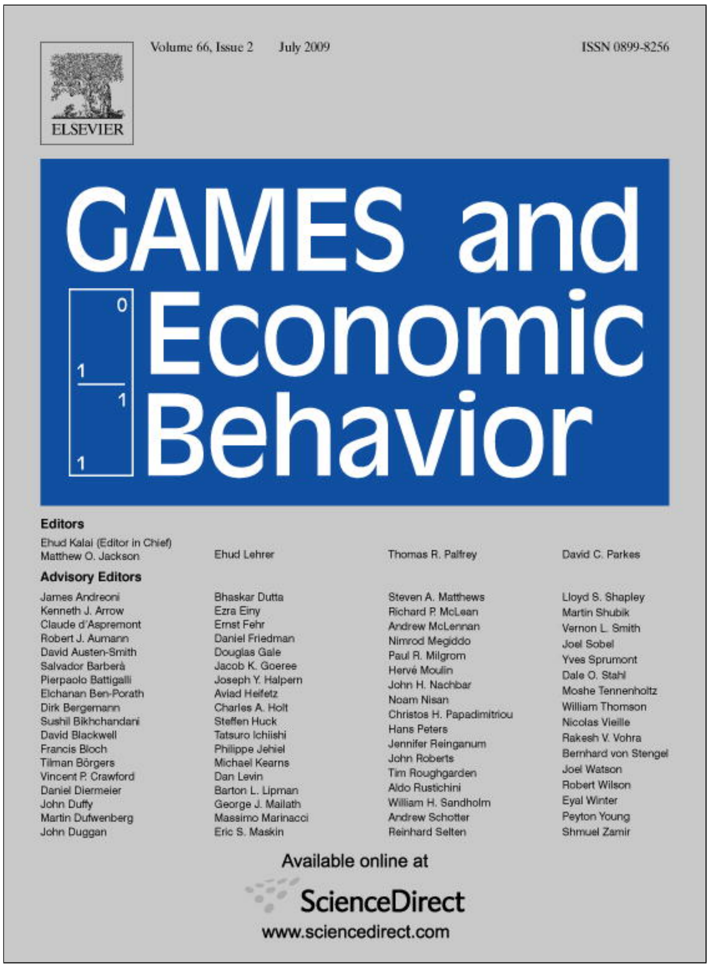
Volume 66, Issue 2 July 2009
ISSN 0899-8256
ELSEVIER
GAMES and Economic Behavior
Editors
Ehud Kalai (Editor in Chief)
Matthew O. Jackson
Ehud Lehrer
Thomas R. Palfrey
David C. Parkes
Advisory Editors
James Andreoni
Kenneth J. Arrow
Claude d'Aspremont
Robert J. Aumann
David Austen-Smith
Salvador Barberà
Pierpaolo Battigalli
Elchanan Ben-Porath
Dirk Bergemann
Sushil Bikhchandani
David Blackwell
Francis Bloch
Tilman Börgers
Vincent P. Crawford
Daniel Diermeier
John Duffy
Martin Dufwenberg
John Duggan
Bhaskar Dutta
Ezra Einy
Ernst Fehr
Daniel Friedman
Douglas Gale
Jacob K. Goeree
Joseph Y. Halpern
Aviad Heifetz
Charles A. Holt
Steffen Huck
Tatsuro Ichiishi
Philippe Jehiel
Michael Kearns
Dan Levin
Barton L. Lipman
George J. Mailath
Massimo Marinacci
Eric S. Maskin
Steven A. Matthews
Richard P. McLean
Andrew McLennan
Nimrod Megiddo
Paul R. Milgrom
Hervé Moulin
John H. Nachbar
Noam Nisan
Christos H. Papadimitriou
Hans Peters
Jennifer Reinganum
John Roberts
Tim Roughgarden
Aldo Rustichini
William H. Sandholm
Andrew Schotter
Reinhard Selten
Lloyd S. Shapley
Martin Shubik
Vernon L. Smith
Joel Sobel
Yves Sprumont
Dale O. Stahl
Moshe Tennenholtz
William Thomson
Nicolas Vieille
Rakesh V. Vohra
Bernhard von Stengel
Joel Watson
Robert Wilson
Eyal Winter
Peyton Young
Shmuel Zamir
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
Games and Economic Behavior 66 (2009) 979–994
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Informational externalities and emergence of consensus ☆

Dinah Rosenberg [a,b], Eilon Solan [c,*], Nicolas Vieille [d]

[a] *Laboratoire d'Analyse Géométrie et Applications, Institut Galilée, Université Paris Nord, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France*
[b] *Laboratoire d'Econométrie de l'Ecole Polytechnique, 1, rue Descartes 75005 Paris, France*
[c] *The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel*
[d] *Département Finance et Economie, HEC, 1, rue de la Libération, 78 351 Jouy-en-Josas, France*

Received 28 November 2007
Available online 2 October 2008

Abstract

We study a general model of dynamic games with purely informational externalities. We prove that eventually all motives for experimentation disappear, and provide the exact rate at which experimentation decays. We also provide tight conditions under which players eventually reach a consensus. These results imply extensions of many known results in the literature of social learning and getting to agreement.

JEL classification: C72; C73; D83; C11

1. Introduction

The dissemination of private information, or knowledge, in a population has attracted much interest, first among sociologists and geographers (see references in Chamley, 2004), and more recently among economists and computer scientists. A question that has attracted a lot of attention is whether as time passes, information spreads through the entire population, and as beliefs become more precise, whether consensus of some sort eventually arises.

Within economics, this work has developed independently in different directions, and several strands of literature can be recast under that heading. In the literature on *getting to agreement*, agents are endowed with private information over the underlying parameter, and exchange information according to some communication protocol. The main purpose of this literature, starting with Geanakoplos and Polemarchakis (1982), is to provide some dynamic foundation for agreement theorems (Aumann, 1976, see also Nielsen et al., 1990 and the references therein). Information is usually publicly broadcasted, with the notable exception of Parikh and Krasucki (1990). Consensus is here *agreement* say, e.g., on the posterior belief assigned to some event. Players are non-strategic, in that they make no attempt at manipulating the protocol to gain more information.

☆ This research was supported by a grant from the Ministry of Science and Technology, Israel, and the Ministry of Research, France. We thank Ehud Lehrer for the comments he provided.
* Corresponding author.
E-mail addresses: dinah@zeus.math.univ-paris13.fr (D. Rosenberg), eilons@post.tau.ac.il (E. Solan), vieille@hec.fr (N. Vieille).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2008.09.027

In the literature on *learning in social networks* (see Goyal, 2005 and the references therein), identical players are identified with the vertices of a directed graph, with the interpretation that each player observes her neighbors, and only them. Players are endowed with private information, and adapt their behavior through time, according to the observed behavior of their neighbors. Here, consensus means *conformism* through the network. Players are myopic, in that they play in every stage an action that maximizes their current expected payoff, in the light of the information received so far.

In *strategic experimentation* models, by contrast, players are non-myopic. Each player faces a statistical dynamic decision problem, such as a multi-arm bandit, and benefits from the "experimentations" performed by her fellow players. Stylized two-arm bandit problems have been considered, see Bolton and Harris (1999), Keller et al. (2005), Rosenberg et al. (2007a) and Murto and Välimäki (2006).

All the models above share the feature that there are no payoff externalities among players. We study a general model of information dissemination, that includes the above models as special cases, and study the limit behavior of the players. Our main findings are the following. (1) Consensus needs not arise. Even if we face a connected network, so that each pair of players are connected with a directed path, and each player on the path observes the actions of the player next to him, asymptotically one player may play actions which are perceived sub-optimal by another player. (2) Nevertheless, if the network is connected then each player believes that her neighbors play asymptotically optimal.

We will now describe our model and then present our results in more details. The game involves finitely many players. Before play starts, a parameter is drawn from some general measurable space of parameters, endowed with a common prior. At each stage $n \in \mathbf{N}$, each player first receives a private signal, then chooses an action, and finally receives a payoff. We assume that the payoff of a player depends only on the parameter and the player's own action. Thus, the interaction among the players is purely informational. We assume that the players are symmetric, in that they share the same action set and the same payoff function. By contrast, we make no *a priori* assumption on the degree of informativeness of signals. In particular, we will allow for cases where (i) payoffs may or may not be observed, publicly or privately, (ii) information is broadcasted, as in Geanakoplos and Polemarchakis (1982), (iii) neighbors' actions are observed, as in Bala and Goyal (1998) and Gale and Kariv (2003), (iv) players observe a random sample from the set of players, as in Ellison and Fudenberg (1995), and Banerjee and Fudenberg (2004), or (v) any combination of these, and beyond. From a game-theoretic viewpoint, we thus allow for general information and monitoring structures, at the cost of restrictive assumptions on the payoff structure. Player i *observes* player j if at every stage the signal that player i receives reveals the action that player j chose at the previous stage (a significantly weaker definition is given below, see Definition 2.2). In this case we say that player j is a *neighbor* of player i. We say that the population is *connected* if for every players i and j, player j is a neighbor of a neighbor of ... of a neighbor of player j.

In games with informational externalities, as soon as discount factors are positive, the agents face a trade-off between optimizing and experimenting – sacrificing current payoffs for the sake of future informational benefits. We show that if the population is connected, then in equilibrium the actions of one's neighbors are eventually optimal according to one's own information. In other words, in the light of the information available to a player A, the actions that her neighbor B plays infinitely often are myopically optimal.

Somewhat surprisingly, as we show by means of two examples, the above result does *not* extend to neighbors of neighbors. That is, in equilibrium, player A may think that her neighbor B is using myopically optimal actions, know that B thinks that her neighbor C is using myopically optimal actions, and know that C thinks that A is using myopically optimal actions, yet, if A was asked whether C's actions are myopically optimal in her (player A) own eyes, the answer may have been negative. This remains true *even* when players observe their own payoffs and their neighbors' payoffs. Our examples are non-generic, and it is possible that in generic games such phenomena do not arise.

Plainly, a player always has the option of mimicking the behavior of her neighbor (with a one-period delay) – the so-called *imitation principle*. This principle is usually interpreted as implying that player A's limit payoff is always at least as high as her neighbor B's limit payoff and hence, since all players are connected, that the limit payoffs of all players do coincide. Such a statement is however ambiguous. We prove that if the network is connected, the *expected* limit payoffs of all players coincide. That is, when computed at the beginning of the game the average limit payoff that players expect to receive, all expectations coincide. On the other hand, the *actual* limit payoffs need not be equal, as we show with an example. We also identify a sufficient condition under which the actual limit payoffs are equal.

Casual intuition suggests that our limit results hold as soon as each player observes her neighbors in infinitely many periods. This is not so, even in two-player games, as we show with an example. Instead, all of our results still hold if the network is connected under the following definition of observability: player i *observes* player j if there is a subset $B^{i,j}$ of the actions that player j plays infinitely often that satisfies the following two conditions: (i) at the end of the game player 1 knows that all the actions in $B^{i,j}$ were played infinitely often, and (ii) at the end of the game player 2 knows that player 1 knows that all the actions in $B^{i,j}$ were played infinitely often. Knowledge of higher level is not needed. This condition is satisfied, e.g., if at every stage each player randomly chooses a neighbor, and observes the action of that neighbor (as in Ellison and Fudenberg, 1995, and Banerjee and Fudenberg, 2004), and in social networks where players occasionally visit their neighbors according to a pre-selected mechanism, as long as each player visits her neighbors infinitely often.

The model and the basic results are presented in Section 2. In Section 3 we discuss the implication of our results to the three strands of literature mentioned above. Examples appear in Section 4, and proofs appear in Section 5.

2. Model and main results

2.1. Setup

We consider games with incomplete information, in which identical players repeatedly choose an action, and receive a payoff that depends on their own action, and on the underlying parameter.

The set of players is a finite set I. The set of parameters is a measurable space $(\Omega, \mathcal{A})$, endowed with a common prior $\mathbf{P}$. Time is discrete, and the set of stages is the set $\mathbf{N}$ of positive integers. At each stage n, each player i first receives a private signal s_n^i from some signal set S^i, then chooses an action a_n^i from her action set A^i, and obtains a utility $u^i(\omega, a_n^i)$. Players discount future payoffs at the common rate $\delta \in [0, 1)$. Players are identical, only in that they share the same action set $A := A^i$, the same signal set $S := S^i$, and the same utility function[1] $u : \Omega \times A \to \mathbf{R}$. However, different players may receive different signals.

We impose the following technical assumptions:

- The common action set A is a compact metric space, endowed with the Borel σ-field.
- The common utility function $u : \Omega \times A \to \mathbf{R}$ is (jointly) measurable, and continuous over A for every fixed $\omega \in \Omega$. In addition, it satisfies the following boundedness condition: the highest payoff $\bar{u} : \omega \mapsto \max_{a \in A} u(\omega, a)$ and the lowest payoff $\underline{u} : \omega \mapsto \min_{a \in A} u(\omega, a)$ are L_2-integrable.
- The signal set S is a measurable set. The signalling function maps past histories into probability distributions over S^I, the space of signal profiles. The past history at stage n is the complete list of the parameter, and of the actions and signals of all players in all previous stages, hence it lies in $H_n := \Omega \times (S^I \times A^I)^{n-1}$. Technically, the signalling function at stage n is *any* transition probability[2] from H_n to S^I.

A few remarks are in order. First, we emphasize that each player's utility function only depends on the underlying parameter, and on her own action, but does not depend on other players' actions. In that sense, the strategic interaction between players is purely informational: actions of player i may provide some information on player i's signals, and hence on the parameter. Therefore, actions of player i are relevant to player j.

We now discuss an important issue of interpretation. We assume in this paper that the payoff to a player is a *deterministic* function $u(\omega, a)$ of the parameter ω and of one own's action a, and that a player may only receive a noisy signal about this payoff. In some applications, such as strategic experimentation models, the payoff to a player is random, with an expectation that depends on ω and a. In such applications, it is typically assumed that the payoff is observed. Our model accommodates such situations by setting $u(\omega, a)$ to be the expectation of the payoff, and setting the signal of the player to include her actual (random) payoff. We discuss this issue in Section 3.2 for multi-arm bandit games.

[1] Here and in the sequel, a product of measure spaces is endowed with the product topology.

[2] A transition probability from X to Y is a function f that assigns for every $x \in X$ a probability distribution $f(x)$ over Y, such that for every measurable subset B of Y, the probability $f(x)[B]$ assigned to B is measurable in x.

The previous point illustrates why it is critical here to assume that payoffs may not be observed. Indeed, observing the payoff in, say, multi-arm bandit games, amounts to observing the *expected* payoff associated with the arm, an assumption which is overly restrictive.

Finally, the assumption that the set of possible signals is independent of the stage, and is the same for all players, is without loss of generality. Indeed, one may otherwise define S as the union of all the signal sets of all players in all stages.

2.2. Information and strategies

The space of plays is $H_\infty := \Omega \times (S^I \times A^I)^{\mathbf{N}}$. A private history of player i at stage n is an element of $H_n^i := (S \times A)^{n-1} \times S$, and $\mathcal{H}_n^i$ is the corresponding σ-algebra over H_∞.

A (behavior) strategy of player i is a sequence (σ_n^i), where σ_n^i assigns to every private history in H_n^i a probability distribution[3] over A. We denote by $\mathcal{H}_\infty^i$ the information of player i at the end of the game. It is the σ-algebra spanned by $(\mathcal{H}_n^i)_{n\in\mathbf{N}}$.

Any strategy profile σ, together with the common prior $\mathbf{P}$ on Ω, induces a probability distribution, $\mathbf{P}_\sigma$, over the set of plays H_∞. Expectation w.r.t. $\mathbf{P}_\sigma$ is denoted by $\mathbf{E}_\sigma$.

Given a strategy profile σ, and a stage $n \in \mathbf{N}$, we denote by q_n^i the conditional distribution over Ω given player i's information $\mathcal{H}_n^i$ at stage n. For a fixed (measurable) subset $F \subseteq \Omega$, the sequence $q_n^i(F)$ is a bounded martingale, which converges, by the martingale convergence theorem, to $\mathbf{E}_\sigma[1_F \mid \mathcal{H}_\infty^i]$, $\mathbf{P}_\sigma$-a.s. We set $q_\infty^i(F) := \lim_{n\to\infty} q_n^i(F)$. It is a probability distribution, to be interpreted as the limit belief of player i.

2.3. Main results

We focus on the asymptotic equilibrium behavior. For non-myopic players ($\delta > 0$), we use the Nash equilibrium notion. If $\delta = 0$, the Nash equilibrium criterion puts no restriction on the player's behavior beyond the first stage. In order to get asymptotic results for myopic players, we require that if $\delta = 0$ each player plays at *every* stage a myopically optimal action. In both cases, we will simply speak of equilibria and best-replies.

Given a strategy profile σ, a stage $n \geqslant 1$, and an action $a \in A$, we let $u(q_n^i, a) := \mathbf{E}_\sigma[u(\cdot, a) \mid \mathcal{H}_n^i]$ denote the expected payoff at stage n (conditional on available information),[4] when playing a. We also set $u_*(q_n^i) = \min_{a\in A} \mathbf{E}_\sigma[u(\cdot, a) \mid \mathcal{H}_n^i]$; similarly, $u_*(q_\infty^i) = \max_{a\in A} \mathbf{E}[u(\cdot, a) \mid \mathcal{H}_\infty^i]$. This is the lowest (expected) payoff the agent may obtain at stage n.

Given a belief q, that is, a probability distribution over Ω, the set of *myopically optimal* actions w.r.t. q is[5]:

$$BR(q) := \operatorname{argmax}_{a\in A} \int u(\omega, a) q(\mathrm{d}\omega) = \operatorname{argmax}_{a\in A} \mathbf{E}_q\big[u(\cdot, a)\big].$$

When playing optimally, a player faces a trade-off between *optimizing* – playing an action which is myopically optimal in the light of the information accumulated so far, and *experimenting*, with the purpose of obtaining further information on ω. This information refines the player's current information, and may be used in subsequent stages to increase the player's myopic payoff, or she can transfer this information to other players, who will afterwards reveal information that she needs.

An action $a \in A$ is a *limit action* if it is a limit point of the sequence $(a_n^i)_{n\in\mathbf{N}}$ of the actions played by the agent along the game.[6] We denote A_*^i the set of limit actions. Since A is compact metric, A_*^i is compact and non-empty. Since the actions of the agent depend on her information, A_*^i is a random variable,[7] measurable w.r.t. the information of the agent at infinity, $\mathcal{H}_\infty^i$.

We first note that in equilibrium, the players eventually stop experimenting.

[3] Formally, it is a transition probability from $(H_\infty, \mathcal{H}_n^i)$ to A.

[4] This notation is motivated by the observation that, $\mathbf{E}_\sigma[u(\cdot, a) \mid \mathcal{H}_n^i]$ is the expectation of $u(\cdot, a)$ under the belief q_n^i of player i.

[5] By dominated convergence, the map $a \mapsto \mathbf{E}_q[u(\cdot, a)]$ is continuous. Hence, $BR(q)$ is non-empty for each q.

[6] That is, there exists an increasing sequence $(n_k)_{k\in\mathbf{N}}$ of stages such that $a = \lim_{k\to\infty} a_{n_k}^i$.

[7] The set of compact subsets of A is endowed with the usual, Hausdorff, distance.

Proposition 2.1. *Let σ be an equilibrium. Then* $\mathbf{P}_\sigma(A^i_* \subseteq BR(q^i_\infty)) = 1$*, for each player* i.

In the traditional literature of social networks, player i observes player j if the signal that player i receives at every stage n reveals the action that player j chose at the previous stage $n-1$. We will use a weaker definition of observability: player i observes player j if she can identify a subset of the limit actions of player j, and player j knows which of her limit actions were identified.

Definition 2.2. Let σ be a strategy profile. Player i *observes* player j w.r.t. σ if for $\mathbf{P}_\sigma$-almost-every ω there is a non-empty and compact set $B^{ij}_*(\omega) \subseteq A^j_*(\omega)$ such that B^{ij}_* is both $\mathcal{H}^i_\infty$-measurable and $\mathcal{H}^j_\infty$-measurable.

Few examples of observability are:

(1) Players are vertices of a connected directed graph. At every stage each player observes the action of her neighbors. In this case $B^{ij}_* = A^j_*$.
(2) Consider the previous example, but at every stage each player observes the action of a random sample of her neighbors that is drawn independently of previous choices (see Ellison and Fudenberg, 1995; Banerjee and Fudenberg, 2004). In this case, by the independence assumption, $B^{ij}_* = A^j_*$.
(3) Consider still the setup in (1), and adopt the monitoring setup of Cripps et al. (2007): at each stage a player does not observe her neighbor's action, but rather a noisy signal of the neighbor's action, that depends on her action and on her neighbor's action. The signals are assumed to be sufficiently revealing, in the sense that with sufficiently many observations, the player can correctly identify, from the frequencies of the signals, any fixed stage-game action of her neighbor (assumption 2 in Cripps et al., 2007). The identifiability assumption implies that the empirical frequencies of signals reveal the limit actions of a player, and therefore in this case i observes j with $B^{ij}_* = A^j_*$.
(4) There are finitely many locations, and in every stage each player randomly chooses a location for that stage. The player determines which action to choose after observing who are the players in her location, and she observes the actions of everyone in her location. In this case B^{ij}_* is the set of limit actions of player j in all stages in which she shared the same location as player i. Note that B^{ij}_* may be different than B^{kj}_* for $i \neq k$.

The compactness requirement in Definition 2.2 is w.l.o.g. It is crucial that player j knows the set of her limit actions that are observed by player i; without this requirement our results do not hold, as the example in Section 4.4 shows.

The definition of observability does not apply to the primitives of the game. That is, whether player i observes player j depends on the strategy profile σ, as well as on the signalling function. Indeed, if player j uses, e.g., a constant strategy, she is observed by any other player i. One can strengthen the definition of observation by requiring that the condition in Definition 2.2 holds for every strategy profile σ. The resulting definition would be intrinsic.[8]

We let G denote the directed graph with vertex set I, that contains an edge $i \to j$ if and only if i observes j. The graph G is *connected* if for every two players i and j there is a directed path from i to j.[9]

Theorem 2.3. *Let σ be an equilibrium, and assume that G is connected. Then*

P1. $\mathbf{E}_\sigma[u_*(q^i_\infty)] = \mathbf{E}_\sigma[u_*(q^j_\infty)]$ *for every two players i and j.*
P2. $\mathbf{P}_\sigma(A^j_* \subseteq BR(q^i_\infty)) = 1$*, provided player i observes player j.*

If G is not connected, **P1** and **P2** still hold provided i and j belong to the same connected component of G.

Since the graph is connected, according to **P1** all players eventually perform equally well in *expected* terms. Even if information is not divided equally, the relevant information spreads along the graph and guarantees asymptotic equality. The intuition behind **P1** relies on the so-called *imitation principle*: there is a strategy for player i that mimics

8 It would still be un-satisfactory, to the extent that checking this definition would involve considering all strategy profiles. Variants can be devised, that are better in this respect, at the cost of some notational complexity.
9 By exchanging the roles of the two players, there is also a directed path from j to i.

in the long run the behavior of her neighbor j: the play of player i converges to some limit action of player j. Since the players play an equilibrium, according to player i's information, the expected payoff of each limit action she uses is at least as much as the expected payoff of each action that player j uses. This observation, together with the connectedness assumption, still does not prove **P1**, since it is not clear that the expected payoff of the action a according to i's information is the same as its expected payoff according to j's information. The definition of observability guarantees that in expected terms **P1** would hold. As we show by means of an example in Section 4.1, it may happen that $u_*(q^i_\infty) \neq u_*(q^j_\infty)$ with positive probability (and even with probability 1), even in generic networks.

According to **P2**, player i eventually thinks that player j is playing in an optimal way: every limit action of player j is optimal according to player i's information. The intuition is that by the imitation principle, every limit action of j yields, according to player i's information, at most as much as player i's own limit action yields. If with positive probability the limit action of the neighbor is strictly sub-optimal according to player i's information, the expectation $\mathbf{E}[u_*(q^j_\infty)]$ would be strictly below $\mathbf{E}[u_*(q^i_\infty)]$, which would violate **P1**.

P1 holds for every pair of players, but **P2** holds only for neighbors. Indeed, it may well be that a limit action of a player who is not player i's neighbor is not optimal according to player i's information. We stress that this negative result is not an artefact of strategic behavior. Indeed, we provide counter examples that hold for every discount factor (see Sections 4.2 and 4.3). Our counter examples are non-generic, and it is possible that in generic games **P2** holds also for non-neighbors.

It is natural to wonder when the equality $u_*(q^i_\infty) = u_*(q^j_\infty)$ would hold path-wise (with probability 1). As Theorem 2.4 below states, it is sufficient to assume that each player (eventually) observes her *own* payoffs.

Theorem 2.4. *Let σ be an equilibrium, and assume that G is connected and that the realized payoff $u(\omega, a^i_n)$ is $\mathcal{H}^i_\infty$-measurable for every player i and every stage n. Then*

P3. $u_*(q^i_\infty) = u_*(q^j_\infty)$ *for every two players i and j.*

Whether or not the measurability condition in Theorem 2.4 is satisfied may depend upon the strategy profile σ. Plainly, it is satisfied as soon as the payoff $u(\omega, a^i_n)$ is part of the signal s^i_{n+1}. However, as we discussed earlier, in some models this is an extremely restrictive assumption. Even under this additional restrictive assumption, **P2** does not hold for neighbors of neighbors, as we show in Section 4.3.

3. Applications and related literature

3.1. Social networks

Many models of learning in social networks have been proposed, see e.g. Bala and Goyal (1998), Gale and Kariv (2003), DeMarzo et al. (2003), Ellison and Fudenberg (1995), Banerjee and Fudenberg (2004), Goyal (2005) and the references therein. We here provide a brief discussion of how our results relate to this literature. We fix the probability space $(\Omega, \mathbf{P})$, the set of players I, the common action set A, and the payoff function u. Let G be a directed graph over the set of players I. It is assumed that each player i observes (at least) the actions of her neighbors.

By Theorem 2.3 if player j is a neighbor of player i, then every action that player j plays infinitely often is optimal according to player i's information, and the asymptotic expected stage payoff is the same for all players. This result was proved by Bala and Goyal (1998) under the additional assumptions that (i) A is finite, (ii) players observe the signals received by their neighbors, and (iii) the players disregard the information revealed by their neighbors' actions. Since the updating of player's beliefs is not Bayesian, Bala and Goyal's (1998) result does not follow from our results. This conclusion was also stated by Gale and Kariv (2003) under the additional assumptions that (i) A is finite and Ω compact, (ii) player i's information only consists of some initial signal and of her and her neighbors' previous actions, and (iii) players are myopic and Bayesian.

Our results imply that these conclusions hold in more complicated social networks, e.g., when

- players are strategic rather than myopic;
- players receive a signal at every stage; the signals of the players may be correlated, and may depend on past actions and signals;

- players only observe a random sample of their neighbors' choices, such as in Ellison and Fudenberg (1995) and Banerjee and Fudenberg (2004);
- the network changes along the play as a function of past play (as long as the underlying graph, that is defined by observation, is connected);
- A is compact metric and Ω is general.

3.2. Strategic experimentation

Our results also apply to pure strategic experimentation models. We here introduce a stylized multi-player bandit problem, which is already more general than the games considered in Bolton and Harris (1999) and in Keller et al. (2005).[10]

Consider a finite set I of players, each of whom operates a K-arm bandit machine. Each of the K arms is of one of several types, which is determined once and for all at the beginning of the game, and the random type θ_k of arm k is common to all $|I|$ machines. Conditional on its type, the kth arm yields a sequence of payoffs, which are identically distributed and independent (across time, players, and arms). At every stage $n \in \mathbf{N}$, each player chooses which arm to operate, and receives the realized payoff. Players are here strategic, and discount future payoffs at the rate δ. For concreteness, we assume that no two types of two arms yield the same expected payoff, hence we may identify a type with the corresponding expected payoff. Finally we assume that there are given subsets of players $(Q_n^i, R_n^i)_{i\in I}^{n\in\mathbf{N}}$ (some of these sets may be empty), and at stage n player i observes the payoffs of all players in Q_n^i, and the actions of all players in R_n^i.

We first discuss how to embed this model in the general model of Section 2. Denote by $X_k(i,n)$ the random payoff generated by arm k if it is operated by player i at stage n. Our basic assumption is thus that the r.v.s $(X_k(i,n))_{k,i,n}$ are conditionally independent given θ, with $\mathbf{E}[X_k(i,n) \mid \theta] = \theta_k$.

We define an auxiliary game as follows. The parameter space $\Omega \subseteq \mathbf{R}^K$ contains all possible vector types, and the action set $A = \{1, 2, \ldots, K\}$ coincides with the set of arms. We denote by $\theta = (\theta_1, \ldots, \theta_K)$ a generic element of Ω. Define $u(\theta, k) = \theta_k$; the payoff upon selecting action k is the expected payoff of the kth arm.[11] The signal to player i at stage $n+1$ contains the actions chosen at stage n by all the players in R_n^i, and in addition it contains $(X_k(j,n))_{j\in Q_n^i}$, the realized payoffs of all the players in Q_n^i.

Observe that the strategy set of each player in the model of strategic experimentation coincides with her strategy set in the auxiliary game we just defined. Since the expectation of the discounted sum is the discounted sum of expectations, one can verify that the expected payoff of every strategy profile in the two models coincide. In particular, the set of equilibria in the model of strategic experimentation coincides with that in the auxiliary game.

Assume, as in Bolton and Harris (1999), that arm choices and payoffs are publicly observed: $Q_n^i = R_n^i = I$ for every i and n. By Theorem 2.4 all players have asymptotically the same payoff in the auxiliary game. Since the payoff in the auxiliary game is the expected payoff of the chosen arm, and since no two arms yield the same expected payoff, it follows that all players end up using the same arm.

Assume now that players are organized along a directed, connected graph. Each player privately observes her payoffs, and she also observes the actions of her neighbors. Using Theorem 2.3 again, all players end up using the same arm.

3.3. Getting to agreement

Finally we relate our results to interactive epistemology. We limit ourselves to showing how a number of existing results can be deduced from our results. We specialize our model as follows. Let $(\Omega, \mathbf{P})$ be the set of parameters, A the action set, and $u : \Omega \times A \to \mathbf{R}$ the payoff function. We will assume that players are myopic, and are endowed with private information over the parameter. In addition, we will assume that along the play, the signals only provide information about the moves chosen earlier by the player's neighbors. Formally, for every player i and every stage n

[10] Except that both these models are continuous-time games.

[11] Since the payoff in the auxiliary game should be a deterministic function of the parameter and the action, we define it as the expected payoff of the chosen arm.

there is a (deterministic) set $R_i^n \subseteq I \setminus \{i\}$; this is the set of neighbors of i at stage n. The signal player i receives at stage n coincides with the list of actions chosen by the players in R_i^n at stage $n-1$.

Player i observes player j if $j \in R_i^n$ for infinitely many n's. Assume that the underlying graph G is connected. Such communication protocols are called *fair* in Parikh and Krasucki (1990).

Let σ be an equilibrium. Since player i is myopic, she plays at each stage n an action which maximizes $\mathbf{E}_\sigma[u(\cdot,a) \mid \mathcal{H}_n^i]$. Here are a few examples:

- Let $A = [0,1]$, and $u(\omega,a) = -(1_E(\omega) - a)^2$, for some fixed event $E \subset \Omega$. Then $\mathbf{E}_\sigma[u(\cdot,a) \mid \mathcal{H}_n^i]$ is uniquely maximized at $a = p_n^i = \mathbf{P}_\sigma(E \mid \mathcal{H}_n^i)$. Thus, in equilibrium, every player "plays" her current posterior belief over E. Therefore $A_*^i = \{q_\infty^i\}$ for every player i, so that by Theorem 2.3, whenever i observes j, $\{q_\infty^j\} = A_*^j \subseteq BR(q_\infty^i) = \{q_\infty^i\}$. Since the population is connected, all posterior beliefs eventually coincide. This result was first proven by Geanakoplos and Polemarchakis (1982) under the assumption that Ω is finite and $R_i^n = N$ for all n and i. It was extended by Nielsen (1984) to general Ω, still assuming $R_i^n = N$. Theorem 2.3 yields a simple generalization to the case of a fair protocol. Setting $\delta > 0$, it yields a strategic version of that result, in which players mis-represent their beliefs, to prompt other players to reveal more information.
- Let $A = \{0,1\}$, $I = \{i,j\}$, and $u(\omega,a) = a(1_E(\omega) - \pi)$, where $\pi \in (0,1)$ is given. The optimal action at stage n is 1 or 0 depending on whether $p_n^i \geqslant \pi$. By Theorem 2.3, both players eventually agree whether the probability of E is higher than π or not. This is the result in Sebenius and Geanakoplos (1983).
- We here let both Ω and A be finite sets. Each player is endowed with private information, described by a partition $\mathcal{P}_i$ of Ω. We moreover assume that $a \mapsto \mathbf{E}_\sigma[u(\cdot,a) \mid B]$ has a unique maximum, for any event B in the join[12] of the partitions $\mathcal{P}_i$, $i \in I$. Player i first considers any parameter in the atom of $\mathcal{P}_i$ that contains ω, $P_i(\omega)$, to be possible, and plays the action that maximizes $\mathbf{E}_\sigma[u(\cdot,a) \mid P_i(\omega)]$. With time, she may observe actions of her neighbors that rule out some parameters in $P_i(\omega)$, and consequently she updates the set of parameters that she views as possible. By the assumption that $a \mapsto \mathbf{E}_\sigma[u(\cdot,a) \mid B]$ has a unique maximum it follows that the set A_*^i of limit actions is a singleton. By Theorem 2.3, $A_*^i = A_*^j$ for any two players $i, j \in I$. This is the result in Ménager (2006a).
- In Parikh and Krasucki (1990), the message sent by player i to her neighbors at stage k is $f_k^i = f(\Omega_k^i)$ where $\Omega_k^i \subseteq \Omega$ is the set of parameters that player i considers possible at stage k, and $f : 2^\Omega \to \mathbf{R}$ is given. It is shown that under a so-called convexity assumption on f the sequence of messages is eventually constant; the map f is convex if, for every two disjoint subsets S, T of Ω, $f(S \cup T)$ is a proper convex combination of $f(S)$ and $f(T)$. Denoting by A the range of f, Parikh and Krasucki's convergence result follows from Ménager (2006a), hence from Theorem 2.3, if there is a function $u : \Omega \times A \to \mathbf{R}$ such that

  $$\sum_{\omega \in S} u(\omega,a) > \sum_{\omega \in S} u(\omega,b), \quad \forall a \in A, b \in A \setminus \{a\},\ \forall S \text{ s.t. } f(S) = a. \tag{1}$$

  There are non-convex functions for which a function u that satisfies (1) exists. For example, when $|\Omega| = 2$ the function f that is defined by $f(\{1\}) = f(\{1,2\}) = 1$, $f(\{2\}) = 0$ is not convex, but there is a function u that satisfies (1) for this f. Conversely, one can show that when the range of a convex function f contains at most five values, or if $f(S)$ only depends on the number of elements in S, there is a function u that satisfies (1). Ménager (2006b) shows that there are convex functions f for which no such function u exists.

4. Examples

We analyze four examples to illustrate the tightness of our results. The example in Section 4.1 shows that **P1** does not hold path-wise. The example in Section 4.2 shows that **P2** does not extend to neighbors of neighbors: a limit action of a player who is not one's neighbor may be sub-optimal according to one's information. The example in Section 4.3 shows that even if each player observes her own payoff as well as her neighbors' payoffs, and all players receive the same limit payoff, **P2** does not extend to neighbors of neighbors.[13] The example in Section 4.4 shows that neither

[12] The *join* of the partitions $\mathcal{P}_1, \ldots, \mathcal{P}_k$ is the coarsest partition that refines $\mathcal{P}_1, \ldots, \mathcal{P}_k$.

[13] The examples in Sections 4.2 and 4.3 challenge the assertion after Theorem 2 in Gale and Kariv (2003), according to which all players are using the same limit actions.

P1 nor **P2** need hold when a player only observes her neighbors infinitely often. That is, it is important that B_*^{ij} be $\mathcal{H}_\infty^j$-measurable (in addition to being $\mathcal{H}_\infty^i$-measurable).

Whereas example in Section 4.1 is generic, the other three examples are not, and their conclusion hinges on the fact that the players are indifferent between certain actions. For myopic players, in generic games with countably many actions there are countably many possible posteriors that can arise along the play, and for each such posterior there is a unique optimal action. Therefore all the players converge to the same limit action, and **P2** holds also for players who do not observe each other. For non-myopic players it is plausible that in generic games uniformity of behavior arises as well, though we do not have a proof for such a result.

4.1. *No convergence of payoffs*

Our first example is a two player example who have a single action a. There are two parameters ω_1 and ω_2, player 1 knows the chosen parameter while player 2 receives no information. The payoff function is

$$u(\omega_1, a) = 1, \qquad u(\omega_2, a) = 2.$$

Since there is only one action, in the unique strategy profile (which is the unique equilibrium) both players play a in all stages. Unless $\mathbf{P}(\omega_1) \in \{0, 1\}$ the expected payoff of the two players, given their information, differs.

4.2. *Neighbors of neighbors*

Our second example is a three-player example. There are two equally likely parameters, ω_1 and ω_2. At stage 1, both players 2 and 3 receive an informative signal in $\{s_1, s_2\}$. The signal to player 2 reveals the parameter with probability $2/3$: $\mathbf{P}(s_k \mid \omega_k) = 2/3$, for $k = 1, 2$, so that $\mathbf{P}(\omega_k \mid s_k) = 2/3$ as well. The signal to player 3 reveals the parameter with probability $5/6$. No further information about ω is provided.

There are three actions, $A = \{a, b, c\}$. Denoting p the belief assigned to ω_1, the utility function u is such that action a is myopically optimal for $p \in [2/7, 5/7]$, action b is myopically optimal for $p \in [0, 2/7]$, and action c is myopically optimal for $p \in [5/7, 1]$. An example for such a payoff function is (see Fig. 1):

$$u(\omega_1, a) = -2/7, \qquad u(\omega_1, b) = 2/7, \qquad u(\omega_1, c) = -1,$$
$$u(\omega_2, a) = 5/7, \qquad u(\omega_2, b) = -5/7, \qquad u(\omega_2, c) = 1.$$

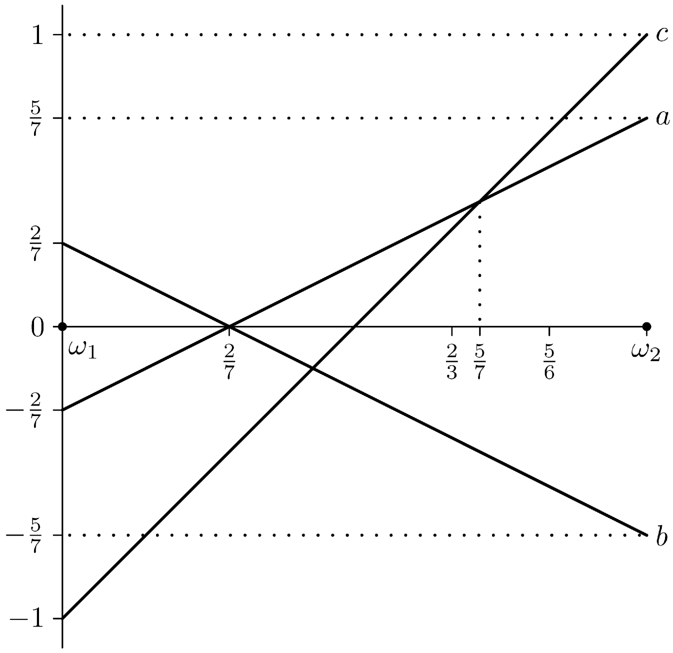

Fig. 1. The utility function.

signal	stage 1			stage 2			stage 3			stage 4		
	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
s_1s_1	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	$\frac{1}{2}$	$\frac{10}{11}$	$\frac{5}{6}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{5}{6}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{10}{11}$
	a	a	c	a	c	c	c	c	c	c	c	c
s_1s_2	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{7}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{7}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{7}$	$\frac{2}{7}$
	a	a	b	a	a	b	a	a	b	a	a	b
s_2s_1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{5}{6}$	$\frac{1}{2}$	$\frac{5}{7}$	$\frac{5}{6}$	$\frac{1}{2}$	$\frac{5}{7}$	$\frac{5}{6}$	$\frac{1}{2}$	$\frac{5}{7}$	$\frac{5}{7}$
	a	a	c	a	a	c	a	a	c	a	a	c
s_2s_2	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{11}$	$\frac{1}{6}$	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{1}{6}$	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{1}{11}$
	a	a	b	a	b	b	b	b	b	b	b	b

Fig. 2. The strategies of the players, and the evolution of the beliefs.

At each stage $n > 1$, each player i observes only the action of player $i + 1$ (modulo 3) in the previous stage. We assume that players are myopic, and describe below one equilibrium profile (see Fig. 2).[14]

Stage 1. Player 1's prior belief assigns probability $1/2$ to ω_1, hence she plays a. Player 2's posterior probability is either $1/3$ or $2/3$, depending on whether her signal is s_1 or s_2, hence she plays a. Player 3's posterior belief is either $1/6$ or $5/6$, hence she plays either b or c, depending on whether her signal is s_1 or s_2.

Stage 2. Players 1 and 3 hold the same belief as at stage 1, and therefore repeat their action. Player 2 infers player 3's signal from her action at stage 1, and she revises her belief accordingly.

If both signals are equal to s_1 (resp. s_2), player 2's posterior belief becomes

$$\frac{2}{3}\frac{5}{6}\Big/\left(\frac{2}{3}\frac{5}{6}+\frac{1}{3}\frac{1}{6}\right)=\frac{10}{11}>\frac{5}{7}$$

(resp. equal to $1/11 < 2/7$). Hence player 2 switches to c (resp. to b).

If the signals of players 2 and 3 mismatch, player 2's new posterior belief is $5/7$ if she received s_1, and $2/7$ if she received s_2. In the former case, she is indifferent between a and c, whereas in the latter she is indifferent between a and b. In our equilibrium she plays a.

Stage 3. Players 2 and 3 hold the same belief as at stage 2. If the signals of players 2 and 3 match, the action of player 2 at stage 2 reveals the common signal, and player 1 revises her belief accordingly. If the two signals mismatch, the belief of player 1 remains $1/2$.

Stage 4. Now only the beliefs of player 3 may change, but actions remain as at stage 3. After stage 3 beliefs and actions do not change.

In Fig. 2, the signals received by players 2 and 3 appear in the left-most column. Subsequent columns describe the belief of each player at each stage and the players' actions.

Observe that the limit action of player 3 is either b or c. If the signals of players 2 and 3 differ, the limit belief of player 1 is $1/2$. Hence, player 3's limit action is *not* optimal in the eyes of player 1. Moreover, in this case player 1's limit conditional payoff, $u_*(q^1_\infty)$, is $3/14$, while the limit conditional payoff of players 2 and 3 is $3/7$ (if the signals are s_1s_2) or 0 (if the signals are s_2s_1). Thus, $u_*(q^1_\infty) \neq u_*(q^2_\infty)$ with probability 1: the players do *not* agree about their limit payoff.

This phenomenon is due to the fact that player 2 may be indifferent between two actions. For games in which no player may ever be indifferent, Ménager (2006a) has shown that all players eventually play the same action.

[14] It can be checked that this profile is actually an equilibrium for every discount factor.

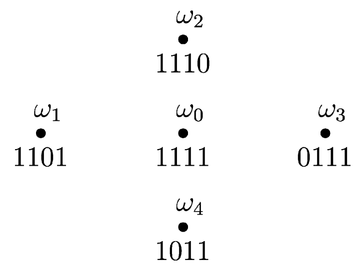

Fig. 3. The parameters and the payoff functions.

This example is non-generic. Nevertheless, if we slightly perturb the prior distribution as well as the probabilities by which the signals are chosen, one can find a utility function for which the same equilibrium behavior will hold.

4.3. Observed payoffs

Here we assume that each player observes her own payoffs. There are four players $N = \{1, 2, 3, 4\}$, four actions $A = \{a, b, c, d\}$, and five parameters $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3, \omega_4\}$. Each player i observes her own payoffs, as well as the actions of player $i + 1$ (modulo 4).[15] The discount factor is arbitrary. In Fig. 3 we graphically describe the five parameters. Below each parameter appear four numbers – the payoffs of the four actions at that parameter (from left to right). Thus, for example,

$$u(\omega_1, a) = 1, \qquad u(\omega_1, b) = 1, \qquad u(\omega_1, c) = 0, \qquad u(\omega_1, d) = 1.$$

Fig. 4 describes the information of the four players, as well as a stationary strategy for each player. The information of the players is described by a partition of the parameter space; each player has three information sets. The action each player plays is written below the parameter.

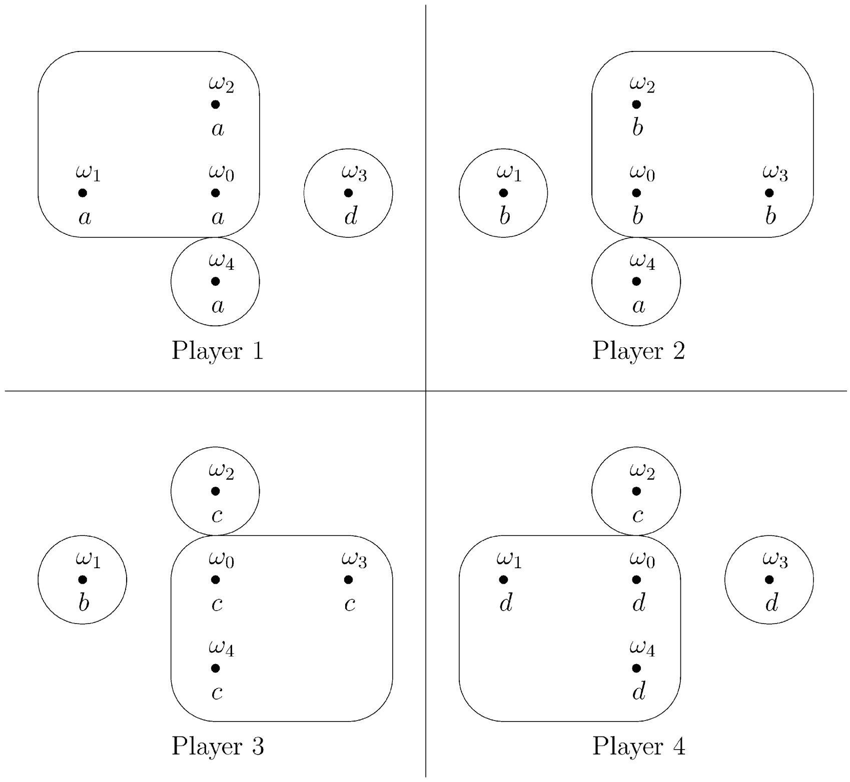

Fig. 4. The partitions of the players, and their strategies.

[15] Our argument will be valid if player i observes the payoff of player $i + 1$ (modulo 4) as well.

Since the payoffs of a player are measurable w.r.t. her information, and since the strategy of each player is measurable w.r.t. the information of the player who observes her, the players do not learn anything along the game.

Under the strategy described in Fig. 4, the payoff of all players is 1 regardless of the parameter. Since 1 is the maximal possible payoff, these strategies form an equilibrium. Nevertheless, in all parameters there is at least one player whose limit action is sub-optimal in the eyes of some other player. For example, in ω_0 player 3's limit actions is sub-optimal in the eyes of player 1 (and vice versa), and in ω_1 players 3 and 4's limit action is sub-optimal in the eyes of player 1.

4.4. *Unknown observed stages*

In this example, there are two players, two actions $A = \{T, B\}$, and four equally likely parameters $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$.

The payoff function is given by:

$$u(\omega_1, T) = u(\omega_3, T) = 1, \qquad u(\omega_1, B) = u(\omega_3, B) = 0,$$

$$u(\omega_2, T) = u(\omega_4, T) = 0, \qquad u(\omega_2, B) = u(\omega_4, B) = 1.$$

Thus, if the parameter is ω_1 or ω_3 one would like to play T, while if the parameter is ω_2 or ω_4 one would like to play B.

At stage 1, the two players receive some information about ω. This information is described by the two partitions $\mathcal{F}_1 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3, \omega_4\}\}$ and $\mathcal{F}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$. Thus, player 1 knows when ω_1 is drawn, knows also when ω_2 is drawn, but cannot distinguish ω_3 from ω_4. No further information about ω is given.

At each stage $n > 1$, player 1 is informed of the action played by player 2 at the previous stage. By contrast, player 2 observes player 1 either in odd stages (if the parameter is ω_1 or ω_2) or in even stages (if the parameter is ω_3 or ω_4).

The tables in Fig. 5 describe *one* strategy profile. The left-hand side (resp. right-hand side) table contains the sequence of moves of player 1 (resp. player 2) in every possible parameter. The size of the letters T and B represents whether they are observed by the other player: actions that appear in large (resp. small) letters are observed (resp. not observed) by the other player.

According to this profile, no player ever refines her initial information. This is obvious for player 1 since player 2 always plays B. On the other hand, if the parameter is in $\{\omega_1, \omega_2\}$ player 2 observes player 1 in odd stages, and in those stages player 1 plays T. If the parameter is in $\{\omega_3, \omega_4\}$ player 2 observes player 1 in even stages, and in those stages player 1 plays B. Thus, player 2 does not gain any information along the play either.

The actions played by the players are myopically optimal, hence this profile is an equilibrium when players are myopic.[16]

We observe that B is the only limit action of player 2, but in ω_1 player 1 believes that B is sub-optimal.

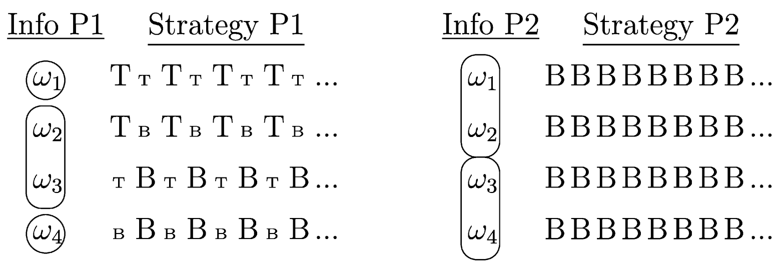

Fig. 5. The partitions, signals, and strategies of the players.

16 As in the previous examples, it turns out that it is an equilibrium for every discount factor.

5. Proofs

5.1. Proof of Proposition 2.1

We start with two technical lemmas.

Lemma 5.1. *The sequence $(u_*(q_n^i))$ is a submartingale. It converges to $u_*(q_\infty^i)$, $\mathbf{P}_\sigma$-a.s.*

Proof. Let a stage $n \in \mathbf{N}$ be given. For each action $a \in A$, one has

$$u\big(q_n^i, a\big) = \mathbf{E}_\sigma\big[u(\cdot, a) \mid \mathcal{H}_n^i\big] = \mathbf{E}_\sigma\big[\mathbf{E}_\sigma\big[u(\cdot, a) \mid \mathcal{H}_{n+1}^i\big] \mid \mathcal{H}_n^i\big] \leqslant \mathbf{E}_\sigma\big[u_*\big(q_{n+1}^i\big) \mid \mathcal{H}_n^i\big].$$

Taking the supremum over a countable dense subset of A, one obtains $u_*(q_n^i) \leqslant \mathbf{E}_\sigma[u_*(q_{n+1}^i) \mid \mathcal{H}_n^i]$, hence $(u_*(q_n^i))$ is a submartingale.

By the integrability condition on u, $(u_*(q_n^i))$ is uniformly integrable, hence it converges, both $\mathbf{P}_\sigma$-a.s. and in L^1.

It remains to prove that the limit is $u_*(q_\infty^i)$. Let $\varepsilon > 0$ be given. For given $\omega \in \Omega$, and by the compactness of A, the map $a \mapsto u(\omega, a)$ is uniformly continuous. By dominated convergence, there exists $\eta > 0$ such that

$$\mathbf{E}_\sigma\Big[\max_{a,b \in A, d(a,b) \leqslant \eta} \big|u(\omega, a) - u(\omega, b)\big|\Big] < \varepsilon. \tag{2}$$

Since A is compact, there is a finite subset $A_f \subseteq A$, such that every action $a \in A$ lies at distance at most η from some action in A_f.

For each $a \in A_f$, the martingale $(u(q_n^i, a))$ converges to $u(q_\infty^i, a)$ ($\mathbf{P}_\sigma$-a.s. and in L^1). Hence, and since A_f is finite, there exists N, such that

$$\mathbf{E}_\sigma\Big[\max_{a \in A_f} \big|u\big(q_n^i, a\big) - u\big(q_\infty^i, a\big)\big|\Big] < \varepsilon, \quad \text{for all } n \geqslant N. \tag{3}$$

Observe next that

$$\mathbf{E}_\sigma\big[\big|u_*\big(q_n^i\big) - u_*\big(q_\infty^i\big)\big|\big] \leqslant \mathbf{E}_\sigma\Big[\max_{a \in A} \big|u\big(q_n^i, a\big) - u\big(q_\infty^i, a\big)\big|\Big]. \tag{4}$$

On the other hand, for a given n, one has by the definition of $u(q_n^i, a)$, by the law of iterated expectations, and by (2):

$$\begin{aligned}
&\mathbf{E}_\sigma\Big[\max_{a,b \in A, d(a,b) \leqslant \eta} \big|u\big(q_n^i, a\big) - u\big(q_n^i, b\big)\big|\Big] \\
&\quad = \mathbf{E}_\sigma\Big[\max_{a,b \in A, d(a,b) \leqslant \eta} \big|\mathbf{E}_\sigma\big[u(\cdot, a) - u(\cdot, b) \mid \mathcal{H}_n^i\big]\big|\Big] \\
&\quad \leqslant \mathbf{E}_\sigma\Big[\mathbf{E}_\sigma\Big[\max_{a,b \in A, d(a,b) \leqslant \eta} \big|u(\cdot, a) - u(\cdot, b)\big| \mid \mathcal{H}_n^i\Big]\Big] \\
&\quad = \mathbf{E}_\sigma\Big[\max_{a,b \in A, d(a,b) \leqslant \eta} \big|u(\omega, a) - u(\omega, b)\big|\Big] < \varepsilon.
\end{aligned} \tag{5}$$

A similar inequality holds with $n = +\infty$.

By (3) and by definition of A_f, it follows that

$$\mathbf{E}_\sigma\Big[\max_{a \in A} \big|u\big(q_n^i, a\big) - u\big(q_\infty^i, a\big)\big|\Big] < 3\varepsilon, \quad \text{for every } n \geqslant N. \tag{6}$$

It follows from (4) and (3) that $(u_*(q_n^i))$ converges to $u_*(q_\infty^i)$.

Let a_*^i be a measurable selection[17] of A_*^i. For every n let $m(n) \leqslant n$ satisfy

$$d\big(a_{m(n)}^i, a_*^i\big) = \min_{m \leqslant n} d\big(a_m^i, a_*^i\big).$$

[17] A *selection* f of a set valued function $F : X \to Y$ is a function $f : X \to Y$ such that $f(x) \in F(x)$ for every $x \in X$.

$a^i_{m(n)}$ is the action that is closest to a^i_* among the actions chosen by player i up to stage n. Since a^i_* is a limit action of player i it follows that $\lim_{n\to\infty} a^i_{m(n)} = a^i_*$. By the triangle inequality,

$$\begin{aligned}
&\mathbf{E}_\sigma\big[\big|u\big(q^i_{m(n)}, a^i_{m(n)}\big) - u\big(q^i_\infty, a^i_*\big)\big|\big] \\
&\quad \leqslant \mathbf{E}_\sigma\big[\big|u\big(q^i_{m(n)}, a^i_{m(n)}\big) - u\big(q^i_{m(n)}, a^i_*\big)\big| + \big|u\big(q^i_{m(n)}, a^i_*\big) - u\big(q^i_\infty, a^i_*\big)\big|\big] \\
&\quad = \mathbf{E}_\sigma\big[\big|u\big(q^i_{m(n)}, a^i_{m(n)}\big) - u\big(q^i_{m(n)}, a^i_*\big)\big|\big] + \mathbf{E}_\sigma\big[\big|u\big(q^i_{m(n)}, a^i_*\big) - u\big(q^i_\infty, a^i_*\big)\big|\big].
\end{aligned}$$

By (5) the first term goes to 0, and by (6) the second term goes to 0. It follows that

$$\liminf_{n\to\infty} u\big(q^i_n, a^i_n\big) = u\big(q^i_\infty, a^i_*\big), \quad \mathbf{P}_\sigma\text{-a.s.} \qquad \square \tag{7}$$

Lemma 5.2. *The sequence* $\max_{a\in A} |u(q^i_n, a) - u(q^i_\infty, a)|$ *converges to zero,* $\mathbf{P}_\sigma$*-a.s.*

Proof. Set $X_n := \max_{a\in A} |u(q^i_n, a) - u(q^i_\infty, a)|$. As mentioned at the end of the previous proof, the sequence (X_n) converges to zero in the L^1 sense. Hence, it has a subsequence, which we denote $(X_{\phi(n)})$, that converges $\mathbf{P}_\sigma$-a.s. to zero.

Let $\varepsilon > 0$ be given. Given N, let F stand for the event $\{\sup_{n\geqslant N} X_{\phi(n)} \geqslant \varepsilon\}$. Provided N is large enough, one has $\mathbf{P}_\sigma(F) \geqslant 1 - \varepsilon$.

Let $n \geqslant \phi(N)$ be an arbitrary stage, and choose $m \geqslant N$ with $\phi(m) \leqslant n$. For each action $a \in A$, one has

$$\begin{aligned}
&\big|u\big(q^i_n, a\big) - u\big(q^i_\infty, a\big)\big| \\
&\quad = \big|\mathbf{E}_\sigma\big[u\big(q^i_\infty, a\big) \mid \mathcal{H}^i_n\big] - u\big(q^i_\infty, a\big)\big| \\
&\quad \leqslant \big|\mathbf{E}_\sigma\big[u\big(q^i_{\phi(m)}, a\big)1_F \mid \mathcal{H}^i_n\big] - u\big(q^i_{\phi(m)}, a\big)1_F\big| + 2\mathbf{E}\big[1_{\bar F}|\bar u| \mid \mathcal{H}^i_n\big] + \varepsilon\big(\mathbf{P}_\sigma\big(F \mid \mathcal{H}^i_n\big) + 1_F\big) \\
&\quad = \big|u\big(q^i_{\phi(m)}, a\big)\mathbf{P}_\sigma\big(F \mid \mathcal{H}^i_n\big) - u\big(q^i_{\phi(m)}, a\big)1_F\big| + 2\mathbf{E}\big[1_{\bar F}|\bar u| \mid \mathcal{H}^i_n\big] + \varepsilon\big(\mathbf{P}_\sigma\big(F \mid \mathcal{H}^i_n\big) + 1_F\big) \\
&\quad \leqslant \mathbf{E}_\sigma\big[|\bar u| \mid \mathcal{H}^i_n\big]\big(\mathbf{P}_\sigma\big(F \mid \mathcal{H}^i_n\big) - 1_F\big) + 2\mathbf{E}\big[1_{\bar F}|\bar u| \mid \mathcal{H}^i_n\big] + \varepsilon\big(\mathbf{P}_\sigma\big(F \mid \mathcal{H}^i_n\big) + 1_F\big)
\end{aligned}$$

where $\bar F$ stands for the complement of F. The first inequality holds by definition of F, and the following equality holds since $n \geqslant \phi(m)$.

Since the final right-hand side is independent of $a \in A$, it follows that

$$X_n \leqslant \mathbf{E}_\sigma\big[|\bar u| \mid \mathcal{H}^i_n\big]\big|\mathbf{P}_\sigma\big(F \mid \mathcal{H}^i_n\big) - 1_F\big| + 2\mathbf{E}\big[1_{\bar F}|\bar u| \mid \mathcal{H}^i_n\big] + \varepsilon\big(\mathbf{P}_\sigma\big(F \mid \mathcal{H}^i_n\big) + 1_F\big), \tag{8}$$

for every $n \geqslant \phi(N)$. On the event F, the right-hand side of (8) converges $\mathbf{P}_\sigma$-a.s. to 2ε, hence $\limsup X_n \leqslant 2\varepsilon$ in that case. It follows that $\mathbf{P}_\sigma(\limsup X_n \leqslant 2\varepsilon) \geqslant 1 - \varepsilon$. Since ε is arbitrary, this implies that X_n converges to 0, $\mathbf{P}_\sigma$-a.s., as desired. $\square$

Proof of Proposition 2.1. Let (ε_k) be a positive sequence that converges to 0, and let an equilibrium σ be given. Given σ^{-i}, player i faces a sequential decision problem. An implication of Theorem 1 in Rosenberg et al. (2007b) is that $u_*(q^i_n) - u(a^i_n, q^i_n)$ converges to zero, $\mathbf{P}_\sigma$-a.s.

In particular, for every $\varepsilon > 0$ there exists a (random) time $N(\varepsilon)$ such that

$$u\big(q^i_n, a^i_n\big) \geqslant u_*\big(q^i_n\big) - \varepsilon, \quad \forall n \geqslant N(\varepsilon). \tag{9}$$

Let a^i_* be a measurable selection of A^i_*. Then by (7), (9) and since $(u_*(q^i_n))$ converges to $u_*(q^i_\infty)$, $\mathbf{P}_\sigma$-a.s. one has:

$$u\big(q^i_\infty, a^i_\infty\big) = \liminf_{n\to\infty} u\big(q^i_n, a^i_n\big) \geqslant \lim_{n\to\infty} u_*\big(q^i_n\big) - \varepsilon = u_*\big(q^i_\infty\big) - \varepsilon.$$

Since this holds for every $\varepsilon > 0$, the result follows. $\square$

5.2. *Proof of Theorem 2.3*

For simplicity, we write $i \to j$ whenever player i observes player j.

We start with a simple observation. Given any σ-field $\mathcal{F}$ over H_∞, the map $(\omega, a) \mapsto \mathbf{E}_\sigma[u(\cdot, a) \mid \mathcal{F}](\omega)$ is $\mathcal{F}$-measurable. Hence, for any $\mathcal{F}$-measurable function $\omega \mapsto a(\omega)$, the composition $\omega \mapsto u(q_{\mathcal{F}}(\omega), a(\omega)) := \mathbf{E}_\sigma[u(\cdot, a(\omega)) \mid \mathcal{F}](\omega)$ is also $\mathcal{F}$-measurable.

The proof of the following lemma is standard, hence omitted.

Lemma 5.3. *Assume that a is $\mathcal{F}$-measurable. Then*

$$\mathbf{E}_\sigma\big[u\big(q_{\mathcal{F}}(\omega), a(\omega)\big)\big] = \mathbf{E}_\sigma\big[u\big(\omega, a(\omega)\big)\big].$$

We now prove Theorem 2.3. Let two players $i, j \in I$ be given, with $i \to j$. By assumption, the set-valued function $\omega \mapsto B_*^{ij}$ is both $\mathcal{H}_\infty^j$- and $\mathcal{H}_\infty^i$-measurable.

Plainly, the map $\omega \mapsto B_*^{ij}(\omega)$ is B_*^{ij}-measurable, with non-empty and compact values. Fix a B_*^{ij}-measurable selection $\omega \mapsto a^j(\omega)$; its existence is guaranteed since B_*^{ij} has compact values, see Kuratowski and Ryll-Nardzewski (1965). In particular, $a^j(\omega)$ is both $\mathcal{H}_\infty^i$-measurable and $\mathcal{H}_\infty^j$-measurable.

From Lemma 5.3 we deduce that

$$\mathbf{E}\big[u\big(q_\infty^j(\omega), a^j(\omega)\big)\big] = \mathbf{E}\big[u\big(\omega, a^j(\omega)\big)\big] = \mathbf{E}\big[u\big(q_\infty^i(\omega), a^j(\omega)\big)\big]. \tag{10}$$

Consider now a path $i_0 \to i_1 \to \cdots \to i_K$ in G, that visits all vertices at least once, and such that $i_K = i_0$. Summing (10) over all pairs $i_k \to i_{k+1}$, we deduce that

$$0 = \sum_{k=0}^{K-1} \mathbf{E}\big[u\big(q_\infty^{i_{k+1}}(\omega), a^{i_{k+1}}(\omega)\big)\big] - \sum_{k=0}^{K-1} \mathbf{E}\big[u\big(q_\infty^{i_k}(\omega), a^{i_{k+1}}(\omega)\big)\big] \tag{11}$$

$$= \sum_{k=0}^{K-1} \mathbf{E}\big[u\big(q_\infty^{i_k}(\omega), a^{i_k}(\omega)\big) - u\big(q_\infty^{i_k}(\omega), a^{i_{k+1}}(\omega)\big)\big]. \tag{12}$$

By Proposition 2.1, each summand in (12) is non-negative, and therefore

$$\mathbf{E}\big[u\big(q_\infty^{i_k}(\omega), a^{i_k}(\omega)\big) - u\big(q_\infty^{i_k}(\omega), a^{i_{k+1}}(\omega)\big)\big] = 0, \quad \forall k. \tag{13}$$

By (10) and (13) we have

$$\mathbf{E}\big[u\big(q_\infty^{i_{k+1}}(\omega), a^{i_{k+1}}(\omega)\big)\big] = \mathbf{E}\big[u\big(q_\infty^{i_k}(\omega), a^{i_{k+1}}(\omega)\big)\big] = \mathbf{E}\big[u\big(q_\infty^{i_k}(\omega), a^{i_k}(\omega)\big)\big].$$

Since this equality holds for every k, and since the path visits all players, **P1** is proven.

By Proposition 2.1 we moreover have that $u(q_\infty^i(\omega), a^i(\omega)) - u(q_\infty^i(\omega), a^{i+1}(\omega))$ is non-negative with probability 1, and therefore with probability 1

$$u\big(q_\infty^{i_k}(\omega), a^{i_k}(\omega)\big) = u\big(q_\infty^{i_k}(\omega), a^{i_{k+1}}(\omega)\big).$$

P2 is proven as well.

5.3. *Proof of Theorem 2.4*

We will use the following observation. Let an arbitrary measure space $(\Omega, \mathcal{A}, \mathbf{P})$ be given, together with a σ-algebra $\mathcal{B} \subseteq \mathcal{A}$, and a random variable $X \in L^2(\mathbf{P})$. Since the conditional expectation operator is a projection operator (in L^2), one has

$$\big\|\mathbf{E}[X \mid \mathcal{B}]\big\|_2 \leqslant \|X\|_2, \tag{14}$$

with equality if and only X is $\mathcal{B}$-measurable.

We proceed with the proof of Theorem 2.4. Let $i, j \in I$ be any two players such that $i \to j$. As in the proof of Theorem 2.3, let $a^j(\omega)$ be a selection of B_*^{ij} which is both $\mathcal{H}_\infty^i$- and $\mathcal{H}_\infty^j$-measurable.

By **P2**, one has $u_*(q^i_\infty) = \mathbf{E}_\sigma[u(\omega, a^j(\omega)) \mid \mathcal{H}^i_\infty]$. Since player j observes her own payoffs, $u(\omega, a^j_n)$ is $\mathcal{H}^j_\infty$-measurable. By Proposition 2.1 this implies that $u(\cdot, a^j(\cdot)) = u_*(q^j_\infty)$, $\mathbf{P}_\sigma$-a.s.

Therefore, by (14),

$$\left\|u_*\left(q^i_\infty\right)\right\|_2 = \left\|\mathbf{E}_\sigma\left[u_*\left(q^j_\infty\right) \mid \mathcal{H}^i_\infty\right]\right\|_2 \leqslant \left\|u_*\left(q^j_\infty\right)\right\|_2. \tag{15}$$

Consider now any path $i_0 \to i_1 \to \cdots \to i_K$ that visits all vertices at least once, and such that $i_K = i_0$. When applying (15) to all edges in this path, one obtains

$$\begin{aligned}\left\|u_*\left(q^{i_0}_\infty\right)\right\|_2 &= \left\|\mathbf{E}_\sigma\left[u_*\left(q^{i_1}_\infty\right) \mid \mathcal{H}^{i_0}_\infty\right]\right\|_2 \leqslant \left\|u_*\left(q^{i_1}_\infty\right)\right\|_2 \\ &= \left\|\mathbf{E}_\sigma\left[u_*\left(q^{i_2}_\infty\right) \mid \mathcal{H}^{i_1}_\infty\right]\right\|_2 \leqslant \cdots \leqslant \left\|u_*\left(q^{i_0}_\infty\right)\right\|_2.\end{aligned}$$

Thus, all inequalities hold with an equality. This implies that $u_*(q^{i_1}_\infty)$ is $\mathcal{H}^{i_0}_\infty$-measurable – player i_0 knows the limit payoffs of player i_1. Since player i_0 observes the actions of player i_1, and her payoffs, and since she plays an optimal strategy, we must have $u_*(q^{i_0}_\infty) \geqslant u_*(q^{i_1}_\infty)$. Since $\|u_*(q^{i_0}_\infty)\|_2 = \|u_*(q^{i_1}_\infty)\|_2$, this implies that $u_*(q^{i_0}_\infty) = u_*(q^{i_1}_\infty)$. Since the path visits all players, the result follows.

References

Aumann, R.J., 1976. Agreeing to disagree. Ann. Statist. 4, 1236–1239.

Bala, V., Goyal, S., 1998. Learning from neighbors. Rev. Econ. Stud. 65, 595–621.

Banerjee, A., Fudenberg, D., 2004. Word-of-mouth learning. Games Econ. Behav. 46, 1–22.

Bolton, P., Harris, C., 1999. Strategic experimentation. Econometrica 67, 349–374.

Chamley, C.P., 2004. Rational Herds. Economic Models of Social Learning. Cambridge University Press, Cambridge.

Cripps, M.W., Mailath, G.J., Samuelson, L., 2007. Disappearing private reputations in long-run relationships. J. Econ. Theory 134, 287–316.

DeMarzo, P.M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and uni-dimensional opinions. Quart. J. Econ. 118, 909–968.

Ellison, G., Fudenberg, D., 1995. Word-of-mouth communication and social learning. Quart. J. Econ. 110 (1), 93–125.

Gale, D., Kariv, S., 2003. Bayesian learning in social networks. Games Econ. Behav. 45, 329–346.

Geanakoplos, J., Polemarchakis, H., 1982. We cannot disagree forever. J. Econ. Theory 28, 192–200.

Goyal, S., 2005. Learning in networks. In: Demange, G., Wooders, M. (Eds.), Group Formation in Economics. Cambridge University Press.

Keller, G., Rady, S., Cripps, M.W., 2005. Strategic experimentation with exponential bandits. Econometrica 73, 39–68.

Kuratowski, K., Ryll-Nardzewski, C., 1965. A general theorem on selectors. Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. 13, 397–403.

Ménager, L., 2006a. Consensus, communication and knowledge: An extension with Bayesian agents. Math. Soc. Sci. 51 (3), 274–279.

Ménager L., 2006b. Communication, Connaissance Commune et Consensus. Ph.D. Thesis. Université Paris 1 Panthéon Sorbonne.

Murto, P., Välimäki, J., 2006. Learning in a model of exit. Preprint.

Nielsen, L.T., 1984. Common knowledge, communication, and convergence of beliefs. Math. Soc. Sci. 8, 1–14.

Nielsen, L.T., Brandenburger, A., Geanakoplos, J., McKelvey, R., Page, T., 1990. Common knowledge of an aggregate of expectations. Econometrica 58, 1235–1239.

Parikh, R., Krasucki, P., 1990. Communication, consensus and knowledge. J. Econ. Theory 52, 178–189.

Rosenberg, D., Solan, E., Vieille, N., 2007a. Social learning in one-arm bandit problems. Econometrica 75, 1591–1611.

Rosenberg, D., Solan, E., Vieille, N., 2007b. On the optimal amount of experimentation in sequential decision problems. Preprint.

Sebenius, J., Geanakoplos, J., 1983. Don't bet on it: A note on contingent agreements with asymmetric information. J. Amer. Statistical Assoc. 78, 224–226.