Network Effects in Information Acquisition*

Tommaso Denti†

March 31, 2017

Abstract

This paper studies endogenous information acquisition in network games. Players, connected via commonly known network, are uncertain about state of fundamentals. Before taking actions, they can acquire costly information to reduce this uncertainty. The basic idea is that network effects in action choice induce externalities in information acquisition: players' information choice depends on neighbors' information choice, which depends on neighbors' neighbors' information choice, and so forth. The analysis shows these externalities can be measured by Bonacich centralities and provide new sources of multiple equilibria. Cost of information is proportional to entropy reduction, as in rational inattention. A representation theorem provides foundation to this functional form in terms of primitive monotonicity properties of cost of information.

1 Introduction

This paper studies games played on networks. Players are connected via a fixed, commonly known network. The network represents the structure of interaction in the game. Players, for instance, may be firms competing a la Cournot; the network corresponds to the pattern of complementarity and substitutability among firms' goods.

In Cournot markets, as in many other network games, state of fundamentals is often uncertain. State affects payoffs but not links among players. While in traditional

*I thank Dilip Abreu, Faruk Gul, Tibor Heumann, Muhamet Yildiz, and, especially, Stephen Morris for very helpful discussions.

†Princeton University, tdenti@princeton.edu

analysis information is exogenous, players typically are not passive to the uncertainty they face. For instance, to reduce demand uncertainty, firms spend considerable resources on market research, surveys, focus groups, ...

In this paper, information is endogenous: players can reduce the uncertainty they face by acquiring costly information. Before taking action, each player observes the realization of a signal, a random variable. Traditional analysis takes a given profile of signals as primitive. In this paper, instead, each player can choose (at a cost) her own signal from some feasible set.

I investigate how the network of relations shapes the endogenous information structure. The basic idea is that network effects in action choice induce externalities in information acquisition. For instance, Cournot competition forces firms to take into account both uncertain market fundamentals and the quantities chosen by firms' direct competitors, their "neighbors" in the network. But neighbors' quantities are uncertain as well, since they are chosen on the basis of neighbors's information. The incentive to reduce this uncertainty make firms' information choice depend on neighbors' information choice, which in turn depends on neighbors' neighbors' information choice, and so forth.

My analysis shows that network effects in information acquisition can be measured by the centralities of Bonacich (1987). When strong, these externalities provide a new source of multiple equilibria for network games. They generate convexities in value of information and coordination problems in information acquisition. This happens regardless of strategic motives for actions, that is, even if actions are strategic substitutes. Uniqueness is restored under stronger contraction assumptions that the ones needed with exogenous information.

The main analytical tool I use is entropy (Shannon 1948): each player pays a cost to acquire information that is proportional to the reduction in her uncertainty as measured by the entropy of her beliefs. This was introduced by Sims (2003) for his theory of rational inattention, and it has become probably the most prominent specification for cost of information in economics. A basic challenge for this literature is to go beyond the functional form of entropy and identify predictions that depend only on primitive assumptions on cost of information.

The methodological innovation of this paper is a representation theorem for entropy. The theorem provides a foundation for the functional form of entropy in terms of primitive monotonicity properties of cost of information. With entropy, cost of information is monotone in the sufficiency ordering of Blackwell (1951): the more information players acquire, the higher the cost they pay. Blackwell monotonicity is a natural property for cost of information. The representation theorem states that, for equilibrium analysis, assuming a mild strengthening of Blackwell monotonicity is equivalent to assuming entropy.

To model strategic interaction on networks, I consider games with linear best responses and Gaussian uncertainty. Players want their action to match an unknown target, a linear combination of actions of others and state. The marginal effects of actions of others on targets depend on the relation between players in the network. Feasible signals and state are normally distributed.

Linear best responses are standard model for network games. Most of the literature has focused on the case of no uncertainty, complete information.[1] Central results pertain how network effects relate to centrality measures (Ballester et al. 2006) and may lead to multiple equilibria (Bramoulle et al. 2014). This paper extends these conclusions to endogenous information acquisition.

Beyond network economics, linear best-response games have been used to study asymmetric information in a variety of settings, e.g., team theory (Radner 1962), oligopoly (Vives 1984), Keynesian beauty contests (Morris and Shin 2002), quadratic economies (Angeletos and Pavan 2007). In most of these applications, players are symmetric and identical. This can be interpreted as if a trivial network is in place, where everyone is connected to everyone else.

Symmetric identical players have also been considered by a recent literature on information acquisition in linear best-response games, e.g., Hellwig and Veldkamp (2009), Myatt and Wallace (2012), and Colombo et al. (2014). These works also discuss the issue of multiplicity, mostly in the case of complementarity in actions. This paper points out that equilibrium determinacy in information acquisition is somehow orthogonal to strategic motives for actions. This is done by considering not only richer network structures, but also richer technologies for information acquisition.

To endogenize information, I use the model of fully flexible or "unrestricted" information acquisition I developed in an earlier paper (Denti 2016). Feasible sets of signals are very rich: any player can choose a signal that is arbitrarily correlated with signals of others and state. This is a natural assumption to study network effects in information choice. Rich correlation possibilities are needed to understand what

[1] See Bramoulle and Kranton (2016) for an up-to-date survey.

players want to learn about state, about neighbors' information, about neighbors' neighbors' information, and so forth.[2]

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 motivates the assumption of entropy for cost of information through a representation theorem. Assuming entropy, Section 4 carries out equilibrium analysis and studies network effects in information acquisition. Sections 3 and 4 are mostly independent of each other and the reader who is uninterested in the representation theorem can safely skip Section 3. The Appendix contains omitted proofs.

2 Model

This section describes the model, a game of information acquisition played on a network. The description is organized in two parts: basic game (network structure, actions, states, utilities) and information acquisition technology (feasible signals, cost of information). The maintained assumptions are linear best responses, Gaussian uncertainty, and unrestricted information acquisition.

2.1 Basic game

Let N be a finite set of n *players* with typical elements i and j. Players are connected via a fixed, commonly known network. The network of relations is represented by a $n \times n$ symmetric matrix $G = [g_{ij}]$. The normalization $g_{ii} = 0$ is adopted.

Each player i has set of *actions* A_i. Actions are real numbers: $A_i = \mathbb{R}$. As usual, A_{-i} and A denote the Cartesian products $\times_{j \neq i} A_j$ and $A_i \times A_{-i}$, respectively. Players' utilities depend on one another's action and an uncertain state. Let $\Theta = \mathbb{R}^n$ be the space of *states* with Gaussian distribution P_Θ. Player i's *utility* $u_i : A \times \Theta \to \mathbb{R}$ is continuously differentiable in a_i and measurable in (a_{-i}, θ).

In this paper, best responses are linear. Set

$$y_i(a_{-i}, \theta) = \theta_i + \sum_{j \neq i} g_{ij} a_j$$

for every $a_{-i} \in A_{-i}$ and $\theta \in \Theta$.

[2] Costly communication provides an alternative approach to endogenous information on networks. See Calvo-Armengol et al. (2015), Herskovic and Ramos (2015), and references therein.

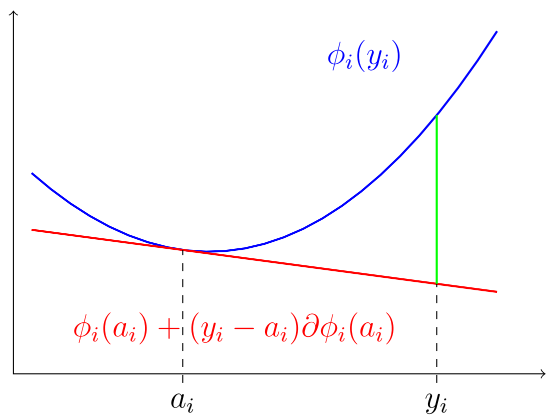

Figure 1: The green line is the Bregman distance of y_i and a_i.

Definition 1. Best responses are *linear* if, for every distribution $P_{A_{-i}\times\Theta}$ on $A_{-i}\times\Theta$ with finite support, $\int y_i dP_{A_{-i}\times\Theta}$ is the unique maximizer of $\int u_i dP_{A_{-i}\times\Theta}$ over $a_i \in A_i$.

We can think of y_i as player i's *target*, a linear combination of state and actions of others. The weight g_{ij} is the link between players i and j, it captures the marginal effect of j's action choice on i's target, hence on i's action choice.

Example 1. Quadratic loss function is standard example for linear best responses:

$$-u_i(a,\theta) = (a_i - y_i)^2 .$$

A generalization is given by *Bregman loss functions*: for $\phi_i : \mathbb{R} \to \mathbb{R}$ strictly convex differentiable function,

$$-u_i(a,\theta) = \phi_i(y_i) - (\phi_i(a_i) + (y_i - a_i)\partial\phi_i(a_i)).$$

The right-hand side is the Bregman distance of y_i and a_i (see Figure 1). The quadratic loss function corresponds to the case $\phi_i(a_i) = a_i^2$.

A result of Banerjee et al. (2005) implies that *all* loss functions that generate linear best responses are of Bregman type:

Fact 1 (Banerjee et al. 2005, Theorem 3)**.** *For every player i, there exists a strictly convex differentiable function $\phi_i : \mathbb{R} \to \mathbb{R}$ such that*

$$u_i(y_i, a_{-i}, \theta) - u_i(a_i, a_{-i}, \theta) = \phi_i(y_i) - (\phi_i(a_i) + (y_i - a_i)\partial\phi_i(a_i))$$

for all $a \in A$ *and* $\theta \in \Theta$*.*

By Fact 1, player i's utility can be identified with a loss function ϕ_i, up to strategically irrelevant terms. To ease the exposition, from now on I assume that

- $u_i(y_i, a_{-i}, \theta) = 0$ for every $a_{-i} \in A_{-i}$ and $\theta \in \Theta$
- ϕ_i is integrable with respect to the standard Gaussian distribution on $\mathbb{R}$.

The first assumption is just a normalization; the second assumption is satisfied if, for instance, ϕ_i is Lipschitz (Bogachev 1998, Corollary 1.7.4).

It is becoming standard to call the object

$$\mathcal{G} = \langle N, \Theta, P_\Theta, (A_i, u_i)_{i \in N} \rangle$$

basic game. In this paper, a basic game is identified by the network of players G, the state distribution P_Θ, and the loss functions $\phi_1, \ldots, \phi_n$.

Remark 1. The covariance matrix of P_Θ may be singular. For instance, if all entries of the matrix are equal, then players have "common values." On the other hand, if the covariance matrix is diagonal, then players are affected only by idyonsicratic shocks. Any situation in between can be modeled as well.

2.2 Information acquisition

In traditional analysis, players' information is exogenously given and represented by an information structure. An *information structure*

$$\mathcal{I} = \langle \Omega, P, \boldsymbol{\theta}, (X_i, \boldsymbol{x}_i)_{i \in N} \rangle$$

consists of an underlying probability space (Ω, P); a random variable $\boldsymbol{\theta} : \Omega \to \Theta$ whose distribution is P_Θ; and, for every player i, a Euclidean space X_i of *messages* and a random variable $\boldsymbol{x}_i : \Omega \to X_i$ called *signal.* Signals represent players' information about state and information of others. Basic game $\mathcal{G}$ and information structure $\mathcal{I}$ define a standard game of incomplete and exogenous information.

In this paper, to endogenize players' information, the information structure is replaced by an information acquisition technology. There are no predetermined signals. Instead, player i has a set $\boldsymbol{X}_i$ of feasible signals she can choose from: $\boldsymbol{X}_i$ is a nonempty set of random variables, measurable functions, from Ω into X_i.

Throughout, uncertainty is Gaussian and signals are normally distributed: for every $\boldsymbol{x} \in \boldsymbol{X}$, the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is Gaussian. Covariance matrices may be singular. For instance, signals may be degenerate and carry no information.

Information is costly and cost of information depends on the joint distribution of messages and states. Player i's *cost of information* is represented by a function C_i from the set of Gaussian distributions on $X \times \Theta$ into $[0, \infty]$. Cost of information may be heterogenous across players. To avoid trivialities, I assume that, for every $\boldsymbol{x}_{-i} \in \boldsymbol{X}_{-i}$, there is $\boldsymbol{x}_i \in \boldsymbol{X}_i$ such that

$$C_i(P_{(\boldsymbol{x},\boldsymbol{\theta})}) < \infty,$$

where $P_{(\boldsymbol{x},\boldsymbol{\theta})}$ stands for the distribution of the random vector $(\boldsymbol{x}, \boldsymbol{\theta})$. To ease notation, from now on I write $C_i(\boldsymbol{x})$ instead of $C_i(P_{(\boldsymbol{x},\boldsymbol{\theta})})$.

Overall, the object

$$\mathcal{T} = \langle \Omega, P, \boldsymbol{\theta}, (X_i, \boldsymbol{X}_i, C_i)_{i \in N} \rangle$$

is called *information acquisition technology*, or more simply, *technology.*

Basic game $\mathcal{G}$ and technology $\mathcal{T}$ define a *game of information acquisition*, denoted by $\langle \mathcal{G}, \mathcal{T} \rangle$. In the game, each player first chooses a signal, then takes an action after observing the realization of her signal, without knowing the signals chosen by others.

The interaction among players can be represented in strategic form as follows. Each player i has a set S_i of *contingency plan*s containing all measurable functions from X_i into A_i. A strategy of hers consists of a signal $\boldsymbol{x}_i$ and a contingency plan s_i. Given profiles $\boldsymbol{x}$ and s of signals and contingency plans, her payoff is

$$E[u_i(s(\boldsymbol{x}), \boldsymbol{\theta})] - C_i(\boldsymbol{x}),$$

expected utility minus cost of information.

The solution concept I consider is pure-strategy Nash equilibrium: signals $\boldsymbol{x}^*$ and contingency plans s^* form an *equilibrium* of $\langle \mathcal{G}, \mathcal{T} \rangle$ if

$$E[u_i(s^*(\boldsymbol{x}^*), \boldsymbol{\theta})] - C_i(\boldsymbol{x}^*) \geq E[u_i(s_i(\boldsymbol{x}_i), s^*_{-i}(\boldsymbol{x}^*_{-i}), \boldsymbol{\theta})] - C_i(\boldsymbol{x}_i, \boldsymbol{x}^*_{-i})$$

for every $i \in N$, $\boldsymbol{x}_i \in \boldsymbol{X}_i$, and $s_i \in S_i$.

The analysis will focus on *linear equilibria* where contingency plans are affine functions, that is, translations of linear functions. Using standard results from team theory (Radner 1962), it is easy to check that linearity is without loss of generality if the largest eigenvalue of G is strictly less than one, the leading case for this paper.

In this paper, information acquisition technologies are very rich, as described by the following assumption:

Assumption 1. *Let $\boldsymbol{x}_{-i}$ be a profile of signals of i's opponents and $P_{X\times\Theta}$ a Gaussian distribution on $X \times \Theta$. If the distribution of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ coincides with the marginal of $P_{X\times\Theta}$ on $X_{-i} \times \Theta$, there there is $\boldsymbol{x}_i \in \boldsymbol{X}_i$ such that $(\boldsymbol{x}, \boldsymbol{\theta})$ has distribution $P_{X\times\Theta}$.*

By Assumption 1, any player can choose a signal that is arbitrarily correlated with signals of others and state. It reflects the idea players can acquire information not only about state, but also about information of others in a flexible way. It can be seen as a natural benchmark case, especially for games on networks. By dropping any exogenous restriction on what players can learn, it allows to study how the network of relations influences what players choose to learn about what others know, and how this affects what information is acquired about the state.

Example 2. There are many technologies that satisfy Assumption 1. For instance, let $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots$ be an infinite sequence of i.i.d. standard Gaussian random variables independent of $\boldsymbol{\theta}$. Suppose that, for every player i, $\boldsymbol{X}_i$ is the set of all random variables $\boldsymbol{x}_i : \Omega \to X_i$ such that

$$\boldsymbol{x}_i = f(\boldsymbol{\theta}, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_k) + t$$

for some positive integer k, linear function $f : \Theta \times \mathbb{R}^k \to X_i$, and constant $t \in \mathbb{R}$. It is easy to check that Assumption 1 is satisfied.

I introduced this model of fully flexible or "unrestricted" information acquisition in Denti 2016. There, I go beyond Gaussianity and consider arbitrary distributions, but focus on games that have a potential structure. Here, on the other hand, linear best responses imply only best-response equivalence to a potential game (quadratic utilities). The difference is thin with exogenous information, but substantial with information acquisition, since the shape of utilities affects the marginal value of information.

3 Representation theorem for entropy

This section introduces a representation theorem for cost of information: for equilibrium analysis, cost of information can be represented by entropy if it satisfies a mild strengthening of Blackwell monotonicity. Beyond identifying a key basic property of cost of information, this result provides a foundation for the equilibrium analysis of next section.

3.1 Entropy

Denote by $Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ the covariance matrix of random vector $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$:

Definition 2. The *entropy* of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ is $\ln \det[Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta})]$.

Entropy is a measure of uncertainty. Originated from information theory (Shannon 1948), it reflects the idea that higher volatility means higher uncertainty. For instance, in the unidimensional case, entropy is an increasing function of variance (the logarithm). More broadly, in the multivariate case, entropy is increasing with respect to the Loewner order on covariance matrices.[3]

In economics, entropy has become a leading specification for cost of information. Following Sims (2003), cost of information is proportional to the expected reduction in uncertainty as measured by entropy of beliefs, that is, to mutual information.

Denote by $Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}_i)$ the conditional covariance matrix of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ given $\boldsymbol{x}_i$:

Definition 3. Let $\det[Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta})] > 0$. The quantity

$$I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}) = \ln \det[Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta})] - \ln \det[Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}_i)]$$

is called the *mutual information* of $\boldsymbol{x}_i$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$.[4]

The mutual information of $\boldsymbol{x}_i$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ is an intuitive measure of the information that i's signal carries about state and information of others. With the idea that more informative signals are more costly, mutual information can be used to specify cost of information:

[3] The Loewner order is the order induced on symmetric matrices by the cone of positive semi-definite matrices.

[4] If the covariance matrix of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ is singular, mutual information is defined with respect to any maximal linearly independent sub-vector of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$.

Example 3. Given scale factor $\mu_i \geq 0$,

$$C_i(\boldsymbol{x}) = \mu_i I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta})$$

for every $\boldsymbol{x} \in \boldsymbol{X}$. The scalar μ_i parametrizes marginal cost of information.

Example 3 is the leading example of cost of information for this paper. The main equilibrium analysis will be carried out for this case. The aim of the rest of this section is to provide a foundation to this functional form.

3.2 Blackwell monotonicity

A basic property satisfied by entropy is Blackwell monotonicity: cost of information is increasing in the sufficiency order of Blackwell (1951). To characterize the implications of entropy, however, the sufficiency order is too coarse and Blackwell monotonicity too weak. In fact, *any* information structure can be endogenized when cost of information is assumed *only* Blackwell monotone, as now I illustrate.

For normally distributed signals, the sufficiency order coincides with the Loewner order on conditional covariance matrices:

Definition 4. Signal $\boldsymbol{x}_i$ is *sufficient* for $\boldsymbol{x}'_i$ with respect to $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ if the matrix

$$Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}'_i) - Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}_i) \text{ is positive semi-definite.}$$

Intuitively, sufficiency means that $\boldsymbol{x}'_i$ provides no more information than $\boldsymbol{x}_i$ about $\boldsymbol{x}_{-i}$ and $\boldsymbol{\theta}$. An equivalent definition of sufficiency would be that

$$Var(f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})|\boldsymbol{x}'_i) \geq Var(f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})|\boldsymbol{x}_i)$$

for every linear function $f : X_{-i} \times \Theta \to \mathbb{R}$. If $\boldsymbol{x}_i$ is sufficient, then it is more correlated than $\boldsymbol{x}'_i$ with any linear function of state and signals of others.

Assumption 2 (Blackwell monotonicity)**.** *Let $\boldsymbol{x}_i, \boldsymbol{x}'_i \in \boldsymbol{X}_i$ and $\boldsymbol{x}_{-i} \in \boldsymbol{X}_{-i}$. If the matrix $Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}'_i) - Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}_i)$ is positive semi-definite, then*

$$C_i(\boldsymbol{x}'_i, \boldsymbol{x}_{-i}) \leq C_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}).$$

The displayed inequality is strict if, in addition, $C_i(\boldsymbol{x})$ is finite and $Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}_i') - Var(\boldsymbol{x}_{-i}, \boldsymbol{\theta}|\boldsymbol{x}_i)$ is positive definite.

By Assumption 2, the more information (in the sense of Blackwell) players acquire about state and information of others, the higher the cost they pay. This is a natural and common property for cost of information, satisfied notably by entropy:

Example 3 (Continued)**.** If $\boldsymbol{x}_i$ is sufficient for $\boldsymbol{x}_i'$ with respect to $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$, then the conditional entropy of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ given $\boldsymbol{x}_i$ is lower than the conditional entropy of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ given $\boldsymbol{x}_i'$. Hence, the expected reduction in entropy is larger for $\boldsymbol{x}_i$ rather than for $\boldsymbol{x}_i'$, that is, $I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}) \geq I(\boldsymbol{x}_i'; \boldsymbol{x}_{-i}, \boldsymbol{\theta})$.[5]

To endogenize information, however, Assumption 2 has limited bite. Let $\langle \mathcal{G}, \mathcal{I} \rangle$ be the game of incomplete information corresponding to basic game $\mathcal{G}$ and information structure $\mathcal{I}$. Say that $s^* \in S$ is an *equilibrium* of $\langle \mathcal{G}, \mathcal{I} \rangle$ if

$$E[u_i(s^*(\boldsymbol{x}), \boldsymbol{\theta})] \geq E[u_i(s_i(\boldsymbol{x}_i), s^*_{-i}(\boldsymbol{x}_{-i}), \boldsymbol{\theta})]$$

for every $i \in N$ and $s_i \in S_i$.

Theorem 0. *Fix basic game $\mathcal{G}$. For every distribution $P_{A\times\Theta}$ on $A \times \Theta$, the following statements are equivalent:*

(i) $P_{A\times\Theta}$ is the action-state distribution in some linear equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$ for some technology $\mathcal{T}$ that satisfies Assumptions 1 and 2.

(ii) $P_{A\times\Theta}$ is the action-state distribution in some linear equilibrium of $\langle \mathcal{G}, \mathcal{I} \rangle$ for some information structure $\mathcal{I}$.

By Theorem 0, any action-state distribution that can arise in equilibrium in a game of incomplete and exogenous information, can also arise in equilibrium in a game of information acquisition (and vice versa). And this is true even if we ask cost of information to be Blackwell monotone.

There is a simple intuition behind this result. By Assumption 2, signals that are more informative in the sense of Blackwell must also be more costly. But more informative signals are also more valuable. Therefore, since Blackwell monotonicity is an ordinal property, marginal cost of information can be adjusted to marginal value to make any desired information structure optimal.

[5] Beyond the Gaussian case, Blackwell monotonicity is a consequence of the data processing inequality (Cover and Thomas 2006, pp. 34-37).

3.3 Strengthening Blackwell

In the same spirit of Blackwell monotonicity, the following assumption provides a complement to Assumption 2:

Assumption 3. *Let $\boldsymbol{x}_i, \boldsymbol{x}'_i \in \boldsymbol{X}_i$ and $\boldsymbol{x}_{-i} \in \boldsymbol{X}_{-i}$. Suppose there is a linear function f from $A_{-i} \times \Theta$ into some Euclidean space such that*

- *the distributions of $(\boldsymbol{x}_i, f(\boldsymbol{x}_{-i}, \boldsymbol{\theta}))$ and $(\boldsymbol{x}'_i, f(\boldsymbol{x}_{-i}, \boldsymbol{\theta}))$ coincide*
- *$\boldsymbol{x}'_i$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ are conditionally independent given $f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$.*

Then $C_i(\boldsymbol{x}'_i, \boldsymbol{x}_{-i}) \le C_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$ and the inequality is strict if, in addition, $C_i(\boldsymbol{x})$ is finite and $\boldsymbol{x}_i$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ are not conditionally independent given $f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$.

By Assumption 3 players pay less for signals that are correlated only with a statistic of signals of others and state. It reflects the same idea of Blackwell monotonicity: the more information players acquire about state and information of others, the higher the cost they pay.

Assumption 3 may be hard to digest at first read. What matters, however, is its implication for the game, which is almost immediate and very intuitive:

Lemma 1. *Let Assumptions 1 and 3 hold. If $\boldsymbol{x}^*_i$ and s^*_i are a best reply to $\boldsymbol{x}_{-i}$ and affine s_{-i}, then $\boldsymbol{x}^*_i$ is conditionally independent of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ given target $\boldsymbol{y}_i$.*

Lemma 1 states that, at the optimum, the correlation of signals with one another and state is explained by the correlation with targets. Players acquire information about state and information of others only if valuable to predict their target.

Assumption 3 is satisfied by entropy, since mutual information, being symmetric, is Blackwell monotone in *both* arguments:

Example 3 (Continued)**.** Let $\boldsymbol{x}_i$, $\boldsymbol{x}'_i$, $\boldsymbol{x}_{-i}$, and f be as in Assumption 3. Since $\boldsymbol{x}'_i$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ are conditionally independent given $f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$, the statistic $f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ is sufficient for $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ with respect to $\boldsymbol{x}'_i$. As a consequence,

$$I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}) \ge I(\boldsymbol{x}_i; f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})) = I(\boldsymbol{x}'_i; f(\boldsymbol{x}_{-i}, \boldsymbol{\theta})) = I(\boldsymbol{x}'_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}),$$

where first inequality and last equality hold since mutual information in increasing in the second argument with respect to the Blackwell order, and the equality in the middle holds since the distributions of $(\boldsymbol{x}_i, f(\boldsymbol{x}_{-i}, \boldsymbol{\theta}))$ and $(\boldsymbol{x}'_i, f(\boldsymbol{x}_{-i}, \boldsymbol{\theta}))$ coincide.

Assumption 3 is neither stronger nor weaker than Assumption 2, the two properties strengthen one another.

3.4 Representation theorem

Now I present the main result of this section: Assumptions 2 and 3 not only are satisfied by entropy, but also fully characterize its implications for equilibrium behavior. This representation theorem will be given for quadratic utilities. This is without loss of generality, as implied by the next lemma:

Lemma 2. *Let $\mathcal{G}$ and $\mathcal{G}'$ be basic games with same network of players and same state distribution. For every distribution $P_{A\times\Theta}$ on $A \times \Theta$, the following statements are equivalent:*

(i) $P_{A\times\Theta}$ is the action-state distribution in some linear equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$ for some technology $\mathcal{T}$ that satisfies Assumptions 1–3.

(ii) $P_{A\times\Theta}$ is the action-state distribution in some linear equilibrium of $\langle \mathcal{G}', \mathcal{T} \rangle$ for some technology $\mathcal{T}$ that satisfies Assumptions 1–3.

By Lemma 2, if network of players and state distribution are fixed, then Assumptions 2 and 3 have the same equilibrium implications for any specification of utilities. As a result, it is without loss of generality to analyze the quadratic case.

If information was exogenous, Lemma 2 would be a consequence of best response equivalence (Morris and Ui 2004). With information acquisition, on the other hand, the curvature of utilities may matter: it affects the marginal value of information. However, Assumptions 2 and 3 impose only ordinal restrictions on cost of information. Therefore, any change in marginal value is offset by a change in marginal cost.

Theorem 1 (Representation theorem)**.** *Fix basic game $\mathcal{G}$ with quadratic utilities:*

$$-u_i(a, \theta) = (a_i - y_i)^2, \quad \forall i \in N,\ a \in A,\ \text{and } \theta \in \Theta.$$

For every distribution $P_{A\times\Theta}$ on $A \times \Theta$, the following statements are equivalent:

(i) $P_{A\times\Theta}$ is the action-state distribution in some linear equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$ for some technology $\mathcal{T}$ that satisfies Assumptions 1–3.

(ii) $P_{A\times\Theta}$ *is the action-state distribution in some linear equilibrium of* $\langle \mathcal{G}, \mathcal{T}\rangle$ *for some technology* $\mathcal{T}$ *and scalars* $\mu_1, \ldots, \mu_n \geq 0$ *that satisfy Assumption 1 and*

$$C_i(\boldsymbol{x}) = \mu_i I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}), \quad \forall i \in N \text{ and } \boldsymbol{x} \in \boldsymbol{X}.$$

If, in addition, $g_{ij} \neq 0$ *and* $\int a_i^2 dP_{A\times\Theta}(a_i) \int a_j^2 dP_{A\times\Theta}(a_j) > 0$*, then* $\mu_i = \mu_j$*.*

Theorem 1 has two main implications. First, it provides a foundation for equilibrium analysis under entropy. Any action-state distribution that can arise in equilibrium under Assumptions 2 and 3, can also arise in equilibrium when cost of information is proportional to mutual information.

The second main implication of Theorem 1 is that, if degenerate situations are disregarded, then neighbors' scale factors can be assumed equal. By induction, the same is true for players that are path-connected. Overall, cost of information can be assumed homogenous within each component of the network.[6] Since there is no strategic interaction across components, ultimately this allows to assume that *all* players share the same cost of information.

The intuition behind Theorem 1 is as follows. Let $\boldsymbol{x}^*$ and s^* be a linear equilibrium under Assumptions 2 and 3. By linearity of best responses and Lemma 1,

$$s_i^*(\boldsymbol{x}_i^*) = E[\boldsymbol{y}_i^* | s_i^*(\boldsymbol{x}_i^*)] \quad \text{and} \quad E[s_i^*(\boldsymbol{x}_i^*) | \boldsymbol{y}_i^*] = E[s_i^*(\boldsymbol{x}_i^*) | s_{-i}^*(\boldsymbol{x}_{-i}^*), \boldsymbol{\theta}].$$

The displayed moment restrictions identify the entire distribution of $(s^*(\boldsymbol{x}^*), \boldsymbol{\theta})$ from just n numbers, $\rho^2_{(s_1^*(\boldsymbol{x}_1^*), \boldsymbol{y}_1^*)}, \ldots, \rho^2_{(s_n^*(\boldsymbol{x}_n^*), \boldsymbol{y}_n^*)}$, the correlations of actions and targets. These correlations correspond to the precision of players' information. With entropy, any level of precision $\rho^2_{(s_i^*(\boldsymbol{x}_i^*), \boldsymbol{y}_i^*)}$ can be made optimal for player i by appropriately setting its marginal cost μ_i.

What μ_i is needed to endogenize $\rho^2_{(s_i^*(\boldsymbol{x}_i^*), \boldsymbol{y}_i^*)}$ typically depend on the marginal value of information, that is, on the curvature of player i's utility. The peculiarity of the quadratic case is that value of information is linear in $\rho^2_{(s_i^*(\boldsymbol{x}_i^*), \boldsymbol{y}_i^*)}$:

$$\min_{s_i \in S_i} E[(s_i(\boldsymbol{x}_i^*) - \boldsymbol{y}_i^*)^2] = Var(\boldsymbol{y}_i^* | \boldsymbol{x}_i^*) = Var(\boldsymbol{y}_i^*) \left(1 - \rho^2_{(s_i^*(\boldsymbol{x}_i^*), \boldsymbol{y}_i^*)}\right).$$

This makes possible to choose homogenous marginal cost for connected players.

[6] The components of a network are its distinct maximal path-connected subgraphs.

Theorem 1 identifies key primitive properties for cost of information. A basic challenge for the rational inattention literature, and more broadly for the study of endogenous information acquisition in economic settings, it is to go beyond functional forms and investigate more primitive assumptions on cost of information. This representation theorem addresses this challenge in settings with linear best responses and Gaussian uncertainty. While these conditions are certainly restrictive, they are nonetheless extensively applied in practice. In particular, most of the rational inattention literature still focuses on the linear-Gaussian case.[7]

4 Network effects in information acquisition

This section studies network effects in information acquisition, assuming quadratic utilities and entropy. Foundation for these functional forms are provided in the previous section on the basis of primitive properties of cost of information and a representation theorem.

Throughout, let $\mathcal{G}$ be a basic game such that

$$-u_i(a, \theta) = (a_i - y_i)^2, \quad \forall i \in N,\ a \in A, \text{ and } \theta \in \Theta.$$

Given scalar $\mu \geq 0$, denote by $\mathcal{T}_\mu$ a technology that satisfies Assumption 1 and

$$C_i(\boldsymbol{x}) = \mu I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}), \quad \forall i \in N \text{ and } \boldsymbol{x} \in \boldsymbol{X}.$$

The quantity $I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta})$ is the Shannon mutual information of $\boldsymbol{x}_i$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$, that is, the expected reduction in the entropy of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ due to the knowledge of $\boldsymbol{x}_i$. The case $\mu = 0$ of costless information is considered for benchmark.

Revelation principle. To ease notation, I assumes $X = A$ and look at equilibria $\boldsymbol{x}^*$ and s^* of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$ such that $s^*(\boldsymbol{x}^*) = \boldsymbol{x}^*$. Messages, therefore, are interpreted as action recommendations. A standard revelation-principle argument justifies this restriction. For short, the reference to the identity functions will be omitted.

Largest eigenvalue of G. For this section, the largest eigenvalue of G is assumed to be strictly less than one. This is a well-known sufficient and somewhat neces-

[7] Sims (2006) provides an early discussion of this issue.

sary condition for equilibrium uniqueness if information is exogenous. Here, with information acquisition, the condition guarantees equilibrium uniqueness if state is degenerate, that is, if $Var(\boldsymbol{\theta}) = 0$. It rules out sunspot equilibria where players coordinate on buying correlation devices.

4.1 Bonacich centralities

The first main result on network effects relates endogenous information acquisition to the centrality measures of Bonacich (1987). These centralities are based on the idea that powerful players have more connections, more "walks" to other players:

Definition 5. Let $\mathbb{I}$ be the identity matrix and define $B(G) = (\mathbb{I} - G)^{-1}$. Given $w \in \mathbb{R}^n$, the vector of *weighted Bonacich centralities* relative to G is $B(G)w$.

The Bonacich centrality of a player refers is the sum of *all* walks in the network emanating to her. To illustrate, let the network matrix be a contraction.[8] Then

$$B(G) = (\mathbb{I} - G)^{-1} = \sum_{k=0}^{\infty} G^k.$$

The ij entry of G^k counts the walks of length k from i to j; e.g., the ij entry of G^2 is

$$g_{i1}g_{1j} + \ldots + g_{in}g_{nj}.$$

Therefore, the ij entry of $B(G)$ counts the walks of any length from i to j. Overall, the weighted Bonacich centrality of player i counts all walks emanating from i; walks towards different players are weighted in different ways according to w.

When information is complete, Ballester et al. (2006) show that network effects can be measured by Bonacich centralities. This key result on network games can be replicated here in the case of costless information:

Example 4. The game $\langle \mathcal{G}, \mathcal{T}_0 \rangle$ has a unique equilibrium $\boldsymbol{x}^* = B(G)\boldsymbol{\theta}$.

If information is costless, the signal of each player correspond to her weighted Bonacich centralities relative to G. The state determines the weights. The matrix $B(G)$ intuitively captures all possible feedback among action choices:

[8] A symmetric matrix is a contraction if all eigenvalues are less than one in absolute value.

Example 4 (Continued)**.** Since $\mu = 0$, in equilibrium signals coincide with targets. But targets depend on signals of others, which are endogenous too. Iterating:

$$\boldsymbol{x}^* = \boldsymbol{\theta} + G\boldsymbol{x}^* = \boldsymbol{\theta} + G\boldsymbol{\theta} + G^2\boldsymbol{x}^* = \ldots = \sum_{k=0}^{l} G^k\boldsymbol{\theta} + G^l\boldsymbol{x}^*.$$

To extend Example 4 to the general case of costly information, the key object is the matrix

$$\left[\rho_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)}\rho_{(\boldsymbol{x}_j^*,\boldsymbol{y}_j^*)}g_{ij}\right].$$

This can be interpreted as an *information-adjusted network* relative to $\boldsymbol{x}^*$, an equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$. In this network, original links are re-weighted by the correlation coefficients of signals and targets. These coefficients can be seen as overall measures of the precision of players' information.

The information-adjusted network can be decomposed into two parts: the given network of relations G and an endogenous network of information precisions

$$R_{\boldsymbol{x}^*} = \left[\rho_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)}\rho_{(\boldsymbol{x}_j^*,\boldsymbol{y}_j^*)}\right].$$

In this network, connections depend on how precise players information is. The more information two players have, the stronger their tie.

The information-adjusted network is the entrywise product of $R_{\boldsymbol{x}^*}$ and G:

$$R_{\boldsymbol{x}^*} \circ G = \left[\rho_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)}\rho_{(\boldsymbol{x}_j^*,\boldsymbol{y}_j^*)}g_{ij}\right].$$

Throughout, the symbol $\circ$ is used for the entrywise product between matrices.

The information-adjusted network provides the right notion of centrality to generalize Example 4. Define the matrix $B(R_{\boldsymbol{x}^*}, G)$ such that

$$B(R_{\boldsymbol{x}^*}, G) = R_{\boldsymbol{x}^*} \circ (\mathbb{I} - R_{\boldsymbol{x}^*} \circ G)^{-1}.$$

Proposition 1. *If $\boldsymbol{x}^*$ is an equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$, then*

- $E[\boldsymbol{x}^*|\boldsymbol{\theta}] = B(G)E[\boldsymbol{\theta}] + B(R_{\boldsymbol{x}^*}, G)(\boldsymbol{\theta} - E[\boldsymbol{\theta}])$
- $Var(\boldsymbol{x}^*|\boldsymbol{\theta}) = \mu B(R_{\boldsymbol{x}^*}, G)$.

Proposition 1 states that, in expectation, the signal of each player corresponds to her weighted Bonacich centralities relative to the information-adjusted network. For instance, if $E[\boldsymbol{\theta}] = 0$, then

$$E[\boldsymbol{x}^*|\boldsymbol{\theta}] = B(R_{\boldsymbol{x}^*}, G)\boldsymbol{\theta} = R_{\boldsymbol{x}^*} \circ (\mathbb{I} - R_{\boldsymbol{x}^*} \circ G)^{-1}\boldsymbol{\theta}.$$

Moreover, the extra correlation among signals the is unexplained by the state, it is explained by the information-adjusted walks between players.[9]

The basic intuition behind Proposition 1 generalizes the one for Example 4. Since information is costly, signals do not coincide exactly with targets, they are only noisy versions of them:

$$\boldsymbol{x}_i^* = \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)}(\boldsymbol{y}_i^* + \boldsymbol{\epsilon}_i) \quad \text{with} \quad \boldsymbol{\epsilon}_i = \frac{\boldsymbol{x}_i^* - E[\boldsymbol{x}_i^*|\boldsymbol{y}_i^*]}{\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)}}.$$

But targets depend on signals of others, which are endogenous too. Iterating

$$\begin{aligned}
\boldsymbol{x}^* &= (R_{\boldsymbol{x}^*} \circ \mathbb{I})(\boldsymbol{\theta} + \boldsymbol{\epsilon}) + (R_{\boldsymbol{x}^*} \circ I)G\boldsymbol{x}^* \\
&= (R_{\boldsymbol{x}^*} \circ \mathbb{I})(\boldsymbol{\theta} + \boldsymbol{\epsilon}) + R_{\boldsymbol{x}^*} \circ (R_{\boldsymbol{x}^*} \circ G)(\boldsymbol{\theta} + \boldsymbol{\epsilon}) + R_{\boldsymbol{x}^*} \circ (R_{\boldsymbol{x}^*} \circ G)G\boldsymbol{x}^* \\
&= \dots \\
&= R_{\boldsymbol{x}^*} \circ \sum_{k=0}^{l} (R_{\boldsymbol{x}^*} \circ G)^k(\boldsymbol{\theta} + \boldsymbol{\epsilon}) + R_{\boldsymbol{x}^*} \circ (R_{\boldsymbol{x}^*} \circ G)^l G\boldsymbol{x}^*.
\end{aligned}$$

The matrix $B(R_{\boldsymbol{x}^*}, G)$ captures all this noisy feedback.

Proposition 1 pins down equilibrium behavior up to the precisions of players' information. The next proposition closes the equilibrium characterization:

Proposition 2. *Suppose $\mu > 0$. For every $\rho^2 = (\rho_1^2, \dots, \rho_n^2) \in [0,1)^n$, the following statements are equivalent:*

(i) There is an equilibrium $\boldsymbol{x}^*$ of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$ such that $\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} = \rho_i^2$ for all players i.

(ii) ρ^2 is a Nash equilibrium of common-interest game $V_\mu : [0,1)^n \to \mathbb{R}$ such that

$$V_\mu(\rho^2) = \mathrm{tr}[Var(\boldsymbol{\theta})B(R, G)] - \mu \ln \frac{\det[\mathbb{I} - R \circ G]}{\det[\mathbb{I} - R \circ \mathbb{I}]}$$

[9] In the definition of $B(R_{\boldsymbol{x}^*}, G)$, the additional entrywise product is mere accounting. If instead $(\mathbb{I} - R_{\boldsymbol{x}^*} \circ G)^{-1}$ was used, a walk of length one from i to j would count $\rho_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)}\rho_{(\boldsymbol{x}_j^*, \boldsymbol{y}_j^*)}g_{ij}$ instead of $\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)}\rho^2_{(\boldsymbol{x}_j^*, \boldsymbol{y}_j^*)}g_{ij}$.

where R stands for the matrix $[\rho_i \rho_j]$.

By Proposition 2, players choose the precision of information *as if* they played an auxiliary complete-information game. In the game, each player chooses a number ρ_i^2 between zero and one. This is interpreted as the precision of the player's signal in the original game of information acquisition. The auxiliary game is of common interest and players share the same payoff function V_μ.

This auxiliary game can be used, for instance, to determine existence of equilibria for the original game of information acquisition:

Corollary 1. *The game $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$ has an equilibrium.*

In earlier work, I proved more abstract versions of Propositions 1 and 2 for general potential games and arbitrary distributions of uncertainty. Indeed, a possible way to derive the propositions is to start from the results of Denti (2016) and specialize to this setting of quadratic utilities and Gaussian uncertainty.

Here I follow a different route. The more specific setting allows to use different, more revealing arguments. As hinted above, the proof of Proposition 1 is an intuitive extension of the complete-information argument of Example 4. On the other hand, the proof of Proposition 2 is based on a closed-form characterization of best replies in the auxiliary game. Such characterization, which I now illustrate, is not available in the general case and substantially simplifies the analysis.

Consider the following factorizations of R, G, and $\mathbb{I}$ in block matrices:

$$R = \begin{bmatrix} \rho_i^2 & \rho_i \rho_{-i}^\intercal \\ \rho_i \rho_{-i} & R_{-i} \end{bmatrix}, \quad G = \begin{bmatrix} 0 & g_i^\intercal \\ g_i & G_{-i} \end{bmatrix}, \quad \text{and} \quad \mathbb{I} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbb{I}_{-i} \end{bmatrix}.$$

In the factorizations, ρ_{-i} and g_{-i} are column vectors, $\rho_{-i}^\intercal$ and $g_i^\intercal$ their transpose. Define $B(R_{-i}, G_{-i})$ for the network of players without player i:

$$B(R_{-i}, G_{-i}) = R_{-i} \circ (\mathbb{I}_{-i} - R_{-i} \circ G_{-i})^{-1}.$$

Lemma 3. *In the auxiliary game V_μ, a necessary and sufficient condition for ρ_i^2 to be a best reply to ρ_{-i}^2 is*

$$\mu \frac{1 - g_i^\intercal B(R_{-i}, G_{-i}) g_i}{1 - \rho_i^2} \geq \frac{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i}) \boldsymbol{\theta}_{-i})}{1 - \rho_i^2 g_i^\intercal B(R_{-i}, G_{-i}) g_i}$$

and equality holds if $\rho_i^2 > 0$.

The optimality condition of Lemma 3 has a simple interpretation. For $\boldsymbol{x}^*$ equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$, it is easy to check from Proposition 2 that

$$Var(\boldsymbol{y}_i^*) = \frac{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{\boldsymbol{x}_{-i}^*}, G_{-i})\boldsymbol{\theta}_{-i})}{(1 - \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} g_i^\intercal B(R_{\boldsymbol{x}_{-i}^*}, G_{-i}) g_i)^2} + \mu \frac{g_i^\intercal B(R_{\boldsymbol{x}_{-i}^*}, G_{-i}) g_i}{1 - \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} g_i^\intercal B(R_{\boldsymbol{x}_{-i}^*}, G_{-i}) g_i}.$$

Then the optimality condition becomes

$$Var(\boldsymbol{y}_i^*)(1 - \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)}) \leq \mu$$

and equality holds if $\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} > 0$. This pins down the optimal correlation coefficient for the information acquisition problem

$$\max_{\boldsymbol{x}_i \in \boldsymbol{X}_i} -E[(\boldsymbol{x}_i - \boldsymbol{y}_i^*)^2] - \mu I(\boldsymbol{x}_i; \boldsymbol{y}_i^*).$$

4.2 Multiplicity and contraction

Externalities in information acquisition provide new source of multiple equilibria for network games. Before presenting general results, I use a simple example to illustrate the main intuition:

Example 5. Let $n = 2$ and $|g_{ij}| < 1$. Suppose marginal cost of information is relatively high: $\mu \geq Var(\boldsymbol{\theta}_i), Var(\boldsymbol{\theta}_j)$. Then an equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$ is $\boldsymbol{x}^* = 0$.

In this two-player example, not acquiring information is an equilibrium. Player i's incentive to acquire information depends on the volatility of her target. If j's signal is deterministic, the only source of uncertainty is the state:

$$Var(\boldsymbol{y}_i) = Var(\boldsymbol{\theta}).$$

Therefore, if variance of the state is sufficiently lower than marginal cost of information, then the player is happy with no learning.

If information was exogenously given, the condition $|g_{ij}| < 1$ would guarantee contraction in best replies and unique equilibrium. A similar channel to uniqueness can be seen here in the extreme case of degenerate state:

Example 6 (Continued)**.** If $Var(\boldsymbol{\theta}) = 0$, then $\boldsymbol{x}^* = 0$ is the unique equilibrium.

If state is degenerate, then not acquiring information is the unique equilibrium. Player i's incentive to acquire information comes solely from the volatility of j's signal:

$$Var(\boldsymbol{y}_i) = g_{ij}^2 Var(\boldsymbol{x}_j).$$

Since $|g_{ij}| < 1$, the effect of j's behavior on the volatility of i's target is capped. This "contraction property" is enough to prevent equilibria where players watch sunspots or buy correlation devices simply because they expect the opponent to do the same.

Moving away from the extreme case of degenerate state, network effects in information acquisition kick in and differentiate endogenous from exogenous information:

Example 7 (Continued)**.** Let $Var(\boldsymbol{\theta}_i) = Var(\boldsymbol{\theta}_j) > 0$ and $\rho^2_{(\boldsymbol{\theta}_i,\boldsymbol{\theta}_j)} < 1$. There is $\epsilon > 0$ such that, if $|1 - g_{ij}| < \epsilon$, then there is another equilibrium $\boldsymbol{x}^{**}$ with $Var(\boldsymbol{x}^{**}) > 0$.

If the state is not degenerate and the link between players sufficiently strong, then there is also an equilibrium with positive information acquisition. If the state is uncertain, the condition $|g_{ij}| < 1$ no longer guarantees contraction in best replies:

$$Var(\boldsymbol{y}_i) = Var(\boldsymbol{\theta}_i) + 2g_{ij}Cov(\boldsymbol{\theta}_i, \boldsymbol{x}_j) + g_{ij}^2 Var(\boldsymbol{x}_j).$$

The additional covariance term amplifies the effect of j's behavior on the volatility of i's target. If $|g_{ij}|$ is sufficiently close to one, then the players can convince one another to acquire information about the state.

The auxiliary game of Proposition 2 provides the proper framework to generalize Example 5. By Proposition 1, equilibrium behavior in the game of information acquisition is pinned down by the precision of players' information. The auxiliary game allows to study in isolation how players choose the precision of their information.

Equilibria in the auxiliary game correspond to inflection points of the map

$$(\rho_1^2, \ldots, \rho_n^2) = \rho^2 \mapsto V_\mu(\rho^2) = \mathrm{tr}[Var(\boldsymbol{\theta})B(R,G)] - \mu \ln \frac{\det[\mathbb{I} - R \circ G]}{\det[\mathbb{I} - R \circ \mathbb{I}]}.$$

The next proposition illustrates the structure of this common payoff function:

Proposition 3. *The functions*

$$\rho^2 \mapsto \mathrm{tr}[Var(\boldsymbol{\theta})B(R,G)] \quad \textit{and} \quad \rho^2 \mapsto \ln \frac{\det[\mathbb{I} - R \circ G]}{\det[\mathbb{I} - R \circ \mathbb{I}]}$$

are increasing and convex.

Proposition 3 states that V_μ is the difference of two increasing convex functions. The two terms can be seen as value and cost of information in the auxiliary game:

$$\underbrace{\mathrm{tr}[Var(\boldsymbol{\theta})B(R,G)]}_{\text{value of information}} - \mu \underbrace{\ln \frac{\det[\mathbb{I} - R \circ G]}{\det[\mathbb{I} - R \circ \mathbb{I}]}}_{\text{cost of information}}.$$

The interpretation is suggested by the presence of the scale factor, and it is ratified by the property of monotonicity. Intuitively, the more precise information is, the higher its value but also its cost.

More interestingly, both value and cost of information in the auxiliary game are convex. This means that not only marginal cost, but also marginal return of information is increasing. Overall, the common payoff function is not concave and this may lead to multiple inflection points and multiple equilibria, as in Example 5.

Convexities in value of information are a unique manifestation of network effects in information acquisition. If there is no strategic interaction, that is, if $G = 0$, then

$$\mathrm{tr}[Var(\boldsymbol{\theta})B(R,G)] = \sum_{i=1}^{n} Var(\boldsymbol{\theta}_i)\rho_i^2$$

and marginal return to information is constant. This observation disentangles Proposition 3 from classic single-agent results on the nonconcavity of the value of information (Radner and Stiglitz 1989).

The proof of convexity is illustrated in Figure 2. The Hessian matrix is hard to compute for more than two players. The idea, therefore, is to decompose the horizontal black arrow into more elementary functions. It is quite easy to see that both vertical blue arrows are linear and increasing. On the other hand, the horizontal red arrow requires some matrix analysis. In particular, both monotonicity and convexity are defined with respect to the Loewner order on symmetric matrices.

Example 5 suggests that, if links are not too strong or state is not too volatile, network effects in information acquisition can be controlled and contraction in best replies restored. The next proposition formalizes this intuition:

Proposition 4. *Let BR_i be i's best reply function in the auxiliary game V_μ. For*

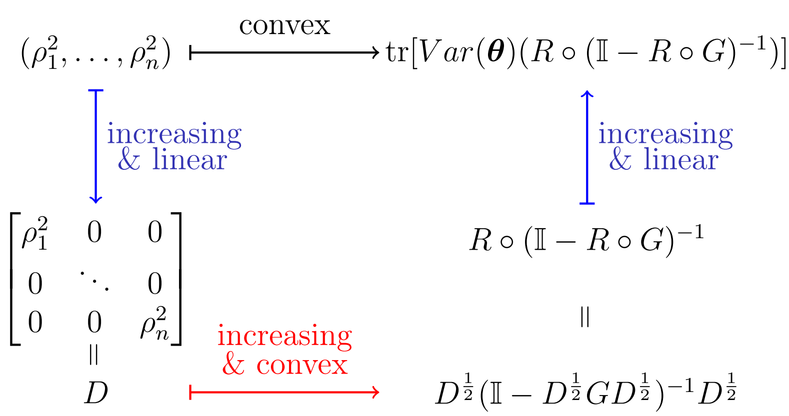

Figure 2: Proof of Proposition 3

every ρ_{-i}^2 *and* $\tilde{\rho}_{-i}^2$ *belonging to* $[0,1)^{n-1}$,

$$\left|BR_i(\rho_{-i}^2) - BR_i(\tilde{\rho}_{-i}^2)\right| \leq \frac{\|G\|\|Var(\boldsymbol{\theta})\|}{\mu}\left(\frac{2}{(1-\|G\|)^+}\right)^{4(n+1)} \max_{j\neq i}\left|\rho_j^2 - \tilde{\rho}_j^2\right|$$

where $\|G\|$ *and* $\|Var(\boldsymbol{\theta})\|$ *are the spectral norms of* G *and* $Var(\boldsymbol{\theta})$.

The proof of Proposition 4 starts from the characterization of best replies from Lemma 3 and perform a series of majorization. The Lipschitz constant, therefore, is not tight. The relevant implication is that, if $\|G\|$ or $\|Var(\boldsymbol{\theta})\|$ are sufficiently low, then best replies in the auxiliary game are contraction mappings. This implies that, if links are not too strong or state is not too volatile, then the game of information acquisition has a unique equilibrium:

Corollary 2. *There is* $\epsilon > 0$ *such that, if* $\|G\| \leq \epsilon$ *or* $\|Var(\boldsymbol{\theta})\| \leq \epsilon$*, then any two equilibria of* $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$ *induce the same action-state distribution.*

Propositions 3 and 4, as well as Example 5, are independent of strategic motives for actions. The literature has mostly focused on the channel "from complementarity in actions to complementarity in information acquisition" for multiple equilibria (Hellwig and Veldkamp 2009, Yang 2015, ...) and convexities in value of information (Amir and Lazzati 2016). Here the analysis discusses general network effects and applies to any pattern of complementarity and substitutability among actions.

5 Conclusion

This paper investigated how network effects in action choice induce externalities in information acquisition. It showed these externalities can be measured by Bonacich centralities and provide new source for multiple equilibria. The results were derived under broad assumptions on information acquisition. Entropy was used as tool for the analysis, not taken as primitive of the model.

Beyond the application to network economics, the framework developed by this paper can be seen as a laboratory to study endogenous information acquisition. A central question to the literature is how structure of payoffs shapes endogenous information structure. If best responses are linear, the structure of payoffs can be summarized by a matrix. This paper showed how characteristics of this matrix naturally translates into properties of information in equilibrium. Moreover, a key challenge is to identify what primitive properties of cost of information drive predictions. If uncertainty is Gaussian, this paper provided simple monotonicity properties that are sufficient to derive sharp predictions.

Appendix

Lemma 4. *Let $\boldsymbol{x}_{-i} \in \boldsymbol{X}_{-i}$ and $s_{-i} \in S_{-i}$ such that s_{-i} is affine. For every $\boldsymbol{x}_i, \boldsymbol{x}_i' \in \boldsymbol{X}_i$, $Var(\boldsymbol{y}_i|\boldsymbol{x}_i) \leq Var(\boldsymbol{y}_i|\boldsymbol{x}_i')$ if and only if*

$$\max_{s_i \in S_i} E[u_i(s_i(\boldsymbol{x}_i), s_{-i}(\boldsymbol{x}_{-i}), \boldsymbol{\theta})] \geq \max_{s_i \in S_i} E[u_i(s_i(\boldsymbol{x}_i'), s_{-i}(\boldsymbol{x}_{-i}), \boldsymbol{\theta})].$$

Proof of Lemma 4. Let ϕ_i be as in Fact 1. Since best responses are linear, the statement to prove is equivalent to

$$Var(\boldsymbol{y}_i|\boldsymbol{x}_i) \leq Var(\boldsymbol{y}_i|\boldsymbol{x}_i') \quad \Leftrightarrow \quad E[\phi_i(E[\boldsymbol{y}_i|\boldsymbol{x}_i])] \geq E[\phi_i(E[\boldsymbol{y}_i|\boldsymbol{x}_i'])].$$

Suppose first that $Var(\boldsymbol{y}_i|\boldsymbol{x}_i) \leq Var(\boldsymbol{y}_i|\boldsymbol{x}_i')$ and $Var(\boldsymbol{y}_i|\boldsymbol{x}_i') = Var(\boldsymbol{y}_i)$. Since ϕ_i is strictly convex, then by Jensen inequality

$$E[\phi_i(E[\boldsymbol{y}_i|\boldsymbol{x}_i])] \geq \phi_i(E[\boldsymbol{y}_i]) = E[\phi_i(E[\boldsymbol{y}_i|\boldsymbol{x}_i'])]$$

and the inequality is strict if $Var(\boldsymbol{y}_i|\boldsymbol{x}_i) < Var(\boldsymbol{y}_i)$.

Suppose now that $Var(\boldsymbol{y}_i|\boldsymbol{x}_i) \leq Var(\boldsymbol{y}_i|\boldsymbol{x}'_i)$ and $Var(\boldsymbol{y}_i|\boldsymbol{x}'_i) < Var(\boldsymbol{y}_i)$. Letting

$$t = \frac{\sqrt{Var(\boldsymbol{y}_i) - Var(\boldsymbol{y}_i|\boldsymbol{x}'_i)}}{\sqrt{Var(\boldsymbol{y}_i) - Var(\boldsymbol{y}_i|\boldsymbol{x}_i)}},$$

it is easy to check that $E[\boldsymbol{y}_i|\boldsymbol{x}'_i]$ has the same distribution of

$$tE[\boldsymbol{y}_i|\boldsymbol{x}_i] + (1-t)E[\boldsymbol{y}_i].$$

Since ϕ_i is convex and $t \in [0,1]$,

$$E[\phi_i(E[\boldsymbol{y}_i|\boldsymbol{x}'_i])] \leq tE[\phi_i(E[\boldsymbol{y}_i|\boldsymbol{x}_i])] + (1-t)\phi_i(E[\boldsymbol{y}_i]) \leq E[\phi_i(E[\boldsymbol{y}_i|\boldsymbol{x}_i])]$$

and the last inequality is strict if if $t < 1$, that is, if $Var(\boldsymbol{y}_i|\boldsymbol{x}_i) < Var(\boldsymbol{y}_i|\boldsymbol{x}'_i)$. ■

Proof of Theorem 0. It is clear that (i) implies (ii). To check the opposite implication, let $P_{A\times\Theta}$ be a distribution over $A \times \Theta$ that satisfies (ii).

Consider a technology $\mathcal{T}$ such that $X = A$ and Assumption 1 holds. Regarding cost of information, for every player i and signal profile $\boldsymbol{x} \in \boldsymbol{X}$, set

$$C_i(\boldsymbol{x}) = \max_{s_i \in S_i} E[u_i(s_i(\boldsymbol{x}_i), \boldsymbol{x}_{-i}, \boldsymbol{\theta})] - E[u_i(E[\boldsymbol{y}_i], \boldsymbol{x}_{-i}, \boldsymbol{\theta})]$$

where $\boldsymbol{y}_i = \boldsymbol{\theta}_i + \sum_{j \neq i} g_{ij}\boldsymbol{x}_j$. By Lemma 4, Assumption 2 holds as well.

By Assumption 1, we can choose $\boldsymbol{x}^* \in \boldsymbol{X}$ such that $P_{A\times\Theta}$ is the distribution of $(\boldsymbol{x}^*, \boldsymbol{\theta})$. Letting s^* be the identity function, we claim that $\boldsymbol{x}^*$ and s^* are an equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$. Indeed, for every player i and deviation $\boldsymbol{x}_i \in \boldsymbol{X}_i$,

$$E[u_i(\boldsymbol{x}^*, \boldsymbol{\theta})] - C_i(\boldsymbol{x}^*) = E[u_i(E[\boldsymbol{y}_i^*], \boldsymbol{x}^*_{-i}, \boldsymbol{\theta})] = \max_{s_i \in S_i} E[u_i(s_i(\boldsymbol{x}_i), \boldsymbol{x}^*_{-i}, \boldsymbol{\theta})] - C_i(\boldsymbol{x}_i, \boldsymbol{x}^*_{-i})$$

where the first equality holds since $P_{A\times\Theta}$ satisfies (ii). ■

Proof of Lemma 1. By Assumption 1, we can choose $\boldsymbol{x}_i \in \boldsymbol{X}_i$ such that the distributions of $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ and $(\boldsymbol{x}^*_i, \boldsymbol{y}_i)$ coincide, but $\boldsymbol{x}_i$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ are conditionally independent given $\boldsymbol{y}_i$. Since $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ and $(\boldsymbol{x}^*_i, \boldsymbol{y}_i)$ are equally distributed, by Lemma 4

$$\max_{s_i \in S_i} E[u_i(s_i(\boldsymbol{x}_i), s_{-i}(\boldsymbol{x}_{-i}), \boldsymbol{\theta})] \geq E[u_i(s^*_i(\boldsymbol{x}^*_i), s_{-i}(\boldsymbol{x}_{-i}), \boldsymbol{\theta})].$$

Therefore, since $\boldsymbol{x}_i^*$ and s_i^* are a best reply to $\boldsymbol{x}_{-i}$ and s_{-i}, $C_i(\boldsymbol{x}_i^*, \boldsymbol{x}_{-i}) \leq C_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$. By Assumption 3, also the opposite inequality holds and $\boldsymbol{x}_i^*$ and $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ are conditionally independent given $\boldsymbol{y}_i$. ■

Proof of Lemma 2. Denote by u_i and u_i' the utility of player i in $\mathcal{G}$ and $\mathcal{G}'$, respectively. We show that (i) implies (ii): the proof of the other direction is analogous.

Let $P_{A\times\Theta}$ be a distribution over $A \times \Theta$ that satisfies (i). Consider a technology $\mathcal{T}$ such that $X = A$ and Assumption 1 holds.

For $t \in [0,1]$ and $\boldsymbol{x}_{-i} \in \boldsymbol{X}_{-i}$, define

$$f_{(\boldsymbol{x}_{-i},\boldsymbol{\theta})}(t) = \max_{s_i \in S_i} E[u_i'(s_i(\boldsymbol{x}_i), \boldsymbol{x}_{-i}, \boldsymbol{\theta})] - E[u_i'(E[\boldsymbol{y}_i], \boldsymbol{x}_{-i}, \boldsymbol{\theta})]$$

where $\boldsymbol{y}_i = \boldsymbol{\theta}_i + \sum_{j\neq i} g_{ij}\boldsymbol{x}_j$ and $\boldsymbol{x}_i$ is any signal of player i such that

$$tVar(\boldsymbol{y}_i) = Var(\boldsymbol{y}_i|\boldsymbol{x}_i).$$

By Lemma 4, the function $f_{(\boldsymbol{x}_{-i},\boldsymbol{\theta})}$ from $[0,1]$ into $[0,\infty)$ is well defined. Moreover, if $Var(\boldsymbol{y}_i) > 0$, then the function is strictly decreasing.

Cost of information in $\mathcal{T}$ for player i is defined as follows:

$$C_i(\boldsymbol{x}) = \begin{cases} f_{(\boldsymbol{x}_{-i},\boldsymbol{\theta})}\left(\frac{Var(\boldsymbol{x}_i|\boldsymbol{x}_{-i},\boldsymbol{\theta})}{Var(\boldsymbol{x}_i)}\right) & \text{if } Var(\boldsymbol{y}_i) > 0, \\ I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}) & \text{else.} \end{cases}$$

It is easy to check that Assumptions 2 and 3 hold.

By Assumption 1, we can choose $\boldsymbol{x}^* \in \boldsymbol{X}^*$ such that $P_{A\times\Theta}$ is the distribution of $(\boldsymbol{x}^*, \boldsymbol{\theta})$. Letting s^* be the identity contingency plans, we claim that $\boldsymbol{x}^*$ and s^* are an equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$.

To verify the claim, consider first the case $Var(\boldsymbol{y}_i^*) = 0$. Since $P_{A\times\Theta}$ satisfies (ii) and best responses are linear, $\boldsymbol{x}_i^* = E[\boldsymbol{y}_i^*|\boldsymbol{x}_i^*] = E[\boldsymbol{y}_i^*]$. Therefore, for every $\boldsymbol{x}_i \in \boldsymbol{X}_i$,

$$\max_{s_i \in S_i} E[u_i'(s_i(\boldsymbol{x}_i), \boldsymbol{x}_{-i}^*, \boldsymbol{\theta})] = E[u_i'(E[\boldsymbol{y}_i^*], \boldsymbol{x}_{-i}^*, \boldsymbol{\theta})] = E[u_i'(\boldsymbol{x}^*, \boldsymbol{\theta})]$$

where the first equality holds by Lemma 4 and $Var(\boldsymbol{y}_i^*) = 0$. Hence, since $C_i(\boldsymbol{x}^*) = 0$, the pair $\boldsymbol{x}_i^*$ and s_i^* is a best reply to $\boldsymbol{x}_{-i}^*$ and s_{-i}^*.

Consider now the case $Var(\boldsymbol{y}_i^*) > 0$. Since $P_{A\times\Theta}$ satisfies (ii) and best responses are linear, $\boldsymbol{x}_i^* = E[\boldsymbol{y}_i^*|\boldsymbol{x}_i^*]$. In addition, $\boldsymbol{x}_i^*$ is conditionally independent of $(\boldsymbol{x}_{-i}^*, \boldsymbol{\theta})$

given $\boldsymbol{y}_i^*$ by Lemma 1. Therefore, for every $\boldsymbol{x}_i \in \boldsymbol{X}_i$,

$$\begin{aligned}
E[u_i'(\boldsymbol{x}^*, \boldsymbol{\theta})] - C_i(\boldsymbol{x}^*) &= \max_{s_i \in S_i} E[u_i'(s_i(\boldsymbol{x}_i^*), \boldsymbol{x}_{-i}^*, \boldsymbol{\theta})] - C_i(\boldsymbol{x}^*) \\
&= E[u_i'(E[\boldsymbol{y}_i^*], \boldsymbol{x}_{-i}^*, \boldsymbol{\theta})] \\
&= \max_{s_i \in S_i} E[u_i'(s_i(\boldsymbol{x}_i), \boldsymbol{x}_{-i}^*, \boldsymbol{\theta})] - f_{(\boldsymbol{x}_{-i}^*, \boldsymbol{\theta})}\left(\frac{Var(\boldsymbol{x}_i|\boldsymbol{y}_i^*)}{Var(\boldsymbol{x}_i)}\right) \\
&\geq \max_{s_i \in S_i} E[u_i'(s_i(\boldsymbol{x}_i), \boldsymbol{x}_{-i}^*, \boldsymbol{\theta})] - C_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}^*),
\end{aligned}$$

where first equality holds since $\boldsymbol{x}_i^* = E[\boldsymbol{y}_i^*|\boldsymbol{x}_i^*]$ and second equality since $\boldsymbol{x}_i^*$ is conditionally independent of $(\boldsymbol{x}_{-i}^*, \boldsymbol{\theta})$ given $\boldsymbol{y}_i^*$. ■

Lemma 5. *Suppose that, for every $a \in A$ and $\theta \in \Theta$,*

$$u_i(a, \theta) = -(a_i - y_i)^2.$$

Given $\mu_i \geq 0$, assume also that, for every $\boldsymbol{x} \in \boldsymbol{X}$,

$$C_i(\boldsymbol{x}) = \mu_i I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}).$$

The pair $\boldsymbol{x}_i^*$ and s_i^* are a best reply to $\boldsymbol{x}_{-i}$ and affine s_{-i} if

(i) $E[\boldsymbol{y}_i|\boldsymbol{x}_i^*] = s_i^*(\boldsymbol{x}_i^*)$

(ii) $\boldsymbol{x}_i^*$ *is conditionally independent of* $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ *given* $\boldsymbol{y}_i$

(iii) $Var(\boldsymbol{y}_i|\boldsymbol{x}_i^*) \leq \mu_i$ *and equality holds if* $\rho^2_{(s_i^*(\boldsymbol{x}_i^*), \boldsymbol{y}_i)} > 0$.

Proof of Lemma 5. Deviating to any pair $\boldsymbol{x}_i$ and s_i is not profitable since

$$\begin{aligned}
-E[(s_i(\boldsymbol{x}_i) - \boldsymbol{y}_i)^2] - \mu_i I(\boldsymbol{x}_i|\boldsymbol{x}_{-i}, \boldsymbol{\theta}) &\leq -E[(s_i(\boldsymbol{x}_i) - \boldsymbol{y}_i)^2] - \mu_i I(\boldsymbol{x}_i|\boldsymbol{y}_i) \\
&\leq -Var(\boldsymbol{y}_i|\boldsymbol{x}_i) - \mu_i I(\boldsymbol{x}_i|\boldsymbol{y}_i) \\
&\leq -Var(\boldsymbol{y}_i|\boldsymbol{x}_i^*) - \mu_i I(\boldsymbol{x}_i^*|\boldsymbol{y}_i) \\
&= -E[(s_i^*(\boldsymbol{x}_i^*) - \boldsymbol{y}_i)^2] - \mu_i I(\boldsymbol{x}_i^*|\boldsymbol{y}_i) \\
&= -E[(s_i^*(\boldsymbol{x}_i^*) - \boldsymbol{y}_i)^2] - \mu_i I(\boldsymbol{x}_i^*|\boldsymbol{x}_{-i}, \boldsymbol{\theta}),
\end{aligned}$$

where first inequality holds since $\boldsymbol{y}_i$ is a statistic of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$, second inequality by linearity of best responses, third inequality by (iii), fourth equality by (i), and last equality by (ii). ■

Lemma 6. *If, for every player i, $E[\boldsymbol{y}_i|\boldsymbol{x}_i] = s_i(\boldsymbol{x}_i)$ and $\boldsymbol{x}_i$ is conditionally independent of $(\boldsymbol{x}_{-i}, \boldsymbol{\theta})$ given $\boldsymbol{y}_i$, then*

$$Cov(E[s_i(\boldsymbol{x}_i)|\boldsymbol{y}_i] - s_i(\boldsymbol{x}_i), s_j(\boldsymbol{x}_j) - E[s_j(\boldsymbol{x}_j)|\boldsymbol{y}_j]) = g_{ij}\rho^2_{(s_i(\boldsymbol{x}_i),\boldsymbol{y}_i)}\rho^2_{(s_j(\boldsymbol{x}_j),\boldsymbol{y}_j)}Var(\boldsymbol{y}_j|\boldsymbol{x}_j).$$

for all pairs of opponents i and j.

Proof of Lemma 6. The result follows from the chain of equalities

$$\begin{aligned} Cov(E[s_i(\boldsymbol{x}_i)|\boldsymbol{y}_i] - s_i(\boldsymbol{x}_i), s_j(\boldsymbol{x}_j) - E[s_j(\boldsymbol{x}_j)|\boldsymbol{y}_j]) &= \\ Cov(E[s_i(\boldsymbol{x}_i)|\boldsymbol{y}_i], s_j(\boldsymbol{x}_j) - E[s_j(\boldsymbol{x}_j)|\boldsymbol{y}_j]) &= \\ \rho^2_{(s_i(\boldsymbol{x}_i),\boldsymbol{y}_i)}Cov(\boldsymbol{y}_i, s_j(\boldsymbol{x}_j) - E[s_j(\boldsymbol{x}_j)|\boldsymbol{y}_j]) &= \\ g_{ij}\rho^2_{(s_i(\boldsymbol{x}_i),\boldsymbol{y}_i)}Cov(s_j(\boldsymbol{x}_j), s_j(\boldsymbol{x}_j) - E[s_j(\boldsymbol{x}_j)|\boldsymbol{y}_j]) &= \\ g_{ij}\rho^2_{(s_i(\boldsymbol{x}_i),\boldsymbol{y}_i)}\rho^2_{(s_j(\boldsymbol{x}_j),\boldsymbol{y}_i)}Var(\boldsymbol{y}_j|\boldsymbol{x}_j) \end{aligned}$$

where first and third equality hold since $E[s_j(\boldsymbol{x}_j)|\boldsymbol{y}_j] = E[s_j(\boldsymbol{x}_j)|\boldsymbol{x}_{-j}, \boldsymbol{\theta}]$, second equality since $E[\boldsymbol{y}_i|\boldsymbol{x}_i] = s_i(\boldsymbol{x}_i)$, and fourth equality since $E[\boldsymbol{y}_j|\boldsymbol{x}_j] = s_j(\boldsymbol{x}_j)$. ■

Proof of Theorem 1. We proceed by steps.

Step 1. (i) implies (ii).

Proof of the Step. Assume $P_{A\times\Theta}$ satisfies (i). Let $\mathcal{T}$ be a technology such that $X = A$ and Assumption 1 holds.

By Assumption 1, we can pick $\boldsymbol{x}^* \in \boldsymbol{X}$ such that $P_{A\times\Theta}$ is the distribution of $(\boldsymbol{x}^*, \boldsymbol{\theta})$. Given $\boldsymbol{y}^* = \boldsymbol{\theta} + G\boldsymbol{x}^*$, we set for cost of information

$$\mu_i = Var(\boldsymbol{y}^*_i|\boldsymbol{x}^*_i) \quad \text{and} \quad C_i(\boldsymbol{x}) = \mu_i I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta})$$

for all $i \in N$ and $\boldsymbol{x} \in \boldsymbol{X}$.

Letting s^* be the identity function, we claim that $\boldsymbol{x}^*$ and s^* are an equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$. Indeed, since $P_{A\times\Theta}$ satisfies (i), then

- $E[\boldsymbol{y}^*_i|\boldsymbol{x}^*_i] = \boldsymbol{x}^*_i$ by linearity of best responses

- $\boldsymbol{x}_i^*$ is conditionally independent of $(\boldsymbol{x}_{-i}^*, \boldsymbol{\theta})$ given $\boldsymbol{y}_i^*$ by Lemma 1.

Moreover, $\mu_i = Var(\boldsymbol{y}_i^*|\boldsymbol{x}_i^*)$ by construction. Therefore, by Lemma 5, $\boldsymbol{x}^*$ and s^* are an equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$.

To conclude, let i and j such that $g_{ij} \neq 0$ and $Var(\boldsymbol{x}_i^*)Var(\boldsymbol{x}_j^*) > 0$. Since

$$Cov(E[\boldsymbol{x}_i^*|\boldsymbol{y}_i^*] - \boldsymbol{x}_i^*, \boldsymbol{x}_j^* - E[\boldsymbol{x}_j^*|\boldsymbol{y}_j^*])$$

is symmetric in i and j, then Lemma 6 implies that

$$g_{ij}\rho^2_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)}\rho^2_{(\boldsymbol{x}_j^*,\boldsymbol{y}_j^*)}\mu_j = g_{ji}\rho^2_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)}\rho^2_{(\boldsymbol{x}_j^*,\boldsymbol{y}_j^*)}\mu_i.$$

By linearity of best responses, $Var(\boldsymbol{x}_i^*)Var(\boldsymbol{x}_j^*) > 0$ implies $\rho^2_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)}\rho^2_{(\boldsymbol{x}_j^*,\boldsymbol{y}_j^*)} > 0$, which implies $\mu_i = \mu_j$. □

Step 2. (ii) implies (i).

Proof of the Step. Assume $P_{A\times\Theta}$ satisfies (ii). Let $\mathcal{T}$ be a technology such that $X = A$ and Assumption 1 holds.

By Assumption 1, we can pick $\boldsymbol{x}^* \in \boldsymbol{X}$ such that $P_{A\times\Theta}$ is the distribution of $(\boldsymbol{x}^*, \boldsymbol{\theta})$. Given $\boldsymbol{y}^* = \boldsymbol{\theta} + G\boldsymbol{x}^*$, we set for cost of information

$$C_i(\boldsymbol{x}) = \begin{cases} Var(\boldsymbol{y}_i)\left(1 - \frac{Var(\boldsymbol{x}_i|\boldsymbol{x}_{-i},\boldsymbol{\theta})}{Var(\boldsymbol{x}_i)}\right) & \text{if } Var(\boldsymbol{y}_i^*) > \mu_i = 0, \\ I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}) & \text{if } Var(\boldsymbol{y}_i^*) = \mu_i = 0, \\ \mu_i I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}, \boldsymbol{\theta}) & \text{otherwise.} \end{cases}$$

Note that, for $\mu_i = 0$, cost of information is defined as in the proof of Lemma 2.

Letting s^* be the identity function, it is easy to check that $\boldsymbol{x}^*$ and s^* are an equilibrium of $\langle \mathcal{G}, \mathcal{T} \rangle$ since $P_{A\times\Theta}$ satisfies (ii). □

■

Lemma 7. *If $\boldsymbol{x}^*$ is an equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$, then*

$$Var(\boldsymbol{y}_i^*)(1 - \rho^2_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)}) \leq \mu_i$$

and equality holds if $\rho^2_{(\boldsymbol{x}_i^*,\boldsymbol{y}_i^*)} > 0$.

Proof of Lemma 7. By Assumption 1, for every $t \in [0, Var(\boldsymbol{y}_i^*)]$ we can pick $\boldsymbol{x}_i \in \boldsymbol{X}_i$ such that $E[\boldsymbol{y}_i|\boldsymbol{x}_i] = \boldsymbol{x}_i$, $E[\boldsymbol{x}_i|\boldsymbol{y}_i] = E[\boldsymbol{x}_i|\boldsymbol{x}_{-i}, \boldsymbol{\theta}]$, and $Var(\boldsymbol{y}_i^*|\boldsymbol{x}_i) = t$. We chose $\boldsymbol{x}_i$ so that

$$E[u_i(, \boldsymbol{x}_{-i}^*, \boldsymbol{\theta})] - \mu I(\boldsymbol{x}_i; \boldsymbol{x}_{-i}^*, \boldsymbol{\theta}) = -t - \mu \ln \frac{t}{Var(\boldsymbol{y}_i^*)}.$$

Since $\boldsymbol{x}^*$ is an equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$, then $Var(\boldsymbol{y}_i^*|\boldsymbol{x}_i^*)$ must be a maximizer of

$$-t - \mu \ln \frac{t}{Var(\boldsymbol{y}_i^*)} \quad \text{over} \quad t \in [0, Var(\boldsymbol{y}_i^*)].$$

The statement is the optimality condition of this maximization problem. ■

Proof of Proposition 1. Without loss of generality, assume $\mu > 0$. By linearity of best responses, for all players i

$$E[\boldsymbol{x}_i^*] = E[E[\boldsymbol{y}_i^*|\boldsymbol{x}_i^*]] = E[\boldsymbol{y}_i^*] = E[\boldsymbol{\theta}_i] + \sum_{j \neq i} g_{ij} E[\boldsymbol{x}_j^*].$$

Equivalently, $(\mathbb{I} - G)E[\boldsymbol{x}^*] = E[\boldsymbol{\theta}]$, that is, $E[\boldsymbol{x}^*] = (\mathbb{I} - G)^{-1} E[\boldsymbol{\theta}]$.

We are left to show that

$$E[\boldsymbol{x}^* - E[\boldsymbol{x}^*]|\boldsymbol{\theta}] = B(R_{\boldsymbol{x}^*}, G)(\boldsymbol{\theta} - E[\boldsymbol{\theta}]) \tag{1}$$

$$Var(\boldsymbol{x}^*|\boldsymbol{\theta}) = \mu B(R_{\boldsymbol{x}^*}, G). \tag{2}$$

The equalities are trivial for all $i \in N$ such that $\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} = 0$. To ease the exposition, we assume there is none of them: from now on, let $\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} > 0$ for all $i \in N$.

By linearity of best responses,

$$Cov(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*) = \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} Var(\boldsymbol{y}_i^*). \tag{3}$$

Setting $\boldsymbol{\epsilon}_i = \boldsymbol{x}_i^* - E[\boldsymbol{x}_i^*|\boldsymbol{y}_i^*]$, we get that

$$\boldsymbol{x}_i^* - E[\boldsymbol{x}_i^*] = (R_{\boldsymbol{x}^*} \circ \mathbb{I})(\boldsymbol{y}^* - E[\boldsymbol{y}^*]) + \boldsymbol{\epsilon}$$

for all players i. Equivalently

$$(\mathbb{I} - R_{\boldsymbol{x}^*} \circ G)(R_{\boldsymbol{x}^*} \circ \mathbb{I})^{-\frac{1}{2}}(\boldsymbol{x}^* - E[\boldsymbol{x}^*]) = (R_{\boldsymbol{x}^*} \circ \mathbb{I})^{\frac{1}{2}}(\boldsymbol{\theta} - E[\boldsymbol{\theta}]) + (R_{\boldsymbol{x}^*} \circ \mathbb{I})^{-\frac{1}{2}}\boldsymbol{\epsilon}.$$

Since the largest eigenvalue of G is strictly less than one, so is the largest eigenvalue of $R_{\boldsymbol{x}^*} \circ G$. Therefore the matrix $\mathbb{I} - R_{\boldsymbol{x}^*} \circ G$ is invertible and

$$\boldsymbol{x}^* - E[\boldsymbol{x}^*] = B(R_{\boldsymbol{x}^*}, G)(\boldsymbol{\theta} - E[\boldsymbol{\theta}]) + B(R_{\boldsymbol{x}^*}, G)(R_{\boldsymbol{x}^*} \circ \mathbb{I})^{-1}\boldsymbol{\epsilon}. \tag{4}$$

By Lemma 1, $\boldsymbol{\epsilon}$ is independent of $\boldsymbol{\theta}$. From (4) we conclude that (1) holds.

By linearity of best responses and Lemma 1, we can apply Lemma 6 and see that

$$Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = -g_{ij}\rho^2_{(\boldsymbol{x}^*_i,\boldsymbol{y}^*_i)}\rho^2_{(\boldsymbol{x}^*_j,\boldsymbol{y}^*_j)}Var(\boldsymbol{y}^*_j|\boldsymbol{x}^*_j)$$

for all pairs of opponents i and j. By Lemma 7, $Var(\boldsymbol{y}^*_j|\boldsymbol{x}^*_j) = \mu$. Hence

$$Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = -\mu g_{ij}\rho^2_{(\boldsymbol{x}^*_i,\boldsymbol{y}^*_i)}\rho^2_{(\boldsymbol{x}^*_j,\boldsymbol{y}^*_j)}.$$

Notice also that

$$Var(\boldsymbol{\epsilon}_i) = \rho^2_{(\boldsymbol{x}^*_i,\boldsymbol{y}^*_i)}Var(\boldsymbol{y}^*_i)(1 - \rho^2_{(\boldsymbol{x}^*_i,\boldsymbol{y}^*_i)}) = \mu\rho^2_{(\boldsymbol{x}^*_i,\boldsymbol{y}^*_i)},$$

where the first equality holds by (3), and the second equality by Lemma 7. Overall,

$$Var(\boldsymbol{\epsilon}) = \mu(R_{\boldsymbol{x}^*} \circ \mathbb{I})B(R_{\boldsymbol{x}^*}, G)^{-1}(R_{\boldsymbol{x}^*} \circ \mathbb{I}).$$

By (4), this implies that (2) holds. ■

Proof of Lemma 3. Necessity will come by differentiating V_μ with respect to ρ_i^2. To do so, notice that we can factorize $\mathbb{I} - R \circ G$ as

$$\mathbb{I} - R \circ G = \begin{bmatrix} 1 & -\rho_i(\rho_{-i} \circ g_i) \\ -\rho_i(\rho_{-i} \circ g_i) & \mathbb{I}_{-i} - R_{-i} \circ G_{-i} \end{bmatrix}.$$

Since $\mathbb{I} - R \circ G$ is positive definite, then $\mathbb{I}_{-i} - R_{-i} \circ G_{-i}$ is positive definite and

$$\frac{\det[\mathbb{I} - R \circ G]}{\det[\mathbb{I}_{-i} - R_{-i} \circ G_{-i}]} = 1 - \rho_i^2 g_i^\intercal B(R_{-i}, G_{-i})g_i > 0.$$

This immediately implies that

$$\frac{\partial}{\partial \rho_i^2} \ln \det[\mathbb{I} - R \circ G] = -\frac{g_i^\intercal B(R_{-i}, G_{-i}) g_i}{1 - \rho_i^2 g_i^\intercal B(R_{-i}, G_{-i}) g_i} \tag{5}$$

From the factorization of $\mathbb{I} - R \circ G$, we can also invert blockwise and obtain that

$$B(R, G) - \begin{bmatrix} 0 & 0 \\ 0 & B(R_{-i}, G_{-i}) \end{bmatrix} \quad \text{is equal to}$$

$$\frac{\rho_i^2}{1 - \rho_i^2 g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i}} \begin{bmatrix} 1 & g_{-i}^\intercal B(R_{-i}, G_{-i}) \\ B(R_{-i}, G_{-i}) g_{-i} & B(R_{-i}, G_{-i}) g_{-i} g_{-i}^\intercal B(R_{-i}, G_{-i}) \end{bmatrix}.$$

From here, it is easy to derive that

$$\frac{\partial}{\partial \rho_i^2} \operatorname{tr}[Var(\boldsymbol{\theta}) B(R, G)] = \frac{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i}) \boldsymbol{\theta}_{-i})}{(1 - \rho_i^2 g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i})^2}. \tag{6}$$

Necessity comes from combining (5) and (6).

For sufficiency, observe that a best reply always exists. Indeed, since $V_\mu(\rho_i^2, \rho_{-i}^2) \to -\infty$ as $\rho_i^2 \to 1$, there exists $t \in [0, 1)$ such that

$$\arg \max_{\rho_i^2 \in [0,1)} V_\mu(\rho_i^2, \rho_{-i}^2) = \arg \max_{\rho_i^2 \in [0,t]} V_\mu(\rho_i^2, \rho_{-i}^2) \neq \varnothing.$$

As a result, if we prove that the first-order condition is satisfied by a unique ρ_i^2, we can conclude that the condition is not only necessary, but also sufficient.

We consider two cases. Assume first that

$$Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i}) \boldsymbol{\theta}_{-i}) > \mu(1 - g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i}). \tag{7}$$

Then $\rho_i^2 = 0$ does not satisfy the first-order condition; the best reply must be

$$\rho_i^2 = \frac{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i}) \boldsymbol{\theta}_{-i}) - \mu(1 - g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i})}{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i}) \boldsymbol{\theta}_{-i}) - \mu g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i} (1 - g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i})}.$$

On the other hand, assume (7) does not hold. Since $g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i} < 1$, then

$$\mu \frac{1 - g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i}}{1 - \rho_i^2} < \frac{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i}) \boldsymbol{\theta}_{-i})}{1 - \rho_i^2 g_{-i}^\intercal B(R_{-i}, G_{-i}) g_{-i}}$$

for all $\rho_i^2 > 0$; the best reply must be $\rho_i^2 = 0$. ■

Proof of Proposition 2. First we show that (i) implies (ii). Let $\boldsymbol{x}^*$ be an equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$. By Lemma 7,

$$Var(\boldsymbol{y}_i^*)(1 - \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)}) \leq \mu_i$$

and equality holds if $\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} > 0$. Moreover, by Proposition 1

$$Var(\boldsymbol{y}_i^*) = \frac{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{\boldsymbol{x}^*_{-i}}, G_{-i})\boldsymbol{\theta}_{-i})}{(1 - \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} g_i^\intercal B(R_{\boldsymbol{x}^*_{-i}}, G_{-i}) g_i)^2} + \mu \frac{g_i^\intercal B(R_{\boldsymbol{x}^*_{-i}}, G_{-i}) g_i}{1 - \rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)} g_i^\intercal B(R_{\boldsymbol{x}^*_{-i}}, G_{-i}) g_i}.$$

Hence, by Lemma 3, $\rho^2_{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)}$ is a best reply to $(\rho^2_{(\boldsymbol{x}_j^*, \boldsymbol{y}_j^*)} : j \neq i)$ in the auxiliary game. This implies that (ii) holds.

Now we show that (ii) implies (i). Let ρ^2 be an equilibrium of the auxiliary game. By Assumption 1, we can choose $\boldsymbol{x}^* \in \boldsymbol{X}$ such that

- $E[\boldsymbol{x}^* | \boldsymbol{\theta}] = B(G)E[\boldsymbol{\theta}] + B(R, G)(\boldsymbol{\theta} - E[\boldsymbol{\theta}])$
- $Var(\boldsymbol{x}^* | \boldsymbol{\theta}) = \mu B(R, G)$.

It can be checked that $R_{\boldsymbol{x}^*} = R$ and the conditions of Lemma 5 are met. Hence $\boldsymbol{x}^*$ is an equilibrium of $\langle \mathcal{G}, \mathcal{T}_\mu \rangle$ and (i) holds. ■

Lemma 8. *Let H be a symmetric, invertible matrix whose largest eigenvalue belongs to $(0, 1)$. Then the function $\rho^2 \mapsto B(R, H)$ is increasing and convex.*[10]

[10]Monotonicity and convexity are defined with respect to the Loewner order. More precisely, let $\rho^2, \tilde{\rho}^2 \in [0, 1)^n$. Monotonicity entails that, if $\rho_i^2 \geq \tilde{\rho}_i^2$ for all players i, then the matrix

$$\mathbb{B}(R, H) - \mathbb{B}(\tilde{R}, H)$$

is positive semidefinite. For $\alpha \in [0, 1]$, define the matrix R_α such that

$$R_\alpha = [(\alpha\rho_i^2 + (1 - \alpha)\tilde{\rho}_i^2)(\alpha\rho_j^2 + (1 - \alpha)\tilde{\rho}_j^2) : i, j \in N].$$

Convexity entails that the matrix

$$\alpha\mathbb{B}(R, H) + (1 - \alpha)\mathbb{B}(\tilde{R}, H) - \mathbb{B}(R_\alpha, H)$$

is positive semidefinite.

Proof of Lemma 8. By continuity, it is sufficient to prove the result for the restriction to $(0,1)^n$, that is, it is enough to show that the function

$$(0,1)^n \in \rho^2 \mapsto ((R \circ \mathbb{I})^{-1} - H)^{-1}$$

is increasing and convex. Moreover, since the function $\rho^2 \mapsto R \circ \mathbb{I}$ is increasing and affine, we can focus on the map $R \circ \mathbb{I} \mapsto ((R \circ \mathbb{I})^{-1} - H)^{-1}$.

Monotonicity is clear from the decomposition into decreasing functions

$$(R \circ \mathbb{I}) \mapsto (R \circ \mathbb{I})^{-1} \mapsto ((R \circ \mathbb{I})^{-1} - H)^{-1}.$$

For convexity, key is the easily checked identity

$$((R \circ \mathbb{I})^{-1} - H)^{-1} = (R \circ \mathbb{I}) + (R \circ \mathbb{I})(H^{-1} - R \circ \mathbb{I})^{-1}(R \circ \mathbb{I}).$$

From Ando (1979, Theorem 1) we learn that the function

$$(R \circ \mathbb{I}, H^{-1} - R \circ \mathbb{I}) \mapsto (R \circ \mathbb{I})(H^{-1} - R \circ \mathbb{I})^{-1}(R \circ \mathbb{I})$$

is jointly convex. The desired convexity follows. ■

Corollary 3. *The function $\rho^2 \mapsto B(R, G)$ is increasing and convex.*[11]

Proof of Corollary 3. Let Λ be a diagonal matrix and Q a unitary matrix such that $G = Q\Lambda Q^\intercal$. For every $\epsilon > 0$, define the symmetric matrix

$$H_\epsilon = Q(\Lambda + \epsilon \mathbb{I})Q^\intercal.$$

If ϵ is sufficiently low, the matrix H_ϵ is invertible and the largest eigenvalue of H_ϵ belongs to $(0,1)$. Since $\|G - H_\epsilon\| = \epsilon$, by continuity

$$\lim_{\epsilon \to 0} B(R, H_\epsilon) = B(R, G).$$

The desired result follows from Lemma 8. ■

Proof of Proposition 3. The function that associates to a positive definite matrix

[11] See footnote 10 for the precise meaning of "increasing and convex."

H the real number

$$\mathrm{tr}\left[Var(\boldsymbol{\theta})^{\frac{1}{2}} H Var(\boldsymbol{\theta})^{\frac{1}{2}}\right]$$

is increasing (with respect to the Loewner order) and affine. By Corollary 3, the function $\rho^2 \mapsto B(R, G)$ is increasing and convex. Hence, overall, the function

$$\rho^2 \mapsto \mathrm{tr}[Var(\boldsymbol{\theta}) B(R, G)]$$

is increasing and convex.

The rest of the proof is based on the easily checked identity

$$\frac{\det[\mathbb{I} - R \circ G]}{\det[\mathbb{I} - R \circ \mathbb{I}]} = \det[\mathbb{I} - (\mathbb{I} - G)^{\frac{1}{2}} B(R, G)(\mathbb{I} - G)^{\frac{1}{2}}].$$

By Corollary 3, the function $\rho^2 \mapsto B(R, G)$ is increasing and convex. Hence, the function

$$\rho^2 \mapsto \mathbb{I} - (\mathbb{I} - G)^{\frac{1}{2}} B(R, G)(\mathbb{I} - G)^{\frac{1}{2}}$$

is decreasing and concave. The function that associates to a positive definite matrix H the real number $-\log\det H$ is decreasing and convex. Hence, overall, the function

$$\rho^2 \mapsto \ln \frac{\det[\mathbb{I} - R \circ G]}{\det[\mathbb{I} - R \circ \mathbb{I}]}$$

is increasing and convex. ■

Proof of Proposition 4. Without loss of generality, assume $\|G\| < 1$. By Lemma 3, $BR_i(\rho^2_{-i})$ is equal to

$$\frac{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) - \mu(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)}{Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) - \mu g_i^\intercal B(R_{-i}, G_{-i})g_i(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)}$$

whenever the following inequality holds:

$$Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) \geq \mu(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i). \tag{8}$$

Otherwise, $BR_i(\rho^2_{-i}) = 0$. The proof proceeds by steps.

Step 1. For every ρ^2_{-i} and $\tilde{\rho}^2_{-i}$ in $[0,1)^{n-1}$,

$$\|B(R_{-i},G_{-i}) - B(\tilde{R}_{-i},G_{-i})\| \leq \frac{\max_{j\neq i}|\rho_j^2 - \tilde{\rho}_j^2|}{(1-\|G\|)^2}.$$

Proof of the Step. By continuity, it is enough to show the result for ρ^2_{-i} and $\tilde{\rho}^2_{-i}$ in $(0,1)^{n-1}$. The key is the easily checked identity

$$B(R_{-i},G_{-i}) - B(\tilde{R}_{-i},G_{-i}) = (\mathbb{I}-(\tilde{R}_{-i}\circ\mathbb{I})G_{-i})^{-1}((R_{-i}\circ\mathbb{I})-(\tilde{R}_{-i}\circ\mathbb{I}))(\mathbb{I}-G_{-i}(R_{-i}\circ\mathbb{I}))^{-1}.$$

Since $\|G\| < 1$ and $\|G\| \geq \|G_{-i}\|$, then $\|G_{-i}\| < 1$. Hence

$$\|(\mathbb{I} - (R_{-i}\circ\mathbb{I})G_{-i})^{-1}\| \leq \frac{1}{1-\|(R_{-i}\circ\mathbb{I})G_{-i}\|} \leq \frac{1}{1-\|R_{-i}\circ\mathbb{I}\|\|G_{-i}\|} \leq \frac{1}{1-\|G\|},$$

where first inequality holds since $\|G_{-i}\| < 1$, second inequality since operator norm is sub-multiplicative, and third inequality since $\|R_{-i}\circ\mathbb{I}\| \leq 1$ and $\|G_{-i}\| \leq \|G\|$. Clearly the same inequalities also hold if R_{-i} is replaced by $\tilde{R}_{-i}$.

Overall, we conclude that

$$\|B(R_{-i},G_{-i}) - B(\tilde{R}_{-i},G_{-i})\| \leq \frac{\|(R_{-i}\circ\mathbb{I}) - (\tilde{R}_{-i}\circ\mathbb{I})\|}{(1-\|G\|)^2} \leq \frac{\max_{j\neq i}|\rho_j^2 - \tilde{\rho}_j^2|}{(1-\|G\|)^2},$$

where we use the identity $\|(R_{-i}\circ\mathbb{I}) - (\tilde{R}_{-i}\circ\mathbb{I})\| = \max_{j\neq i}|\rho_j^2 - \tilde{\rho}_j^2|$ and the fact that operator norm is sub-multiplicative. □

Step 2. For every ρ^2_{-i} in $[0,1)^{n-1}$, $\|B(R_{-i},G_{-i})\| \leq (1-\|G\|)^{-1}$.

Proof of the Step. Since $B(R_{-i},G_{-i})$ is increasing in ρ^2_{-i} (Corollary 3), then

$$\|B(R_{-i},G_{-i})\| \leq \|B(G_{-i})\| = \|(\mathbb{I}_{-i} - G_{-i})^{-1}\|.$$

Since $\|G_{-i}\| \leq \|G\| < 1$, then $\|(\mathbb{I}_{-i} - G_{-i})^{-1}\| \leq (1-\|G_{-i}\|)^{-1} \leq (1-\|G\|)^{-1}$. □

Step 3. For every ρ^2_{-i} in $[0,1)^{n-1}$, $1 - g_i^\intercal B(R_{-i},G_{-i})g_i \geq 2^{-n}(1-\|G\|)^n$.

Proof of the Step. Since $\mathbb{B}(R_{-i},G_{-i})$ is increasing in ρ^2_{-i} (Corollary 3), then

$$1 - g_i^\intercal B(R_{-i},G_{-i})g_i \geq 1 - g_i^\intercal B(G_{-i})g_i = 1 - g_i^\intercal(\mathbb{I}-G_{-i})^{-1}g_i.$$

To bound $1 - g_i^\intercal(\mathbb{I} - G_{-i})^{-1}g_i$, we use the identity

$$1 - g_i^\intercal(\mathbb{I} - G_{-i})^{-1}g_i = \frac{\det[\mathbb{I} - G]}{\det[\mathbb{I}_{-i} - G_{-i}]}.$$

We obtain the desired conclusion from

$$\begin{aligned}\det[\mathbb{I} - G] &\geq (1 - \lambda_{max}(G))^n \geq (1 - \|G\|)^n, \\ \det[\mathbb{I} - G_{-i}] &\leq (1 - \lambda_{min}(G_{-i}))^{n-1} \leq (1 + \|G\|)^{n-1} \leq (1 + \|G\|)^n,\end{aligned}$$

and $\|G\| < 1$. □

Step 4. For every ρ_{-i}^2 in $[0, 1)^{n-1}$,

$$Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) \leq \frac{2\|Var(\boldsymbol{\theta})\|}{(1 - \|G\|)^2}.$$

Proof of the Step. From the identity

$$Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) = \begin{pmatrix} 1 \\ B(R_{-i}, G_{-i})g_i \end{pmatrix}^\intercal Var(\boldsymbol{\theta}) \begin{pmatrix} 1 \\ B(R_{-i}, G_{-i})g_i \end{pmatrix},$$

we deduce that

$$Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) \leq \|Var(\boldsymbol{\theta})\|(1 + g_i^\intercal B(R_{-i}, G_{-i})^2 g_i)$$

by Cauchy-Schwarz inequality and definition of operator norm. Moreover

$$g_i^\intercal B(R_{-i}, G_{-i})^2 g_i \leq \|B(R_{-i}, G_{-i})g_i\|^2 \leq \|B(R_{-i}, G_{-i})\|^2\|g_i\|^2 \leq \frac{\|G\|^2}{(1 - \|G\|)^2},$$

where first inequality holds by Cauchy-Schwarz inequality, second inequality by definition of operator norm, and last inequality by Step 3 and $\|g_i\| \leq \|G\|$. Finally

$$1 + \frac{\|G\|^2}{(1 - \|G\|)^2} \leq \frac{2}{(1 - \|G\|)^2}$$

since $\|G\| < 1$. □

Step 5. For every ρ^2_{-i} and $\tilde{\rho}^2_{-i}$ in $[0,1)^{n-1}$,

$$|Var(\boldsymbol{\theta}_i + g_i^\intercal B(\tilde{R}_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) - Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i})| \tag{9}$$

is lower or equal than

$$\frac{3\|G\|\|Var(\boldsymbol{\theta})\|}{(1-\|G\|)^3}\max_{j\neq i}|\rho_j^2 - \tilde{\rho}_j^2|.$$

Proof of the Step. By the triangle inequality, (9) is lower or equal than

$$|g_i^\intercal(B(\tilde{R}_{-i}, G_{-i}) - B(R_{-i}, G_{-i}))Cov(\boldsymbol{\theta}_{-i}, \boldsymbol{\theta}_i)| \tag{10}$$
$$+\, |g_i^\intercal B(R_{-i}, G_{-i})Var(\boldsymbol{\theta}_{-i})(B(R_{-i}, G_{-i}) - B(\tilde{R}_{-i}, G_{-i}))g_i| \tag{11}$$
$$+\, |g_i^\intercal(B(R_{-i}, G_{-i}) - B(\tilde{R}_{-i}, G_{-i}))Var(\boldsymbol{\theta}_{-i})B(\tilde{R}_{-i}, G_{-i})g_i|. \tag{12}$$

Now we bound separately the three terms of the sum.

First, observe that (10) is lower or equal than

$$\|B(\tilde{R}_{-i}, G_{-i}) - B(R_{-i}, G_{-i})\|\|g_i\|\|Cov(\boldsymbol{\theta}_{-i}, \boldsymbol{\theta}_i)\| \leq \frac{\|G\|\|Var(\boldsymbol{\theta})\|}{(1-\|G\|)^2}\max_{j\neq i}|\rho_j^2 - \tilde{\rho}_j^2|,$$

where first inequality uses Cauchy-Schwarz inequality and definition of operator norm, while second inequality holds by $\|g_i\| \leq \|G\|$, $\|Cov(\boldsymbol{\theta}_{-i}, \boldsymbol{\theta}_i)\| \leq \|Var(\boldsymbol{\theta})\|$, and Step 1.

Next, notice that (11) is lower or equal than

$$\begin{aligned}\|B(R_{-i}, G_{-i})Var(\boldsymbol{\theta}_{-i})(B(R_{-i}, G_{-i}) - B(\tilde{R}_{-i}, G_{-i}))\|\|g_i\|^2 \leq \\ \|B(R_{-i}, G_{-i})\|\|Var(\boldsymbol{\theta}_{-i})\|\|B(R_{-i}, G_{-i}) - B(\tilde{R}_{-i}, G_{-i})\|\|g_i\|^2 \leq \\ \frac{\|G\|^2\|Var(\boldsymbol{\theta})\|}{(1-\|G\|)^3}\max_{j\neq i}|\rho_j^2 - \tilde{\rho}_j^2|\end{aligned}$$

where first inequality uses Cauchy-Schwarz inequality and definition of operator norm, second inequality holds since operator norm is sub-multiplicative, and third inequality holds by $\|g_i\| \leq \|G\|$, $\|Var(\boldsymbol{\theta}_{-i})\| \leq \|Var(\boldsymbol{\theta})\|$, Step 2, and Step 1.

It is clear that the bound derived for (11) is valid also for (12). We obtain the desired inequality by combining the bounds derived for (10)–(12) with $\|G\| < 1$. □

Step 6. For every ρ^2_{-i} and $\tilde{\rho}^2_{-i}$ in $[0,1)^{n-1}$ such that (8) holds,

$$|BR_i(\rho^2_{-i}) - BR_i(\tilde{\rho}^2_{-i})| \leq \frac{\|G\|\|Var(\boldsymbol{\theta})\|}{\mu}\left(\frac{2}{1-\|G\|}\right)^{4(n+1)} \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|.$$

Proof of the Step. By (8) and Step 3,

$$Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) - \mu g_i^\intercal B(R_{-i}, G_{-i})g_i(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)$$

is at least as large as $\mu\left(\frac{1-\|G\|}{2}\right)^{2n}$. Clearly the same inequality is true if we replace R_{-i} with $\tilde{R}_{-i}$. Hence, by the triangle inequality, $\mu\left(\frac{1-\|G\|}{2}\right)^{4n}|BR_i(\rho^2_{-i}) - BR_i(\tilde{\rho}^2_{-i})|$ is lower or equal than

$$\begin{aligned}&(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)^2|Var(\boldsymbol{\theta}_i + g_i^\intercal B(\tilde{R}_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) - Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}| \\ &+ Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i})|(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)^2 - (1 - g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i)^2| \\ &+ \mu(1 - g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i)(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)|g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i - g_i^\intercal B(R_{-i}, G_{-i})g_i|.\end{aligned}$$

Now we bound separately the three terms of this sum.

First term. Since $g_i^\intercal B(R_{-i}, G_{-i})g_i$ is between zero and one, then by Step 5

$$(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)^2|Var(\boldsymbol{\theta}_i + g_i^\intercal B(\tilde{R}_{-i}, G_{-i})\boldsymbol{\theta}_{-i}) - Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i}|$$

is lower or equal than

$$\frac{3\|G\|\|Var(\boldsymbol{\theta}\frac{3\|G\|\|Var(\boldsymbol{\theta})\|}{(1-\|G\|)^3})\|}{(1-\|G\|)^3} \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|. \tag{13}$$

Second term. By convexity of the square function,

$$|(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)^2 - (1 - g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i)^2|$$

is at most

$$2|g_i^\intercal(B(R_{-i}, G_{-i}) - B(\tilde{R}_{-i}, G_{-i}))g_i|.$$

Moreover, $|g_i^\intercal(B(R_{-i}, G_{-i}) - B(\tilde{R}_{-i}, G_{-i}))g_i|$ is lower or equal than

$$\|B(R_{-i}, G_{-i}) - B(R_{-i}, G_{-i}))g_i\|\|g_i\| \leq \frac{\|G\|^2}{(1-\|G\|)^2} \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|$$

where first inequality uses Cauchy-Schwarz inequality and definition of operator norm, while second inequality follows from $\|g_i\| \leq \|G\|$ and Step 1. Hence, by Step 4,

$$Var(\boldsymbol{\theta}_i + g_i^\intercal B(R_{-i}, G_{-i})\boldsymbol{\theta}_{-i})|(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)^2 - (1 - g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i)^2|$$

is lower or equal tha**n**

$$\frac{2\|Var(\boldsymbol{\theta})\|\|G\|^2}{(1-\|G\|)^4} \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|. \tag{14}$$

Third term. Bounding the second term, we have already showed that

$$|g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i - g_i^\intercal B(R_{-i}, G_{-i})g_i| \leq \frac{\|G\|^2}{(1-\|G\|)^2} \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|.$$

By $g_i^\intercal B(R_{-i}, G_{-i})g_i \in (0,1)$, (8), and Step 4, we deduce that

$$\mu(1 - g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i)(1 - g_i^\intercal B(R_{-i}, G_{-i})g_i)|g_i^\intercal B(\tilde{R}_{-i}, G_{-i})g_i - g_i^\intercal B(R_{-i}, G_{-i})g_i|$$

is lower or equal than

$$\frac{2\|Var(\boldsymbol{\theta})\|\|G\|^2}{(1-\|G\|)^4} \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|. \tag{15}$$

Conclusion. From (13)–(15), we deduce that $|BR_i(\rho_{-i}^2) - BR_i(\tilde{\rho}_{-i}^2)|$ is at most

$$\left(\frac{2}{1-\|G\|}\right)^{4n} \left(\frac{3\|G\|\|Var(\boldsymbol{\theta})\|}{\mu(1-\|G\|)^3} + \frac{2\|Var(\boldsymbol{\theta})\|\|G\|^2}{\mu(1-\|G\|)^4} + \frac{2\|Var(\boldsymbol{\theta})\|\|G\|^2}{\mu(1-\|G\|)^4}\right) \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|.$$

The desired conclusion follows from simple majorizations. □

Step 7. For every ρ_{-i}^2 and $\tilde{\rho}_{-i}^2$ in $[0,1)^{n-1}$,

$$|BR_i(\rho_{-i}^2) - BR_i(\tilde{\rho}_{-i}^2)| \leq \frac{\|G\|\|Var(\boldsymbol{\theta})\|2^{4(n+1)}}{\mu(1-\|G\|)^{4(n+1)}} \max_{j\neq i} |\rho_j^2 - \tilde{\rho}_j^2|.$$

Proof of the Step. The case where (8) does not hold neither for ρ_{-i}^2 nor $\tilde{\rho}_{-i}^2$ is trivial:

indeed, if so, then $BR_i(\rho^2_{-i}) = BR_i(\tilde{\rho}^2_{-i}) = 0$. The case where (8) holds for both ρ^2 and $\tilde{\rho}^2$ is covered by Step 6.

For the rest of the proof, assume (8) holds for ρ^2_{-i} but not for $\tilde{\rho}^2_{-i}$. By continuity, there is $t \in [0,1]$ such that (8) holds with equality for $\tilde{\tilde{\rho}}^2_{-i} := t\rho^2_{-i} + (1-t)\tilde{\rho}^2_{-i}$. Then

$$|BR_i(\rho^2_{-i}) - BR_i(\tilde{\rho}^2_{-i})| = |BR_i(\rho^2_{-i}) - BR_i(\tilde{\tilde{\rho}}^2_{-i})| \leq \frac{\|G\|\|Var(\boldsymbol{\theta})\|2^{4(n+1)}}{\mu(1-\|G\|)^{4(n+1)}} \max_{j\neq i} |\rho^2_j - \tilde{\tilde{\rho}}^2_j|$$
$$\leq \frac{\|G\|\|Var(\boldsymbol{\theta})\|2^{4(n+1)}}{\mu(1-\|G\|)^{4(n+1)}} \max_{j\neq i} |\rho^2_j - \tilde{\rho}^2_j|$$

where first equality holds since $BR_i(\tilde{\rho}^2_{-i}) = Br_i(\tilde{\tilde{\rho}}^2_{-i}) = 0$, second inequality follows from Step 6, and third inequality holds since $\tilde{\tilde{\rho}}^2_{-i}$ is a convex combination of ρ^2_{-i} and $\tilde{\rho}^2_{-i}$. □

■

References

Amir, R. and N. Lazzati (2016): "Endogenous Information Acquisition in Bayesian Games with Startegic Complementarities." *Journal of Economic Theory*, vol. 163, pp. 684–698. (Cited on p. 23.)

Ando, T. (1979): "Concavity of Certain Maps on Positive Definite Matrices and Applications to Hadamard Products." *Linear Algebra and Its Applications*, vol. 26, pp. 203–241. (Cited on p. 34.)

Angeletos, G.-M. and A. Pavan (2007): "Efficient Use of Information and Social Value of Information." *Econometrica*, vol. 75, pp. 1103–1142. (Cited on p. 3.)

Ballester, C., A. Calvo-Armengol, and Y. Zenou (2006): "Who's Who in Netwoks. Wanted: The Key Player." *Econometrica*, vol. 74, pp. 1403–1417. (Cited on pp. 3 and 16.)

Banerjee, A., X. Guo, and H. Wang (2005): "On the Optimality of Conditional Expectation as a Bregman Predictor." *IEEE Transactions on Information Theory*, vol. 51, pp. 2664–2669. (Cited on p. 5.)

Blackwell, D. (1951): "Comparison of Experiments." In "Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability," pp. 93–102. University of California Press. (Cited on pp. 3 and 10.)

Bogachev, V. I. (1998): *Gaussian Measures*. American Mathematical Society. (Cited on p. 6.)

Bonacich, P. (1987): "Power and Centrality: A Family of Measures." *American Journal of Sociology*, vol. 92, pp. 1170–1182. (Cited on pp. 2 and 16.)

Bramoulle, Y. and R. Kranton (2016): "Games Played on Networks." In Y. Bramoulleé, A. Galeotti, and B. Rogers, eds., "The Oxford Handbook of the Economics of Networks," Oxford University Press. (Cited on p. 3.)

Bramoulle, Y., R. Kranton, and M. D'Amours (2014): "Startegic Interaction and Networks." *The American Economic Review*, vol. 104, pp. 898–930. (Cited on p. 3.)

Calvo-Armengol, A., J. de Marti, and A. Prat (2015): "Communication and Influence." *Theoretical Economics*, vol. 10, pp. 649–690. (Cited on p. 4.)

Colombo, L., G. Femminis, and A. Pavan (2014): "Information Acquisition and Welfare." *The Review of Economic Studies*, vol. 81, pp. 1438–1483. (Cited on p. 3.)

Cover, T. M. and J. A. Thomas (2006): *Elements of Information Theory*. 2nd edn. Wiley-Interscience. (Cited on p. 11.)

Denti, T. (2016): "Unrestricted Information Acquisition." (Cited on pp. 3, 8, and 19.)

Hellwig, C. and L. Veldkamp (2009): "Knowing What Others Know: Coordination Motives in Information Acquisition." *The Review of Economic Studies*, vol. 76, pp. 223–251. (Cited on pp. 3 and 23.)

Herskovic, B. and J. Ramos (2015): "Acquiring Information Through Peers." (Cited on p. 4.)

Morris, S. and H. S. Shin (2002): "Social Value of Public Information." *The American Economic Review*, vol. 92, pp. 1521–1534. (Cited on p. 3.)

Morris, S. and T. Ui (2004): "Best Response Equivalence." *Games and Economic Behavior*, vol. 49, pp. 260–287. (Cited on p. 13.)

Myatt, D. P. and C. Wallace (2012): "Endogenous Information Acquisition in Coordination Games." *The Review of Economic Studies*, vol. 79, pp. 340–374. (Cited on p. 3.)

Radner, R. (1962): "Team Decision Problems." *The Annals of Mathematical Statistics*, vol. 33, pp. 857–881. (Cited on pp. 3 and 8.)

Radner, R. and J. Stiglitz (1989): "A Nonconcavity in the Value of Information." In M. Boyer and R. Kihlstrom, eds., "Bayesian models in economic theory," pp. 33–52. Elsevier. (Cited on p. 22.)

Shannon, C. E. (1948): "A Mathematical Theory of Communication." *Bell System Technical Journal*, vol. 27, pp. 379–423. (Cited on pp. 2 and 9.)

Sims, C. A. (2003): "Implications of Rational Inattention." *Journal of Monetary Economics*, vol. 50, pp. 665–690. (Cited on pp. 2 and 9.)

Sims, C. A. (2006): "Rational Inattention: Beyond the Linear-Quadratic Case." *The American Economic Review*, vol. 6, pp. 158–163. (Cited on p. 15.)

Vives, X. (1984): "Duopoloy Information Equilibrium: Cournot and Bertrand." *Journal of Economic Theory*, vol. 34, pp. 71–94. (Cited on p. 3.)

Yang, M. (2015): "Coordination with Flexible Information Acquisition." *Journal of Economic Theory*, vol. 158, pp. 721–738. (Cited on p. 23.)