Searching for Information*

Jungsuk Han[†] and Francesco Sangiorgi[‡]

October 30, 2015

Abstract

This paper provides a search-based information acquisition framework using an urn model with an asymptotic approach. The underlying intuition of the model is simple: when the scope of information search is more limited, marginal search efforts produce less useful information due to redundancy, but commonality of information among different agents increases. Consequently, limited information searchability induces a trade-off between an information source's precision and its commonality. In a "beauty contest" game with endogenous information acquisition, this precision-commonality trade-off generates non-fundamental volatility through the channel of information acquisition.

JEL Classification Codes: C65, D80, D81, D83

Keywords: information acquisition, information search, commonality of information, urn model, coordination games

*We thank Tomas Björk, James Dow, Simon Gervais, Johan Walden and Victoria Vanasco for helpful comments and suggestions. We also thank seminar participants at Stockholm School of Economics, Sveriges Riksbank, Arne Ryde Workshop in Financial Economics, Barcelona GSE Summer Forum Workshop on Information Frictions and Learning, and 11th World Congress of the Econometric Society.

[†]Corresponding author. Stockholm School of Economics, Sveavägen 65, Stockholm, Sweden, SE 113 83, E-mail: Jungsuk.Han@hhs.se.

[‡]Stockholm School of Economics, Sveavägen 65, Stockholm, Sweden, SE 113 83, E-mail: Francesco.Sangiorgi@hhs.se

1 Introduction

It is well known that equilibrium outcomes depend on what information agents act upon, and that this information itself is (at least in part) also the result of a choice. Less is known, however, about *how* agents come to learn their information and whether this process influences the type of information that agents choose to observe. In our paper, we address the following questions: How do agents learn when relevant information has to be searched for? How does an agent's search effort translate into the information he learns? How similar is the information that agents learn through search? We provide a microfounded information acquisition technology that answers these questions. By means of applications, we illustrate its usefulness and tractability in applied theory modeling.

Our framework is based on the following simple intuition: the more an agent searches for new information, the more likely he is to encounter information that overlaps–and that is therefore redundant–with what he has already discovered from past searching activities. This increasing redundancy results in diminishing returns to scale in information search, or concavity in an agent's learning curve. This concavity is more pronounced when the total amount of potentially observable information (henceforth, "searchable information") is more limited. Similarly, multiple agents who search from the same source of information will also face increasing redundancy as more information is collected. Thus, the information they learn becomes increasingly more similar. Such commonality of acquired information is more pronounced when the searchable information is more limited.

We formalize the aforementioned intuition by employing an urn model with an asymptotic approach. Consider drawing balls with replacement from an urn containing a finite number of balls. Drawing a ball is interpreted as collecting a costly signal through search. Because the collected balls are replaced into the urn, the odds of drawing a previously collected ball increase each time an agent draws a ball from the urn. In the context of information acquisition, drawing a previously collected ball means collecting a redundant (thus uninformative) signal. On average, each additional draw provides less information due to redundancy. Hence, the expected overall informativeness is concave in the number of collected signals; such concavity is more pronounced the more limited is the searchable information (or the smaller is the urn).

In spite of these intuitive properties of the expected precision, the actual informativeness of the acquired information is uncertain; it inherits the randomness in the number of non-redundant signals. This randomness will impair tractability in economic applications and is therefore undesirable. To resolve this shortcoming, we consider the limiting case in which each signal becomes infinitesimally small. This asymptotic approach allows us to obtain a smooth and deterministic mapping from the search effort (i.e., the inputs of the resources used in the information search) into the precision of the information that is collected by an agent (Theorem 2.1). The curvature of this asymptotic precision function decreases with the amount of searchable information.

We also formalize the case with multiple agents using the same logic. Imagine that multiple agents are independently drawing balls from a single urn that is interpreted as a shared source of information. Because the number of balls in the urn is finite, they tend to gather a more similar set of balls as they increase the number of drawings from the urn. Furthermore, this tendency is more pronounced when the number of balls in the urn is smaller. In the context of information acquisition, the shared component of agents' collected signals grows larger as more information is acquired by each agent. Consequently, the correlation of their acquired information increases in the information acquisition, and the correlation is more pronounced when the information searchability is lower (Theorem 2.2). The main result in our paper is Theorem 2.3, which derives the joint distribution of the agents' information acquired through search.

Our framework provides a microfoundation for a commonly used setup in which private signals are imperfectly correlated among agents. In our model, the correlation of the agents' signals is determined by the interaction between the agents' endogenous choices of effort and the exogenous searchability of information. When the number of agents is large, we can further show that an agent's acquired information, S_i, can be decomposed as

$$S_i = \theta + \mu + \eta_i, \tag{1}$$

where θ is the variable of interest, μ is an error term common to all agents, and η_i is an idiosyncratic error term; all three components on the right hand side of Eq. (1) are independent of each other (see Corollary 2.1). By investing more resources in information search, an agent can influence the distribution of η_i but not that of μ, which is endogenous to the information search process and related to the searchability of information.

A distinct feature of the information structure in Eq. (1) that emerges from our microfoundation is the trade-off between an information source's precision and its publicity. The idea is that the smaller is the information content that is available from an information source, the more common is the information of the agents who search from it. In equilibrium models in which the commonality of information matters–such as beauty contest games, global games and speculative trading models–this trade-off between precision and publicity can have a significant impact on the incentives to acquire information and the resulting outcomes.

As an application, we study a "beauty contest" coordination game with endogenous information acquisition. Our setup follows the standard two-period setup in the literature: Agents' final payoffs depend on the quadratic distance of actions from an unobserved fundamental value and the average action. To acquire information prior to taking actions, agents can allocate their efforts (or resources) among different information sources to maximize their ex ante utility. When agents care about forecasting other agents' information, we show how information search leads to non-concavities that can result in multiple equilibria. Our findings contribute to the discussion of equilibrium determinacy in coordination games with endogenous information (see Section 3.4).

We then specialize to a setup with only two information sources. One of them is superior to the other in the sense that it offers more precise information about the fundamental given the same level of inputs. On the other hand, because of its lower searchability, the inferior source provides information that is more correlated among agents. Therefore, if other agents are learning from this source, it gives more precise information on what the other agents will do. When the coordination motive is sufficiently strong, there exists an equilibrium in which all agents choose to focus on the inferior information source. Because less searchable information leads to more covariance, this equilibrium outcome becomes more "likely" (i.e., it exists on a larger set of parameters) precisely when the inferior information source becomes more inefficient. This outcome may not be socially optimal because agents are acting based on information from a less efficient source. For instance, the inferior information equilibrium is associated with an average action that is more volatile and less correlated with the fundamental. Our results can be applied to situations with strategic complementarity, such as bank runs, analysts' herding behavior, etc. For example, agents may decide to run on a healthy bank based on less accurate information (e.g., rumors) instead of investigating more accurate sources of information, because information from a less accurate source is more likely to be correlated due to imperfect information searchability.

There have been various approaches to modeling information choices put forward in the literature, such as rational inattention (e.g., Sims (2003)), costly information (e.g., Grossman and Stiglitz (1980)) and markets for information (e.g., Admati and Pfleiderer (1986)).[1] Each approach has different advantages that may be useful in different situations. For example, the approach of rational inattention quantifies the amount of collected or processed information based on entropy theory. In his seminal paper, Sims (2003) connected information theory to agents' utility maximization problem using entropy as the measure of information. Due to its elegance and practical usefulness, there have been numerous applications in macroeconomics (e.g., Woodford (2009); Mackowiak and Wiederholt (2009)) and finance (e.g., Peng (2005); Peng and Xiong (2006); Van Nieuwerburgh and Veldkamp (2010)). While rational inattention is very useful for modeling the allocation of resources to information that is publicly available, it is not designed to address learning when information has to be searched for, nor to deliver implications for the commonality of agents' information.[2] Therefore, we see our approach as complementary.

On the technical side, we employ an urn model in contrast to other approaches. Urn models, which developed as a branch of probability theory, have been popularized in many fields such as biology, engineering, operations research, and mathematical psychology for their usefulness in applications.[3] Information search naturally lends itself to modeling with the urn approach that we use in our paper; this approach enables us to relate search frictions to the features of the underlying information environment (such as the availability of information). It allows us to provide a microfoundation for important features of information acquisition such as the quality and commonality of information with a greater degree of mathematical tractability. As our examples illustrate, our model can be fruitfully used in applications.

The organization of the paper is as follows. Section 2 develops the framework of information

[1] See, for example, Veldkamp (2011) for an excellent survey on this topic.

[2] As was pointed by Marschak (1974), entropy is more relevant to the cost of communicating than to the cost of searching for and collecting information. In Shannon (1948), entropy measures the quantity of transferred information under the optimal coding scheme. The optimal coding scheme does not reflect frictions arising from using natural languages.

[3] See, for example, Johnson and Kotz (1977) for a textbook treatment of urn problems and detailed discussions about their applications. In economics, urn models are used to model reinforcement learning in game theory (e.g., Beggs (2005); Hopkins and Posch (2005)), and the path dependence of technological innovations (e.g., David (1985)) based on the "Pólya urn scheme".

acquisition under imperfect information searchability. Section 3 studies a coordination game with complementarities as an example of possible applications. Section 4 concludes.

2 Information Search

In this section, we develop our methodology and characterize endogenous information under imperfect information searchability. We begin by describing the basic setup, then derive the asymptotic precision function (Theorem 2.1) and the asymptotic covariance function (Theorem 2.2). Then we derive our main result–the joint distribution of agents' information after information search (Theorem 2.3). Finally, we provide a public-private decomposition of the resulting information structure (Corollary 2.1).

2.1 The Setup

2.1.1 Basic signals

Consider an economic agent who is interested in acquiring information in order to resolve uncertainties that are relevant to his payoffs. There is a random variable of interest, θ, which follows a normal distribution with mean $\bar{\theta}$ and precision τ_θ.[4,5] For example, θ could be the payoff of an investment opportunity such as the liquidation value of a tradable asset. Suppose the underlying source of information on θ is given by a set $\mathbf{L}$ of "basic signals" that consists of L distinct signals on θ. Each basic signal $m \in \{1, 2, \ldots, L\}$ in $\mathbf{L}$ is given by

$$s^m = \theta + \epsilon^m, \tag{2}$$

where $\epsilon^m \sim i.i.d.\ \mathcal{N}\left(0, \tau_\epsilon^{-1}\right)$ is a noise that is independent of θ. We refer to τ_ϵ as the precision of the basic signal s^m.

[4] The precision and the inverse of variance are used interchangeably throughout this paper.

[5] We can alternatively assume that θ follows a uniform distribution on the real line (i.e., the agent has an improper prior on θ). Our results are unaffected by this alternative assumption.

2.1.2 Information searchability

We construct a formal model of information search using an urn model. Consider the set of signals $\mathbf{L}$ to be an urn, and the basic signals to be balls in the urn. We capture the idea of impediments to information search by allowing for redundancy among acquired signals. Imagine that the agent is randomly drawing balls *with* replacement from the urn. The agent can identify the index of each signal after acquiring it; ex post, he knows whether a signal is redundant or not given the set of collected signals. Formally, we have:

Assumption 2.1. *Signals are drawn with replacement from* $\mathbf{L}$.

This assumption plays a critical role in our model because it gives a foundation for the concept of information searchability. If the number of balls in the urn is limited, the chance of drawing a ball that is distinct from the balls drawn in the previous trials will get smaller as the agent draws more balls from the urn.[6]

2.1.3 Precision function

Because redundant signals are completely uninformative, the informativeness of a set of acquired signals only depends on the distinct signals among the set. Let H denote the set of distinct signals among those acquired by the agent, and let h denote the number of signals in H. Let $S(h)$ denote the mean of the signals $s^1, s^2, \ldots, s^h$ in H as follows:

$$S(h) = \frac{1}{h}\sum_{m \in H} s^m = \theta + \frac{1}{h}\sum_{m \in H} \epsilon^m. \tag{3}$$

Notice that $S(h)$ is a sufficient statistic for all the signals acquired by the agent because they are i.i.d. normally distributed. By the standard Bayesian belief update formula, the precision of the posterior belief about θ conditional on $S(h)$ is given by

$$Var(\theta|S(h))^{-1} = \underbrace{\tau_\theta}_{\text{precision of prior belief}} + \underbrace{\tau_\epsilon h}_{\text{signal precision}}. \tag{4}$$

[6] The opposite case to Assumption 2.1 would be sampling without replacement, in which case acquired signals would never be redundant. In this case, any additional draw would be directly translated into a greater amount of information (or a greater resolution of uncertainty). See Section 2.2 for further discussion.

That is, the set of h i.i.d. signals is equivalent to having a single signal with precision that is h times higher than that of each individual signal in the set. These observations lead to the following definition of the precision function, given the number of distinct signals that are collected:

Definition 2.1. *The precision function $\Phi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is defined by*

$$\Phi(h) = \tau_\epsilon h, \tag{5}$$

where h is the number of distinct basic signals drawn from $\mathbf{L}$, and τ_ϵ is the precision of each basic signal.

Suppose that l signals are drawn with replacement from a finite set $\mathbf{L}$ that consists of L distinct signals. We let $\tilde{h}$ denote the (random) number of distinct signals among l collected signals. The following lemma derives the expected number of distinct signals $E(\tilde{h})$.

Lemma 2.1. *Suppose that signals are drawn l times from a set of L distinct signals. Then, the expected number of distinct signals among the collected signals is given by*

$$E(\tilde{h}) = L\left[1 - \left(1 - \frac{1}{L}\right)^l\right]. \tag{6}$$

Proof. A more general proof for this can be found in Stadje (1990). For each $m \in \{1, 2, \ldots, L\}$, we define $\tilde{h}^m$ to be one if signal s^m is collected eventually, and zero otherwise. Then, it is immediate that $Pr(\tilde{h}^m = 1) = 1 - \left(\frac{L-1}{L}\right)^l$. Because $\tilde{h} = \sum_{m=1}^{L} \tilde{h}^m$, we get

$$E(\tilde{h}) = \sum_{m=1}^{L}\left[1 - \left(\frac{L-1}{L}\right)^l\right] = L\left[1 - \left(1 - \frac{1}{L}\right)^l\right]. \tag{7}$$

□

Notice that $E(\tilde{h})$ is monotone increasing and concave in l.[7] That is, when more signals are drawn from the urn, the expected number of distinct signals increases, but it does so at

[7] The monotonicity and concavity can easily be verified from the following:

$$\frac{\partial E(\tilde{h})}{\partial l} = -L\left(1 - \frac{1}{L}\right)^l \log\left(1 - \frac{1}{L}\right) > 0, \quad \text{and} \quad \frac{\partial^2 E(\tilde{h})}{\partial l^2} = -L\left(1 - \frac{1}{L}\right)^l \left[\log\left(1 - \frac{1}{L}\right)\right]^2 < 0. \tag{8}$$

a decreasing rate as more and more signals are collected. Furthermore, $E(\tilde{h})$ is monotone increasing in L. Intuitively, the more independent signals are in the urn, the higher will be the expected number of distinct signals for a given number of draws. Hence, the number of signals in $\mathbf{L}$ reflects the degree of information searchability. We explore these ideas in the next subsection by connecting the precision function with the amount of resources spent on information collection.

2.1.4 Resources and precision

In this subsection, we introduce a set of assumptions that allow us to study an asymptotic limit of the precision function. To exploit the law of large numbers, we consider the case where the signals (or balls) in the urn become infinitesimally small so that the number of signals grows to infinity. That is, information acquisition becomes continuous in the limit rather than discrete. This continuous limit yields a smooth and deterministic precision function with desirable properties that can be applied to various economic applications.

To acquire necessary information, the agent needs to use his endowed resources. Let $c \in (0, \infty)$ be the unit of resources required to collect one signal on θ (i.e., the cost of one draw from the urn). We assume that any amount of resources less than c cannot be utilized to acquire a signal. Hence, an input of k units of resources would enable the agent to collect $\lfloor \frac{k}{c} \rfloor$ signals.[8]

If the agent could observe all signals in $\mathbf{L}$, the agent's posterior precision in (4) would be $\Phi(L)$. Therefore, $\Phi(L)$ is the upper bound on the precision of information that can be learned from $\mathbf{L}$. We will consider the behavior of the precision function as the cost c becomes small, while keeping this upper bound fixed. Accordingly, in the next two assumptions we relate the number and precision of basic signals to the cost c.

Assumption 2.2. *For some $\mathcal{L} \in [0, \infty]$, the number of basic signals in $\mathbf{L}$ is given by $L = \lfloor \frac{\mathcal{L}}{c} \rfloor$.*

Note that $\mathcal{L}$ determines the number of signals in $\mathbf{L}$**,** and therefore parameterizes the degree of information searchability. All the collected signals will be distinct (thus informative) in the absence of impediments to information search ($\mathcal{L} = \infty$). On the other hand, some of

[8] $\lfloor x \rfloor = \max\{z \in \mathbb{Z} | z \leq x\}$.

the collected signals may be redundant in the presence of impediments to information search ($\mathcal{L} < \infty$).

Assumption 2.3. *For some $\tau \in [0, \infty)$, the precision of each basic signal $s^m \in \mathbf{L}$ equals $\tau_\epsilon = \tau c$.*

The parameter τ captures the efficiency of each basic signal per unit of cost. For given values of $\mathcal{L}$ and τ, Assumption 2.2 and Assumption 2.3 imply that the total amount of information available to the agent is in fact independent of c. Intuitively, we are keeping the total amount of information available from the urn fixed and dividing it into more signals as the cost c decreases. For example, when the required input of resources for one signal is halved, the number of basic signals available in the population doubles but the precision of each basic signal decreases is halved.

We define $\tilde{h}(k; c)$ to be the number of distinct signals drawn from $\mathbf{L}$ given an input of k units of resources, when the cost per draw is c. Then, the precision function according to Definition 2.1 under Assumption 2.3 is given by

$$\Phi(\tilde{h}(k; c)) = \tau c \tilde{h}(k; c). \tag{9}$$

There are two major problems in using the precision function in Eq. (9). First, the precision of information given an input of k units of resources is random because the number of distinct signals is random. Second, the function is not smooth in k because the number of distinct signals is given by discrete numbers. These shortcomings make the precision function defined in Eq. (9) unattractive in most economic applications. To resolve these shortcomings, we will consider the limiting case in which the cost c tends to zero, and rely on the following notion:

Definition 2.2. *The asymptotic precision function $\phi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is defined to be a function that satisfies*[9]

$$\Phi(\tilde{h}(k; c)) \to \phi(k) \;\; a.s. \;\; as\ c \to 0. \tag{10}$$

We demonstrate below that the asymptotic precision function $\phi(k)$ in Eq. (10) exists and resolves both problems with Eq. (9).

[9] One may alternatively state that $Pr\left[\lim_{c \to 0} \left|\Phi(\tilde{h}(k; c)) - \phi(k)\right| < \alpha\right] = 1$ for each $\alpha > 0$.

2.2 Asymptotic Precision Functions

Here we derive the asymptotic precision function in the presence of impediments to information search.

As a benchmark, consider the case of perfect searchability in which signals are drawn from $\mathbf{L}$ *without* replacement. Since there cannot be redundant signals in this case, the number of distinct signals drawn from $\mathbf{L}$ given an input of resources k such that $\left\lfloor \frac{k}{c} \right\rfloor \leq L$ is trivially equal to the number of collected signals, $\left\lfloor \frac{k}{c} \right\rfloor$. Hence, the corresponding precision function is given by

$$\Phi(\tilde{h}(k;c)) = \tau c \left\lfloor \frac{k}{c} \right\rfloor = \tau \left[k - g(c) \right], \tag{11}$$

where $g(c) \leq c$. By taking the limit as c goes to zero, we immediately obtain the asymptotic precision in the case of perfect information searchability:

$$\phi(k) = \tau k. \tag{12}$$

Now, we turn to the case of imperfect information searchability, i.e., when signals are drawn from $\mathbf{L}$ with replacement and $\mathcal{L} < \infty$. Using Lemma 2.1, we can derive the expected number of distinct signals given the resource input k as follows:

$$E(\tilde{h}(k;c)) = \lfloor \frac{\mathcal{L}}{c} \rfloor \left(1 - \left(1 - \frac{1}{\lfloor \frac{\mathcal{L}}{c} \rfloor} \right)^{\lfloor \frac{k}{c} \rfloor} \right). \tag{13}$$

Multiplying by c and taking the limit as c goes to zero in Eq. (13) yields

$$E(c\tilde{h}(k;c)) \to \mathcal{L} \left(1 - \exp\left(-\frac{k}{\mathcal{L}} \right) \right), \quad \text{as } c \to 0. \tag{14}$$

Therefore, the expectation of the precision function $\Phi(\tilde{h}(k;c))$ becomes smooth in the limit as c approaches zero. However, it is not clear that the precision function itself will be a deterministic function: proving this result is a non-trivial task because the number of collected signals grows large as c approaches zero but so does the number of redundant signals. Intuitively, proving that uncertainty in $c\tilde{h}(k;c)$ disappears as c approaches zero requires us to show that the fraction of redundant signals converges to its expectation, or, more formally, that $c\tilde{h}(k;c)$ can only deviate from $E(c\tilde{h}(k;c))$ in measure zero cases as c approaches zero. This result is provided by the following lemma:

Lemma 2.2. *As $c \to 0$, the difference between $c\tilde{h}(k;c)$ and $E(c\tilde{h}(k;c))$ converges to zero almost surely.*

Proof. See Appendix A. □

Because $\Phi(\tilde{h}(k;c)) = \tau c\tilde{h}(k;c)$, Lemma 2.2 gives the main argument in the proof of the following theorem.

Theorem 2.1. *In the case of imperfect information searchability, the asymptotic precision is given by*

$$\phi(k) = \tau\mathcal{L}\left(1 - \exp\left(-\frac{k}{\mathcal{L}}\right)\right) \tag{15}$$

As mentioned above, the asymptotic precision function $\phi(k)$ overcomes the two major difficulties that exist in the case of $\Phi(\tilde{h}(k;c))$. First, $\phi(k)$ is a deterministic function in k. Second, $\phi(k)$ is a smooth function in k, i.e., $\phi(k)$ is continuous in k and is also infinitely differentiable with respect to k. Furthermore, it has the following standard properties that are frequently assumed in the information economics literature:
(i) non-negativity: $\phi \geq 0$;
(ii) monotonicity: $\partial\phi/\partial k \geq 0$;
(iii) concavity: $\partial^2\phi/\partial k^2 \leq 0$; and
(iv) curvature: $-\frac{\partial^2\phi/\partial k^2}{\partial\phi/\partial k} = 1/\mathcal{L}$.

These properties fit the intuition quite well. As one learns more about a subject, the probability of encountering redundant material goes up. One realizes that the collected materials overlap with those that were previously acquired only after searching. Such concavity is associated with a negatively accelerated learning curve, which has been repeatedly reported in cognitive science and psychology. A large body of literature with empirical and experimental evidence finds learning data showing a rapid improvement followed by lesser improvements to be best fitted by an exponential function.[10]

The curvature decreases in $\mathcal{L}$, that is, worse information searchability would make the asymptotic precision function more concave. On the other hand, as information searchability

[10] See Ritter and Schooler (2001) for surveys on the "power law" of learning curve that has been widely observed in cognitive psychology.

improves (i.e., $\mathcal{L} \to \infty$), the asymptotic precision function in Eq. (15) converges to the linear function in Eq. (12) that is obtained when signals are drawn without replacement. We remark that Eq. (15) implies $\phi'(0) = \tau$, that is, the precision obtained from the first unit of input only depends on the precision of the underlying information and is unaffected by information searchability–the very first unit of information cannot be redundant, regardless of $\mathcal{L}$.

From Eq. (15), it can immediately be verified that the precision $\phi(k)$ increases in both parameters τ and $\mathcal{L}$. This is intuitive: at any positive level of input k, higher values of τ increase the precision of each non-redundant signal while higher values of $\mathcal{L}$ increase the number of non-redundant signals.

In the literature, information acquisition technologies are often specified in terms of a cost function $k : \phi \mapsto k(\phi)$ that maps the given level of precision ϕ to the amount of required resources k. In our setup, a cost function is readily obtained as the inverse of the precision function in (15), $k(\cdot) = \phi^{-1}(\cdot)$, as

$$k(\phi) = -\mathcal{L} \log \left(1 - \frac{\phi}{\tau \mathcal{L}}\right). \tag{16}$$

The cost function $k(\phi)$ has the following properties: it is non-negative, monotone increasing and convex, with curvature decreasing in the information searchability parameter, $\mathcal{L}$. Note that $k(\phi)$ becomes infinite as ϕ approaches $\tau\mathcal{L}$, which represents the upper bound on the information precision. Finally, we remark that $k'(0) = \tau^{-1} > 0$, that is, the marginal cost of the first unit of information is bounded away from zero.

Figure 1 provides a graphical illustration of $\phi(k)$ and $k(\phi)$ for different values of the information searchability parameter $\mathcal{L}$. The curvature of both functions decreases for larger values of $\mathcal{L}$, the precision function becoming less concave and the cost function becoming less convex. The functions are linear in the case of perfect searchability ($\mathcal{L} = \infty$).

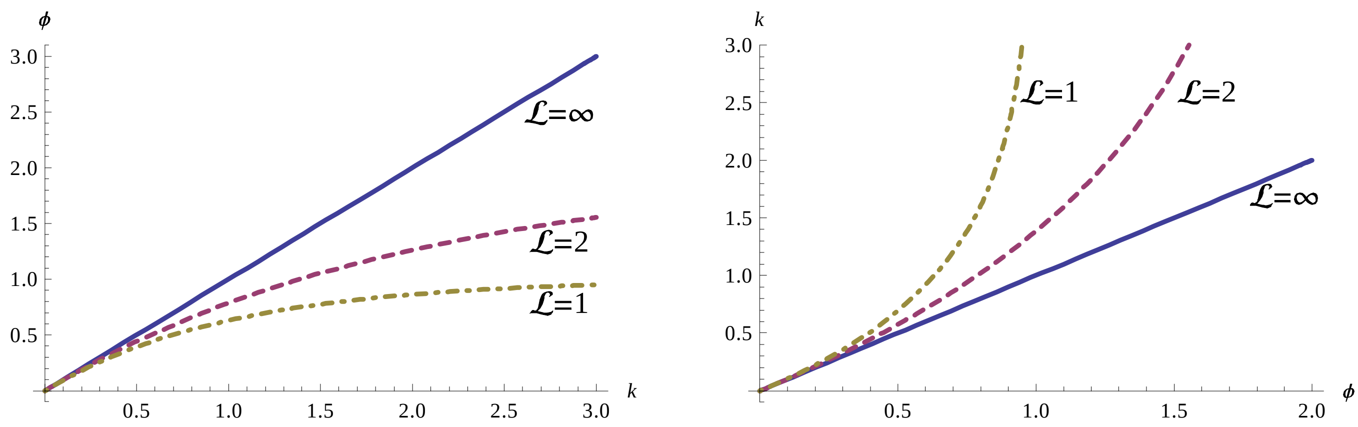

FIGURE 1. Left panel: precision function $\phi(k)$. Right panel: cost function $k(\phi)$. Parameter values: $\tau = 1$ and $\mathcal{L} \in \{1, 2, \infty\}$.

Finally, we call an information source "superior" to another source only if it provides more precise information than the other source given the same level of inputs. We give a more formal definition of superiority as follows:

Definition 2.3. *Information source i is superior to information source j if i is both more efficient and more searchable, i.e., $\tau^i \geq \tau^j$ and $\mathcal{L}^i \geq \mathcal{L}^j$ with at least one inequality being strict.*

A superior information source will always have higher precision given the same level of inputs. That is, suppose that i is superior to j (equivalently, j is inferior to i); then $\phi^i(k) > \phi^j(k)$ for all $k > 0$. One may imagine that a superior source will always be preferred to an inferior source, but later in the paper (Section 3) we show that this is not necessarily the case when agents' actions are strategic complements. This result builds on the multiple agent framework that we develop next.

2.3 Multiple Agents

In this section, we extend our framework to the case of multiple agents. In particular, we focus on the covariance of the acquired signals at the given level of information searchability. The same intuition about drawing balls with replacement from an urn still applies to the case with multiple agents; when the number of balls in the urn gets smaller, the possibility of different agents collecting overlapping information becomes higher. That is, more severe impediments to information search would induce higher covariance of errors among the acquired signals across different agents.

Suppose that there are I agents in the economy, and let $\mathbf{I}$ denote the set of agents. Adapting the notation introduced in Section 2.1.3, we denote by H_i the set of distinct signals acquired by agent i, and by h_i the number of signals in H_i. Let $S_i(h_i)$ be the mean of the distinct signals acquired by agent i. Then, $S_i(h_i)$ and $S_j(h_j)$ are sufficient statistics for the information acquired by agents i and j:

$$S_i(h_i) = \frac{1}{h_i} \sum_{m \in H_i} s^m = \theta + \frac{1}{h_i} \sum_{m \in H_i} \epsilon^m, \tag{17}$$

$$S_j(h_j) = \frac{1}{h_j} \sum_{n \in H_j} s^n = \theta + \frac{1}{h_j} \sum_{n \in H_j} \epsilon^n. \tag{18}$$

Therefore, the covariance between $S_i(h_i)$ and $S_j(h_j)$ is given by

$$Cov(S_i(h_i), S_j(h_j)) = \frac{1}{\tau_\theta} + Cov\left(\frac{1}{h_i} \sum_{m \in H_i} \epsilon^m, \frac{1}{h_j} \sum_{n \in H_j} \epsilon^n\right). \tag{19}$$

Let $H_{i,j}$ denote the set of indices of signals that belong to both H_i and H_j. Then, it is immediate that

$$Cov\left(\frac{1}{h_i} \sum_{m \in H_i} \epsilon^m, \frac{1}{h_j} \sum_{n \in H_j} \epsilon^n\right) = \frac{1}{h_i h_j} Var\left(\sum_{m \in H_{i,j}} \epsilon^m\right) = \frac{h_{i,j}}{\tau c h_i h_j}, \tag{20}$$

where $h_{i,j}$ denotes the number of distinct signals in $H_{i,j}$.

Suppose that agents i and j use amounts of resource k_i and k_j, respectively, when the cost of each signal is set to be c. Denote by $\tilde{h}_i(k_i; c)$ and $\tilde{h}_j(k_i; c)$ the resulting number of distinct signals collected by agents i and j and let $\tilde{h}_{i,j}(k_i, k_j; c)$ denote the number of distinct signals among the commonly collected signals. Of course, for any positive value of c, $\tilde{h}_i(k_i; c)$, $\tilde{h}_j(k_i; c)$ and $\tilde{h}_{i,j}(k_i, k_j; c)$ are random, and so is the covariance among the error terms in the signals $S_i(\tilde{h}_i(k_i; c))$ and $S_j(\tilde{h}_i(k_i; c))$ (see Eq. (20)). To restore tractability, we will again consider the limit in which c goes to zero and rely on the following definition:

Definition 2.4. *The asymptotic covariance σ_{ij} of the error terms in the signals $S_i(\tilde{h}_i(k_i;c))$ and $S_j(\tilde{h}_i(k_i;c))$ satisfies*

$$\frac{\tilde{h}_{i,j}(k_i,k_j;c)}{\tau c\tilde{h}_i(k_i;c)\tilde{h}_j(k_j;c))} \to \sigma_{ij} \ \ a.s. \ \ as \ c \to 0. \tag{21}$$

Using an argument similar to Lemma 2.2, we can show that randomness in $c\tilde{h}_{i,j}(k_i,k_j;c)$ disappears in the limit in which the cost c tends to zero. We have:

Lemma 2.3. *As $c \to 0$, $c\tilde{h}_{i,j}(k_i,k_j;c)$ converges to a deterministic function in k_i and k_j almost surely, i.e.,*

$$c\tilde{h}_{i,j}(k_i,k_j;c) \to \mathcal{L}\left(1-\exp\left(-\frac{k_i}{\mathcal{L}}\right)\right)\left(1-\exp\left(-\frac{k_j}{\mathcal{L}}\right)\right) \ \ a.s., \ as \ c \to 0. \tag{22}$$

Proof. See Appendix A. □

Then, Lemma 2.2 and Lemma 2.3 together with Eq. (20) provide the proof of the following theorem:

Theorem 2.2. *For each agent pair $i,j \in \mathbf{I}$, the asymptotic covariance of the error terms in the signals $S_i(\tilde{h}_i(k_i;c))$ and $S_j(\tilde{h}_i(k_i;c))$ satisfies*

$$\sigma_{ij} = \frac{1}{\tau\mathcal{L}}. \tag{23}$$

Notice that the asymptotic covariance σ_{ij} is constant and monotone decreasing in $\tau\mathcal{L}$. The latter result confirms our initial intuition that worse information searchability would increase the covariance of acquired information across different agents.

Finally, using Theorem 2.1 and Theorem 2.2, we can obtain the asymptotic correlation of the error terms between the two signals given the input of resources k_i and k_j as follows:

$$\rho(k_i,k_j) = \lim_{c\to 0} Corr\left(\frac{1}{h_i}\sum_{m\in H_i}\epsilon^m, \frac{1}{h_j}\sum_{n\in H_j}\epsilon^n\right) = \left[\left(1-\exp\left(-\frac{k_i}{\mathcal{L}}\right)\right)\left(1-\exp\left(-\frac{k_j}{\mathcal{L}}\right)\right)\right]^{\frac{1}{2}}. \tag{24}$$

2.4 Asymptotic Normality

In Theorem 2.1 and Theorem 2.2, we have obtained the second moments of the error terms in the asymptotic signals. In the following theorem, we derive their exact joint asymptotic distribution.

Theorem 2.3. *For each agent pair $i, j \in \mathbf{I}$ using inputs k_i and k_j, respectively, as c goes to zero the information acquired by agent $i, j \in \mathbf{I}$ is equivalent to the asymptotic signals*

$$S_i(k_i) = \theta + \epsilon_i, \tag{25}$$

$$S_j(k_j) = \theta + \epsilon_j, \tag{26}$$

where ϵ_i and ϵ_j are jointly normally distributed with mean zero and variance-covariance matrix $\Sigma_{i,j}$, where

$$\Sigma_{i,j} = \begin{pmatrix} \phi(k_i)^{-1} & \frac{1}{\tau\mathcal{L}} \\ \frac{1}{\tau\mathcal{L}} & \phi(k_j)^{-1} \end{pmatrix},$$

and the function $\phi(\cdot)$ is as in Eq. (15).

Proof. See Appendix A. □

2.5 A Public-Private Decomposition of Signals' Noise Terms

In the multiple agent case, we can decompose the noise term in agent i's signal from Theorem 2.3 in terms of the average noise across agents:

$$\epsilon_i = \mu + \eta_i, \tag{27}$$

where μ and η_i are defined as

$$\mu = \frac{1}{I}\sum_{i=1}^{I} \epsilon_i, \qquad \eta_i = \epsilon_i - \mu. \tag{28}$$

The next corollary provides a characterization of the representation in Eq. (27) for an economy with a large number of agents.

Corollary 2.1. *Consider an economy with I agents. As $I \to \infty$, each signal in Theorem 2.3 can be decomposed as*

$$S_i(k_i) = \theta + \mu + \eta_i, \tag{29}$$

where μ and the η_i's are independent of each other for all $i \in \mathbf{I}$ and

$$\mu \sim N\left(0, (\tau\mathcal{L})^{-1}\right); \qquad \eta_i \sim N\left(0, [\tau\mathcal{L}\left(\exp\left(k_i/\mathcal{L}\right) - 1\right)]^{-1}\right). \tag{30}$$

Proof. See Appendix A. □

The representation of agents' signals in Eq. (29) decomposes the original individual error term in each signal into two independent parts: one component that is common across all agents, μ, and an idiosyncratic component, η_i. Agent i can reduce the idiosyncratic variance of his signal by increasing the amount of resources k_i used for information acquisition. However, agent i cannot reduce the variance of the common component μ, which is determined by the underlying features of the information source, τ and $\mathcal{L}$.

Corollary 2.1 gives a microfoundation for a signal structure that blends together two common assumptions used in the literature, in which the error terms in the signals are either fully private (i.e., purely idiosyncratic noise) or fully public (i.e., purely public noise). A signal structure as in Eq. (29) is employed in various contexts: finance (e.g., Grundy and McNichols (1989); Manzano and Vives (2011)), political science (e.g., Dewan and Mayatt (2008)), information economics (e.g., Myatt and Wallace (2012)), and industrial organization (e.g., Mayatt and Wallace (2015)). In contrast to this literature, the precision of the common term (μ) and the precision of the idiosyncratic term (η_i) are jointly determined in our framework. In particular, limited information searchability induces a trade-off between an information source's precision and its publicity, as we explain next.

While the overall noise $Var(\epsilon_i)$ is a decreasing function of both τ and $\mathcal{L}$ (see Section 2.2), these parameters have different effects on the private and public components of $Var(\epsilon_i)$ at any given level of k_i. From Eqs. (30) it can immediately be verified that $Var(\mu)$ decreases in $\mathcal{L}$ whereas $Var(\eta_i)$ increases in $\mathcal{L}$. To further understand these comparative statics, consider the degree of "publicity" of the agents' information as measured by the fraction of a signal's noise that is attributed to the common part. By Eqs. (30), this ratio is

$$\frac{Var(\mu)}{Var(\epsilon_i)} = 1 - \exp\left(-\frac{k_i}{\mathcal{L}}\right). \tag{31}$$

Eq. (31) illustrates that the searchability parameter $\mathcal{L}$ affects the composition of the agents' information: lower values of $\mathcal{L}$ make the signals more public in nature. This property of the model is intuitive: the smaller the information content that is available from an information source, the more common the information of the agents who search from it. By contrast, the efficiency parameter τ scales all variances in the same way, leaving the overall composition of $Var(\epsilon_i)$ unchanged. Figure 2 illustrates these findings.

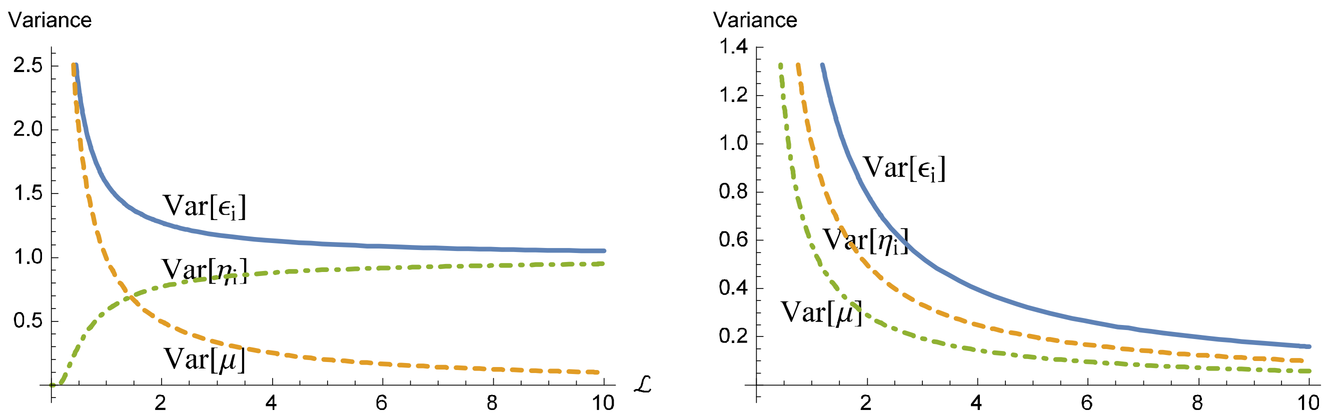

FIGURE 2. Left panel: variances across $\mathcal{L}$ (Parameter values: $\tau = 1, k_i = 1$). Right panel: variances across τ (Parameter values: $\mathcal{L} = 1, k_i = 1$).

These properties of the model have implications for the agents' incentives to search for information. For instance, suppose that agents care about knowing what other agents know because of strategic motives in their actions. In this case, agents might be willing to trade off precision for publicity in their information choice and favor a dominated information source. These information choices will have consequences for aggregate outcomes. In the next section, we confirm this intuition by exploring an equilibrium model with coordination motives and endogenous information.

3 Application: Endogenous Information in Coordination Games

We consider endogenous information choice in a beauty contest coordination game of the type popularized by Morris and Shin (2002). Our analysis builds on the existing literature (e.g., Hellwig and Veldkamp (2009); Myatt and Wallace (2012); Hellwig, Kohls, and Veldkamp (2012)) and complements it by adopting the information technology derived in the previous section. Our contribution is twofold. First, we show that our information acquisition technology leads to qualitatively different implications regarding the nature of the information structure and the existence of multiple equilibria. Second, we provide comparative statics on the different equilibria and searchability of information that are unique to our framework.

3.1 The Setup

There is a continuum of agents indexed by $i \in [0,1]$ who play a simultaneous move game with the following stages. First, each agent i gathers information, in a way that we specify below, on an aggregate state variable θ. Second, each agent i chooses an action $a_i \in \mathbb{R}$ that is based on the information he has observed. Agent i's payoff depends on how well his action does at matching the state variable θ as well as the average action $\overline{a} = \int_0^1 a_h dh$. Agent i's payoff function is assumed to be quadratic:

$$u_i = -(1-\delta)(\theta - a_i)^2 - \delta(\overline{a} - a_i)^2, \tag{32}$$

where the parameter $\delta \in (-1,1)$ measures the intensity of agents' coordination motive relative to the fundamental motive. For $\delta > 0$ ($\delta < 0$), agents' choices are strategic complements (substitutes). Furthermore, more positive (more negative) values of δ reflect greater desire to choose an action that is as close (distant) as possible to (from) the average action. For $\delta = 0$, the coordination motive plays no role.

We assume θ is normally distributed with mean $\bar{\theta}$ and variance τ_θ^{-1}. To acquire information about θ, each agent in the model allocates a fixed amount of resources K to $J > 1$ independent information sources. Each information source $j \in \{1, ..., J\}$ is characterized by its own efficiency parameter τ^j and searchability parameter $\mathcal{L}^j$. Each agent i chooses an allocation of his resources across information sources $k_i = \left(k_i^1, ..., k_i^J\right)$ such that $\Sigma_j k_i^j \leq K$. The mapping from resources to information is based on the information technology derived in the previous section. When agent i allocates a positive amount of resources to information source j (i.e., $k_i^j > 0$), the information obtained through this source is equivalent to a signal of the form

$$S_i^j = \theta + \epsilon_i^j; \qquad \epsilon_i^j \sim N\left(0, \phi^j\left(k_i^j\right)^{-1}\right), \tag{33}$$

where the precision function $\phi^j(\cdot)$ is as specified in Eq. (15) in Theorem 2.1,

$$\phi^j\left(k_i^j\right) = \tau^j \mathcal{L}^j \left(1 - \exp\left(-\frac{k_i^j}{\mathcal{L}^j}\right)\right). \tag{34}$$

When agent i does not allocate any resource to information source j (i.e., $k_i^j = 0$), the signal S_i^j is pure noise.

Corollary 2.1 implies that we can represent each signal S_i^j by decomposing the error term as follows:

$$S_i^j = \theta + \mu^j + \eta_i^j, \tag{35}$$

where μ^j's and the η_i^j's are jointly independent and normally distributed for all i and j:[11]

$$\mu^j \sim N\left(0, \left(\tau^j \mathcal{L}^j\right)^{-1}\right); \qquad \eta_i^j \sim N\left(0, \exp\left(-k_i^j/\mathcal{L}^j\right)\phi^j\left(k_i^j\right)^{-1}\right). \tag{36}$$

3.2 Equilibrium

In line with the literature, we focus on equilibria in which actions are affine functions of the signals, i.e., agent i's action takes the form $a_i = \gamma_i^0 + \Sigma_j \gamma_i^j S_i^j$.[12] We denote $\gamma_i = \left(\gamma_i^0, ..., \gamma_i^J\right)$ and let Δ be the set of feasible resource allocations, $\Delta = \left\{k \in \mathbb{R}_+^J \mid \Sigma_j k^j \leq K\right\}$. The strategy space is $\Gamma = \Delta \times \mathbb{R}^{J+1}$. An agent's strategy is a pair $(k, \gamma) \in \Gamma$.

We focus on symmetric equilibria in which all agents play the same strategy. When all other agents play the same strategy $\left(\hat{k}, \hat{\gamma}\right)$, agent i's ex ante utility from a strategy (k_i, γ_i) is as follows:

$$E(u_i) = -L_1(k_i, \gamma_i) - L_2(\gamma_i, \hat{\gamma}), \tag{37}$$

[11] Using Eq. (34), it can immediately be verified that Eqs. (36) is equivalent to Eqs. (30).

[12] See Myatt and Wallace (2012) for a discussion of sufficient conditions on the strategy space that ensure this assumption to be made without loss of generality.

where $L_1(k_i, \gamma_i)$ is the quadratic loss experienced by an agent when all players play the same strategy, and $L_2(\gamma_i, \hat{\gamma})$ is the quadratic loss (for $\delta > 0$) or gain (for $\delta < 0$) experienced by an agent when he deviates from other players' strategy (see Eqs. (B.1)-(B.2) in Appendix B for details of the derivations). Then, a Symmetric Bayesian Nash Equilibrium (SBNE) is a strategy $\left(\hat{k}, \hat{\gamma}\right)$ such that

$$\left(\hat{k}, \hat{\gamma}\right) \in \underset{(k_i,\gamma_i)\in\Gamma}{\operatorname{argmin}}\, L_1(k_i, \gamma_i) + L_2(\gamma_i, \hat{\gamma}). \tag{38}$$

We can show that a global minimizer of $L_1(k_i, \gamma_i)$ in Eq. (38) is a SBNE (see Lemma B.3 in Appendix B); because the term $L_2(\gamma_i, \hat{\gamma})$ vanishes when agent i plays $\gamma_i = \hat{\gamma}$, a global minimizer of $L_1(k_i, \gamma_i)$ is in fact a *payoff-maximizing equilibrium*.[13] In Appendix B (see Lemma B.1), we also show that finding a strategy that minimizes $L_1(k_i, \gamma_i)$ reduces to finding an allocation of resources among information sources k^* that satisfies

$$k^* \in \underset{k\in\Delta}{\operatorname{argmax}}\, G(k), \tag{39}$$

where

$$G(k) = \Sigma_{j=1}^{J} g_j(k^j); \qquad g_j(k^j) = \left[\frac{1-\delta}{\phi^j(k^j)} + \frac{\delta}{\exp\left(k^j/\mathcal{L}^j\right)\phi^j(k^j)}\right]^{-1}. \tag{40}$$

The optimization problem in Eq. (39) can be interpreted as a planner's problem of finding an allocation of resources that maximizes agents' payoffs in a symmetric equilibrium. The objective function $G(k)$ in Eq. (40) has an intuitive interpretation. Each function $g_j(k^j)$ is an (weighted, harmonic) average of the precisions of the overall error term ϵ_i^j of the signal in Eq. (33) and the idiosyncratic error term η_i^j in Eq. (35). When agents do not care about other agents' actions (i.e., $\delta = 0$), forecasting θ is all that matters for agents' utility. Then, $G(k)$ becomes simply a sum of signals' precisions (i.e., $g_j(\cdot) = \phi^j(\cdot)$), and k^* is chosen to maximize the precision of information about θ. Because the precision functions $\phi^j(\cdot)$'s are concave, $G(k)$ is concave if the coordination motive is sufficiently weak, in which case the solution to Eq. (39) is the unique equilibrium.[14]

[13] We remark that a payoff-maximizing equilibrium may not coincide with the first-best if coordination has no social value. See Colombo, Femminis, and Pavan (2014) for a welfare analysis of information acquisition.

[14] See Lemma B.1, Lemma B.2 and Lemma B.3 in Appendix B for the details of the argument.

On the other hand, for $\delta \neq 0$, agents' coordination motives introduce a distortion. Consider the case of strategic complementarities in actions, $\delta > 0$. In a symmetric equilibrium, the average action equals $\bar{a} = \hat{\gamma}^0 + \Sigma_j \hat{\gamma}^j (\theta + \mu^j)$, and what matters for predicting $\bar{a}$ is not just learning about the fundamental, but learning about the fundamental-plus-common-noise terms $(\theta + \mu^j)$'s. For instance, in the extreme case in which δ approaches one, $G(k)$ becomes simply a sum of the idiosyncratic precisions (i.e., $g_j(k_i^j) = var\left(\eta_i^j\right)^{-1}$ for $\delta = 1$), in which case k^* is chosen to maximize the precision of information about the terms $(\theta + \mu^j)$'s. The precision about $\theta + \mu^j$ is convex in the input k_i^j because uncertainty about $\theta + \mu^j$ decreases faster than it does for θ. As a result, a strong enough coordination motive makes the problem in Eq. (39) non-concave, which can lead to multiple equilibria. Intuitively, an allocation of resources that constitutes a local (but not global) maximum of $G(k)$ is a SBNE if the cost an agent incurs when moving away from other agents' actions is sufficient to deter deviation.

This discussion suggests that the interplay between the coordination motive and the nature of the information is the key determinant of information choices and equilibrium uniqueness. We formalize this idea in the following proposition:

Proposition 3.1. *(SBNE)(i) When actions are either strategic substitutes or strategic complements with a weak coordination motive (i.e., for $\delta \in (-1, 1/2)$), there exists a unique equilibrium. In equilibrium, information acquisition satisfies*

$$\hat{k}^j(\lambda) = \begin{cases} \mathcal{L}^j \log\left(\frac{\tau^j - 2\lambda\delta(1-\delta) + \sqrt{\tau^j(\tau^j - 4\lambda\delta(1-\delta))}}{2(1-\delta)^2\lambda}\right) & \text{for } 0 < \lambda < \tau^j; \\ 0 & \text{for } \lambda \geq \tau^j, \end{cases} \tag{41}$$

for some $\lambda > 0$ that is the unique solution to $\Sigma_{j=1}^J \hat{k}^j(\lambda) = K$.
(ii) When the coordination motive is strong (i.e., for $\delta \in [1/2, 1)$), an equilibrium exists but there may be multiple equilibria.

Proof. See Appendix B. □

This proposition confirms the intuition from the previous discussion. In Proposition 3.1-(i), the equilibrium is unique if the coordination motive is not sufficiently strong (i.e., $\delta < 1/2$).[15] By contrast, Proposition 3.1-(ii) reveals the possibility of multiple equilibria when the

[15] In the equilibrium described in Proposition 3.1-(i), the following properties can easily be shown to be true.

coordination motive is sufficiently strong (i.e., $\delta \geq 1/2$). The next proposition examines this case in a simplified environment with only two information sources.

Proposition 3.2. *(Inferior information equilibrium when coordination motive is strong) Assume there are only two information sources, A and B, such that A has perfect searchability and is superior to B in the sense of Definition (2.3). Then,*
(i) $\hat{k}^A = K$ is an equilibrium; and
(ii) there exists a threshold $\bar{\mathcal{L}} > 0$ such that, for all $\mathcal{L}^B < \bar{\mathcal{L}}$, $\hat{k}^B = K$ is the payoff-maximizing equilibrium if the coordination motive is sufficiently strong.

Proof. See Appendix B. □

Devoting all resources to the superior technology (i.e., $k_i^A = K$ for all i) is an equilibrium regardless of the coordination motive (Proposition 3.2-(i)).[16] In this equilibrium, agents act upon signals that are private in nature and the average action only depends on the fundamental. Hence, precise information on the fundamental reveals other agents' actions, thus, it facilitates coordination. Because information source B is inferior, an agent has no incentive to deviate from $k_i^A = K$.

The equilibrium in Proposition 3.2-(ii) is in stark contrast with the $\delta < 1/2$ case. If all agents are learning from the inferior information source (i.e., $k_i^B = K$ for all i), agents obtain information that is less precise but highly correlated among them. This correlation facilitates coordination in agents' actions. When coordination is valuable and searchability is sufficiently poor (i.e., $\mathcal{L}^B < \bar{\mathcal{L}}$), choosing the inferior information source is individually optimal because the benefit from higher correlation outweighs the benefit from higher precision.[17]

First, agents devote attention to an information source only if this source is sufficiently efficient (i.e., $\lambda < \tau^j$). Second, if information source j is superior to information source i according to Definition (2.3), information source j gets more resources than i. Third, if an information source is superior to all other information sources and has perfect searchability, it gets all the resources. Fourth, as agents care more about other agents' actions (i.e., as δ increases), the number of information sources to which agents allocate their resources decreases weakly (i.e., $\frac{d\lambda}{d\delta} > 0$). These properties are consistent with those of Myatt and Wallace (2012).

[16] Because information source A is superior to B and $\mathcal{L}^A = \infty$, Proposition 3.1 implies that this is the unique equilibrium if $\delta < 1/2$.

[17] Proposition 3.2-(ii) further shows that $k_i^B = K$ is the payoff-maximizing equilibrium. The intuition is as

As explained in Section 2, lower values of $\mathcal{L}^B$ make the signal from source B less precise but also more public in nature. As a result, it can be easier for agents to coordinate on the inferior source when this information is less precise, i.e., the set of parameter values for which Proposition 3.2-(ii) holds expands as $\mathcal{L}^B$ decreases. The following corollary states this result formally:

Corollary 3.1. *(Comparative statics) As $\mathcal{L}^B$ decreases from the threshold $\bar{\mathcal{L}}$, a less strong coordination motive is needed for $\hat{k}^B = K$ to be the payoff-maximizing equilibrium.*

Proof. See Appendix B. □

The left panel of Figure 3 provides an illustration of Corollary 3.1. The figure further shows that the same comparative statics holds when $\hat{k}^B = K$ is a SBNE, without the further requirement that it is the payoff-maximizing SBNE.

Equilibrium information choices have implications for aggregate volatility, as illustrated by the right panel in Figure 3. The aggregate action is perfectly correlated with the fundamental when resources are fully invested in the superior information. In contrast, this correlation is significantly lower than one in the inferior information equilibrium because of the common noise μ^B. As a result, the volatility of the average action is significantly higher in the $\hat{k}^B = K$ equilibrium than it is in the $\hat{k}^A = K$ equilibrium, and this difference is more pronounced for lower values of the searchability parameter $\mathcal{L}^B$.

follows. When $\mathcal{L}^B < \bar{\mathcal{L}}$, allocating all resources to information source B leaves less uncertainty about $\theta+\mu^B$ than does allocating all resources to information source A about θ (i.e., $Var(\theta + \mu^B|S_i^B)\Big|_{k_i^B=K} < Var(\theta|S_i^A)\Big|_{k_i^A=K}$ for $\mathcal{L}^B < \bar{\mathcal{L}}$). If all agents are learning from the superior information source A, agents' payoffs only depend on θ, while if all agents are learning from the inferior information source B, the average action depends on $\theta + \mu^B$. As a result, for large enough values of δ (so that agents' payoffs crucially depend on how well they can predict the average action), agents' payoffs are larger in the $\hat{k}^B = K$ equilibrium that in the $\hat{k}^A = K$ equilibrium. Furthermore, there is no interior equilibrium because the objective function in Eqs. (39) and (40) is strictly convex under the conditions in the proposition.

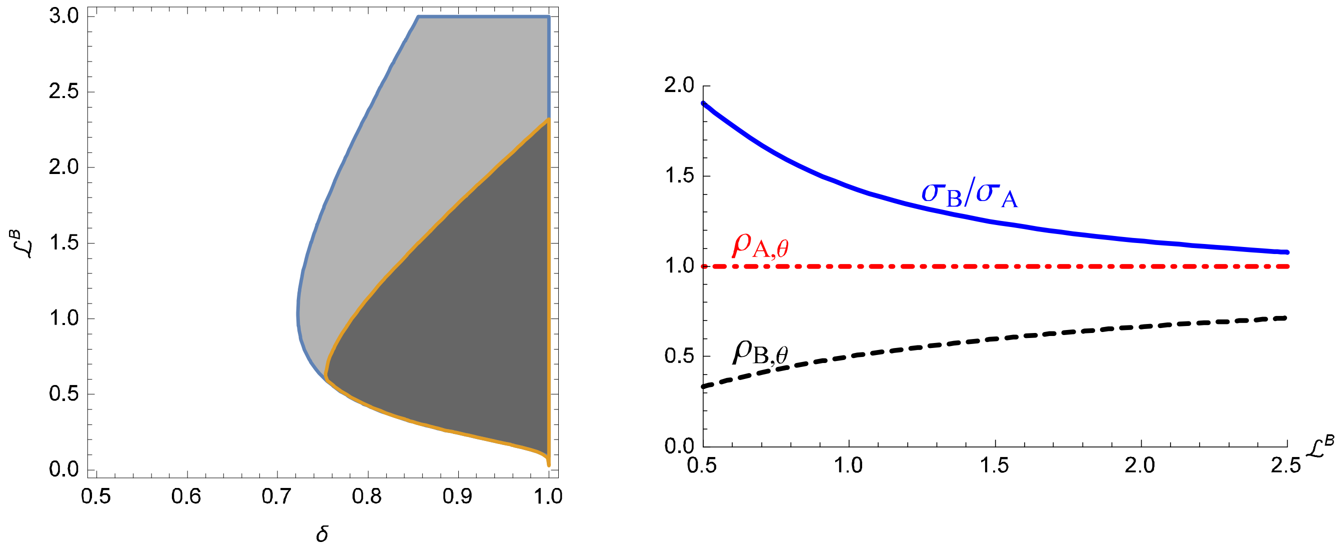

FIGURE 3. Left panel: light gray area: $\hat{k}^B = K$ is a SBNE; dark gray area: $\hat{k}^B = K$ is the payoff-maximizing SBNE. Right panel: thick line, σ_B/σ_A: relative volatility of the aggregate action across $\hat{k}^B = K$ and $\hat{k}^A = K$ equilibria; dashed (dot-dashed) line, $\rho_{B,\theta}$ ($\rho_{A,\theta}$): correlation coefficient between the average action $\overline{a}$ and θ in the $\hat{k}^B = K$ ($\hat{k}^A = K$) equilibrium. Parameter values: $\tau_\theta = \tau^A = K = 1$, $\tau^B = 0.8$ in both panels and $\delta = 0.8$ in the right panel.

3.3 Interpretation

Our results in this section have interesting implications for many economic situations with coordination motives (e.g., price-setting models, bank runs, analysts' forecasting, fund managers' portfolio choices). For example, our results shed light on the well-documented financial (or macroeconomic) analysts' herding behavior in forecasting or recommendations (e.g., Hong, Kubik, and Solomon (2000); Welch (2000); Lamont (2002); Clement and Tse (2005)). In that context, each individual agent's action in our model can be interpreted as an analyst's forecast whereas the average action can be interpreted as the consensus forecast among analysts. It is well-known that analysts' incentives do not only depend on their forecast accuracy but also on their career concerns.[18] That is, there is an element of strategic complementarities in the incentives behind analysts' forecasting, and, therefore, the payoff function in Eq. (32) can be interpreted as an approximation of such career concerns.

We can interpret the superior source in Proposition 3.2 as information based on firms' fundamentals while we can interpret the inferior source as rumors, information from social media or web-based discussions, and news covered in the media. Our results show that there is an equilibrium in which analysts gather all their information from the inferior source (see Proposition 3.2).

As a measure of herding behavior, we can consider the non-fundamental part of the average correlation among two analysts' forecasts: $Corr\left(a_i, a_{i'}|\theta\right)$. In a symmetric equilibrium in which agents invest all resources in information source j, this measure is equivalent to

$$Corr\left(a_i, a_{i'}|\theta\right) = 1 - \exp\left(-\frac{K}{\mathcal{L}^j}\right). \tag{42}$$

Notice that this is the same as the degree of publicity defined in Eq. (31). The publicity measure becomes more pronounced with lower searchability; thus, we can predict a higher tendency for analysts' herding when information searchability is lower. Although searchability itself is difficult to observe, the resulting forecasts are easily observed. Empirical evidence indeed suggests analysts' tendency for herding creates (serial) correlations of forecasts that may show no relevance to fundamentals (e.g., Chan, Jegadeesh, and Lakonishok (1996); Welch (2000); Zitzewitz (2001)).

We can also assess the impact of herding behavior on informational efficiency, which we measure with the conditional precision of the fundamental value θ given the consensus forecast $\bar{a}$. It can be shown that

$$Var(\theta|\bar{a})^{-1} = \frac{\tau_\theta}{1 - Corr\left(\theta, \bar{a}\right)^2}, \tag{43}$$

where $Corr(\theta, \bar{a})$ denotes the correlation coefficient between θ and $\bar{a}$. Hence, the correlation coefficients in Figure 3 measure the informational efficiency of the equilibrium outcome. The consensus forecast reveals the fundamental value perfectly in the $\hat{k}^A = K$ equilibrium but not in the $\hat{k}^B = K$ equilibrium. Therefore, our results imply that financial analysts may herd on

[18] For example, Hong, Kubik, and Solomon (2000) document that analysts are more likely to be terminated for bold forecasts that deviate from consensus.

less precise information sources due to career concerns and this can create negative effects for informational efficiency.[19]

3.4 Relation to the Literature

Our analysis in this section shows that information search can be a source of aggregate volatility when agents' coordination motives are sufficiently strong. Agents may collectively choose to search for imprecise but highly correlated information; these information choices lead to non-fundamental correlation in individual actions as well as aggregate outcomes that are dislocated from fundamentals.

Methodologically, our analysis is closely related to two recent papers by Hellwig and Veldkamp (2009), and Myatt and Wallace (2012), who study beauty contest games with endogenous information acquisition.

A key message in Hellwig and Veldkamp (2009) is that the endogenous choice of public information generates multiple equilibria. The idea is that a public signal is more valuable than a private signal because it carries information both about the fundamental and about what other agents have learned (and, hence, about what other agents will do). However, this second effect depends on whether the public signal has been acquired by others or not, and this leads to multiple equilibria.[20]

A very different message emerges from Myatt and Wallace (2012), who assume a signal structure equivalent to Eq. (35), in which costly information acquisition from an information source reduces the idiosyncratic noise (but not the common noise). In their setup, the equilibrium is unique. The key difference is that the correlation in public information is bounded away from zero in Hellwig and Veldkamp (2009), while in Myatt and Wallace (2012) the publicity of a signal depends on agents' information choices and the first bits of information are effectively private. As Myatt and Wallace (2012) put it,"this smooths out the first step of the information acquisition process and eliminates multiple equilibria, even though the informative

[19] For example, Jegadeesh and Kim (2009) empirically find that the impact of forecast revisions on market prices is smaller when they are driven by analysts' herding behavior.

[20] This result also holds when acquisition of public information is "near continuous," in which case the value of information is kinked at the point where other agents have stopped learning from the public source. Hence, there can be many equilibria (a continuum, in fact).

signals actually acquired in equilibrium may be relatively public in nature."

In our model, uniqueness of equilibrium is guaranteed only if the coordination motive is not too strong. Key to this result is the publicity-precision trade-off induced by limited information searchability, which provides multiple ways for coordination to be achieved among agents (see the discussion following Proposition 3.2). This mechanism is absent in Myatt and Wallace (2012). It is the source of multiplicity in our model when the coordination motive is strong.[21]

4 Conclusion

In our paper, we develop a microfounded framework of information acquisition in which information is acquired through search. Our framework is based on a simple intuition: as an agent keeps searching for new information, it is likely that he will encounter some pieces of information that overlap with findings from his past searching activities. Furthermore, other agents searching for information from the same source will face the same difficulty in collecting new information; thus, they are more likely to have similar information if the amount of searchable information is smaller. We formalize this idea by employing an urn model, where signals are drawn with replacement. This allows us to develop a framework in which both the concavity of signal precision and the correlation among signals increase as information becomes less searchable. Using an asymptotic approach, we construct a tractable mapping from resource allocations to the precision and the correlation of agents' acquired information under varying degrees of searchable information.

Our analysis highlights that limited information searchability induces a trade-off between the precision and the publicity of information that is acquired. To illustrate the potential implications of this trade-off for equilibrium outcomes, we embed our information acquisition technology in a beauty contest coordination game with endogenous information. We find that agents may collectively prefer an inferior information source due to coordination motives.

[21] Myatt and Wallace (2012) also show that uniqueness of equilibrium is not robust in a version of their model that employs a cost function derived through rational inattention constraints on information transmission (see Sections 8 and 9 in their paper). This cost function is concave and introduces an exogenous element of convexity into the problem, which is independent of the agents' coordination motives. By contrast, the precision function in our model is concave (equivalently, the cost function for precision is convex); in our setup, non-concavities arise from the interaction between the commonality of information and the coordination motive.

As these information choices influence aggregate outcomes, information search is a potential source of aggregate volatility.

Our model has a broad spectrum of potential applications. It provides a tractable information acquisition technology for modeling situations in which acquiring information is costly. In particular, the model's implications for the commonality of information among agents provide a useful benchmark for information acquisition in a multitude of economic situations that exhibit payoff externalities. Examples include price setting models, experts' recommendations (such as financial analysts), bank runs, and trading in the financial markets. In these situations, the commonality of information among agents plays an important role for equilibrium outcomes. In summary, our framework provides a tool for modeling the endogenous formation of agents' information in a tractable way, thus, allowing us to better analyze and understand the resulting equilibrium outcomes.

Appendix A

Proof of Lemma 2.2: We prove this lemma in a similar fashion to standard proofs of the strong law of large numbers.[22] The major difference from the proof of the standard case is that, here, samples of the random variables from the population allow redundancy at varying rates as the number of samples increases.

Let L denote the number of distinct signals in $\mathbf{L}$, i.e., $L = \lfloor \frac{\mathcal{L}}{c} \rfloor$. We also denote as l the number of signals collected from $\mathbf{L}$, i.e., $l = \lfloor \frac{k}{c} \rfloor$. For each $m \in \{1, 2, \ldots, L\}$, we define $\tilde{h}^m(k;c)$ to be one if signal s^m is collected eventually, and zero otherwise. Then, we have $\tilde{h}(k;c) = \sum_{m=1}^{L} \tilde{h}^m(k;c)$, and

$$E(\tilde{h}^m(k;c)) = Pr(\tilde{h}^m(k;c) = 1) = 1 - \left(\frac{L-1}{L}\right)^l. \tag{A.1}$$

By Markov's inequality, we have

$$Pr\left[\left|c\tilde{h}(k;c) - E(c\tilde{h}(k;c))\right| \geq \alpha\right] \leq \frac{c^4 E\left[\left|\tilde{h}(k;c) - E[\tilde{h}(k;c)]\right|^4\right]}{\alpha^4}. \tag{A.2}$$

We first prove that $c^2 E\left[\left|\tilde{h}(k;c) - E[\tilde{h}(k;c)]\right|^4\right]$ converges as $c \to 0$. That is, the r.h.s. will be less than $\frac{c^2 M}{\alpha^4}$ for sufficiently small c for some positive constant M. This will allow us to have the desired result.

We now drop the arguments in $\tilde{h}(k;c)$ and $\tilde{h}^m(k;c)$ for notational convenience throughout this proof. Observe that

$$E\left[\left|\tilde{h} - E(\tilde{h})\right|^4\right] = E\left(\tilde{h}^4\right) - 4E\left(\tilde{h}^3\right)E\left(\tilde{h}\right) + 6E\left(\tilde{h}^2\right)E\left(\tilde{h}\right)^2 - 4E\left(\tilde{h}\right)^4 + E\left(\tilde{h}\right)^4. \tag{A.3}$$

Then, we can obtain the exact expression for Eq. (A.3) by obtaining each element in it separately as follows:

$$E\left(\tilde{h}^2\right) = LE\left[(\tilde{h}^m)^2\right] + L(L-1)E\left(\tilde{h}^m\tilde{h}^n\right), \tag{A.4}$$

$$E\left(\tilde{h}^3\right) = LE\left[(\tilde{h}^m)^3\right] + \binom{3}{2}L(L-1)E\left[(\tilde{h}^m)^2\tilde{h}^n\right] + L(L-1)(L-2)E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\right), \tag{A.5}$$

$$\begin{aligned} E\left(\tilde{h}^4\right) = {} & LE\left[(\tilde{h}^m)^4\right] + \binom{4}{3}L(L-1)E\left[(\tilde{h}^m)^3\tilde{h}^n\right] + \frac{1}{2}\binom{4}{2}L(L-1)E\left[(\tilde{h}^m)^2(\tilde{h}^n)^2\right] \\ & + \binom{4}{2}L(L-1)(L-2)E\left[(\tilde{h}^m)^2\tilde{h}^n\tilde{h}^x\right] + L(L-1)(L-2)(L-3)E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\tilde{h}^y\right). \end{aligned} \tag{A.6}$$

[22] See, for example, Billingsley (1995) for the standard proofs of the strong law of large numbers.

Because $(h^m)^r = h^m$ for all $r \in \mathbb{N}$, we have $E\left[(\tilde{h}^m)^r\right] = E\left(\tilde{h}^m\right)$, $E\left[(\tilde{h}^m)^r(\tilde{h}^n)^q\right] = E\left(\tilde{h}^m\tilde{h}^n\right)$, and $E\left[(\tilde{h}^m)^r(\tilde{h}^n)^q(\tilde{h}^x)^s\right] = E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\right)$ for any $r, q, s \in \mathbb{N}$. By substituting Eqs. (A.4), (A.5) and (A.6) into Eq. (A.3), we have

$$\begin{aligned} E\left[\left|\tilde{h} - E(\tilde{h})\right|^4\right] &= LE\left(\tilde{h}^m\right) + (3L^2 - 7L)E\left(\tilde{h}^m\tilde{h}^n\right) - 6(L^2 + 2L)E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\right) \\ &\quad + 3(L^2 - 2L)E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\tilde{h}^y\right). \end{aligned} \tag{A.7}$$

We denote by S the set of outcomes from drawing of l signals from the set $\mathbf{L}$ (i.e., the urn). Then, we have $|S| = L^l$ because there are L signals in the set $\mathbf{L}$.[23] We define A_m to be an event where signal i is not drawn within l trials (i.e., $\tilde{h}^m$ is equal to zero). Then, the expectation of the product of the random variables $\tilde{h}^m$ and $\tilde{h}^n$ is given by

$$E\left(\tilde{h}^m\tilde{h}^n\right) = Pr(\tilde{h}^m\tilde{h}^n = 1) = Pr(A_m^c \cap A_n^c) = \frac{|A_m^c \cap A_n^c|}{|S|}. \tag{A.8}$$

Using the inclusion–exclusion principle, we obtain[24]

$$\begin{aligned} E\left(\tilde{h}^m\tilde{h}^n\right) &= \frac{|S| - 2|A_m^c| + |A_m^c \cap A_n^c|}{|S|}. \\ &= 1 - \frac{2(L-1)^l - (L-2)^l}{L^l} = 1 - 2\left(1 - \frac{1}{L}\right)^l + \left(1 - \frac{2}{L}\right)^l. \end{aligned} \tag{A.10}$$

Therefore, taking the limit of c in Eqs. (A.1) and (A.10) yields

$$\lim_{c \to 0} E\left(\tilde{h}^m\right) = 1 - \exp\left(-\frac{k}{\mathcal{L}}\right), \tag{A.11}$$

$$\lim_{c \to 0} E\left(\tilde{h}^m\tilde{h}^n\right) = \left(1 - \exp\left(-\frac{k}{\mathcal{L}}\right)\right)^2. \tag{A.12}$$

In a similar fashion as in Eq. (A.10), we obtain the following using the inclusion–exclusion principle:

$$E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\right) = Pr(\tilde{h}^m\tilde{h}^n\tilde{h}^x) = 1 - \frac{3(L-1)^l - 3(L-2)^l + (L-3)^l}{L^l}, \tag{A.13}$$

[23] $|A|$ indicates the cardinality of a set A.

[24] Suppose that there are finite sets $A_1, A_2, \ldots, A_M$ that belong to a set S. Then, the inclusion–exclusion principle states that

$$|\cap_{m=1}^M A_m^c| = |S| - \sum_{m=1}^{M} |A_m| + \sum_{1 \le m < n \le M} |A_m \cap A_n| - \sum_{1 \le m < n < r \le M} |A_m \cap A_n \cap A_r| + \ldots + (-1)^M |\cap_{m=1}^M A_m|. \tag{A.9}$$

$$E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\tilde{h}^y\right) = Pr(\tilde{h}^m\tilde{h}^n\tilde{h}^x\tilde{h}^y) = 1 - \frac{4(L-1)^l - 6(L-2)^l + 4(L-3)^l - (L-4)^l}{L^l}. \tag{A.14}$$

Then, taking the limit as c tends to zero in Eqs. (A.13) and (A.14) yields the followings:

$$\lim_{c\to 0} E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\right) = \left(1 - \exp\left(-\frac{k}{\mathcal{L}}\right)\right)^3, \tag{A.15}$$

$$\lim_{c\to 0} E\left(\tilde{h}^m\tilde{h}^n\tilde{h}^x\tilde{h}^y\right) = \left(1 - \exp\left(-\frac{k}{\mathcal{L}}\right)\right)^4. \tag{A.16}$$

Multiplying Eq. (A.7) by c^2 and taking the limit as c tends to zero yields

$$\lim_{c\to 0} c^2 E\left[\left|\tilde{h} - E[\tilde{h}]\right|^4\right] = 3\mathcal{L}^2 \exp\left(-\frac{2k}{\mathcal{L}}\right)\left(1 - \exp\left(-\frac{k}{\mathcal{L}}\right)\right)^2.$$

Given a positive real number δ, let $\bar{\mathcal{L}}$ denote $3\mathcal{L}^2 \exp\left(-\frac{2k}{\mathcal{L}}\right)\left(1 - \exp\left(-\frac{k}{\mathcal{L}}\right)\right)^2 + \delta$. Then, there exists $\bar{c}$ such that $c^2 E\left[\left|\tilde{h} - E(\tilde{h})\right|^4\right] < \bar{\mathcal{L}}$ for all $c < \bar{c}$. Therefore, there exists $N > 0$ such that for all $n \geq N$ and $n \in \mathbb{N}$ we have

$$Pr\left[\left|\frac{1}{n}\tilde{h} - \frac{1}{n}E\left[\tilde{h}\right]\right| \geq \alpha\right] < \frac{\bar{\mathcal{L}}}{n^2\alpha^4}.$$

Then, the first Borel-Cantelli lemma implies that

$$Pr\left[\lim_{n\to\infty}\left|\Phi\left(\tilde{h}(k;\frac{1}{n})\right) - \phi(k)\right| < \alpha\right] = 1,$$

or equivalently

$$Pr\left[\lim_{c\to 0}\left|\Phi\left(\tilde{h}(k;c)\right) - \phi(k)\right| < \alpha\right] = 1. \quad \blacksquare$$

Proof of Lemma 2.3: This proof is parallel to that of Lemma 2.2. Let L be the number of distinct signals in $\mathbf{L}$, i.e., $L = \lfloor\frac{\mathcal{L}}{c}\rfloor$. We also denote by l^i and l^j the number of signals collected by agents i and j from $\mathbf{L}$, respectively, i.e., $l^i = \lfloor\frac{k^i}{c}\rfloor$ and $l^j = \lfloor\frac{k^j}{c}\rfloor$. For each $m \in \{1, 2, \ldots, L\}$, we define $\tilde{h}^m_{i,j}(k_i;c)$ to be one if signal s^m belongs to the group of commonly collected signals $H_{i,j}$, and zero otherwise. We also define $\tilde{h}^m_i(k_i;c)$ (or $\tilde{h}^m_j(k_j;c)$) to be one if signal s^m is collected by agent i (or j), and zero otherwise. Then, we have

$$\tilde{h}_{i,j}(k_i, k_j; c) = \sum_{m=1}^{L} \tilde{h}^m_{i,j}(k_i, k_j; c) = \sum_{m=1}^{L} \tilde{h}^m_i(k_i;c)\tilde{h}^m_j(k_j;c).$$

Because $\tilde{h}_i^m(k_i;c)$ and $\tilde{h}_j^m(k_j;c)$ are independent, we get

$$
\begin{aligned}
E[\tilde{h}_{i,j}(k_i,k_j;c)] &= \sum_{m=1}^{L} Pr(\tilde{h}_i^m(k_i;c)\tilde{h}_j^m(k_j;c)=1) = \sum_{m=1}^{L} Pr(\tilde{h}_i^m(k_i;c)=1)Pr(\tilde{h}_j^m(k_j;c)=1) \\
&= L\left[\left(1-\left(\frac{L-1}{L}\right)^{l_i}\right)\left(1-\left(\frac{L-1}{L}\right)^{l_j}\right)\right].
\end{aligned}
\tag{A.17}
$$

We can represent Eq. (A.17) given k_i and k_j as follows:

$$
E[\tilde{h}_{i,j}(k_i,k_j;c)] = \lfloor\frac{\mathcal{L}}{c}\rfloor\left[\left(1-\left(1-\frac{1}{\lfloor\frac{\mathcal{L}}{c}\rfloor}\right)^{\lfloor\frac{k_i}{c}\rfloor}\right)\left(1-\left(1-\frac{1}{\lfloor\frac{\mathcal{L}}{c}\rfloor}\right)^{\lfloor\frac{k_j}{c}\rfloor}\right)\right]. \tag{A.18}
$$

Multiplying Eq. (A.18) by c and taking the limit as c tends to zero yields

$$
\lim_{c\to 0} E[c\tilde{h}_{i,j}(k_i,k_j;c)] \to \mathcal{L}\left(1-\exp\left(-\frac{k_i}{\mathcal{L}}\right)\right)\left(1-\exp\left(-\frac{k_j}{\mathcal{L}}\right)\right). \tag{A.19}
$$

We now drop the arguments in $\tilde{h}_{i,j}(k_i,k_j;c)$ and $\tilde{h}_{i,j}^m(k_i,k_j;c)$ for notational convenience throughout this proof.

By Markov's inequality, we have

$$
Pr\left[\left|c\tilde{h}_{i,j}-E[c\tilde{h}_{i,j}]\right|\geq\alpha\right]\leq\frac{c^4E\left[\left|\tilde{h}_{i,j}-E[\tilde{h}_{i,j}]\right|^4\right]}{\alpha^4}. \tag{A.20}
$$

We aim to prove Eq. (A.20) by showing that $c^2E\left[\left|\tilde{h}_{i,j}-E[\tilde{h}_{i,j}]\right|^4\right]$ converges as $c\to 0$. The rest of the proof of Lemma 2.3 is identical to the proof of Lemma 2.2 up to Eq. (A.7).

We denote $\varphi_1(z),\varphi_2(z)$ and $\varphi_3(z)$ to be

$$
\varphi_1(z) = \left(\frac{L-1}{L}\right)^z, \tag{A.21}
$$

$$
\varphi_2(z) = \left(\frac{L-2}{L}\right)^z, \tag{A.22}
$$

$$
\varphi_3(z) = \left(\frac{L-3}{L}\right)^z. \tag{A.23}
$$

Using the inclusion–exclusion principle (which is analogous to Eq. (A.10)), we have[25]

$$\begin{aligned} E\left(\tilde{h}^m_{i,j}\tilde{h}^n_{i,j}\right) &= Pr(\tilde{h}^m_{i,j}\tilde{h}^n_{i,j} = 1) \\ &= 1 - Pr(\tilde{h}^m_{i,j} = 0) - Pr(\tilde{h}^n_{i,j} = 0) + Pr(\tilde{h}^m_{i,j} = 0 \wedge \tilde{h}^n_{i,j} = 0). \end{aligned} \tag{A.25}$$

Using the inclusion–exclusion principle again, we derive

$$Pr(\tilde{h}^m_{i,j} = 0) = Pr(\tilde{h}^n_{i,j} = 0) = \varphi_1(l_i) + \varphi_1(l_j) - \varphi_1(l_i + l_j), \tag{A.26}$$

and

$$\begin{aligned} Pr(\tilde{h}^m_{i,j} = 0 \wedge \tilde{h}^n_{i,j} = 0) &= Pr((\tilde{h}^m_i = 0 \vee \tilde{h}^m_j = 0) \wedge (\tilde{h}^n_i = 0 \vee \tilde{h}^n_j = 0)) \\ &= 2\varphi_1(l_i)(1 - \varphi_1(l_i))\varphi_1(l_j)(1 - \varphi_1(l_j)) \\ &+ \varphi_2(l_i) + \varphi_2(l_j) - \varphi_2(l_i)\varphi_2(l_j). \end{aligned} \tag{A.27}$$

Substituting Eqs. (A.26) and (A.27) into Eq. (A.25), and taking the limit as c tends to zero yields

$$\lim_{c\to 0} E\left(\tilde{h}^m_{i,j}\tilde{h}^n_{i,j}\right) = \left(1 - \exp\left(-\frac{k_m}{\mathcal{L}}\right)\right)^2 \left(1 - \exp\left(-\frac{k_n}{\mathcal{L}}\right)\right)^2. \tag{A.28}$$

Similarly as in Eq. (A.28), we obtain the expectation of the cross product of the three variables $\tilde{h}^m, \tilde{h}^n$ and $\tilde{h}^x$ as follows:

$$\begin{aligned} E\left(\tilde{h}^m_{i,j}\tilde{h}^n_{i,j}\tilde{h}^x_{i,j}\right) &= Pr(\tilde{h}^m_{i,j}\tilde{h}^n_{i,j}\tilde{h}^x_{i,j} = 1) \\ &= 1 - 3\left[\varphi_1(l_i) + \varphi_1(l_j) - \varphi_1(m+n)\right] \\ &+ 3\left[\varphi_1(l_i)(1 - \varphi_1(l_i))\varphi_1(l_j)(1 - \varphi_1(l_j)) + \varphi_2(l_i) + \varphi_2(l_j) - \varphi_2(l_i)\varphi_2(l_j)\right] \\ &- [3(\varphi_1(l_i)^2(1 - \varphi_1(l_i))\varphi_1(l_j)(1 - \varphi_1(l_j))^2 + \varphi_1(l_i)(1 - \varphi_1(l_i))^2\varphi_1(l_j)^2(1 - \varphi_1(l_j))) \\ &+ 6\varphi_2(l_i)(1 - \varphi_1(l_i))\varphi_2(l_j)(1 - \varphi_1(l_j)) + \varphi_3(l_i) + \varphi_3(l_j) - \varphi_3(l_i)\varphi_3(l_j)]. \end{aligned} \tag{A.29}$$

Taking the limit as c tends to zero in Eq. (A.29) yields

$$\lim_{c\to 0} E\left(\tilde{h}^m_{i,j}\tilde{h}^n_{i,j}\tilde{h}^x_{i,j}\right) = \left(1 - \exp\left(-\frac{k_m}{\mathcal{L}}\right)\right)^3 \left(1 - \exp\left(-\frac{k_n}{\mathcal{L}}\right)\right)^3. \tag{A.30}$$

[25] In this case, we use the inclusion–exclusion principle in the following form:

$$|\cup_{m=1}^M A_m| = \sum_{m=1}^M |A_m| - \sum_{1\le m<n\le M} |A_m \cap A_n| + \sum_{1\le m<n<r\le M} |A_m \cap A_n \cap A_r| + \ldots + (-1)^{M+1}|\cap_{m=1}^M A_m|. \tag{A.24}$$

We can repeat the same exercise as in Eq. (A.29) for the expectation of the cross-product of the four variables $\tilde{h}^m, \tilde{h}^n, \tilde{h}^x$ and $\tilde{h}^y$ to obtain the following:

$$\lim_{c \to 0} E\left(\tilde{h}^m_{i,j} \tilde{h}^n_{i,j} \tilde{h}^x_{i,j} \tilde{h}^y_{i,j}\right) = \left(1 - \exp\left(-\frac{k_m}{\mathcal{L}}\right)\right)^4 \left(1 - \exp\left(-\frac{k_n}{\mathcal{L}}\right)\right)^4.$$

The rest of the proof is again identical to the proof of Lemma 2.2 and is omitted. ■

Proof of Theorem 2.3 Let $n = 1/c$. With a slight modification of the notation in the main text, let H^i_n denote the set of distinct signals among those acquired by agent i for a fixed k^i and c, and let h^i_n denote the number of signals in H^i_n. Similarly, denote by $H^{i,j}_n$ the set of distinct signals among the overlapping signals acquired by agent i and agent j for fixed k^i, k^j and c, and let $h^{i,j}_n$ denote the number of signals in $H^{i,j}_n$. Further, let $\mathbf{L}_n$ be the set of signals in the urn when the cost of each draw is c,[26] and let L_n be the cardinality of $\mathbf{L}_n$. Then, let S^i_n denote the mean of the signals $s^1, s^2, \ldots, s^{h^i_n}$ in H^i_n as follows:[27]

$$S^i_n = \frac{1}{h^i_n} \sum_{m \in H^i_n} s^m = \theta + \frac{1}{h^i_n} \sum_{m \in H^i_n} \epsilon^m,$$

and let $\widetilde{\epsilon}^i_n$ denote $S^i_n - \theta$, that is,

$$\widetilde{\epsilon}^i_n = \frac{1}{h^i_n} \sum_{m \in H^i_n} \epsilon^m. \tag{A.31}$$

Outline of the proof. We will prove the joint asymptotic normality of $\widetilde{\epsilon}^i_n, \widetilde{\epsilon}^j_n$ by showing that, as n goes to infinity,

$$a\widetilde{\epsilon}^i_n + b\widetilde{\epsilon}^j_n \xrightarrow{d} N\left(0, \frac{a^2}{\phi(k^i)} + \frac{b^2}{\phi(k^j)} + 2\frac{ab}{\tau \mathcal{L}}\right) \text{ for all } a, b \in \mathbb{R}^2. \tag{A.32}$$

The plan of the proof is as follows. As a first step, starting from $\widetilde{\epsilon}^i_n, \widetilde{\epsilon}^j_n$, we construct two alternative random variables, $\widehat{\epsilon}^i_n, \widehat{\epsilon}^j_n$ say, whose distribution is unaffected by the randomness in h^i_n, h^j_n and $h^{i,j}_n$. As a second step, we use the Central Limit Theorem to prove the asymptotic normality of $a\widehat{\epsilon}^i_n + b\widehat{\epsilon}^j_n$ as c goes to zero. As a third step, we prove that $a\widetilde{\epsilon}^i_n + b\widetilde{\epsilon}^j_n$ converges in probability to $a\widehat{\epsilon}^i_n + b\widehat{\epsilon}^j_n$ as c goes to zero. The fourth step combines the previous results and completes the proof.

First step. Let $\tilde{\mathbf{L}}_n$ be a set of L_n signals of the form $\theta + \tilde{\epsilon}^m$, where each $\tilde{\epsilon}^m$ is independently and identically distributed. Furthermore, each $\tilde{\epsilon}^m$ has the same distribution of the ϵ^m's in the signals in $\mathbf{L}_n$ and is independent of the ϵ^m's in the signals in $\mathbf{L}_n$.

[26] In terms of the notation in the main text we have $h_n = \tilde{h}(k; \frac{1}{n})$ and $L_n = \lfloor n\mathcal{L} \rfloor$.

[27] The signals and error terms s^m, ϵ^m should also have a n subscript to highlight that the distribution depends on n (i.e., c). We will omit such additional notation in the rest of the proof.

Let the random variable $z_n^{i,j}$ be defined as $z_n^{i,j} = h_n^{i,j} - \left\lceil E\left(h_n^{i,j}\right)\right\rceil$.[28] Then, let the set $Z_n^{i,j}$ be defined as follows. If $z_n^{i,j} > 0$, let $Z_n^{i,j}$ be a set of $z_n^{i,j}$ random draws (without replacement) from $H_n^{i,j}$. If $z_n^{i,j} < 0$, let $Z_n^{i,j}$ be a set of $\left|z_n^{i,j}\right|$ random draws (without replacement) from $\tilde{\mathbf{L}}_n$. If $z_n^{i,j} = 0$, let $Z_n^{i,j}$ be the null set. Then, let the set $\hat{H}_n^{i,j}$ be defined as follows:

$$\hat{H}_n^{i,j} = \begin{cases} H_n^{i,j} \diagdown Z_n^{i,j} & \text{if } z_n^{i,j} > 0 \\ H_n^{i,j} \cup Z_n^{i,j} & \text{if } z_n^{i,j} \leq 0 \end{cases}.$$

By construction, the cardinality of $\hat{H}_n^{i,j}$ equals $\left\lceil E\left(h_n^{i,j}\right)\right\rceil$.

Let the random variable z_n^i be defined as $z_n^i = h_n^i - \left\lceil E\left(h_n^i\right)\right\rceil - z_n^{i,j}$. Then, let the set Z_n^i be defined as follows. If $z_n^i > 0$, let Z_n^i be a set of z_n^i random draws (without replacement) from $H_n^i \diagdown H_n^{i,j}$. If $z_n^i < 0$, let Z_n^i be a set of $\left|z_n^{i,j}\right|$ random draws (without replacement) from $\tilde{\mathbf{L}}_n$. If $z_n^i = 0$, let Z_n^i be the null set. Then, let the set $\hat{H}_n^i$ be defined as follows:

$$\hat{H}_n^i = \begin{cases} \left(H_n^i \diagdown H_n^{i,j}\right) \diagdown Z_n^i & \text{if } z_n^i > 0 \\ \left(H_n^i \diagdown H_n^{i,j}\right) \cup Z_n^i & \text{if } z_n^i \leq 0 \end{cases}.$$

By construction, the cardinality of $\hat{H}_n^i$ equals $\left\lceil E\left(h_n^i\right)\right\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil$. Define the random variable $\widehat{\epsilon}_n^i$ as

$$\widehat{\epsilon}_n^i = \frac{1}{\left\lceil E\left(h_n^i\right)\right\rceil}\left[\sum_{m \in \hat{H}_n^i} \epsilon^m + \sum_{m \in \hat{H}_n^{i,j}} \epsilon^m\right]. \tag{A.33}$$

By construction, $\widehat{\epsilon}_n^i$ is therefore the sample average of $\left\lceil E\left(h_n^i\right)\right\rceil$ i.i.d. error terms, while $\widetilde{\epsilon}_n^i$ is the sample average of h_n^i i.i.d. error terms.

Finally, let the random variable $\widehat{\epsilon}_n^j$ be constructed in an equivalent manner to $\widehat{\epsilon}_n^i$ but for agent j.

<u>Second step</u>. Let $r_n = \left\lceil E\left(h_n^i\right)\right\rceil + \left\lceil E\left(h_n^j\right)\right\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil$. By construction, $a\widehat{\epsilon}_n^i + b\widehat{\epsilon}_n^j$ can be written out as the sum of r_n independent terms as follows:

$$a\widehat{\epsilon}_n^i + b\widehat{\epsilon}_n^j = \sum_{k=1}^{r_n} X_{nk},$$

where a number $\left\lceil E\left(h_n^i\right)\right\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil$ of the X_{nk} terms are of the form $X_{nk} = \frac{a}{\left\lceil E\left(h_n^i\right)\right\rceil}\epsilon^k$, a number $\left\lceil E\left(h_n^j\right)\right\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil$ of the X_{nk} terms are of the form $X_{nk} = \frac{b}{\left\lceil E\left(h_n^j\right)\right\rceil}\epsilon^k$ and a number $\left\lceil E\left(h_n^{i,j}\right)\right\rceil$ of the X_{nk} terms are of the form $X_{nk} = \left(\frac{a}{\left\lceil E\left(h_n^i\right)\right\rceil} + \frac{b}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)\epsilon^k$. Since $E\left(\epsilon^k\right) = 0$, $E\left(X_{nk}\right) = 0$.

[28] $\lceil x \rceil = \min\{z \in \mathbb{Z} | z \geq x\}$.

Letting V_n^2 denote the variance of $a\widehat{\epsilon}_n^i + b\widehat{\epsilon}_n^j$, we have

$$
\begin{aligned}
V_n^2 &= \sum_{k=1}^{r_n} Var(X_{nk}) \\
&= Var\left(\epsilon^k\right)\left[\left(\lceil E\left(h_n^i\right)\rceil - \lceil E\left(h_n^{i,j}\right)\rceil\right)\left(\frac{a}{\lceil E\left(h_n^i\right)\rceil}\right)^2 + \left(\lceil E\left(h_n^j\right)\rceil - \lceil E\left(h_n^{i,j}\right)\rceil\right)\left(\frac{b}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)^2\right. \\
&\left. + \lceil E\left(h_n^{i,j}\right)\rceil\left(\frac{a}{\lceil E\left(h_n^i\right)\rceil} + \frac{b}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)^2\right] \\
&= Var\left(\epsilon^k\right)\left[\frac{a^2}{\lceil E\left(h_n^i\right)\rceil} + \frac{b^2}{\left\lceil E\left(h_n^j\right)\right\rceil} + \frac{2ab}{\lceil E\left(h_n^i\right)\rceil\left\lceil E\left(h_n^j\right)\right\rceil}\lceil E\left(h_n^{i,j}\right)\rceil\right] \qquad \text{(A.34)} \\
&= \frac{n}{\tau}\left[\frac{a^2}{n\mathcal{L}\left[1-\left(1-\frac{1}{n\mathcal{L}}\right)^{nk^i}\right] + g_i(n)} + \frac{b^2}{n\mathcal{L}\left[1-\left(1-\frac{1}{n\mathcal{L}}\right)^{nk^j}\right] + g_j(n)} + \frac{2ab}{n\mathcal{L} + g_{i,j}(n)}\right],
\end{aligned}
$$

for some deterministic functions $g_i(c)$, $g_j(c)$ and $g_{i,j}(c)$ that all vanish as $n \to \infty$. Hence, we have

$$\lim_{n\to\infty} V_n^2 = \frac{a^2}{\phi\left(k^i\right)} + \frac{b^2}{\phi\left(k^j\right)} + 2\frac{ab}{\tau\mathcal{L}}. \qquad \text{(A.35)}$$

The Lindeberg condition requires that, for all $\delta > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{V_n^2} E\left(X_{nk}^2 \mathbf{1}_{\{|X_{nk}| \geq \delta V_n\}}\right) = 0. \qquad \text{(A.36)}$$

We can write

$$\sum_{k=1}^{r_n} \frac{1}{V_n^2} E\left(X_{nk}^2 \mathbf{1}_{\{|X_{nk}| \geq \delta V_n\}}\right) = \lambda_n^i + \lambda_n^j + \lambda_n^{i,j},$$

where

$$
\begin{aligned}
\lambda_n^i &= \frac{\left(\lceil E\left(h_n^i\right)\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil\right)}{V_n^2} E\left(\left(\frac{a\epsilon^k}{\lceil E\left(h_n^i\right)\rceil}\right)^2 \mathbf{1}_{\left\{\left|\frac{a\epsilon^k}{\lceil E\left(h_n^i\right)\rceil}\right| \geq \delta V_n\right\}}\right); \\
\lambda_n^j &= \frac{\left(\left\lceil E\left(h_n^j\right)\right\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil\right)}{V_n^2} E\left(\left(\frac{b\epsilon^k}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)^2 \mathbf{1}_{\left\{\left|\frac{b\epsilon^k}{\left\lceil E\left(h_n^j\right)\right\rceil}\right| \geq \delta V_n\right\}}\right);
\end{aligned}
$$

$$\lambda_n^{i,j} = \frac{\left\lceil E\left(h_n^{i,j}\right)\right\rceil}{V_n^2} E\left(\left[\left(\frac{a}{\lceil E\left(h_n^i\right)\rceil} + \frac{b}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)\epsilon^k\right]^2 \mathbf{1}_{\left\{\left|\left(\frac{a}{\lceil E\left(h_n^i\right)\rceil}+\frac{b}{\lceil E\left(h_n^j\right)\rceil}\right)\epsilon^k\right| \geq \delta V_n\right\}}\right).$$

Using the expression for V_n^2 in (A.34) and simplifying, we can write

$$\lambda_n^i = \alpha_n^i \beta_n^i,$$

where

$$\alpha_n^i = \left(1 - \frac{\left\lceil E\left(h_n^{i,j}\right)\right\rceil}{\lceil E\left(h_n^i\right)\rceil}\right) \frac{a^2}{\left[a^2 + b^2 \frac{\lceil E\left(h_n^i\right)\rceil}{\lceil E\left(h_n^j\right)\rceil} + 2ab \frac{\lceil E\left(h_n^{i,j}\right)\rceil}{\lceil E\left(h_n^j\right)\rceil}\right]},$$

and

$$\beta_n^i = E\left(\left(\frac{\epsilon^k}{\sqrt{Var\left(\epsilon^k\right)}}\right)^2 \mathbf{1}_{\left\{\left|\frac{a\epsilon^k}{\lceil E\left(h_n^i\right)\rceil}\right| \geq \delta V_n\right\}}\right).$$

Furthermore, note that we can write (assuming $a \neq 0$)

$$\left|\frac{a\epsilon^k}{\lceil E\left(h_n^i\right)\rceil}\right| \geq \delta V_n \Leftrightarrow \left|\frac{\epsilon^k}{\sqrt{Var\left(\epsilon^k\right)}}\right| \geq \frac{\delta}{|a|}\sqrt{\frac{\lceil E\left(h_n^i\right)\rceil^2 V_n^2}{Var\left(\epsilon^k\right)}} \Leftrightarrow \left|y^k\right| \geq \gamma_n^i,$$

where we define

$$y^k = \frac{\epsilon^k}{\sqrt{Var\left(\epsilon^k\right)}}, \tag{A.37}$$

and

$$\gamma_n^i = \frac{\delta}{|a|}\sqrt{a^2 \lceil E\left(h_n^i\right)\rceil + b^2 \frac{\lceil E\left(h_n^i\right)\rceil^2}{\left\lceil E\left(h_n^j\right)\right\rceil} + 2ab \frac{\left\lceil E\left(h_n^{i,j}\right)\right\rceil}{\left\lceil E\left(h_n^j\right)\right\rceil} \lceil E\left(h_n^i\right)\rceil}.$$

Hence, we can write

$$\lim_{n\to\infty} \lambda_n^i = \lim_{n\to\infty} \alpha_n^i \int_{\left|y^k\right| \geq \gamma_n^i} \left(y^k\right)^2 dP. \tag{A.38}$$

Note that $\lim_{n\to\infty} \gamma_n^i = \infty$ while the distribution of y^k is unaffected by n (see Eq. (A.37)), and therefore $\left[\left|y^k\right| \geq \gamma_n^i\right] \downarrow \varnothing$ as $n \uparrow \infty$.[29] Since α_n^i has a finite limit as $n \to \infty$, then $\lim_{n\to\infty} \lambda_n^i = 0$ in

[29] For notational convenience, we write $\left[\left|y^k\right| \geq \gamma_n^i\right] := \left[\omega \in \Omega : \left|y^k(\omega)\right| \geq \gamma_n^i\right]$ where Ω is the space of events ω.

Eq. (A.38). Similar steps show that $\lim_{n\to\infty}\lambda_n^j = \lim_{n\to\infty}\lambda_n^{i,j} = 0$, so that the Lindeberg condition (A.36) is satisfied. Then, the Lindeberg-Feller Central Limit Theorem implies

$$\frac{a\widehat{\epsilon}_n^i + b\widehat{\epsilon}_n^j}{V_n} \xrightarrow{d} N(0,1),$$

or, equivalently, that

$$a\widehat{\epsilon}_n^i + b\widehat{\epsilon}_n^j \xrightarrow{d} N\left(0, \frac{a^2}{\phi\left(k^i\right)} + \frac{b^2}{\phi\left(k^j\right)} + 2\frac{ab}{\tau\mathcal{L}}\right). \tag{A.39}$$

<u>Third step</u>. Note that we can write $\widetilde{\epsilon}_n^i$ in (A.31) as

$$\widetilde{\epsilon}_n^i = \frac{\lceil E\left(h_n^i\right)\rceil}{h_n^i}\frac{1}{\lceil E\left(h_n^i\right)\rceil}\sum_{m\in H_n^i}\epsilon^m = \frac{\lceil E\left(h_n^i\right)\rceil}{h_n^i}\breve{\epsilon}_n^i,$$

where we define

$$\breve{\epsilon}_n^i = \frac{1}{\lceil E\left(h_n^i\right)\rceil}\sum_{m\in H_n^i}\epsilon^m. \tag{A.40}$$

We will first prove that $\breve{\epsilon}_n^i \xrightarrow{i.p.} \widehat{\epsilon}_n^i$. We need to prove

$$\lim_{n\to\infty}\text{Prob}\left(\left|\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\right| > \alpha\right) = 0. \tag{A.41}$$

By Chebyshev's inequality,

$$\text{Prob}\left(\left|\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\right| > \alpha\right) \le \frac{Var\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\right)}{\alpha^2}. \tag{A.42}$$

By the variance decomposition formula,

$$Var\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\right) = E\left[Var\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\middle| h_n^i, h_n^{i,j}\right)\right] + Var\left[E\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\middle| h_n^i, h_n^{i,j}\right)\right].$$

Since $E\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\middle| h_n^i, h_n^{i,j}\right) = 0$, we are left with

$$E\left[Var\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\middle| h_n^i, h_n^{i,j}\right)\right] = \frac{1}{\lceil E\left(h_n^i\right)\rceil^2}E\left[Var\left(\sum_{m\in H_n^i\setminus H_n^{i,j}}\epsilon^m - \sum_{m\in\hat{H}_n^i}\epsilon^m + \sum_{m\in H_n^{i,j}}\epsilon^m - \sum_{m\in\hat{H}_n^{i,j}}\epsilon^m\right)\right].$$

Note that, by construction, $H_n^{i,j}$ and $\hat{H}_n^{i,j}$ differ by exactly $\left|z_n^{i,j}\right|$ elements, while $H_n^i\setminus H_n^{i,j}$ and $\hat{H}_n^i$ differ by exactly $\left|z_n^i\right|$ elements. Hence, we can write the last expression as

$$E\left[Var\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\middle| h_n^i, h_n^{i,j}\right)\right] = \frac{Var\left(\epsilon^m\right)}{\lceil E\left(h_n^i\right)\rceil^2}E\left[\left|z_n^i\right| + \left|z_n^{i,j}\right|\right]$$

$$
\begin{aligned}
&= \frac{Var\left(\epsilon^m\right)}{\lceil E\left(h_n^i\right)\rceil^2} E\left[\left|h_n^i - \lceil E\left(h_n^i\right)\rceil - z_n^{i,j}\right| + \left|z_n^{i,j}\right|\right] \\
&\leq \frac{Var\left(\epsilon^m\right)}{\lceil E\left(h_n^i\right)\rceil^2} E\left[\left|h_n^i - \lceil E\left(h_n^i\right)\rceil\right| + 2\left|z_n^{i,j}\right|\right] \\
&= \frac{Var\left(\epsilon^m\right)}{\lceil E\left(h_n^i\right)\rceil^2} E\left[\left|h_n^i - \lceil E\left(h_n^i\right)\rceil\right| + 2\left|h_n^{i,j} - \lceil E\left(h_n^{i,j}\right)\rceil\right|\right] \\
&= \frac{Var\left(\epsilon^m\right)}{\lceil E\left(h_n^i\right)\rceil} \left(E\left[\left|\frac{h_n^i}{\lceil E\left(h_n^i\right)\rceil} - 1\right|\right] + 2\frac{\left\lceil E\left(h_n^{i,j}\right)\right\rceil}{\lceil E\left(h_n^i\right)\rceil} E\left[\left|\frac{h_n^{i,j}}{\left\lceil E\left(h_n^{i,j}\right)\right\rceil} - 1\right|\right]\right) \\
&= \frac{Var\left(\epsilon^m\right)}{\lceil E\left(h_n^i\right)\rceil} \left(E\left[\left|w_n^i - 1\right|\right] + 2\frac{\left\lceil E\left(h_n^{i,j}\right)\right\rceil}{\lceil E\left(h_n^i\right)\rceil} E\left[\left|w_n^{i,j} - 1\right|\right]\right),
\end{aligned}
$$

where the second line follows from the definition of z_n^i, the third line follows by the triangle inequality, the fourth line follows from the definition of $z_n^{i,j}$, the fifth line from rearranging terms and the last line uses the following definitions:

$$
w_n^i = \frac{h_n^i}{E\left(h_n^i\right) + g_i(n)}; \qquad w_n^{i,j} = \frac{h_n^{i,j}}{E\left(h_n^{i,j}\right) + g_{i,j}(n)},
$$

for two deterministic functions $g_i\left(n\right)$ and $g_{i,j}\left(n\right)$ that converge to zero as $n \to \infty$.

By Lemma 2.2, $\frac{1}{n} h_n^i \xrightarrow{a.s.} \frac{1}{n} E\left(h_n^i\right)$ and therefore $w_n^i \xrightarrow{a.s.} 1$. Since $\left|w_n^i\right|$ is bounded from above by the constant $\left(1 - e^{-k^i/\mathcal{L}}\right)^{-1}$, the dominated convergence theorem implies that w_n^i converges in the L^1 norm, that is,

$$
\lim_{n\to\infty} E\left[\left|w_n^i - 1\right|\right] = 0.
$$

By Lemma 2.3, $\frac{1}{n} h_n^{i,j} \xrightarrow{a.s.} \frac{1}{n} E\left(h_n^{i,j}\right)$ and therefore $w_n^{i,j} \xrightarrow{a.s.} 1$. Since $\left|w_n^{i,j}\right|$ is bounded from above by the constant $\left[\left(1 - e^{-k^i/\mathcal{L}}\right)\left(1 - e^{-k^j/\mathcal{L}}\right)\right]^{-1}$, the dominated convergence theorem implies that $w_n^{i,j}$ converges in the L^1 norm, that is,

$$
\lim_{n\to\infty} E\left[\left|w_n^{i,j} - 1\right|\right] = 0.
$$

Since $\frac{Var(\epsilon^m)}{\lceil E(h_n^i)\rceil}$ and $\frac{\lceil E(h_n^{i,j})\rceil}{\lceil E(h_n^i)\rceil}$ have finite limits as $n \uparrow \infty$, we have shown that

$$
\lim_{n\to\infty} Var\left(\breve{\epsilon}_n^i - \widehat{\epsilon}_n^i\right) = 0,
$$

which completes the proof of (A.41).

Finally, since $\breve{\epsilon}_n^i \xrightarrow{i.p.} \widehat{\epsilon}_n^i$ and $\frac{\lceil E(h_n^i) \rceil}{h_n^i} \xrightarrow{i.p.} 1$ (which is implied by $\frac{1}{n} h_n^i \xrightarrow{a.s.} \frac{1}{n} E\left(h_n^i\right)$ and the continuous mapping theorem) and $\widetilde{\epsilon}_n^i = \frac{\lceil E(h_n^i) \rceil}{h_n^i} \breve{\epsilon}_n^i$, then we have $\widetilde{\epsilon}_n^i \xrightarrow{i.p.} \widehat{\epsilon}_n^i$. An identical proof shows that $\widetilde{\epsilon}_n^j \xrightarrow{i.p.} \widehat{\epsilon}_n^j$, and therefore

$$a\widetilde{\epsilon}_n^i + b\widetilde{\epsilon}_n^j \xrightarrow{i.p.} a\widehat{\epsilon}_n^i + b\widehat{\epsilon}_n^j. \tag{A.43}$$

Fourth step. By (A.39) and (A.43), Theorem 2.7 in Van der Vaart (2000) implies that (A.32) holds. Hence, by Theorem 29.4 in Billingsley (1995), $\widetilde{\epsilon}_n^i$ and $\widetilde{\epsilon}_n^j$ are jointly normally distributed. ■

Proof of Corollary 2.1: Notice that

$$Var(\mu) = \frac{1}{I^2} \sum_{i=1}^{I} Var(\epsilon_i) + \frac{1}{I^2} \sum_{i=1}^{I} \sum_{j \neq i}^{I} Cov(\epsilon_i, \epsilon_j) \tag{A.44}$$

$$= \frac{1}{I^2} \sum_{i=1}^{I} \phi^{-1}(k_i) + \frac{I-1}{I} \frac{1}{\tau \mathcal{L}}, \tag{A.45}$$

and

$$Cov(\epsilon_i, \mu) = \frac{1}{I} \left(Var(\epsilon_i) + \sum_{j \neq i} Cov(\epsilon_i, \epsilon_j) \right) \tag{A.46}$$

$$= \frac{1}{I} \phi^{-1}(k_i) + \frac{I-1}{I} \frac{1}{\tau \mathcal{L}}. \tag{A.47}$$

Therefore, we have

$$Cov(\eta_i, \mu) = Cov(\epsilon_i, \mu) - Var(\mu) = \frac{1}{I^2} \sum_{j=1}^{I} \phi^{-1}(k_j) - \frac{1}{I} \phi^{-1}(k_i), \tag{A.48}$$

and

$$Var(\eta_i) = Var(\epsilon_i - \mu) = Var(\epsilon_i) - 2Cov(\epsilon_i, \mu) + Var(\mu) \tag{A.49}$$

$$= \left(1 - \frac{2}{I}\right) \phi^{-1}(k_i) - \frac{I-1}{I} \frac{1}{\tau \mathcal{L}} + \frac{1}{I^2} \sum_{j=1}^{I} \phi^{-1}(k_j), \tag{A.50}$$

and

$$Cov(\eta_i, \eta_j) = Cov(\epsilon_i - \mu, \epsilon_j - \mu) = Cov(\epsilon_i, \epsilon_j) - Cov(\epsilon_i, \mu) - Cov(\epsilon_j, \mu) + Var(\mu) \tag{A.51}$$

$$= \frac{1}{\tau\mathcal{L}} - \frac{1}{I}\phi^{-1}(k_i) - \frac{1}{I}\phi^{-1}(k_j) - \frac{I-1}{I}\frac{1}{\tau\mathcal{L}} + \frac{1}{I^2}\sum_{i=1}^{I}\phi^{-1}(k_i). \tag{A.52}$$

Therefore, we have the following results in the limit as I tends to infinity:

$$\lim_{I\to\infty} Var(\eta_i) = \frac{1}{\phi(k_i)} - \frac{1}{\tau\mathcal{L}}, \quad \text{for all } i \in \mathbf{I} \tag{A.53}$$

$$\lim_{I\to\infty} Cov(\eta_i, \mu) = 0, \quad \text{for all } i \in \mathbf{I} \tag{A.54}$$

$$\lim_{I\to\infty} Cov(\eta_i, \eta_j) = 0, \quad \text{for all } i, j \in \mathbf{I} \tag{A.55}$$

$$\lim_{I\to\infty} Var(\mu) = \frac{1}{\tau\mathcal{L}}. \tag{A.56}$$

Using the definition of $\phi(k_i)$ in Theorem 2.1 we can immediately rearrange the r.h.s. of Eq. (A.53) as in the statement of Corollary 2.1. ■

Appendix B

Ex ante utility. Assuming all agents play some strategy $\left(\hat{k}, \hat{\gamma}\right)$, the average action equals $\overline{a} = \hat{\gamma}^0 + \Sigma_j \hat{\gamma}^j \left(\theta + \mu_j\right)$. Agent i's ex ante utility from playing strategy $\left(k_i, \gamma_i\right)$ is

$$E\left(u_i\right) = -\left(1-\delta\right) E\left(\theta - a^i\right)^2 - \delta E\left(\overline{a} - a^i\right)^2,$$

where

$$\begin{aligned} E\left(\theta - a_i\right)^2 &= E\left(\theta - \gamma_i^0 - \Sigma_j \gamma_i^j S_j^i\right)^2 \\ &= E\left(\left(\theta - \overline{\theta}\right)\left(1 - \Sigma_j \gamma_i^j\right) + \overline{\theta}\left(1 - \Sigma_j \gamma_i^j\right) - \gamma_i^0 - \Sigma_j \gamma_i^j \epsilon_j^i\right)^2 \\ &= \left(1 - \Sigma_j \gamma_i^j\right)^2 \tau_\theta^{-1} + \left(\overline{\theta}\left(1 - \Sigma_j \gamma_i^j\right) - \gamma_i^0\right)^2 + \Sigma_j \left(\gamma_i^j\right)^2 \phi^j \left(k_i^j\right)^{-1} \end{aligned}$$

and

$$\begin{aligned} E\left(\overline{a} - a_i\right)^2 &= E\left(\hat{\gamma}^0 + \Sigma_j \hat{\gamma}^j \left(\theta + \mu_j\right) - \gamma_i^0 - \Sigma_j \gamma_i^j S_j^i\right)^2 \\ &= E\left(\hat{\gamma}^0 - \gamma_i^0 + \overline{\theta}\left(\Sigma_j \hat{\gamma}^j - \Sigma_j \gamma_i^j\right) + \left(\Sigma_j \hat{\gamma}^j - \Sigma_j \gamma_i^j\right)\left(\theta - \overline{\theta}\right) + \Sigma_j \left(\hat{\gamma}^j - \gamma_i^j\right)\mu_j - \Sigma_j \gamma_i^j \eta_j^i\right)^2 \\ &= \left(\hat{\gamma}^0 - \gamma_i^0 + \overline{\theta}\left(\Sigma_j \hat{\gamma}^j - \Sigma_j \gamma_i^j\right)\right)^2 + \left(\Sigma_j \hat{\gamma}^j - \Sigma_j \gamma_i^j\right)^2 \tau_\theta^{-1} + \\ &\quad + \Sigma_j \left(\hat{\gamma}^j - \gamma_i^j\right)^2 \left(\tau^j \mathcal{L}^j\right)^{-1} + \Sigma_j \left(\gamma_i^j\right)^2 \left(\phi^j \left(k_i^j\right)^{-1} - \left(\tau^j \mathcal{L}^j\right)^{-1}\right). \end{aligned}$$

We can immediately rearrange terms as in Eq. (37), defining

$$L_1\left(k_i, \gamma_i\right) = \left(1-\delta\right)\left(\tau_\theta^{-1}\left(1 - \Sigma_{j=1}^J \gamma_i^j\right)^2 + \left(\overline{\theta}\left(1 - \Sigma_{j=1}^J \gamma_i^j\right) - \gamma_i^0\right)^2\right) + \Sigma_{j=1}^J \left(\gamma_i^j\right)^2 \left(\phi^j \left(k_i^j\right)^{-1} - \frac{\delta}{\tau^j \mathcal{L}^j}\right), \tag{B.1}$$

and

$$L_2\left(\gamma_i, \hat{\gamma}\right) = \delta\left[\tau_\theta^{-1}\left(\Sigma_{j=1}^J \gamma_i^j - \Sigma_{j=1}^J \hat{\gamma}^j\right)^2 + \Sigma_{j=1}^J \left(\gamma_i^j - \hat{\gamma}^j\right)^2 \frac{1}{\tau^j \mathcal{L}^j} + \left(\overline{\theta}\left(\Sigma_{j=1}^J \gamma_i^j - \Sigma_{j=1}^J \hat{\gamma}^j\right) + \gamma_i^0 - \hat{\gamma}^0\right)^2\right]. \tag{B.2}$$

We remark that Eq. (B.2) implies

$$L_2\left(\hat{\gamma}, \hat{\gamma}\right) = 0; \qquad \frac{\partial}{\partial \gamma_i} L_2\left(\hat{\gamma}, \hat{\gamma}\right) = 0. \tag{B.3}$$

Lemma B.1. *(Necessary conditions for a SBNE)A strategy* $\left(\hat{k}, \hat{\gamma}\right)$ *is a SBNE only if the following hold:*
(i) $\hat{\gamma}$ *satisfies*

$$\hat{\gamma}^j = \tilde{\gamma}^j\left(\hat{k}\right) = \begin{cases} \overline{\theta}\left(1 - \Sigma_{j=1}^J \hat{\gamma}^j\right) & \textit{for } j = 0 \\ \frac{(1-\delta) g_j(\hat{k}^j)}{\tau_\theta + (1-\delta) G\left(\hat{k}\right)}, & \textit{for } j = 1, ..., J, \end{cases} \tag{B.4}$$

where

$$g_j(k_i^j) = \phi^j\left(k_i^j\right)\left[1 - \delta + \delta \exp\left(-k_i^j / \mathcal{L}^j\right)\right]^{-1}, \tag{B.5}$$

and

$$G\left(k_i\right) = \Sigma_{j=1}^J g_j(k_i^j). \tag{B.6}$$

(ii) $\hat{k}$ *satisfies* $\Sigma_{j=1}^J \hat{k}^j = K$ *and is such that* $\nabla G\left(\hat{k}\right) \cdot z \leq 0$ *for any* $z \in \mathbb{R}^J$ *such that* $\Sigma_{j=1}^J z^j \leq 0$ *and* $z^j \geq 0$ *for every* j *with* $\hat{k}^j = 0$.[30]

Proof of part (i). Define

$$\tilde{\gamma}\left(k_i\right) \in \underset{\gamma_i}{\operatorname{argmin}}\ L_1\left(k_i, \gamma_i\right). \tag{B.7}$$

Fixing k_i, we can immediately verify that $L_1\left(k_i, \gamma_i\right)$ is strictly convex in γ_i for all $\delta \in (-1, 1)$. Differentiating $L_1\left(k_i, \gamma_i\right)$ with respect to γ_i and solving the system of first-order conditions for γ_i gives

$$\tilde{\gamma}_0\left(k_i\right) = \overline{\theta}\left(1 - \Sigma_{j=1}^J \tilde{\gamma}^j\left(k_i\right)\right);$$
$$\tilde{\gamma}^j\left(k_i\right) = \frac{(1-\delta)\,\tau_\theta^{-1} \phi^j\left(k_i^j\right)\left[1 - \delta + \delta \exp\left(-k_i^j/\mathcal{L}^j\right)\right]^{-1}}{1 + (1-\delta)\,\tau_\theta^{-1} \Sigma_{n=1}^J \phi_n\left(k_i^n\right)\left[1 - \delta + \delta \exp\left(-k_i^n/\mathcal{L}^n\right)\right]^{-1}} \text{ for } j = 1, ..., J.$$

Note that the expression for $\hat{\gamma}^j$ in Eq. (B.4) equals $\tilde{\gamma}^j\left(\hat{k}\right)$. Then, assume $\left(\hat{k}, \hat{\gamma}\right)$ is a SBNE and $\hat{\gamma}^j \neq \tilde{\gamma}^j\left(\hat{k}\right)$ for some j. Consider an agent deviating locally from $\hat{\gamma}^j$. First-order effects of deviations of γ_i^j from $\hat{\gamma}^j$ are zero for L_2 (see Eq. (B.3)). The strict convexity of L_1 in γ_i implies that, if $\hat{\gamma}^j \neq \tilde{\gamma}^j\left(\hat{k}\right)$, then $\frac{\partial}{\partial \gamma_i} L_1\left(\hat{k}, \hat{\gamma}\right) \neq 0$. Therefore, there is a profitable deviation, contradicting $\left(\hat{k}, \hat{\gamma}\right)$ being a SBNE. ■

[30] $\nabla G\left(\hat{k}\right)$ denotes the vector of partial derivatives of $G\left(\hat{k}\right)$ and "$\cdot$" denotes the dot product.

Proof of part (ii). Assume $\left(\hat{k}, \hat{\gamma}\right)$ is a SBNE. By part (i), it has to be the case that $\hat{\gamma} = \tilde{\gamma}\left(\hat{k}\right)$. Substituting $\tilde{\gamma}(k_i)$ for γ_i in $L_1(k_i, \gamma_i)$, we obtain

$$L_1(k_i, \tilde{\gamma}(k_i)) = \frac{(1-\delta)\,\tau_\theta^{-1}}{1 + (1-\delta)\,\tau_\theta^{-1}\Sigma_{j=1}^{J} g_j(k_i^j)}. \tag{B.8}$$

Writing γ_i as a function of k_i, we can drop the second argument in $L_1(k_i, \gamma_i)$ and represent it as $\tilde{L}_1(k_i) := L_1(k_i, \tilde{\gamma}(k_i))$.

In an equilibrium, local deviations in (k_i, γ_i) starting from $\left(\hat{k}, \tilde{\gamma}\left(\hat{k}\right)\right)$ must not provide a profitable deviation to an agent. Since an agent's local deviation from a symmetric strategy profile has no first-order effect on $L_2(\gamma_i, \hat{\gamma})$ (see Eq. (B.3)), any feasible direction of displacement z from $\hat{k}$ (i.e., any $z \in \mathbb{R}^J$ such that $z^j \geq 0$ for every j with $\hat{k}^j = 0$ and such that $\Sigma_{j=1}^{J} z^j \leq 0$ if $\Sigma_{j=1}^{J} \hat{k}^j = K$) must not create a first-order decrease in the loss function $\tilde{L}_1\left(\hat{k}\right)$, or $\nabla \tilde{L}_1\left(\hat{k}\right) \cdot z \geq 0$ for any feasible direction of displacement z from $\hat{k}$. Equivalently, since k_i enters $\tilde{L}_1(k_i)$ only through $G(k_i)$, and $\tilde{L}_1(k_i)$ is strictly decreasing in $G(k_i)$ for all $\delta \in (-1, 1)$, we must have $\nabla G\left(\hat{k}\right) \cdot z \leq 0$. Inspection of the functions g_j in Eq. (B.5) reveals that $g_j\left(k_i^j\right)$ has strictly positive first derivative for all $k_i^j \geq 0$ and all $\delta \in (-1, 1)$. Therefore, it has to be the case that $\Sigma_{j=1}^{J} \hat{k}^j = K$, since otherwise a marginal increase in any k^j from $\hat{k}$ would create a first-order decrease in the loss function $\tilde{L}_1\left(\hat{k}\right)$ and constitute a profitable deviation. ■

Lemma B.2. *A SBNE exists.*

Proof This lemma is proven using similar steps as in Hellwig, Kohls, and Veldkamp (2012). By Lemma B.1, in an equilibrium $\hat{\gamma}$ must satisfy $\hat{\gamma} = \tilde{\gamma}\left(\hat{k}\right)$. Hence, we can restrict the definition of a SBNE in Eq. (38) as $\left(\hat{\gamma}, \hat{k}\right)$ that satisfy

$$\left(\hat{\gamma}, \hat{k}\right) \in \underset{(\gamma_i, k_i)}{\operatorname{argmin}}\ L_1(\gamma_i, k_i) + L_2\left(\gamma_i, \tilde{\gamma}\left(\hat{k}\right)\right). \tag{B.9}$$

By Eqs. (B.1) and (B.2), it can immediately be verified that the minimand in (B.9) is strictly convex in γ_i, so that there exists a unique $\gamma^*\left(k_i, \hat{k}\right)$ that minimizes $L_1(\gamma_i, k_i) + L_2\left(\gamma_i, \tilde{\gamma}\left(\hat{k}\right)\right)$ with respect to γ_i given k_i and $\hat{k}$. Then, define $k^* : \Delta \rightarrow\rightarrow \Delta$ to be the best response correspondence

$$k^*\left(\hat{k}\right) \in \underset{k_i \in \Delta}{\operatorname{argmin}}\ L_1\left(\gamma^*\left(k_i, \hat{k}\right), k_i\right) + L_2\left(\gamma^*\left(k_i, \hat{k}\right), \tilde{\gamma}\left(\hat{k}\right)\right). \tag{B.10}$$

Since it can be shown that $\gamma^*\left(\hat{k}, \hat{k}\right) = \tilde{\gamma}\left(\hat{k}\right)$, a SBNE exists if the mapping k^* in (B.10) has a fixed point. Note that the minimand in (B.10) is continuous in k_i and Δ is compact, and therefore

$k^*\left(\hat{k}\right) \subset \Delta$ is non-empty. Moreover, it can be shown that the minimand in (B.10) is convex in k_i, and therefore $k^*\left(\hat{k}\right) \subset \Delta$ is convex. Then, the Maximum Theorem implies that $k^*\left(\hat{k}\right)$ is upper-hemicontinuous and compact-valued. Therefore, the Kakutani Fixed Point Theorem implies that k^* has a fixed point. ■

Lemma B.3. *For $\delta \in (-1, 1/2)$, $L_1(\gamma_i, k_i)$ has a unique global minimizer and this is the unique SBNE. For $\delta \in [1/2, 1)$, a global minimizer of $L_1(\gamma_i, k_i)$ is a SBNE.*

Proof Consider the case $\delta \in (-1, 1/2)$. Inspection of the functions g_j in Eq. (B.5) reveals that each g_j is strictly concave for $\delta \in (-1, 1/2)$. Thus, $G(k)$ is strictly concave for $k \in \Delta$. Then, if $\hat{k}$ satisfies part (ii) of Lemma B.1, $\hat{k}$ is the unique global maximum in the following problem:[31]

$$\max_{k_i \in \Delta} \Sigma_{j=1}^J g_j(k^j) \quad \text{s.t.} \quad \Sigma_{j=1}^J k^j = K. \tag{B.11}$$

Hence, $\left(\hat{\gamma}, \hat{k}\right)$ such that $\hat{k}$ is the unique solution to the problem in Eq. (B.11) and $\hat{\gamma} = \tilde{\gamma}\left(\hat{k}\right)$ as in part (i) of Lemma B.1 is the unique global minimizer of $L_1(\gamma_i, k_i)$ and the unique candidate for a SBNE for $\delta \in (-1, 1/2)$. Since a SBNE exists by Lemma B.2, this is the unique equilibrium.

For $\delta \in [1/2, 1)$ we have $L_2(\gamma_i, \hat{\gamma}) > 0$ for all $\gamma_i \neq \hat{\gamma}$. Therefore, by the definition of a SBNE in Eq. (38), a global minimum of $L_1(\gamma_i, k_i)$ is a SBNE. ■

Proof of Proposition 3.1. By Lemma B.1, Lemma B.2 and Lemma B.3, what is left to show for the proof of Proposition 3.1 is that $\hat{k}$ in Eq. (41) solves the problem in Eq. (B.11). We can convert the problem in Eq. (B.11) into the following dual problem:

$$\min_{\lambda} \lambda K - \Sigma_{j=1}^J g_j^*(\lambda),$$

where $g_j^*(\lambda)$ is the conjugate function of $g_j(k^j)$ such that

$$g_j^*(\lambda) = \min_{k^j \geq 0} \left(\lambda k^j - \phi^j\left(k^j\right) \frac{\exp\left(k^j/\mathcal{L}^j\right)}{\exp\left(k^j/\mathcal{L}^j\right)(1-\delta)+\delta} \right).$$

The F.O.C. for this problem implies that each $\hat{k}^j$ solves

$$\lambda - \frac{\tau^j \exp\left(\hat{k}^j/\mathcal{L}^j\right)}{\left[(1-\delta)\exp\left(\hat{k}^j/\mathcal{L}^j\right)+\delta\right]^2} = 0, \tag{B.12}$$

[31] We remark that if $\hat{k}$ is a local maximum of the problem in Eq. (B.11), then $\hat{k}$ clearly satisfies part (ii) of Lemma B.1 regardless of whether $G(k)$ is concave or not. However, if $\hat{k}$ satisfies part (ii) of Lemma B.1 and $G(k)$ is strictly concave, then $\hat{k}$ must be the unique global maximum of the problem in Eq. (B.11).

which has a strictly positive solution if and only if $0 < \lambda < \tau^j$, in which case it can immediately be verified that $\hat{k}^j$ is as in Eq. (41). Finally, λ can be obtained by solving the following equation:

$$\Sigma_{j=1}^{J} \hat{k}^j(\lambda) = K. \tag{B.13}$$

Notice that the l.h.s. is zero when $\lambda = \infty$ and infinity when $\lambda = 0$ and each $\hat{k}^j(\lambda)$ is strictly decreasing in λ for $0 < \lambda < \tau^j$. Therefore, there exists a unique $\lambda > 0$ that solves Eq. (B.13) because the l.h.s. is continuous and monotone decreasing in λ. ■

Proof of Proposition 3.2-(i). We want to prove that $\hat{k}^A = K$, $\hat{k}^B = 0$ is an equilibrium. By Lemma B.1 and Eqs. (B.4), it must be the case that $\hat{\gamma}^0 = (1-\hat{\gamma}^A)\overline{\theta}$, $\hat{\gamma}^A = \frac{(1-\delta)\tau^A K}{\tau\theta + (1-\delta)\tau^A K}$ and $\hat{\gamma}^B = 0$. Now, consider the corresponding problem in Eq. (38) using these values for $\left(\hat{k}, \hat{\gamma}\right)$. Denote $k(\alpha_i) = \left(k^A(\alpha_i), k^B(\alpha_i)\right)$ where $k^A(\alpha_i) = (1-\alpha_i)K$ and $k^B = \alpha_i K$. Fixing $\alpha_i \in [0,1]$ and letting $\gamma^*(\alpha_i) = \arg\min_{\gamma_i} L_1(k(\alpha_i), \gamma_i) + L_2(\gamma_i, \hat{\gamma})$ we obtain

$$\begin{aligned}
\gamma^{*0}(\alpha_i) &= (1 - \gamma_i^A(\alpha_i) - \gamma_i^B(\alpha_i))\overline{\theta}; \\
\gamma^{*A}(\alpha_i) &= \frac{(1-\delta)\left(\phi^A(K) + \tau_\theta\right)}{(\phi^A(K)(1-\delta) + \tau_\theta)\left(\tau_\theta + \phi^A((1-\alpha_i)K) + \phi^B(\alpha_i K)\right)} \phi^A((1-\alpha_i)K); \\
\gamma^{*B}(\alpha_i) &= \frac{(1-\delta)\left(\phi^A(K) + \tau_\theta\right)}{(\phi^A(K)(1-\delta) + \tau_\theta)\left(\tau_\theta + \phi^A((1-\alpha_i)K) + \phi^B(\alpha_i K)\right)} \phi^B(\alpha_i K).
\end{aligned}$$

Substituting these optimal values for γ_i into the problem leaves

$$\begin{aligned}
&L_1(k(\alpha_i), \gamma^*(\alpha_i)) + L_2(\gamma^*(\alpha_i), \hat{\gamma}) \\
&= \frac{(1-\delta)\tau_\theta\left(\phi^A(K)^2(1-\delta)\tau_\theta^{-1} + (2 - \delta(1+\alpha_i))\,\phi^A(K) + \tau_\theta + \delta\phi^B(\alpha_i K)\right)}{(\phi^A(K)(1-\delta) + \tau_\theta)^2\left(\tau_\theta + \phi^A((1-\alpha_i)K + \phi^B(\alpha_i K)\right)}.
\end{aligned}$$

Straightforward algebra shows that the latter expression is increasing in α_i if $\tau^A \geq \tau^B$. ■

Proof of Proposition 3.2-(ii). Define

$$\tilde{\delta}\left(\mathcal{L}^B\right) = \frac{e^{K/\mathcal{L}^B}}{1 + e^{K/\mathcal{L}^B}},$$

and

$$G_\alpha(\alpha) = g_A((1-\alpha)K) + g_B(\alpha K). \tag{B.14}$$

By Lemma B.3, a global minimum of L_1 is a payoff-maximizing SBNE. Equivalently, by Lemma B.1, $\hat{k}^A = (1-\hat{\alpha})K$ and $\hat{k}^B = \hat{\alpha}K$ is a payoff-maximizing equilibrium if $G_\alpha(\hat{\alpha}) \geq G(\alpha)$ for all $\alpha \in [0,1]$.

It can immediately be verified that $G_\alpha(\alpha)$ in Eq. (B.14) is strictly convex in α for all $\alpha \in [0,1]$ if $\delta \in (\tilde{\delta}(\mathcal{L}^B), 1)$. Then, the strict convexity of $G_\alpha(\alpha)$ implies that $G_\alpha(\alpha)$ is maximized when α is either zero or one. Therefore, $\alpha = 1$ is the unique payoff-maximizing SBNE if

$$G_\alpha(1) > G_\alpha(0) \Leftrightarrow \frac{\tau^A}{\tau^B}\frac{K}{\mathcal{L}^B} < \frac{e^{K/\mathcal{L}^B} - 1}{e^{K/\mathcal{L}^B}(1-\delta) + \delta} \tag{B.15}$$

Notice that the r.h.s. of the second inequality in Eq. (B.15) is increasing in δ. As $\delta \to 1$ and $\tau^A \geq \tau^B$, Eq. (B.15) holds for all $\mathcal{L}^B < \overline{\mathcal{L}}$, where $\overline{\mathcal{L}} \in (0, \infty]$ solves

$$\frac{\tau^A}{\tau^B}\frac{K}{\overline{\mathcal{L}}} = e^{K/\overline{\mathcal{L}}_S} - 1. \tag{B.16}$$

Hence, for all $\mathcal{L}^B < \overline{\mathcal{L}}$, Eq. (B.15) holds if $\delta \in (\check{\delta}(\mathcal{L}^B), 1)$, where we define

$$\check{\delta}(\mathcal{L}^B) = \frac{e^{K/\mathcal{L}^B}}{e^{K/\mathcal{L}^B} - 1} - \frac{\mathcal{L}^B\tau^B}{K\tau^A}.$$

Combining these results, $\alpha = 1$ is the unique payoff-maximizing SBNE if $\mathcal{L}^B < \overline{\mathcal{L}}_S$ and $\delta \in (\delta_S(\mathcal{L}^B), 1)$, where $\delta_S(\mathcal{L}^B) = \max\{\tilde{\delta}(\mathcal{L}^B), \check{\delta}(\mathcal{L}^B)\}$. ■

Proof of Corollary 3.1 From the definitions of $\delta_S(\mathcal{L}^B)$, $\tilde{\delta}(\mathcal{L}^B)$ and $\check{\delta}(\mathcal{L}^B)$ in the proof of Proposition 3.2-(ii), we have that $\delta_S(\overline{\mathcal{L}}) = 1$ and $\delta_S(\mathcal{L}^B) < 1$ for all $\mathcal{L}^B < \overline{\mathcal{L}}$. The statement in the corollary follows by the continuity of $\delta_S(\mathcal{L}^B)$, which in turn is implied by the continuity of $\tilde{\delta}(\mathcal{L}^B)$ and $\check{\delta}(\mathcal{L}^B)$. ■

References

Admati, A. R., and P. Pfleiderer, 1986, "A monopolistic market for information," *Journal of Economic Theory*, 39(2), 400–438.

Beggs, A. W., 2005, "On the convergence of reinforcement learning," *Journal of Economic Theory*, 122(1), 1–36.

Billingsley, P., 1995, *Probability and measure*. John Wiley & Sons, Hoboken, NJ, 3rd edn.

Chan, L. K. C., N. Jegadeesh, and J. Lakonishok, 1996, "Momentum Strategies," *Journal of Finance*, 51(5), 1681–1713.

Clement, M. B., and S. Y. Tse, 2005, "Financial Analyst Characteristics and Herding Behavior in Forecasting," *Journal of Finance*, 60(1), 307–341.

Colombo, L., G. Femminis, and A. Pavan, 2014, "Information Acquisition and Welfare," *Review of Economic Studies*, 81(4), 1438–1483.

David, P. A., 1985, "Clio and the Economics of QWERTY," *American Economic Review*, 75(2), 332–337.

Dewan, T., and D. P. Mayatt, 2008, "The Qualities of Leadership: Direction, Communication, and Obfuscation," *American Political Science Review*, 102(3), 351–368.

Grossman, S. J., and J. E. Stiglitz, 1980, "On the Impossiblity of Informationally Efficient Markets," *American Economic Review*, 70(3), 393–408.

Grundy, B. D., and M. McNichols, 1989, "Trade and the revelation of information through prices and direct disclosure," *Review of financial Studies*, 2(4), 495–526.

Hellwig, C., S. Kohls, and L. Veldkamp, 2012, "Information Choice Technologies," *American Economic Review Papers and Proceedings*, 102(3), 25–40.

Hellwig, C., and L. Veldkamp, 2009, "Knowing What Others Know: Coordination Motives in Information Acquisition," *Review of Economic Studies*, 76, 223–251.

Hong, H., J. D. Kubik, and A. Solomon, 2000, "Rational Herding in Financial Economics," *RAND Journal of Economics*, 31(1), 121–144.

Hopkins, E., and M. Posch, 2005, "Attainability of boundary points under reinforcement learning," *Games and Economic Behavior*, 53(1), 110–125.

Jegadeesh, N., and W. Kim, 2009, "Do analysts herd? An analysis of recommendations and market reactions," *Review of Financial Studies*, 23(2), 901–937.

Johnson, N. L., and S. Kotz, 1977, *Urn models and their application: an approach to modern discrete probability theory*. Wiley, New York.

Lamont, O. A., 2002, "Macroeconomic forecasts and microeconomic forecasters," *Journal of Economic Behavior & Organization*, 48(3), 265–280.

Mackowiak, B., and M. Wiederholt, 2009, "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99(3), 769–803.

Manzano, C., and X. Vives, 2011, "Public and private learning from prices, strategic substitutability and complementarity, and equilibrium multiplicity," *Journal of Mathematical Economics*, 47(3), 346–369.

Marschak, J., 1974, "Limited Role of Entropy in Information Economics," *Theory and Decision*, 5(1), 1–7.

Mayatt, D. P., and C. Wallace, 2015, "Cournot competition and the social value of information," *Journal of Economic Theory*, 158(B), 466–506.

Morris, S., and H. S. Shin, 2002, "Social Value of Public Information," *American Economic Review*, 92(5), 1521–1534.

Myatt, D. P., and C. Wallace, 2012, "Information Acquisition and Welfare," *Review of Economic Studies*, 79(1), 340–374.

Peng, L., 2005, "Learning with Information Capacity Constraints," *Journal of Financial and Quantitative Analysis*, 40, 307–329.

Peng, L., and W. Xiong, 2006, "Investor Attention, Overconfidence, and Category Learning," *Journal of Financial Economics*, 80, 563–602.

Ritter, F. E., and L. J. Schooler, 2001, "The Learning Curve," in *International Encyclopedia of the Social and Behavioral Sciences*. Pergamon, Amsterdam, pp. 8602–8605.

Shannon, C. E., 1948, "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27(3), 379–423.

Sims, C. A., 2003, "Implication of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690.

Stadje, W., 1990, "The Collector's Problem with Group Drawings," *Advances in Applied Probability*, 22(4), 866–882.

Van der Vaart, A. W., 2000, *Asymptotic statistics*. Cambridge university press, Camridge.

Van Nieuwerburgh, S., and L. Veldkamp, 2010, "Information Acquisition and Under-Diversification," *Review of Economic Studies*, 2, 779–805.

Veldkamp, L. L., 2011, *Information choice in macroeconomics and finance*. Princeton University Press, Princeton, New Jersey.

Welch, I., 2000, "Herding among security analysts," *Journal of Financial Economics*, 58(3), 369–396.

Woodford, M., 2009, "Information-constrained state-dependent pricing?," *Journal of Monetary Economics*, 56(15), 100–124.

Zitzewitz, E., 2001, "Measuring herding and exaggeration by equity analysts and other opinion sellers," Working paper.