Searching for Information*

Jungsuk Han[†] and Francesco Sangiorgi[‡]

May 8, 2015

Abstract

This paper provides a microfounded information acquisition technology based on a simple framework with information search. When searchable information is limited, an agent encounters increasingly more redundant information in his search for new information. Redundancy slows down the learning process and generates decreasing returns. Furthermore, as multiple agents search for information from the same source, limited searchability leads to covariance as the acquired information becomes increasingly more overlapped among agents. Using an asymptotic approach, we construct a tractable mapping from resource (attention) allocations to the precision and the correlation of agents' information under varying degrees of searchability of information. We study two economic applications with endogenous information acquisition using our model: (i) a "beauty contest" coordination game, and (ii) a portfolio optimization problem.

JEL Classification Codes: C65, D80, D81, D83

Keywords: information processing, concavity, precision, asymptotic analysis, coordination games, portfolio choice.

---

*We thank James Dow and Johan Walden for helpful comments and suggestions. We also thank seminar participants at Sveriges Riksbank.

[†]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 a truism that available information for economic decision making is scarce, and acquiring more relevant information is costly. One of such reasons could be the scarcity of cognitive resources such as limited attentional capacity as was pointed out by Kahneman (1973). There also can be lack of other types of resources such as time, budget, skills or even social network that facilitates collecting useful information. Due to all these frictions, available resources should be allocated optimally in searching and processing information to make a more informed decision. Although the economic literature often employs a set of assumptions that allows quantifying the amount of collected information given an input of resources, there has been a scarcity of theoretical justifications to support them. Our paper provides a microfoundation for an information acquisition technology in the presence of impediments to information search, and investigates its implications in economic and financial decision making.

Our framework on learning under imperfect information searchability is based on the following simple intuition: As an agent keeps searching for new information, it is likely that he would encounter some overlapping–and therefore redundant–pieces of information from the past searching activities. The tendency of increasing redundancy results in diminishing returns to scale in information search, and such concavity of informativeness is more pronounced when the total amount of potentially observable information (henceforth, "searchable information") is more limited. Similarly, multiple agents who search in an identical source of information would also face increasing redundancy as more information is collected, thus, their acquired information becomes increasingly more similar. Such commonality of acquired information is more pronounced when 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 with a finite number of balls; drawing a ball from an urn with replacement is interpreted as collecting a signal through search. Because the collected balls are replaced into the urn, the odds of drawing a previously-collected ball increases 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. Because each new signal becomes increasingly less informative on average due to redundancy, the expected overall informativeness is concave in the number of collected signals. Such concavity is more pronounced as searchable information is more limited (or the urn is smaller). Therefore, the expected precision of acquired information becomes concave in the input of resources. However, one may find it rather difficult to apply such a result to most of economic applications because the resulting precision is random and discrete given the input of resources. Considering the limiting case in which each signal becomes infinitesimally small, we can study the continuous limit of a precision function under which the law of large numbers can be applied. This asymptotic approach allows us to obtain a smooth and deterministic function maps the inputs of resources used into the precision of the information that is eventually collected by an agent (see Theorem 2.1). The curvature of this asymptotic precision function decreases with the amount of searchable information, thus a smaller curvature is associated with a higher efficiency of information search.

We also formalize the case with multiple agents using the same logic; imagine that they are independently drawing balls from a single urn which 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, such tendency is more pronounced as the number of balls in the urn becomes 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 degree of information acquisition, and such tendency is more pronounced when information searchability is lower (Theorem 2.2). This information acquisition technology developed in our paper provides a framework of endogenizing "publicity" of acquired information; the correlation between different agents' information is endogenously determined in terms of the input of resources by each agent as well as searchability of information from the source. In the limit case where the number of agents is large, noise in each agent's acquired information can be decomposed into a public component and an i.i.d. idiosyncratic component where all the components are independent of each other (Corollary 2.1). This framework provides a microfoundation for popular setups in the literature such as private signals with perfect or imperfect correlations among agents and private signals with public component of noise.

We study two applications of our information acquisition technology. In the first application, we study a "beauty contest"coordination game with endogenous information acquisition. Our setup follows a 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 the actions, agents can allocate their efforts (or resources) among different information sources to maximize their ex-ante utility. In the recent debate in the related literature, there are mixed messages about the determinacy (or the possibility of multiple equilibria) for coordination games with endogenous information acquisition. Hellwig and Veldkamp (2009) find that the endogenous choice of how much public information to acquire naturally leads to multiple equilibria. In contrast, Myatt and Wallace (2012) assume an information structure similar to the representation that we obtain in Corollary 2.1 and find a unique equilibrium. The results of our analysis lie somewhere in between these two papers. In our setup, the endogenous publicity of the information that is acquired by the agents is not enough to guarantee a unique equilibrium unless the coordination motive is sufficiently weak. When the mapping from agents' effort choices to information is endogenously derived as in our paper, we show how non-concavities arise that can lead to multiple equilibria. Hence, the microfoundation provided in our paper is relevant because it leads to qualitatively different implications. These implications shed new light on the relation between the information structure and equilibrium determinacy in coordination games with endogenous information.

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 leaning 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 into more accurate sources of information because information from a less accurate source is more likely to be correlated due to imperfect information searchability.

In the second application, we study a speculative investor's problem in which he allocates his resources to acquire information across risky assets before forming an optimal portfolio. Under-diversification arises naturally as the investor concentrates his portfolio on a few assets that allow him to exploit private information best. In particular, he would concentrate more in case there is a larger difference in information searchability among assets. On the other hand, we also find that the investor tends to diversify his portfolio in case he is endowed with a larger resource capacity. Finally, an attention crowding out effect emerges, by which an increase in the supply of information for one asset results in the agent devoting less resources to other assets. Our results can potentially shed light on the empirical findings that retails investors generally hold highly-concentrated portfolio on a few stocks as well as other familiarity biases in individual and institutional portfolio holdings.

In the literature, there have been various formulations that quantify the amount of collected (or processed) information. One of well-known attempt is rational inattention theory 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 practical usefulness, there have been numerous applications in economics and finance based on such rational inattention framework.[1] For example, Peng (2005) shows a financial equilibrium with information capacity constraints. While the entropy is a convenient and useful measure of information, it may not reflect relevant features of learning problems such as impediments to information search in our model. As was pointed by Marschak (1974), entropy is only relevant to the cost of communicating rather than the cost of searching and collecting information. The original idea of Shannon (1948) is not meant for information processing in economic contexts. In that context, the entropy measure is suitable as communicating constraint rather than information acquisition constraint. Shannon's second proposition asserts that we can transfer a signal without errors under certain capacity limit by using an optimal coding scheme. This optimal coding scheme may not apply to natural languages or economic and financial terms. Developments in information technology imply that we bear minimum costs of communicating; however, natural language is far from being an optimal coding scheme. More importantly,

[1] See, for example, Veldkamp (2011) for an excellent survey on applications in macroeconomics and finance.

entropy is silent about searching and collecting information. In contrast, our model provides a microfoundation for measuring information in an environment where information search is important. Our results complement the existing literature (such as those in rational attention), and contribute to it by finding further implications that arise from impediments to information search.

The organization of the paper is as follows. Section 2 develops the framework of information acquisition under imperfect information searchability. Section 3 studies a coordination game with complementarities as an example of possible applications. Section 4 studies a portfolio optimization problem as an example of possible applications. Section 5 concludes.

2 Information Search

In this section, we develop our methodology and characterize endogenous precision and covariance of information under imperfect information searchability. We begin by describing the basic setup, then derive the asymptotic precision function and the asymptotic covariance function. Finally, we provide a public-private decomposition of the resulting information structure.

2.1 The Setup

2.1.1 Basic signals

Consider an economic agent who is endowed with limited information processing resources (henceforth, resources) that allow the agent to search necessary information. The agent acquires information using his resources 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 variance τ_θ^{-1}. 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 $L \in \mathbb{N}$ distinct signals on θ. Each basic signal $m \in \{1, 2, \ldots, L\}$ in $\mathbf{L}$ is given by

$$s^m = \theta + \varepsilon^m, \tag{1}$$

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

2.1.2 Information searchability

We define information search as a process of gathering information by using endowed resources. We say that information search is easier (or harder) if the same input of resources resolves a greater (or smaller) amount of uncertainty about the variable of interest. Therefore, "information searchability" is defined as the degree of resolution of uncertainty with respect to the increase in resource input. Assuming that a fixed amount of resource input gives the agent a signal in Eq. (1), information searchability is inversely related to the possibility of encountering redundant signals in his search process.

We formalize the idea of information searchability 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. Imagine that the agent is sequentially drawing a ball with replacement from the urn. The agent can identify the index of each signal, thus he knows whether a signal is redundant or not given the set of collected signals. 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 would get smaller as the agent draws more balls from the urn.

The following assumption embeds this idea of impediments to information search by allowing redundancy among acquired signals.

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. The opposite case to this assumption is sampling without replacement, in which case acquired signals are never redundant. In this case, any increases in the input of resources would be directly translated into a greater amount of information (or a greater resolution of uncertainty) regardless of the scale of resource investment. On the other hand, the assumption of sampling with replacement in Assumption 2.1 makes it impossible to maintain the constant returns to scale in information search.

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 them. 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:

$$S(h) = \frac{1}{h}\sum_{m\in H} s^m = \theta + \frac{1}{h}\sum_{m\in H}\varepsilon^m. \tag{2}$$

Notice that $S(h)$ is a sufficient statistic for the signals acquired by the agent. By the standard Bayesian belief update formula, the precision of the posterior belief about θ conditional on observing the signal $S(h)$ is given by

$$Var[\theta|S(h)]^{-1} = \underbrace{\tau_\theta}_{\text{precision of prior belief}} + \underbrace{\tau_\varepsilon h}_{\text{signal precision}}. \tag{3}$$

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 $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is defined by*

$$\Phi(h) = \tau_\varepsilon h, \tag{4}$$

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 $\mathbf{L}$ which 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*

$$E[\tilde{h}] = L\left[1 - \left(1 - \frac{1}{L}\right)^l\right]. \tag{5}$$

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

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

□

Notice that $E[\tilde{h}]$ is monotone increasing and concave in l.[2] That is, by drawing more signals from the urn, the expected number of distinct signals increases, but it does at a decreasing rate as more and more signals are being collected. Furthermore, $E[\tilde{h}]$ is monotone increasing in L. Intuitively, the more independent signals are in the urn, the higher the expected number of distinct signals for a given number of draws. Hence, the number of signals in $\mathbf{L}$ reflects the easiness of searching information (or the degree of 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 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 applied to various economic applications with ease.

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 smaller amount of resources which is less than c cannot

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

$$\frac{\partial E[\tilde{h}]}{\partial l} = -L\left(1 - \frac{1}{L}\right)^l \log\left(1 - \frac{1}{L}\right) > 0, \quad \text{and} \quad \frac{\partial^2 E[\tilde{h}]}{\partial l^2} = -L\left(1 - \frac{1}{L}\right)^l \left[\log\left(1 - \frac{1}{L}\right)\right]^2 < 0. \tag{7}$$

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.[3]

If an agent could observe all signals in $\mathbf{L}$, the agent's posterior precision in (3) would be $\tau^* \equiv \tau_\theta + \tau_\varepsilon L$. Therefore, τ^* is the upper bound to an agent's information precision, which is regarded as an exogenous feature of the underlying informational environment. As we elaborate below, we will consider the behavior of the precision function as the cost c becomes small while leaving τ^* unchanged. Accordingly, in the next two assumptions we relate the number and precision of basic signals to the cost c.[4]

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

Because $\mathcal{L}$ determines the number of signals in $\mathbf{L}$, $\mathcal{L}$ parameterizes the degree of information searchability. All the collected signals will be distinct (thus informative) in the absence of impediments to information search (or in case of perfect searchability, i.e., $\mathcal{L} = \infty$). On the other hand, some of the collected signals may be redundant in the presence of impediments to information search (or in case of imperfect searchability, i.e., $\mathcal{L} < \infty$), thus the number of distinct signals can be strictly lower than the number of collected signals.

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

Therefore, 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. For example, when the required input of resources for one signal decreases by half, the number of basic signals available in the population increases by twice but the precision of each basic signal decreases by half.

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 minimum fraction of resource inputs is set to be c. Then, the precision function according to Definition 2.1 under Assumption 2.3 is given by

$$\Phi(\tilde{h}(k; c)) = \tau c \tilde{h}(k; c). \tag{8}$$

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

[4] This following is a technical assumption which obtains smooth extrapolation of discrete choices. Moscarini and Smith (2001) provides a similar extrapolation out of number of sampling.

There are two major problems in using the precision function in Eq. (8): 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. (8) 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 $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is defined to be a function that satisfies the following*[5]

$$\Phi(\tilde{h}(k;c)) \to \phi(k) \ \ a.s. \ \ as \ c \to 0. \tag{9}$$

As we will demonstrate below, the asymptotic precision function $\phi(k)$ in Eq. (9) resolves both problems in Eq. (8). That is, it becomes deterministic and smooth (i.e., continuous in k and also infinitely differentiable with respect to k).

2.2 Asymptotic Precision Functions

Here we derive our main result, 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 **L** *without* replacement. Since there cannot be redundant signals in this case, the number of distinct signals drawn from **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

$$\Phi(\tilde{h}(k;c)) = \tau c \left\lfloor \frac{k}{c} \right\rfloor = \tau \Big[k - g(c)\Big], \tag{10}$$

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

$$\phi(k) = \tau k. \tag{11}$$

[5] 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 $\alpha > 0$.

Now, we turn to the case of imperfect information searchability, i.e., 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:

$$E[\tilde{h}(k;c)] = \lfloor \frac{\mathcal{L}}{c} \rfloor \left(1 - \left(1 - \frac{1}{\lfloor \frac{\mathcal{L}}{c} \rfloor} \right)^{\lfloor \frac{k}{c} \rfloor} \right). \tag{12}$$

Multiplying by c and taking the limit as c goes to zero in Eq. (12) yields

$$E[c\tilde{h}(k;c)] \to \mathcal{L} \left(1 - \exp\left(-\frac{k}{\mathcal{L}} \right) \right), \quad \text{as } c \to 0. \tag{13}$$

Therefore, the expectation of the precision function $\Phi(\tilde{h}(k;c))$ becomes smooth in the limit where 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 showing 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 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 case of imperfect information searchability, the asymptotic precision is given by*

$$\phi(k) = \tau\mathcal{L} \left(1 - \exp\left(-\frac{k}{\mathcal{L}} \right) \right) \tag{14}$$

As mentioned above, the asymptotic precision function $\phi(k)$ overcomes the two major difficulties that exist in 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: $\partial^2\phi/\partial k^2 \leq 0$,
(iv) Curvature: $-\frac{\partial^2\phi/\partial k^2}{\partial\phi/\partial k} = 1/\mathcal{L}$.

The precision increases with more input of resources, but the marginal benefits diminish in scale. Furthermore, such diminishing marginal benefits are larger with worse information searchability. These properties fit intuition quite well. As one learns more about one subject, the probability of encountering redundant materials is going up. He realizes that the collected materials are overlapping with those that are previously acquired only after searching. The concavity of the signal precision function is also associated with 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 are best fitted with an exponential function.[6]

The increasing curvature inverse to $\mathcal{L}$ implies that worse information searchability would make the asymptotic precision function more concave. On the other hand, as information searchability improves (i.e., $\mathcal{L} \to \infty$), the asymptotic precision function (14) converges to the linear function in (11) that is obtained when signals are drawn without replacement. This result is intuitive: as information searchability deteriorates, so does the possibility of drawing redundant signals. We remark that (14) 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 $\mathcal{L}$.

In applications, it is common to specify the information acquisition technology in terms of a cost function $k : \phi \mapsto k(\phi)$ that specifies the amount of resources k that are required

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

to collect information with precision ϕ. In our setup, the asymptotic cost function is readily obtained as the inverse of the precision function in (14), $k(\cdot) = \phi^{-1}(\cdot)$, as

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

The cost function $k(\phi)$ has the following properties: it is non-negative, monotone decreasing and concave, 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 to the information precision. Finally, we remark that $k'(0) = \tau^{-1}$. Differently from what is often assumed in applications, 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 functions are linear in the case of perfect searchability ($\mathcal{L} = \infty$) while the curvature decreases for larger values of $\mathcal{L}$, the precision function becoming less concave and the cost function becoming less convex.

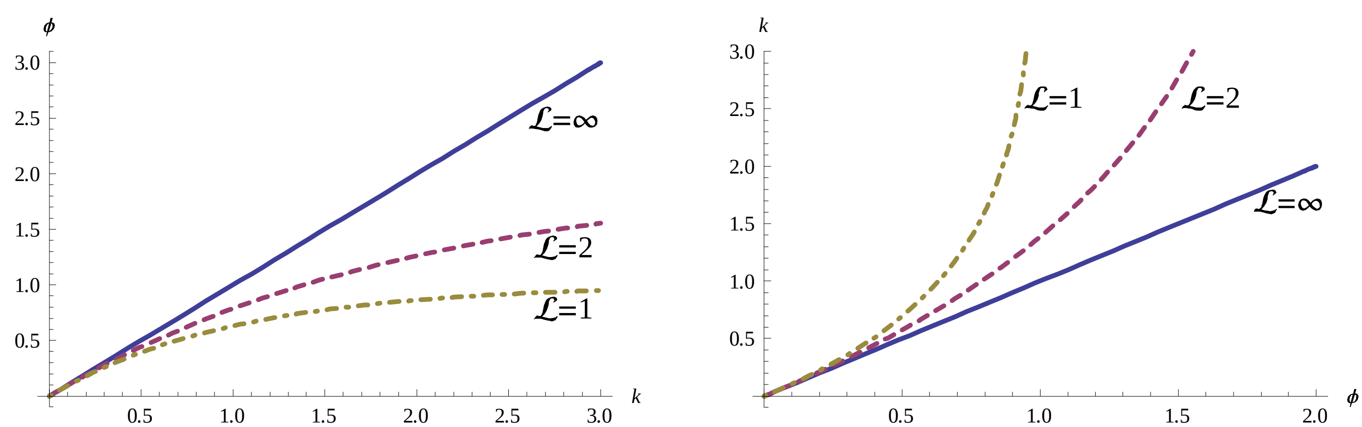

FIGURE 1. Left panel: precision function $\phi(k)$. Right panel: cost function $k(\phi)$. Parameter values: $\tau = 1$ and $\mathcal{L} \in \{1, 2, \infty\}$.

Finally, we call an information source "superior" to the other 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 equality being strict.*

A superior information source will always have higher precision given the same level of inputs, i.e., suppose that i is superior to j, then $\phi^i(k) > \phi^j(k)$ for all $k > 0$. One may imagine that a superior source is always preferred to an inferior source, but later in the paper (Section 3) we show that is 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 collecting overlapping information among different agents 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, let H_i denote the set of distinct signals acquired by agent i, and h_i denote 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 agent i and j:

$$S_i(h_i) = \frac{1}{h_i} \sum_{m \in H_i} s^m = \theta + \frac{1}{h_i} \sum_{m \in H_i} \varepsilon^m, \tag{16}$$

$$S_j(h_j) = \frac{1}{h_j} \sum_{m \in H_j} s^m = \theta + \frac{1}{h_j} \sum_{m \in H_j} \varepsilon^m. \tag{17}$$

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

$$Cov(S_i(h_i), S_j(h_j)) = \frac{1}{\tau_\theta} + Cov\left(\frac{1}{h_i} \sum_{m \in H_i} \varepsilon^m, \frac{1}{h_j} \sum_{n \in H_j} \varepsilon^n\right). \tag{18}$$

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

$$Cov\left(\frac{1}{h_i}\sum_{m\in H_i}\varepsilon^m, \frac{1}{h_j}\sum_{n\in H_j}\varepsilon^n\right) = \frac{1}{h_i h_j}Var\left(\sum_{m\in H_{i,j}}\varepsilon^m\right) = \frac{h_{i,j}}{\tau c h_i h_j}, \tag{19}$$

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

Suppose that agent i and j use an amount of resource k_i and k_j, respectively, when the cost of each signal is set to be c. Denote with $\tilde{h}_i(k_i;c)$ and $\tilde{h}_j(k_i;c)$ the resulting number of distinct signals collected by agent 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. 19). 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*

$$\frac{\tilde{h}_{i,j}(k_i,k_j;c)}{\tau c\tilde{h}_i(k_i;c)\tilde{h}_j(k_j;c))} \to \sigma_{ij} \ \text{a.s.} \ \text{as } c \to 0. \tag{20}$$

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.,*

$$c\tilde{h}_{i,j}(k_i,k_j;c) \to \mathcal{L}\left(1-\exp\left(-\frac{k_i}{\mathcal{L}}\right)\right)\left(1-\exp\left(-\frac{k_j}{\mathcal{L}}\right)\right) \ \text{a.s.,} \ \text{as } c \to 0. \tag{21}$$

Proof. See Appendix A. □

Then, Lemma 2.2 and Lemma 2.3 together with Eq. (19) provide the proof for 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*

$$\sigma_{ij} = \frac{1}{\tau \mathcal{L}}. \tag{22}$$

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:

$$\rho(k_i, k_j) = \lim_{c \to 0} Corr\left(\frac{1}{h_i}\sum_{m \in H_i} \varepsilon^m, \frac{1}{h_j}\sum_{m \in H_j} \varepsilon^m\right) = \left[\left(1 - \exp\left(-\frac{k_i}{\mathcal{L}}\right)\right)\left(1 - \exp\left(-\frac{k_j}{\mathcal{L}}\right)\right)\right]^{\frac{1}{2}}. \tag{23}$$

2.4 Asymptotic Normality

While Theorem 2.1 and Theorem 2.2 show the second moments of the error terms in the asympotic signals, the proof of the next theorem derives their joint asymptotic distribution. We can summarize our results as follows:

Theorem 2.3. *For each agent pair $i, j \in \mathbf{I}$ using input 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*

$$S_i(k_i) = \theta + \varepsilon_i, \tag{24}$$

$$S_j(k_j) = \theta + \varepsilon_j, \tag{25}$$

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

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

where the function $\phi(\cdot)$ is as in Eq. (14).

2.5 A public-private decomposition of signals' noise terms

From Theorem 2.3, we can represent each asymptotic signal $S_i(k_i)$ by decomposing the noise term ε_i into two parts as follows:

$$S_i(k_i) = \theta + \mu + \eta_i, \tag{26}$$

where μ is a common component of noise, and η_i is an idiosyncratic component of noise such that

$$\mu = \frac{1}{I}\sum_{i=1}^{I} \varepsilon_i, \qquad \eta_i = \varepsilon_i - \mu. \tag{27}$$

The next corollary provides a characterization of the representation in Eq. (26) for an economy with a large number of agents.

Corollary 2.1. *When the number of agents I goes to infinity, the noise decomposition in Eq. (26) satisfies the following properties:*

(i) μ and η_i become independent of each other for all $i \in I$,

(ii) η_i and η_j become independent of each other for all $i, j \in I$,

(iii) The precision of η_i approaches $\tau\mathcal{L}\left(\exp\left(k_i/\mathcal{L}\right) - 1\right)$ for all $i \in I$,

(iv) The precision of μ approaches $\tau\mathcal{L}$.

Proof. See Appendix A. □

Parts (i) and (ii) in the corollary show that the representation in Eq. (26) separates out the original individual error term in each signal (the ε_i's in Corollary 2.3) into the sum of one component that is common across all agents, μ, and a truly idiosyncratic component, η_i. By parts (iii) and (iv) in the corollary, 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 extent of information searchability.

Corollary 2.1 gives a microfoundation for a signal structure that blends together the two common assumptions used in the literature, in which the error terms in the signals are typically assumed to be either fully private (i.e., purely idiosyncratic noise) or fully public (i.e., purely public noise). A signal structure with similar properties to ours has been assumed in recent work by Manzano and Vives (2011) and Myatt and Wallace (2012).

For a given input of resources k_i, the public-private nature of a signal can be further described as follows. The fraction of a signal's error term variance that is attributed to the common part is

$$\frac{Var(\mu)}{Var(\varepsilon_i)} = 1 - \exp\left(-\frac{k_i}{\mathcal{L}}\right). \tag{28}$$

Hence, the smaller the searchability parameter $\mathcal{L}$, the more the error terms the signals become "public" across agents. 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.

It is worth noting that the precision of η_i in Corollary 2.1 is a convex function of k_i. To intuitively understand this, one could interpret η_i as the error term of a signal on $\theta + \mu$. Consider the limiting cases in which $\mathcal{L}$ is either very large or very small. As $\mathcal{L}$ becomes very large, $\theta + \mu$ becomes equivalent to θ because the common component μ vanishes. Then, the precision of η_i becomes equivalent to the precision of the error term ε_i in the original signal, which approaches the linear function τk_i as $\mathcal{L} \uparrow \infty$. In contrast, as $\mathcal{L}$ is very small, the error term in the original signal is completely dominated by the common noise (see Eq (28)). In the limit as $\mathcal{L} \downarrow 0$, the content of the information source becomes akin to a single "small" noisy signal of the form $\theta + \mu$, which can be learnt with any positive amount of resources spent on information search. This case coresponds to an extreme form of convexity for the precision function of η_i. Finite values of $\mathcal{L}$ correspond to cases between these two extremes. For all $\mathcal{L} \in (0, \infty)$, the precision function of η_i is then strictly convex, with curvature that increases for lower values of $\mathcal{L}$.

3 Application I: 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 complements the existing literature (e.g., Hellwig and Veldkamp (2009) and Myatt and Wallace (2012)) 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:

$$u_i = -(1-\delta)(\theta - a_i)^2 - \delta(\overline{a} - a_i)^2 . \tag{29}$$

The parameter $\delta \in [0,1]$ in Eq. (29) measures the intensity of agents' coordination motive: larger values of δ reflect larger concerns for an agent to choose an action that is as close as possible to the average action. We assume $\theta \sim N\left(\bar{\theta}, \tau_\theta^{-1}\right)$. To gather 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 $k_i^j > 0$ resources to information source j, the information obtained through this source is equivalent to a signal of the form

$$S_i^j = \theta + \varepsilon_i^j; \qquad \varepsilon_i^j \sim N\left(0, \phi^j\left(k_i^j\right)^{-1}\right), \tag{30}$$

where the precision function $\phi^j(\cdot)$ is as specified in Eq. (14) in Theorem 2.1,

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

Vice versa, agent i ignores information source j whenever $k_i^j = 0$, in which case the signal S_j^i is pure noise.

The error terms in Eq. (30) are assumed to be independently distributed across information sources. On the other hand, within each information source, the decomposition in Corollary 2.1 implies that we can write each signal S_i^j as

$$S_i^j = \theta + \mu^j + \eta_i^j, \tag{31}$$

where μ^j and η_i^j are independent for all j and i and

$$\mu^j \sim N\left(0, \left(\tau^j \mathcal{L}^j\right)^{-1}\right); \qquad \eta_i^j \sim N\left(0, \exp\left(-k^j/\mathcal{L}^j\right)\phi^j \left(k_i^j\right)^{-1}\right). \tag{32}$$

An interpretation of the setup in this section is that of financial analysts or professional forecasters issuing their forecasts on some random variable of interest (e.g., earnings per share on a given stock, a macroeconomic aggregate, a commodity price etc.). In this interpretation, agent i's action is her forecast, and the average action $\bar{a}$ is the consensus forecast. Resources K would be total working hours. Intuitively, an analyst's payoff depends on the accuracy of her forecast. However, the reputational damage from a wrong forecast is more severe if the analyst's forecast is different from the consensus. The payoff function in Eq. (29) is a second order approximation of these concerns.

3.2 Equilibrium

In line with the literature, we focus on equilibria in which actions are affine functions of the signals, that is, in which agent i's action takes the form $a_i = \gamma_i^0 + \Sigma_j \gamma_i^j S_i^j$.[7] 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 some strategy $\left(\hat{k}, \hat{\gamma}\right)$, agent i's ex-ante utility from a strategy (k_i, γ_i) equals

$$E\left(u_i\right) = -L_1\left(k_i, \gamma_i\right) - L_2\left(\gamma_i, \hat{\gamma}\right), \tag{33}$$

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

where $L_1(k_i, \gamma_i)$ and $L_2(\gamma_i, \hat{\gamma})$ are given in Eqs. (B.1)-(B.2) in Appendix B. (Appendix B also contains the derivation of Eq. (33).) $L_1(k_i, \gamma_i)$ is the quadratic loss experienced by an agent when all players play the same strategy. $L_2(\gamma_i, \hat{\gamma})$ is the quadratic loss experienced by an agent when he deviates from other players' strategy. A Symmetric Bayesian Nash Equilibrium (SBNE or equilibrium, hereafter) is a strategy $\left(\hat{k}, \hat{\gamma}\right)$ such that

$$\left(\hat{k}, \hat{\gamma}\right) \in \arg \min_{(k_i, \gamma_i) \in \Gamma} L_1(k_i, \gamma_i) + L_2(\gamma_i, \hat{\gamma}). \tag{34}$$

Since $L_2(\gamma_i, \hat{\gamma})$ vanishes when agent i plays $\gamma_i = \hat{\gamma}$ (and is strictly positive otherwise), a global minimizer of $L_1(k_i, \gamma_i)$ in Eq. (34) is a *payoff maximizing equilibrium*.[8] In Appendix B (see Lemma B.1) we 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

$$k^* \in \arg \max_{k \in \Delta} G(k), \tag{35}$$

where we define

$$G(k) = \Sigma_{j=1}^J g_j(k^j); \qquad g_j(k^j) = \left[\frac{1-\delta}{\phi^j(k^j)} + \frac{\delta}{\exp\left(k^j/\mathcal{L}^j\right)\phi^j(k^j)}\right]^{-1}. \tag{36}$$

The problem in Eqs. (35)-(36) is that of finding an allocation of resources that maximizes agents' payoffs in a symmetric equilibrium. Because an agents' local deviation from a symmetric strategy profile has no first-order effects on $L_2(\gamma_i, \hat{\gamma})$ (see Eq. (B.3)), k^* in Eq. (35) is the unique SBNE if $G(k)$ is strictly concave in k.

The objective function $G(k)$ in Eq. (36) has an intuitive interpretation. Each function $g_j(k^j)$ is an (weighted, harmonic) average of the precisions of the error term ε_i^j of the signal in Eq. (30) and of the idiosyncratic error term η_i^j in the decomposition in Eq. (31). When agents do not care about other agents' actions, forecasting the fundamental θ is all that matters for agents' utility. In this case, $G(k)$ is simply the sum of signals' precisions ($g_j(\cdot) = \phi^j(\cdot)$ for $\delta = 0$, see Eq. (36)) and k^* is the allocation of resources that gives the most precise forecast of θ. Because the precision functions $\phi^j(\cdot)$ are concave, $G(k)$ is concave, and thus the equilibrium unique, if the coordination motive is sufficiently weak.

[8] 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.

On the other hand, coordination motives in actions introduce a distortion. When the coordination motive is strong, agents' utilities depend crucially on whether their actions deviate from the average action. 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 learning about the sums $\theta + \mu^j$. For instance, in the extreme case $\delta = 1$, $G(k)$ is simply the sum of the idiosyncratic precisions ($g_j(k_i^j) = var\left(\eta_i^j\right)^{-1}$ for $\delta = 1$, see Eq. (36) and Eq. (32)) and k^* is the allocation of resources that gives the highest precision on the sums $\theta + \mu^j$. As explained in Section 2.3.1, learning about the sums $\theta + \mu^j$ introduces a source of non-concavity because uncertainty about $\theta + \mu^j$ decreases faster than it does for θ. A strong enough coordination motive makes the problem in Eq. (35) non-concave, which can result in multiple equilibria. For instance, 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 deter deviation.

This discussion suggests that the interplay between the coordination motive and the nature of the information is key for information choices and equilibrium uniqueness. The next propositions analyze this interplay formally.

Proposition 3.1. *(i) When the coordination motive is weak, i.e., for $\delta \in [0, 1/2)$, there exists a unique equilibrium. In equilibrium, information acquisition satisfies*

$$\hat{k}^j(\lambda) = \begin{cases} \mathcal{L}^j \log\left(\frac{\tau^j - 2\lambda\delta(1-\delta) + \sqrt{\tau^j(\tau^j - 4\lambda\delta(1-\delta))}}{2(1-\delta)^2\lambda}\right) & \text{for } 0 < \lambda < \tau^j \\ 0 & \text{for } \lambda \geq \tau^j \end{cases} \tag{37}$$

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 $\delta \in [1/2, 1)$, there may be multiple equilibria. An equilibrium allocation of resources is either a local maximum or a critical point of $G(k) = \Sigma_{j=1}^J g_j(k^j)$.

Proof. See Appendix B. □

Proposition 3.1-(i) confirms the intuition from the previous discussion. The equilibrium is unique if the coordination motive is not sufficiently strong. In this equilibrium, agents devote attention to an information source only if this source is sufficiently efficient. As agents care more about other agents' actions, agents allocate their resources among a (weakly) lower number of information sources (i.e., $\frac{d\lambda}{d\delta} > 0$). If information source j is superior to information source i according to Definition 2.3, then information source j gets more resources than i. If an information source is superior to all other information sources and has perfect searchability, it gets all resources.

Proposition 3.1-(ii) reveals the possibility of multiple equilibria when the coordination motive is sufficiently strong. In such case, an equilibrium allocation of resources may have very different properties than the ones discussed above. For example, an equilibrium allocation of resources may favor an inferior information source. 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 is superior to B in that it is relatively more efficient and has perfect searchability ($\tau^A > \tau^B$ and $\mathcal{L}^A = \infty$). Then:
(i) $\hat{k}^A = K$ is an equilibrium.
(ii) There exists a threshold $\bar{\mathcal{L}}_N < \infty$ such that $\hat{k}^B = K$ is an equilibrium only if $\mathcal{L}^B \leq \bar{\mathcal{L}}_N$.
(iii) There exists a threshold $\bar{\mathcal{L}}_S < \infty$ such that, for all $\mathcal{L}^B < \bar{\mathcal{L}}_S$, $\hat{k}^B = K$ is the payoff maximizing equilibrium if the coordination motive is sufficiently strong.

Proof. See Appendix B. □

Since information source A has perfect searchability, the nature of this information is purely private. Then, when $k_i^A = K$ for all agents, an agent's payoffs only depend on the fundamental θ, and deviating from $k_i^A = K$ is costly because A is a superior to B. Devoting all resources to this superior technology is therefore an equilibrium.

The necessary condition in Proposition 3.2-(ii) requires that the inferior information source must be sufficiently public in nature. The intuition is as follows. For $k^B = K$ to be an equilibrium, it must be a local maximum of $G(k)$ (Proposition 3.1-(ii)). Instead, as $\mathcal{L}^B$ becomes large and information source B becomes private in nature, predicting $\theta + \mu^B$ is the same as predicting θ. Because $\tau^A > \tau^B$, agents would be better off in a symmetric equilibrium in which $\hat{k}^A > 0$.

The intuition for Proposition 3.2-(iii) is the following. If all agents are learning from the superior information source A, then the average action only depends on θ. Instead, if all agents are learning from the inferior information source B, then the average action depends on $\theta + \mu^B$. For $\mathcal{L}^B < \bar{\mathcal{L}}_S$, $k_i^B = K$ gives a more precise source of information about $\theta + \mu^B$ than $k_i^A = K$ does about θ. If agents' concerns about predicting the average action are sufficiently strong, $\hat{k}^B = K$ is an equilibrium that gives higher payoffs to agents than the $\hat{k}^A = K$ equilibrium. (Under the conditions in the proposition, the problem in Eqs. (35)-(36) is strictly convex, and interior equilibria are dominated.)

How strong does the desire for coordination need to be in Proposition 3.2-(iii)? The answer depends on the parameters of the model, including the searchability parameter $\mathcal{L}^B$. The lower $\mathcal{L}^B$, the less precise a B-signal is about θ. At the same time, lower $\mathcal{L}^B$ makes a B-signal a more precise source of information about $\theta + \mu^B$. The following corollary shows that the second effect dominates the first one when $\mathcal{L}^B$ is close to the threshold $\bar{\mathcal{L}}_S$. We have:

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

Proof. See Appendix B. □

Corollary 3.1 underlies the following surprising result. Since a B-signal's precision is increasing in $\mathcal{L}^B$, then, over a range of values, an even lower precision of the inferior information source is associated with a *larger* set of parameters for which agents choose this information in the payoff maximizing equilibrium. The left panel of Figure 2 provides and illustration. The figure further shows that the same comparative statics holds when $\hat{k}^B = K$ is a SBNE, without the further requirement of being the payoff maximizing SBNE.

Equilibrium information choices have implications for aggregate volatility, as illustrated by the right panel in Figure 2. The aggregate action is perfectly correlated with the fundamental when resources are fully invested in the superior information. Instead, 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 $\mathcal{L}^B$.

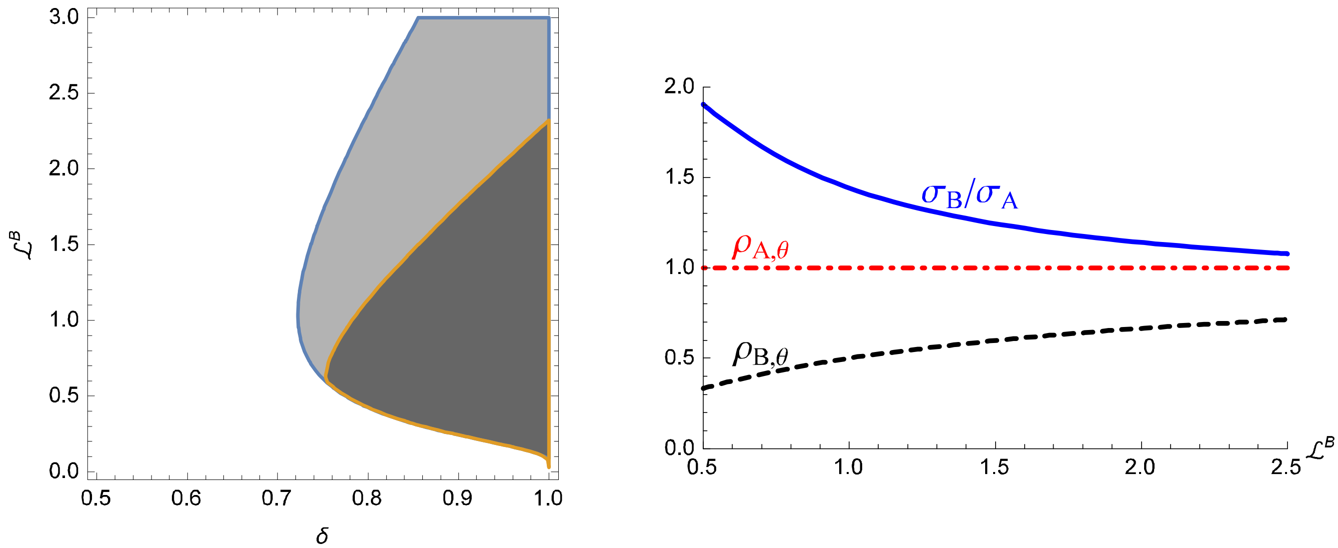

FIGURE 2. Left panel: light grey area: $\hat{k}^B = K$ is a SBNE; dark grey area: $\hat{k}^B = K$ is the payoff maximizing SBNE. Right panel: thick line, σ_B/σ_A: relative volatility of the aggregate action across $\hat{k}^B = K$ and $\hat{k}^A = K$ equilibria; dashed (dot-dashed) line, $\rho_{B,\theta}$ ($\rho_{A,\theta}$): 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.

In the interpretation of this model as one of financial analysts or professional forecasters, the superior information technology could be fundamental analysis, and the inferior information source could be information originating from the media (e.g., social media, traders' online chat rooms, popular business TV shows). Proposition 3.2 and Corollary 3.1 describe an equilibrium in which analysts gather all their information from the inferior source. In this equilibrium, the informational content of the consensus forecast is

$$Var[\theta|\bar{a}]^{-1} = \frac{\tau_\theta}{1 - corr\,(\theta, \bar{a})^2},$$

where $corr(\theta, \bar{a})$ denotes the correlation coefficient between θ and $\bar{a}$. Hence, the correlation coefficients in Figure 2 measure the informational efficiency of the equilibrium outcome. While the consensus forecast reveals the fundamental perfectly in the $\hat{k}^A = K$ equilibrium, the equilibrium outcome informationally inefficient in the $\hat{k}^B = K$ equilibrium.

3.3 Relation to the literature

The analysis in this section relates very closely to Hellwig and Veldkamp (2009) and Myatt and Wallace (2012). Both papers consider endogenous information acquisition in a beauty contest coordination game.

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.[9]

A very different message emerges from Myatt and Wallace (2012), who assume a signal structure equivalent to Eq. (31) 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 signals actually acquired in equilibrium may be relatively public in nature."

The message from Proposition 3.1 is somewhat intermediate between these two papers. In our setup, the endogeneity of the signal structure in Eq. (31) leads to non-concavities that are absent in Myatt and Wallace (2012). The endogenous publicity of signals guarantees a unique equilibrium only if the coordination motive is sufficiently weak.

[9] 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).

4 Application 2: Portfolio Choice with Endogenous Information

In this section, we study an example of resource allocations under imperfect information searchability in the context of portfolio optimization. There has been a large volume of literature on portfolio choice with endogenous information acquisition (e.g., Grossman and Stiglitz (1980), Peress (2004), Peng (2005) , Van Nieuwerburgh and Veldkamp (2009), Van Nieuwerburgh and Veldkamp (2010) and Mondria (2010)). Our results complement those of the existing papers by suggesting an alternative mechanism that affects investors' optimal portfolio decision making through information searchability.

4.1 The setup

Consider a competitive investor who wants to allocate his resources optimally in order to maximize his expected utility from investment in a one-period setup ($t = 0, 1$). There exist N risky assets, and there also exists a riskless asset whose gross return is normalized to one. Each asset i generates a random payoff θ^i at $t = 1$ that follows a normal distribution with mean $\bar{\theta}^i$ with variance $1/\tau^i$. Without loss of generality, the expected payoff of each risky asset is set to be zero, i.e., $\bar{\theta}_i = 0$. For simplicity, we assume that the payoff each asset is independent of each other, i.e., $Cov(\theta^i, \theta^j) = 0$ for all i, j. We denote p^i to be the price of each asset in the market. To focus on the effects of learning from costly information, we assume that the price of each asset i does not provide any additional information to the agent about the asset payoff and set $p^i = E[\theta^i] = 0$.[10] Since asset payoffs are independent, the agent's optimal portfolio weight in asset i is zero unless the agent acquires information on that asset.

The investor is endowed with an initial wealth w_0, and has a CARA utility function such that $u(w_1) = -\frac{1}{\gamma}\exp(-\gamma w_1)$ where w_1 is the trader's wealth at $t = 1$. At $t = 0$, the investor collects information on the risky assets and, once this information is observed, the investor forms a portfolio using his initial wealth w_0. At $t = 1$, he receives the realized payoff of his portfolio.

[10] One can consider a situation where the trader submits orders given prices posted by competitive risk-neutral market makers.

The investor is endowed with a total amount of resources K which can be allocated to gather information about the risky assets:

$$\sum_{i=1}^{N} k^i \leq K, \tag{38}$$

where k^i denotes the amount of resource allocated for asset i. By Corollary 2.3 , for a given an allocation $\{k^1, k^2, ..., k^N\}$ the information collected by the investor for each asset i is equivalent to a noisy signal on the fundamental value of each asset i,

$$S^i = \theta^i + \varepsilon^i, \tag{39}$$

where ε^i follows a normal distribution with mean zero and precision $\phi^i(k^i)$ such that

$$\phi^i(k^i) = \tau^i \mathcal{L}^i \left(1 - \exp\left(-\frac{k^i}{\mathcal{L}^i}\right)\right). \tag{40}$$

For notational convenience, we define $\hat{\tau}^i = \tau^i/\tau^i_\theta$ as the normalized signal efficiency. By solving the resource allocation problem, we find:

Proposition 4.1. *There exists a unique optimal allocation of resources for the risky assets such that the optimal allocation k^i is given by*

$$k^i(\lambda) = \begin{cases} \mathcal{L}^i \log\left(\frac{\frac{\hat{\tau}^i}{\lambda} + \hat{\tau}^i \mathcal{L}^i}{1 + \hat{\tau}^i \mathcal{L}^i}\right) & \text{if } 0 < \lambda < \hat{\tau}^i \\ 0 & \text{if } \lambda \geq \hat{\tau}^i \end{cases} \tag{41}$$

where λ is a constant. Furthermore, the optimal portfolio is given by

$$x^i = \left(\frac{\tau^i \mathcal{L}^i(\hat{\tau}^i - \lambda)}{\gamma \hat{\tau}^i (1 + \lambda \mathcal{L}^i)}\right) S^i, \tag{42}$$

and the expected size of position is given by

$$E\left[|x^i|\right] = \begin{cases} \frac{1}{\gamma(1+\lambda\mathcal{L}^i)} \sqrt{\frac{2\tau^i \mathcal{L}^i (\hat{\tau}^i - \lambda)(1 + \hat{\tau}^i \mathcal{L}^i)}{\pi \hat{\tau}^i}} & \text{if } 0 < \lambda < \hat{\tau}^i \\ 0 & \text{if } \lambda \geq \hat{\tau}^i \end{cases} \tag{43}$$

Proof. See Appendix C. □

Proposition 4.1 shows that the allocation of resources is governed by the endogenously-determined multiplier λ. Notice that k^i is decreasing in λ. This in turn affects the portfolio choice x^i and the average portfolio size $E[|x^i|]$. For example, Eq. (41) shows us that the number of assets with non-negative resource allocation (i.e., $k^i > 0$) will weakly decrease in λ because the investor allocates resources to less assets. Also, Eqs. (42) and (43) imply that the portfolio weight will be zero for those assets that do not get allocated for resources. Therefore, increasing λ will result in the reduction of the number of assets in the portfolio (or portfolio concentration). On the other hand, decreasing λ will result in the increase in the number of assets in the portfolio (or portfolio concentration).

In the next two corollaries, we study comparative statics about how λ changes given the changes in some parameter values such as K and $\mathcal{L}^i$.

Corollary 4.1. *(Portfolio diversification) (i) The number of assets in the portfolio weakly increases in K (i.e., $\frac{d\lambda}{dK} < 0$), (ii) The amount of resource investment on the assets in the portfolio increases in K (i.e., $\frac{dk^i}{dK} > 0$ for all i with $k^i > 0$), (iii) The expected size of position in each asset in the portfolio increases in K (i.e., $\frac{dE[|x^i|]}{dK} > 0$ for all i with $k^i > 0$).*

Proof. See Appendix C. □

Corollary 4.1 implies that the number of assets in the optimal portfolio is non-decreasing (or weakly increasing) as the total resource capacity of the investor increases. That is, if the investor has more resources, the investors will spread out resources across more assets. Therefore, the optimal portfolio becomes more diversified. We can relate this result to empirical observations such that retails investors, who have less informational advantage than institutional investors, tend to concentrate their portfolios relative to institutional investors do.

Using Definition 2.3, we say that asset i is superior to asset j if information source of asset i is both more efficient (i.e., $\tau^i > \tau^j$) and more searchable (i.e., $\mathcal{L}^i > \mathcal{L}^j$).

Corollary 4.2. *(Specialization) The difference of resource allocations between a superior and an inferior asset becomes larger as K increases, i.e., $\frac{d(\frac{k^i}{K} - \frac{k^j}{K})}{dK} > 0$ for any asset i and j where asset i is superior to asset j.*

Proof. See Appendix C. □

Corollary 4.2 implies that differences in efficiency and searchability matters more as the investor gets more resources. This is so because the slope of a precision function $\phi(\cdot)$ of an inferior asset gets flat faster as k increases because the curvature is inverse to the searchability parameter. When investors allocate small amount of resources across assets, the difference between marginal increases in precision between a superior and an inferior asset is smaller. However, as the total amount of available resources increases, such difference becomes larger. Therefore, for investors with very little resource, there is little specialization among those asset in the portfolio. On the other hand, for investors with large amount of resources, the degree of specialization is greater. This suggests that investors with little resource relatively put more efforts on less searchable assets.

Corollary 4.3. *(Attention crowding out) (i) If $k^i > 0$, then λ increases as $\mathcal{L}^i$ increases (i.e., $\frac{d\lambda}{d\mathcal{L}^i} > 0$). (ii) The amount of resource investment on other assets in the portfolio decreases in $\mathcal{L}^i$ (i.e., $\frac{dk^j}{d\mathcal{L}^i} < 0$ for all j with $k^j > 0$), (iii) The expected size of position in each asset in the portfolio decreases in $\mathcal{L}^i$ (i.e., $\frac{dE[|x^j|]}{d\mathcal{L}^i} < 0$ for all j with $k^j > 0$),*

Corollary 4.3 implies that the number of assets in the optimal portfolio is non-increasing (or weakly decreasing) as the potentially available informational for one asset (which is already in the portfolio) increases. This is quite intuitive. If the informational environment for an asset already in the portfolio becomes richer, information about that asset can be acquired more efficiently and the investor will increase attention on the asset to exploit further private information. This attention effect naturally crowds out attention on other assets. Therefore, the portfolio becomes more concentrated.

4.2 Relation to the literature

The empirical evidence shows that investors systematically under-diversify their portfolios. Furthermore, retail investors are also subject to significant home biases in their portfolio choices. This is puzzling because it contradicts the conventional wisdom of minimizing idiosyncratic risks by holding a fully-diversified portfolio such as the market portfolio (e.g., the Capital Asset Pricing Model). What causes such under-diversification? The existing literature has suggested various reasons: (i) informational frictions (e.g., Van Nieuwerburgh and

Veldkamp (2009), Van Nieuwerburgh and Veldkamp (2010)), (ii) institutional barriers (e.g., Black (1974)), and (iii) behavioral biases such as familiarity bias (e.g., Portes and Rey (2005)). For example, Van Nieuwerburgh and Veldkamp (2009) employ the Kreps and Porteus (1978) framework to induce preference for an early resolution of uncertainty, and this creates benefits of specialization among investors. Therefore, those papers can explain why investors want to concentrate their portfolios because specialization benefits induce them to focus on a few assets.

Our analysis is in line with those in the literature that connect informational frictions to under-diversification, but we complement the literature by suggesting an interesting difference from the existing papers. As in other existing papers, increased frictions can cause more concentration portfolios. Unlike the existing papers, however, our paper shows that portfolio concentration can be caused by differences in resource endowments. This fits well-documented empirical observation that retail investors tend to under-diversify their portfolios more than institutional investors, and furthermore they tend to hold more neglected assets. Given the fact that retail investors are likely to have less information searching resources than institutional investors, our model predicts higher portfolio concentration among retail investors relative to institutional investors. Our model also predicts that increasing searchability of one asset crowds out investing in other assets by taking away attentions from the investors. Market segmentation can create home biases for other assets through crowing out. That is, if there are assets traded only in the home market with superior sources, this crowds out investment in other assets. Therefore, our model contributes to the literature by suggesting an alternative mechanism of specialization and portfolio concentration.

5 Conclusion

In our paper, we develop a microfounded framework of information acquisition in which searchable information is limited. Our framework on learning under imperfect information searchability is based on a simple intuition: As an agent keeps searching for new information, it is likely that he would encounter some overlapping pieces of information from the past searching activities. Furthermore, other agents searching for information from the same source would face the same difficulty in collecting new information, thus, they are more likely to end up with similar information in case 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 the concavity of signal precisions and the correlations among signals increase as searchable information becomes smaller. Using an asymptotic approach, we construct a tractable mapping from resource allocations to the precisions and the correlations of agents' acquired information under varying degrees of searchable information. We study two economic applications with endogenous information acquisition using our model. In the first application with "beauty contest" coordination game, we find that it is possible that agents prefer an inferior information source with less searchable information due to coordination motives. 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, such equilibrium outcome becomes more likely precisely when the inferior information source becomes more inefficient. In the second application with portfolio optimization problem, we study an investor's optimal allocation of resources across multiple risky assets. We find that the portfolio gets more concentrated when the investor is more constrained in resources. Furthermore, the investor's attention on other assets can be crowded out by an attention-grabbing information if it becomes more searchable relative to other assets.

Appendix A

Proof of Lemma 2.2: We prove this lemma in a similar fashion as in standard proofs of the strong law of large numbers.[11] The major difference of the proof from the standard case is that 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 l to be the number of collected signals 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. would 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:

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

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

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

[11] See, for example, Billingsley (1979) 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 S to be 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 $\mathbf{L}$.[12] 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 between 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[13]

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

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 followings using the inclusion–exclusion principle:

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

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

[13] 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

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

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

Then, taking the limit of c in Eqs. (A.13) and (A.14) yields the followings:

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

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

Multiplying c^2 to Eq. (A.7) and taking the limit of c yields

$$\lim_{c\to 0} c^2 E\left[\left|\tilde{h} - E[\tilde{h}]\right|^4\right] = 3\mathcal{L}^2 \exp\left(-\frac{2k}{\mathcal{L}}\right)\left(1 - \exp\left(-\frac{k}{\mathcal{L}}\right)\right)^2. \quad \text{(A.17)}$$

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 $c^2 E\left[\left|\tilde{h} - E[\tilde{h}]\right|^4\right] < \bar{\mathcal{L}}$. Therefore, there exists $N > 0$ such that for all $n \geq N$ and $n \in \mathbb{N}$ we have

$$Pr\left[\left|\frac{1}{n}\tilde{h} - \frac{1}{n}E\left[\tilde{h}\right]\right| \geq \alpha\right] < \frac{\bar{\mathcal{L}}}{n^2\alpha^4}. \quad \text{(A.18)}$$

Then, the first Borel-Cantelli lemma implies that

$$Pr\left[\lim_{n\to\infty}\left|\Phi\left(\tilde{h}(k;\frac{1}{n})\right) - \phi(k)\right| < \alpha\right] = 1, \quad \text{(A.19)}$$

or equivalently

$$Pr\left[\lim_{c\to 0}\left|\Phi\left(\tilde{h}(k;c)\right) - \phi(k)\right| < \alpha\right] = 1. \ \blacksquare \quad \text{(A.20)}$$

Proof of Lemma 2.3: The proof is parallel with Lemma 2.2. Let L to be the number of distinct signals in $\mathbf{L}$, i.e., $L = \lfloor\frac{\mathcal{L}}{c}\rfloor$. We also denote l^i and l^j to be the number of signals collected by agent 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 we define $\tilde{h}^m_{i,j}(k_i;c)$ to be one if signal s^m belongs to the group of the 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

$$\tilde{h}_{i,j}(k_i,k_j;c) = \sum_{m=1}^{L}\tilde{h}^m_{i,j}(k_i,k_j;c) = \sum_{m=1}^{L}\tilde{h}^m_i(k_i;c)\tilde{h}^m_j(k_j;c). \quad \text{(A.21)}$$

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

$$
\begin{aligned}
E[\tilde{h}_{i,j}(k_i,k_j;c)] &= \sum_{m=1}^{L} Pr(\tilde{h}_i^m(k_i;c)\tilde{h}_j^m(k_j;c)=1) = \sum_{m=1}^{L} Pr(\tilde{h}_i^m(k_i;c)=1)Pr(\tilde{h}_j^m(k_j;c)=1) \\
&= L\left[\left(1-\left(\frac{L-1}{L}\right)^{l_i}\right)\left(1-\left(\frac{L-1}{L}\right)^{l_j}\right)\right].
\end{aligned}
\tag{A.22}
$$

We can represent Eq. (A.22) given k_i and k_j as follows:

$$
E[\tilde{h}_{i,j}(k_i,k_j;c)] = \lfloor\frac{\mathcal{L}}{c}\rfloor\left[\left(1-\left(1-\frac{1}{\lfloor\frac{\mathcal{L}}{c}\rfloor}\right)^{\lfloor\frac{k_i}{c}\rfloor}\right)\left(1-\left(1-\frac{1}{\lfloor\frac{\mathcal{L}}{c}\rfloor}\right)^{\lfloor\frac{k_j}{c}\rfloor}\right)\right]. \tag{A.23}
$$

Multiplying c to Eq. (A.23) and taking the limit of c yields

$$
\lim_{c\to 0} E[c\tilde{h}_{i,j}(k_i,k_j;c)] \to \mathcal{L}\left(1-\exp\left(-\frac{k_i}{\mathcal{L}}\right)\right)\left(1-\exp\left(-\frac{k_j}{\mathcal{L}}\right)\right). \tag{A.24}
$$

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

$$
Pr\left[\left|c\tilde{h}_{i,j}-E[c\tilde{h}_{i,j}]\right|\geq\alpha\right]\leq\frac{c^4E\left[\left|\tilde{h}_{i,j}-E[\tilde{h}_{i,j}]\right|^4\right]}{\alpha^4}. \tag{A.25}
$$

We aim to prove Eq. (A.25) 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 Lemma 2.2 up to Eq. (A.7).

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

$$
\varphi_1(z) = \left(\frac{L-1}{L}\right)^z, \tag{A.26}
$$

$$
\varphi_2(z) = \left(\frac{L-2}{L}\right)^z, \tag{A.27}
$$

$$
\varphi_3(z) = \left(\frac{L-3}{L}\right)^z \tag{A.28}
$$

Using the inclusion–exclusion principle (which is analogous to Eq. (A.10)), we have[14]

$$\begin{aligned} E\left[\tilde{h}_{i,j}^{m}\tilde{h}_{i,j}^{n}\right] &= Pr(\tilde{h}_{i,j}^{m}\tilde{h}_{i,j}^{n}=1) \\ &= 1 - Pr(\tilde{h}_{i,j}^{m}=0) - Pr(\tilde{h}_{i,j}^{n}=0) + Pr(\tilde{h}_{i,j}^{m}=0 \wedge \tilde{h}_{i,j}^{n}=0). \end{aligned} \tag{A.30}$$

Using the inclusion–exclusion principle again, we derive

$$Pr(\tilde{h}_{i,j}^{m}=0) = Pr(\tilde{h}_{i,j}^{n}=0) = \varphi_1(l_i) + \varphi_1(l_j) - \varphi_1(l_i + l_j). \tag{A.31}$$

and

$$\begin{aligned} Pr(\tilde{h}_{i,j}^{m}=0 \wedge \tilde{h}_{i,j}^{n}=0) &= Pr((\tilde{h}_{i}^{m}=0 \vee \tilde{h}_{j}^{m}=0) \wedge (\tilde{h}_{i}^{n}=0 \vee \tilde{h}_{j}^{n}=0)) \\ &= 2\varphi_1(l_i)(1-\varphi_1(l_i))\varphi_1(l_j)(1-\varphi_1(l_j)) \\ &+ \varphi_2(l_i) + \varphi_2(l_j) - \varphi_2(l_i)\varphi_2(l_j). \end{aligned} \tag{A.32}$$

Substituting Eqs. (A.31) and (A.32) into Eq. (A.30), and taking the limit of c yields

$$\lim_{c\to 0} E\left[\tilde{h}_{i,j}^{m}\tilde{h}_{i,j}^{n}\right] = \left(1 - \exp\left(-\frac{k_m}{\mathcal{L}}\right)\right)^2 \left(1 - \exp\left(-\frac{k_n}{\mathcal{L}}\right)\right)^2. \tag{A.33}$$

Similarly as in Eq. (A.33), we obtain the expectation of the cross product of three variables $\tilde{h}^m, \tilde{h}^n$ and $\tilde{h}^x$ as follows:

$$\begin{aligned} E\left[\tilde{h}_{i,j}^{m}\tilde{h}_{i,j}^{n}\tilde{h}_{i,j}^{x}\right] &= Pr(\tilde{h}_{i,j}^{m}\tilde{h}_{i,j}^{n}\tilde{h}_{i,j}^{x}=1) \\ &= 1 - 3\left[\varphi_1(l_i) + \varphi_1(l_j) - \varphi_1(m+n)\right] \\ &+ 3\left[\varphi_1(l_i)(1-\varphi_1(l_i))\varphi_1(l_j)(1-\varphi_1(l_j)) + \varphi_2(l_i) + \varphi_2(l_j) - \varphi_2(l_i)\varphi_2(l_j)\right] \\ &- [3(\varphi_1(l_i)^2(1-\varphi_1(l_i))\varphi_1(l_j)(1-\varphi_1(l_j))^2 + \varphi_1(l_i)(1-\varphi_1(l_i))^2\varphi_1(l_j)^2(1-\varphi_1(l_j))) \\ &+ 6\varphi_2(l_i)(1-\varphi_1(l_i))\varphi_2(l_j)(1-\varphi_1(l_j)) + \varphi_3(l_i) + \varphi_3(l_j) - \varphi_3(l_i)\varphi_3(l_j)]. \end{aligned} \tag{A.34}$$

Taking the limit of c in Eq. (A.34) yields

$$\lim_{c\to 0} E\left[\tilde{h}_{i,j}^{m}\tilde{h}_{i,j}^{n}\tilde{h}_{i,j}^{x}\right] = \left(1 - \exp\left(-\frac{k_m}{\mathcal{L}}\right)\right)^3 \left(1 - \exp\left(-\frac{k_n}{\mathcal{L}}\right)\right)^3. \tag{A.35}$$

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

$$|\cup_{m=1}^{M} A_m| = \sum_{m=1}^{M} |A_m| - \sum_{1\le m<n\le M} |A_m \cap A_n| + \sum_{1\le m<n<r\le M} |A_m \cap A_n \cap A_r| + \ldots + (-1)^{M+1} |\cap_{m=1}^{M} A_m|. \tag{A.29}$$

We can repeat the same exercise as in Eq. (A.34) for the expectation of the cross product of four variables $\tilde{h}^m, \tilde{h}^n, \tilde{h}^x$ and $\tilde{h}^y$ to obtain the following:

$$\lim_{c \to 0} E\left[\tilde{h}^m_{i,j}\tilde{h}^n_{i,j}\tilde{h}^x_{i,j}\tilde{h}^y_{i,j}\right] = \left(1 - \exp\left(-\frac{k_m}{\mathcal{L}}\right)\right)^4 \left(1 - \exp\left(-\frac{k_n}{\mathcal{L}}\right)\right)^4. \tag{A.36}$$

Then, the rest of the proof is again identical to Lemma 2.2 to finish the proof. ■

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 $H^{i,j}_n$ 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,[15] 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:[16]

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

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

$$\widetilde{\varepsilon}^i_n = \frac{1}{h^i_n}\sum_{m \in H^i_n} \varepsilon^m. \tag{A.37}$$

Theorems 1 and 2 give the asymptotic variances and covariance of $\widetilde{\varepsilon}^i_n, \widetilde{\varepsilon}^j_n$ as c goes to zero (equivalently, as n goes to infinity). Here we will prove joint asymptotic normality of $\widetilde{\varepsilon}^i_n, \widetilde{\varepsilon}^j_n$ by showing that

$$a\widetilde{\varepsilon}^i_n + b\widetilde{\varepsilon}^j_n \xrightarrow{d} N\left(0, \frac{a^2}{\phi\left(k^i\right)} + \frac{b^2}{\phi\left(k^j\right)} + 2\frac{ab}{\tau\mathcal{L}}\right) \text{ for all } a, b \in \mathbb{R}^2 \tag{A.38}$$

The plan of the proof is as follows. As a first step, starting from $\widetilde{\varepsilon}^i_n, \widetilde{\varepsilon}^j_n$, we construct two alternative random variables, $\widehat{\varepsilon}^i_n, \widehat{\varepsilon}^j_n$ 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 CLT to prove asymptotic normality of $a\widehat{\varepsilon}^i_n + b\widehat{\varepsilon}^j_n$ as c goes to zero. As a third step, we prove that $a\widetilde{\varepsilon}^i_n + b\widetilde{\varepsilon}^j_n$ converges in probability to $a\widehat{\varepsilon}^i_n + b\widehat{\varepsilon}^j_n$ as c goes to zero. The fourth step combines the previous results and completes the proof.

First step. Let $\tilde{\mathbf{L}}_n$ be an independent copy of let $\mathbf{L}_n$, that is, a set of L_n signals of the form $\theta + \tilde{\varepsilon}^m$, where each $\tilde{\varepsilon}^m$ is independently and identically distributed to each ε^m in the signals in $\mathbf{L}_n$.

Let the random variable $z^{i,j}_n$ be defined as $z^{i,j}_n = h^{i,j}_n - \left\lceil E\left(h^{i,j}_n\right)\right\rceil$. Then, let the set $Z^{i,j}_n$ be defined as follows. If $z^{i,j}_n > 0$, let $Z^{i,j}_n$ be a set of $z^{i,j}_n$ random draws (without replacement) from $H^{i,j}_n$.

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

[16] 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.

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 $z_n^i = h_n^i - \lceil E\left(h_n^i\right)\rceil - z_n^{i,j}$. 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 $\lceil E\left(h_n^i\right)\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil$. Define the random variable $\widehat{\varepsilon}_n^i$ as

$$\widehat{\varepsilon}_n^i = \frac{1}{\lceil E\left(h_n^i\right)\rceil}\left[\sum_{m \in \hat{H}_n^i} \varepsilon^m + \sum_{m \in \hat{H}_n^{i,j}} \varepsilon^m\right]. \tag{A.39}$$

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

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

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

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

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

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

$$V_n^2 = \sum_{k=1}^{r_n} Var(X_{nk})$$

$$= Var\left(\varepsilon^k\right)\Bigg[\left(\lceil E\left(h_n^i\right)\rceil - \lceil E\left(h_n^{i,j}\right)\rceil\right)\left(\frac{a}{\lceil E\left(h_n^i\right)\rceil}\right)^2 + \left(\lceil E\left(h_n^j\right)\rceil - \lceil E\left(h_n^{i,j}\right)\rceil\right)\left(\frac{b}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)^2$$

$$+ \lceil E\left(h_n^{i,j}\right)\rceil\left(\frac{a}{\lceil E\left(h_n^i\right)\rceil} + \frac{b}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)^2\Bigg]$$

$$= Var\left(\varepsilon^k\right)\left[\frac{a^2}{\lceil E\left(h_n^i\right)\rceil} + \frac{b^2}{\left\lceil E\left(h_n^j\right)\right\rceil} + \frac{2ab}{\lceil E\left(h_n^i\right)\rceil\left\lceil E\left(h_n^j\right)\right\rceil}\lceil E\left(h_n^{i,j}\right)\rceil\right] \tag{A.40}$$

$$= \frac{n}{\tau}\left[\frac{a^2}{n\mathcal{L}\left[1-\left(1-\frac{1}{n\mathcal{L}}\right)^{nk^i}\right] + g_i(n)} + \frac{b^2}{n\mathcal{L}\left[1-\left(1-\frac{1}{n\mathcal{L}}\right)^{nk^j}\right] + g_j(n)} + \frac{2ab}{n\mathcal{L} + g_{i,j}(n)}\right],$$

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

$$\lim_{n\to\infty} V_n^2 = \frac{a^2}{\phi\left(k^i\right)} + \frac{b^2}{\phi\left(k^j\right)} + 2\frac{ab}{\tau\mathcal{L}}. \tag{A.41}$$

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

$$\lim_{n\to\infty}\sum_{k=1}^{r_n}\frac{1}{V_n^2}E\left(X_{nk}^2\mathbf{1}_{\{|X_{nk}|\geq\delta V_n\}}\right) = 0. \tag{A.42}$$

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

$$\lambda_n^i = \frac{\left(\lceil E\left(h_n^i\right)\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil\right)}{V_n^2}E\left(\left(\frac{a\varepsilon^k}{\lceil E\left(h_n^i\right)\rceil}\right)^2\mathbf{1}_{\left\{\left|\frac{a\varepsilon^k}{\lceil E\left(h_n^i\right)\rceil}\right|\geq\delta V_n\right\}}\right)$$

$$\lambda_n^j = \frac{\left(\left\lceil E\left(h_n^j\right)\right\rceil - \left\lceil E\left(h_n^{i,j}\right)\right\rceil\right)}{V_n^2}E\left(\left(\frac{b\varepsilon^k}{\left\lceil E\left(h_n^j\right)\right\rceil}\right)^2\mathbf{1}_{\left\{\left|\frac{b\varepsilon^k}{\lceil E\left(h_n^j\right)\rceil}\right|\geq\delta V_n\right\}}\right)$$

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

Using the expression for V_n^2 in (A.40) and simplifying, we can write

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

where

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

and

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

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

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

where we define

$$y^k = \frac{\varepsilon^k}{\sqrt{Var\left(\varepsilon^k\right)}}$$

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

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

Note that $\lim_{n\to\infty} \gamma_n^i = \infty$ while the distribution of y^k is independent of n (y^k is a standardized version of the original signal error term ε^k in the urn), and therefore $\left[\left|y^k\right| \geq \gamma_n^i\right] \downarrow \varnothing$ as $n \uparrow \infty$. Since α_n^i has a finite limit as $n \to \infty$, then $\lim_{n\to\infty} \lambda_n^i = 0$. 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.42) is satisfied. Then, the Lindeberg-Feller Central Limit Theorem implies

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

or, equivalently, that

$$a\widehat{\varepsilon}_n^i + b\widehat{\varepsilon}_n^j \xrightarrow{d} N\left(0, \frac{a^2}{\phi\left(k^i\right)} + \frac{b^2}{\phi\left(k^j\right)} + 2\frac{ab}{\tau\mathcal{L}}\right). \tag{A.43}$$

Third step. Note that we can write $\widetilde{\varepsilon}_n^i$ in (A.37) as

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

where we define

$$\breve{\varepsilon}_n^i = \frac{1}{\lceil E\left(h_n^i\right)\rceil} \sum_{m \in H_n^i} \varepsilon^m. \tag{A.44}$$

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

$$\lim_{n \to \infty} \text{Prob}\left(\left|\breve{\varepsilon}_n^i - \widehat{\varepsilon}_n^i\right| > \alpha\right) = 0. \tag{A.45}$$

By Chebyshev's inequality,

$$\text{Prob}\left(\left|\breve{\varepsilon}_n^i - \widehat{\varepsilon}_n^i\right| > \alpha\right) \le \frac{Var\left(\breve{\varepsilon}_n^i - \widehat{\varepsilon}_n^i\right)}{\alpha^2}. \tag{A.46}$$

By the variance decomposition formula,

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

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

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

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

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

$$= \frac{Var\left(\varepsilon^{m}\right)}{\lceil E\left(h_{n}^{i}\right)\rceil}\left(E\left[\left|\frac{h_{n}^{i}}{\lceil E\left(h_{n}^{i}\right)\rceil}-1\right|\right]+2\frac{\left\lceil E\left(h_{n}^{i,j}\right)\right\rceil}{\lceil E\left(h_{n}^{i}\right)\rceil}E\left[\left|\frac{h_{n}^{i,j}}{\left\lceil E\left(h_{n}^{i,j}\right)\right\rceil}-1\right|\right]\right)$$

$$= \frac{Var\left(\varepsilon^{m}\right)}{\lceil E\left(h_{n}^{i}\right)\rceil}\left(E\left[\left|w_{n}^{i}-1\right|\right]+2\frac{\left\lceil E\left(h_{n}^{i,j}\right)\right\rceil}{\lceil E\left(h_{n}^{i}\right)\rceil}E\left[\left|w_{n}^{i,j}-1\right|\right]\right)$$

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 in the paper, $\frac{1}{n}h_n^i \xrightarrow{a.s.} \frac{1}{n}E\left(h_n^i\right)$ 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 in the paper, $\frac{1}{n}h_n^{i,j} \xrightarrow{a.s.} \frac{1}{n}E\left(h_n^{i,j}\right)$ 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 $\frac{Var(\varepsilon^m)}{\lceil E(h_n^i)\rceil}$ and $\frac{\left\lceil E\left(h_n^{i,j}\right)\right\rceil}{\lceil E(h_n^i)\rceil}$ have finite limits as $n \uparrow \infty$, we have shown that

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

which completes the proof of (A.45).

Finally, since $\breve{\varepsilon}_n^i \xrightarrow{i.p.} \widehat{\varepsilon}_n^i$ 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 $\widetilde{\varepsilon}_n^i=\frac{\lceil E(h_n^i)\rceil}{h_n^i}\breve{\varepsilon}_n^i$, then $\widetilde{\varepsilon}_n^i \xrightarrow{i.p.} \widehat{\varepsilon}_n^i$. An identical proof shows that $\widetilde{\varepsilon}_n^j \xrightarrow{i.p.} \widehat{\varepsilon}_n^j$, and therefore

$$a\widetilde{\varepsilon}_n^i+b\widetilde{\varepsilon}_n^j \xrightarrow{i.p.} a\widehat{\varepsilon}_n^i+b\widehat{\varepsilon}_n^j. \tag{A.47}$$

Fourth step. By (A.43) and (A.47), Theorem 2.7 in van der Vaart (1998) implies that (A.38) holds. Hence, by Theorem 29.4 in Billingley (1995), $\widetilde{\varepsilon}_n^i$ and $\widetilde{\varepsilon}_n^j$ are jointly normally distributed. ■

Proof of Corollary 2.1: Notice that

$$Var(\mu) = \frac{1}{I^2}\sum_{i=1}^{I} Var(\varepsilon_i) + \frac{1}{I^2}\sum_{i=1}^{I}\sum_{j\neq i}^{I} Cov(\varepsilon_i, \varepsilon_j) \tag{A.48}$$

$$= \frac{1}{I^2}\sum_{i=1}^{I} \phi^{-1}(k_i) + \frac{I-1}{I}\frac{1}{\tau\mathcal{L}}, \tag{A.49}$$

and

$$Cov(\varepsilon_i, \mu) = \frac{1}{I}\left(Var(\varepsilon_i) + \sum_{j\neq i} Cov(\varepsilon_i, \varepsilon_j)\right) \tag{A.50}$$

$$= \frac{1}{I}\phi^{-1}(k_i) + \frac{I-1}{I}\frac{1}{\tau\mathcal{L}}. \tag{A.51}$$

Therefore, we have

$$Cov(\eta_i, \mu) = Cov(\varepsilon_i, \mu) - Var(\mu) = \frac{1}{I^2}\sum_{j=1}^{I}\phi^{-1}(k_j) - \frac{1}{I}\phi^{-1}(k_i), \tag{A.52}$$

and

$$Var(\eta_i) = Var(\varepsilon_i - \mu) = Var(\varepsilon_i) - 2Cov(\varepsilon_i, \mu) + Var(\mu) \tag{A.53}$$

$$= \left(1 - \frac{2}{I}\right)\phi^{-1}(k_i) - \frac{I-1}{I}\frac{1}{\tau\mathcal{L}} + \frac{1}{I^2}\sum_{j=1}^{I}\phi^{-1}(k_j), \tag{A.54}$$

and

$$Cov(\eta_i, \eta_j) = Cov(\varepsilon_i - \mu, \varepsilon_j - \mu) = Cov(\varepsilon_i, \varepsilon_j) - Cov(\varepsilon_i, \mu) - Cov(\varepsilon_j, \mu) + Var(\mu) \tag{A.55}$$

$$= \frac{1}{\tau\mathcal{L}} - \frac{1}{I}\phi^{-1}(k_i) - \frac{1}{I}\phi^{-1}(k_j) - \frac{I-1}{I}\frac{1}{\tau\mathcal{L}} + \frac{1}{I^2}\sum_{i=1}^{I}\phi^{-1}(k_i). \tag{A.56}$$

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

$$\lim_{I\to\infty} Var(\eta_i) = \frac{1}{\phi(k_i)} - \frac{1}{\tau\mathcal{L}}, \text{ for all } i \in \mathbf{I} \tag{A.57}$$

$$\lim_{I\to\infty} Cov(\eta_i, \mu) = 0, \text{ for all } i \in \mathbf{I} \tag{A.58}$$

$$\lim_{I\to\infty} Cov(\eta_i, \eta_j) = 0, \text{ for all } i, j \in \mathbf{I} \tag{A.59}$$

$$\lim_{I\to\infty} Var(\mu) = \frac{1}{\tau\mathcal{L}}. \tag{A.60}$$

Using the definition of $\phi(k_i)$ in Theorem 2.1 it is immediate to rearrange the r.h.s. of Eq. (A.57) 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

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

and

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

It is immediate to rearrange terms as in Eq. (33), where we define

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

and

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

We remark that Eq. (B.2) implies

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

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

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

where

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

and (ii) $\hat{k}$ *satisfies* $\Sigma_{j=1}^J\hat{k}^j = K$.
(iii) In a payoff maximizing equilibrium, the equilibrium resource allocation solves

$$k^* \in \arg\max_{k_i\in\Delta}\Sigma_{j=1}^J g_j(k_j) \quad s.t.\ \Sigma_{j=1}^J k_i^j = K. \tag{B.6}$$

Proof of part (i). Let $\tilde{\gamma}(k_i) \in \arg\min_{(\gamma_i)} L_1(k_i,\gamma_i)$. Fixing k_i, it is immediate to verify that $L_1(k_i,\gamma_i)$ is strictly convex in γ_i for all $\delta\in[0,1]$. Differentiating $L_1(k_i,\gamma_i)$ with respect to γ_i and solving the system of first-order conditions for γ_i gives

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

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)). 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. ■

Proof of part (ii). Assume $\left(\hat{k},\hat{\gamma}\right)$ is a SBNE and $\Sigma_{j=1}^J\hat{k}_j < K$. By part (i), it has to be that $\hat{\gamma} = \tilde{\gamma}\left(\hat{k}\right)$. Substituting $\tilde{\gamma}(k_i)$ for γ_i into $L_1(k_i,\gamma_i)$, we obtain

$$\tilde{L}_1(k_i) = L_1(k_i,\tilde{\gamma}(k_i)) = \frac{(1-\delta)\tau_\theta^{-1}}{1+(1-\delta)\tau_\theta^{-1}\Sigma_{j=1}^J g_j(k_i^j)}. \tag{B.7}$$

Inspection of the functions g_j in Eq (B.5) reveals that each g_j is strictly increasing for $\delta\in[0,1]$. Then, assume $\left(\hat{k},\tilde{\gamma}\left(\hat{k}\right)\right)$ is a SBNE and $\Sigma_{j=1}^J\hat{k}_j < K$. Then, consider agent i deviating locally from $\hat{k}$ by increasing marginally k_i^j from $\hat{k}_j$. The deviation is feasible, it decreases L_1 and it has no first-order effect on L_2 (see Eq. (B.3)). Then, because g_j is strictly increasing we have $\partial\tilde{L}_1(k_i)/\partial k_i^j\big|_{k^i=\hat{k}} > 0$. Therefore, there exists a profitable deviation, contradicting $\left(\hat{k}, \tilde{\gamma}\left(\hat{k}\right)\right)$ being a SBNE. ■

Proof of part (iii). Follows trivially by part (ii) and the fact that $\tilde{L}_1(k_i)$ in Eq. (B.7) is strictly decreasing in $\Sigma_{j=1}^J g_j(k_i^j)$. ■

Proof of Proposition 3.1 . By Lemma B.1 and Eq. (B.7), finding k^* that minimizes L_1 is equivalent to the following problem. Inspection of the functions g_j in Eq (B.5) reveals that each g_j is strictly concave for $\delta \in [0, 1/2)$. Therefore, we can convert the problem in Eq. (B.6) to 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

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

The F.O.C. gives

$$\lambda - \frac{\tau^j \exp\left(k^j/\mathcal{L}^j\right)}{\left[(1-\delta)\exp\left(k^j/\mathcal{L}^j\right)+\delta\right]^2} = 0, \tag{B.8}$$

which has a strictly positive solution for k^j if and only if $0 < \lambda < \tau^j$, in which case it is immediate to verify that k^j is as in Eq. (37). Finally, λ can be obtained by solving the following equation:

$$\Sigma_{j=1}^J \hat{k}^j(\lambda) = K. \tag{B.9}$$

Notice that 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 for Eq. (B.9) because l.h.s. is continuous and monotone decreasing in λ.

We now prove the statements in the proposition regarding uniqueness of equilibrium. By Lemma B.1, in an equilibrium $\hat{\gamma}$ must satisfy $\hat{\gamma} = \tilde{\gamma}\left(\hat{k}\right)$. Therefore, the problem is reduced to whether the equilibrium resource allocation $\hat{k}$ is unique. 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)), an equilibrium resource allocation must be either a local minimum or a critical point of $\tilde{L}_1(k_i)$ in Eq. (B.7). Equivalently, an equilibrium resource allocation must be either a local maximum or a critical point of $G(k) = \Sigma_{j=1}^J g_j(k^j)$. For $\delta \in [0, 1/2)$ the resource allocation in Eq. (37) is the unique maximizer of $G(k)$. Since $G(k)$ is strictly concave in k_i for $\delta \in [0, 1/2)$, the equilibrium resource allocation in Eq. (37)

is unique. This completes the proof of part (i) of the proposition. For $\delta \in [1/2, 1]$, $G(k)$ fails to be strictly concave. Then, a global maximizer of $G(k)$ is clearly an equilibrium, but a critical point and a local maximum of $G(k)$ can also be an equilibrium because the loss term $L_2(\gamma_i, \hat{\gamma}) > 0$ for $\gamma_i \neq \hat{\gamma}$. This completes the proof of part (ii) of the proposition. ■

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 $\hat{\gamma}^0 = (1 - \hat{\gamma}^A)\overline{\theta}$, $\hat{\gamma}^A = \frac{(1-\delta)\tau^A K}{\tau\theta + (1-\delta)\tau^A K}$ and $\hat{\gamma}^B = 0$. Then, consider the corresponding problem in Eq. (34) 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 $\gamma_i(\alpha_i) = \arg\min_{\gamma_i}$ $L_1(k(\alpha_i), \gamma_i) + L_2(\gamma_i, \hat{\gamma})$ we obtain

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

Substituting these optimal values for γ_i into the problem leaves

$$\begin{aligned}
L_3(\alpha_i) &= L_1(k(\alpha_i), \gamma_i(\alpha_i)) + L_2(\gamma_i(\alpha_i), \hat{\gamma}) \\
&= \frac{(1-\delta)\tau_\theta\left(\phi^A(K)^2(1-\delta)\tau_\theta^{-1} + (2 - \delta(1+\alpha_i))\phi^A(K) + \tau_\theta + \delta\phi^B(\alpha_i K)\right)}{\left(\phi^A(K)(1-\delta) + \tau_\theta\right)^2\left(\tau_\theta + \phi^A((1-\alpha_i)K + \phi^B(\alpha_i K)\right)}.
\end{aligned}$$

Straightforward algebra shows that $L_3(\alpha_i)$ is strictly increasing in α_i if $\tau^A > \tau^B$. ■

Proof of Proposition 3.2-(ii). By Proposition 3.2-(ii), an equilibrium $\hat{k}$ must be a local maximum or a critical point of $G(k) = \Sigma_{j=1}^J g_j(k_i^j)$. Equivalently $\alpha = 1$ must be a local maximum or a critical point of $G_a(\alpha)$, where we define

$$G_\alpha(\alpha) = g_A((1-\alpha)K) + g_B(\alpha K). \tag{B.10}$$

That is, for $\alpha = 1$ to be an equilibrium it must be

$$\frac{d}{d\alpha}G_\alpha(1) \geq 0 \Leftrightarrow e^{K/\mathcal{L}^B} \geq \frac{\tau^A}{\tau^B}\left(e^{K/\mathcal{L}^B}(1-\delta) + \delta\right)^2. \tag{B.11}$$

Letting $\overline{\mathcal{L}}_N = \frac{K}{\log(\tau^A/\tau^B)}$, it is immediate to verify that the necessary condition in Eq. (B.11) is satisfied only if $\mathcal{L}^B \leq \overline{\mathcal{L}}_N$. ■

Proof of Proposition 3.2-(iii). Define

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

It is immediate to verify tht $G_\alpha(\alpha)$ in Eq. (B.10) is strictly convex in α for all $\alpha \in [0,1]$ if $\delta \in (\tilde{\delta}\left(\mathcal{L}^B\right), 1]$. Letting $\alpha^* = \arg\max_{\alpha\in[0,1]} G_\alpha(\alpha)$, strict convexity of $G_\alpha(\alpha)$ implies $\alpha^* \in \{0,1\}$. Therefore, $\alpha = 1$ is the unique payoff maximizing SBNE if

$$G_\alpha(1) > G_\alpha(0) \Leftrightarrow \frac{\tau^A}{\tau^B}\frac{K}{\mathcal{L}^B} < \frac{e^{K/\mathcal{L}^B}-1}{e^{K/\mathcal{L}^B}(1-\delta)+\delta} \tag{B.12}$$

Notice that the r.h.s. of the second inequality in Eq. (B.12) is increasing in δ. As $\delta \to 1$, Eq. (B.12) holds if $\mathcal{L}^B < \overline{\mathcal{L}}_S$, where $\overline{\mathcal{L}}_S \in (0,\infty)$ solves

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

Hence, for all $\mathcal{L}^B < \overline{\mathcal{L}}_S$, Eq. (B.12) holds if $\delta \in (\check{\delta}\left(\mathcal{L}^B\right), 1]$, where we define

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

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

Proof of Corollary 3.1 For all for all $\mathcal{L}^B \in (\infty, \overline{\mathcal{L}}_S]$, we have $\tilde{\delta}\left(\mathcal{L}^B\right) < 1$ and $\check{\delta}\left(\mathcal{L}^B\right) \leq 1$, where the equality holds if and only if $\mathcal{L}^B = \overline{\mathcal{L}}_S$. Then, then, we have $\delta_S\left(\overline{\mathcal{L}}_S\right) = 1$ and $\delta_S\left(\mathcal{L}^B\right) < 1$ for all $\mathcal{L}^B < \overline{\mathcal{L}}_S$. The statement in the corollary follows by continuity of $\delta_S\left(\mathcal{L}^B\right)$, which in turn is implied by continuity of $\tilde{\delta}\left(\mathcal{L}^B\right)$ and $\check{\delta}\left(\mathcal{L}^B\right)$. ■

Appendix C

There is a standard formula which computes the certainty equivalence of expected utilities in case of CARA utilities. (For example, see Dow and Rahi (2003))

Lemma C.2. *Suppose A is a symmetric $m \times m$ matrix, b is an m-vector, d is a scalar, and w is an m-dimensional normal variate: $w \sim N(0, \Sigma)$, Σ positive definite. Then, we can find the following certainty equivalence of expected utilities if $(I - 2\Sigma A)$ is positive definite*

$$E\left[\exp(w^\top A w + b^\top w + d)\right] = |I - 2\Sigma A|^{-\frac{1}{2}} \exp\left[\frac{1}{2} b^\top (I - 2\Sigma A)^{-1} \Sigma b + d\right]. \tag{C.1}$$

Proof of Proposition 4.1. For notational convenience, we denote $S = (S^1, S^2, \ldots, S^N)$ to be the vector of all the signals acquired by the trader. The value function for the trader's optimal portfolio choice problem conditional on the acquired signals is given by

$$V(w_0; S) = \max_{x^1, x^2, \ldots, x^N} E\left[-\exp\left(-\gamma\left(w_0 + \sum_{i=1}^{N} (\theta^i - p^i) x^i\right)\right) \Big| S\right] \tag{C.2}$$

where x^i is the unit of holdings of asset i. Then, the demand for each asset i is given by the standard CARA-Gaussian demand:

$$x^i = \frac{E[\theta | S^i] - p^i}{\gamma Var[\theta | S^i]}. \tag{C.3}$$

By Lemma C.2, the expected utility can be obtained from Eqs. (C.2) and (C.3) as follows:

$$E\left[V(w_0; S)\right] = -\prod_{i=1}^{N} \sqrt{\frac{\tau_\theta^i}{\tau_\theta^i + \phi^i(k^i)}} \exp\left(-\gamma w_0 - \frac{1}{2} \sum_{i=1}^{N} \tau_\theta^i {p^i}^2\right). \tag{C.4}$$

Maximizing Eq. (C.4) under the resource constraint in Eq. (38) is equivalent to the following optimal resource allocation problem:[17]

$$\max_{\sum_{i=1}^N k^i = K} \sum_{i=1}^{N} G_i(k^i) \tag{C.5}$$

where $G_i(k^i) = \log\left(\tau_\theta^i + \phi^i(k^i)\right)$.

Notice that each $G_i(k^i)$ is concave in k^i because $\phi^i(k^i)$ is concave. Therefore, we can convert this problem in Eq. (C.5) to the following dual problem:

$$\min_{\lambda} \lambda K - \sum_{i=1}^{N} G_i^*(\lambda) \tag{C.6}$$

[17] Notice that other components in Eq. (C.4) are unaffected by the choice of resource allocations than $\prod_{i=1}^N \sqrt{\frac{\tau_\theta^i}{\tau_\theta^i + \phi^i(k^i)}}$. Using a log-transformation, we get (C.5).

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

$$G_i^*(\lambda) = \min_{k^i \geq 0} \left(\lambda k^i - \log\left(\tau_\theta^i + \phi^i(k^i)\right)\right) \tag{C.7}$$

For all each asset i, the optimal allocation k^i given λ is

$$k^i(\lambda) = \begin{cases} \mathcal{L}^i \log\left(\frac{\frac{\hat{\tau}^i}{\lambda} + \hat{\tau}^i \mathcal{L}^i}{1 + \hat{\tau}^i \mathcal{L}^i}\right) & \text{if } 0 < \lambda < \hat{\tau}^i \\ 0 & \text{if } \lambda \geq \hat{\tau}^i \end{cases} \tag{C.8}$$

Substituting k^i into Eq. (40) yields the precision ϕ^i as a function of λ such that

$$\phi^i = \frac{\tau^i \mathcal{L}^i}{\hat{\tau}^i} \left(\frac{\hat{\tau}^i - \lambda}{1 + \lambda \mathcal{L}^i}\right). \tag{C.9}$$

Finally, λ can be obtained by solving the following equation:

$$\sum k^i(\lambda) = K. \tag{C.10}$$

Notice that l.h.s. is zero when $\lambda = \infty$ and infinity when $\lambda = 0$. Therefore, it is easily verified that there exists a unique solution for Eq. (C.9) because l.h.s. is continuous and monotone decreasing in λ.

By Bayesian updating rule, we obtain the expectation and variance of θ^i conditional on S^i:

$$E[\theta|S^i] = \frac{\phi^i(k^i)}{\tau_\theta^i + \phi^i(k^i)} S^i, \tag{C.11}$$

$$Var[\theta|S^i] = \frac{1}{\tau_\theta^i + \phi^i(k^i)}. \tag{C.12}$$

By substituting Eqs. (C.11) and (C.12) into Eq. (C.3), we have

$$x^i = \frac{\phi(k^i)}{\gamma} S^i. \tag{C.13}$$

Then, Eq. (C.13) implies that

$$x^i = \left(\frac{\tau^i \mathcal{L}^i (\hat{\tau}^i - \lambda)}{\gamma \hat{\tau}^i (1 + \lambda \mathcal{L}^i)}\right) S^i. \tag{C.14}$$

Notice that x^i is normally distributed with mean zero and variance $\frac{\phi(k^i)^2}{\gamma^2}({\tau_\theta}^{-1} + \phi(k^i)^{-1})$. Because $|x^i|$ follows a folded normal distribution, we have

$$E[|x^i|] = \frac{\phi(k^i)}{\gamma} \sqrt{\frac{2}{\pi}\left({\tau_\theta}^{-1} + \phi(k^i)^{-1}\right)}. \tag{C.15}$$

By substituting Eq. (C.9) into Eq. (C.15), we get the desired result. ■

Proof of Corollary 4.1. Let $\mathbb{J}$ be the set of assets that have a non-zero resource allocation for searching information:

$$\mathbb{J} = \{j \in \{1, 2, \ldots, N\} | k^j > 0\}. \tag{C.16}$$

(i) Using the implicit function theorem, we get

$$\frac{\partial(\sum_{j=1}^{N} k^j(\lambda) - K)}{\partial K} dK + \frac{\partial(\sum_{j=1}^{N} k^j(\lambda) - K)}{\partial \lambda} d\lambda = 0, \tag{C.17}$$

which implies that

$$\frac{d\lambda}{dK} = -\frac{1}{\sum_{j \in \mathbb{J}} \mathcal{L}^j \left(\frac{\frac{1}{\lambda^2}}{\frac{1}{\lambda} + \mathcal{L}^j} \right)} < 0. \tag{C.18}$$

(ii) For any asset i with $k^i > 0$, it is immediate from Eq. (41) that

$$\frac{dk^i}{d\lambda} = -\frac{\frac{\mathcal{L}^i}{\lambda^2}}{\frac{1}{\lambda} + \mathcal{L}^i} < 0. \tag{C.19}$$

By the chain rule, we have

$$\frac{dk^i}{dK} = \frac{dk^i}{d\lambda} \frac{d\lambda}{dK} > 0. \tag{C.20}$$

(iii) For any asset i with $k^i > 0$, it is immediate from Eq. (43) that

$$\frac{dE\left[|x^i|\right]}{d\lambda} = -\frac{\mathcal{L}^i}{\gamma(1 + \lambda \mathcal{L}^i)^2} \sqrt{\frac{2\tau_\theta^i \mathcal{L}^i (\hat{\tau}^i - \lambda)(1 + \hat{\tau}^i \mathcal{L}^i)}{\pi}} - \frac{1}{\gamma(1 + \lambda \mathcal{L}^i)} \sqrt{\frac{\pi \tau_\theta^i \mathcal{L}^i (1 + \hat{\tau}^i \mathcal{L}^i)}{2(\hat{\tau}^i - \lambda)}} < 0. \tag{C.21}$$

By the chain rule, we have

$$\frac{dE\left[|x^i|\right]}{dK} = \frac{dE\left[|x^i|\right]}{d\lambda} \frac{d\lambda}{dK} > 0. \tag{C.22}$$

Proof of Corollary 4.3. (i) Using the implicit function theorem, we get

$$\frac{\partial(\sum_{j=1}^{N} k^j(\lambda))}{\partial \mathcal{L}^i} d\mathcal{L}^i + \frac{\partial(\sum_{j=1}^{N} k^j(\lambda))}{\partial \lambda} d\lambda = 0, \tag{C.23}$$

which implies that

$$\frac{d\lambda}{d\mathcal{L}^i} = \frac{\log\left(\frac{\frac{\hat{\tau}^i}{\lambda} + \hat{\tau}^i \mathcal{L}^i}{1 + \hat{\tau}^i \mathcal{L}^i} \right) + \frac{\hat{\tau}^i \mathcal{L}^i (1 - \frac{\hat{\tau}^i}{\lambda})}{(\frac{\hat{\tau}^i}{\lambda} + \hat{\tau}^i \mathcal{L}^i)(1 + \hat{\tau}^i \mathcal{L}^i)}}{\sum_{j \in \mathbb{J}} \mathcal{L}^j \left(\frac{\frac{\hat{\tau}^j}{\lambda^2}}{\frac{\hat{\tau}^j}{\lambda} + \mathcal{L}^j} \right)}. \tag{C.24}$$

Eq. (C.24) implies that $\frac{d\lambda}{d\mathcal{L}^i} > 0 \Leftrightarrow F(\frac{\hat{\tau}^i}{\lambda}, \hat{\tau}^i \mathcal{L}^i) > 0$, where we define

$$F(x,y) = \log\left(\frac{x+y}{1+y}\right) + \frac{y^i(1-x)}{(x+y)(1+y)}.$$

Notice that $F(1,y) = 0$ and $\partial F(x,y)/\partial x = x/(x+y)^2 > 0$ for all $x > 0$, which imply $\frac{d\lambda}{d\mathcal{L}^i} > 0$ for all i that satisfy $\lambda < \hat{\tau}^i$ and, therefore, $\hat{\tau}^i/\lambda > 1$.
(ii) For any asset i with $k^i > 0$, it is immediate from Eq. (41) that

$$\frac{dk^i}{d\lambda} = -\frac{\frac{\mathcal{L}^i}{\lambda^2}}{\frac{1}{\lambda} + \mathcal{L}^i} < 0. \tag{C.25}$$

By the chain rule, we have

$$\frac{dk^i}{d\mathcal{L}^i} = \frac{dk^i}{d\lambda}\frac{d\lambda}{d\mathcal{L}^i} < 0. \tag{C.26}$$

(iii) For any asset i with $k^i > 0$, it is immediate from Eq. (43) that

$$\frac{dE\left[|x^i|\right]}{d\lambda} = -\frac{\mathcal{L}^i}{\gamma(1+\lambda\mathcal{L}^i)^2}\sqrt{\frac{2\tau_\theta^i\mathcal{L}^i(\hat{\tau}^i-\lambda)(1+\hat{\tau}^i\mathcal{L}^i)}{\pi}} - \frac{1}{\gamma(1+\lambda\mathcal{L}^i)}\sqrt{\frac{\pi\tau_\theta^i\mathcal{L}^i(1+\hat{\tau}^i\mathcal{L}^i)}{2(\hat{\tau}^i-\lambda)}} < 0. \tag{C.27}$$

By the chain rule, we have

$$\frac{dE\left[|x^i|\right]}{d\mathcal{L}^i} = \frac{dE\left[|x^i|\right]}{d\lambda}\frac{d\lambda}{d\mathcal{L}^i} < 0. \tag{C.28}$$

References

Billingsley, P., 1979, *Probability and measure*. John Wiley & Sons, Hoboken, NJ.

Black, F., 1974, "International capital market equilibrium with investment barriers," *Journal of Financial Economics*, 1(4), 337–352.

Dow, J., and R. Rahi, 2003, "Informed Trading, Investment and Welfare," *Journal of Business*, 76(3), 439–454.

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

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

Kahneman, D., 1973, *Attention and Effort*. Prentice-Hall, Englewood Cliffs, NJ.

Kreps, D. M., and E. L. Porteus, 1978, "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *ema*, 46(1), 185–200.

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.

Mondria, J., 2010, "Portfolio choice, attention allocation, and price comovement," *Journal of Economic Theory*, 145(5), 1837–1864.

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

Moscarini, G., and L. Smith, 2001, "The Optimal Level of Experimentation," *Econometrica*, 69(6), 1629–1644.

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.

Peress, J., 2004, "Wealth, Information Acquisition and Portfolio Choice," *Review of Financial Studies*, 17(2), 879–914.

Portes, R., and H. Rey, 2005, "The determinants of cross-border equity flows," *Journal of international Economics*, 65(2), 269–296.

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," *The Bell System Technical Journal*, 27, 379–423, 623–656.

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 Nieuwerburgh, S., and L. Veldkamp, 2009, "Information Immobility and the Home Bias Puzzle," *Journal of Finance*, 3, 118–1215.

——— , 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.