Flexible Information Acquisition in Large Coordination Games

Alexandros Rigos*

May 2020

Abstract

This paper studies how large populations of rationally inattentive individuals acquire information about economic fundamentals when, along with the motive to accurately estimate the fundamental, they have coordination motives. Information acquisition is costly but flexible: players freely choose how to correlate their signal with the fundamental, paying costs linear in Shannon mutual information. A characterization of the class of equilibria in continuous strategies is provided, without restricting the fundamental to follow a normal prior. Equilibria where the population-wide average action is an affine function of the fundamental exist only when the fundamental is normally distributed. A novel method that allows the study of non-normal priors is developed. Using this method, the paper demonstrates that small departures from normality can lead to distributions of equilibrium actions that differ significantly from those of Gaussian models. Such "distortions" can potentially explain highly skewed distributions of observed actions (e.g. prices) even if economic fundamentals are almost normally distributed.

Keywords: Coordination games; Beauty contest; Flexible information acquisition; Rational inattention; Non-normal prior; Skew normal distribution
JEL classification: C72, D83

---

*Department of Economics, Lund University. Email: alexandros.rigos@gmail.com. The present paper supersedes the manuscript circulated as "A Beauty Contest with Flexible Information Acquisition." I would like to thank Martin Kaae Jensen and Chris Wallace for their invaluable guidance and advice. I also thank Jens Gudmundsson, Mogens Fosgerau, Erik Mohlin, Heinrich Nax, André Stenzel, Petra Thiemann, Jörgen Weibull, Christoph Wolf, and Ming Yang for their useful comments. Finally, I thank participants at GAMES 2016, the 2016 RES conference, the 2017 Conference on Economic Design, the 2nd SOCRATES Workshop, the Workshop on Limited Attention in Copenhagen, the II Leicester Economics PhD Conference, the Stony Brook 26th International Conference on Game Theory, and seminar participants in Stockholm School of Economics, Copenhagen, Lund, Leicester, and ETH Zürich. All errors are mine.

1 Introduction

A situation that often arises in financial and economic settings is one in which "players wish to do the right thing [...] and do it together" (Myatt and Wallace 2012, p. 340). That is, players have incentives to guess what the value of some economic fundamental is ("do the right thing") and to make guesses close to those of others ("do it together"). Motivating examples come from industrial organization, where firms compete in price-setting oligopolies (see Myatt and Wallace 2012, 2015); financial markets, where traders try to forecast the value of the fundamental while competing with each other (Allen, Morris, and Shin 2006); or investment games (as in Angeletos and Pavan 2004). When accurate and clear signals that reflect economic fundamentals are publicly available, these coordination problems can be resolved, or at least alleviated. Nevertheless, economic agents interact primarily in complex environments where — even though all relevant information about fundamentals is freely available in principle — paying attention (*processing* information) comes at non-negligible costs.

This paper studies such situations where many players are driven by fundamental motives along with coordination motives. The players are rationally inattentive (Sims 2003) and can acquire costly information about the fundamental in a flexible manner (Yang 2015). Solving for equilibrium in a complex environment like that can be challenging without imposing simplifying assumptions. Firstly, despite this difficulty, the present study characterizes equilibria in which players' actions follow continous distributions by providing an easy-to-verify condition without imposing a normal prior (as is often done in existing studies). Secondly, the paper shows that equilibria in which the population-wide average action is a linear function of the fundamental exist only if the fundamental is normally distributed. Thirdly, it demonstrates that even small departures from a Gaussian prior can lead to distributions of equilibrium objects that differ significantly from the ones obtained for normal priors, giving new insights into how markets aggregate information.

These interactions are modeled through a beauty contest game à la Morris and Shin (2002). The economic fundamental that is relevant for players' payoffs (henceforth *the fundamental)* is a random variable that follows a commonly known prior over the real line. Each of many players takes an action (a real number) and loses payoff according to a weighted average between (a) the squared distance of her action from the realization of the fundamental and (b) the squared distance of her action from the population-wide average action. The weight placed on the distance from the average action can vary and is termed the *coordination motive.*

Following the seminal work of Morris and Shin (2002), there is a large and growing literature that studies under which conditions more "public" or more "private" information is socially optimal. Traditionally, authors considered exogenous information structures: the players cannot affect the information they get and can only make decisions based on signals they passively receive. Exceptions include Myatt and Wallace (2012, 2015), Dewan and Myatt (2008) and Hellwig and Veldkamp (2009). In these models, players obtain information endogenously by purchasing more signals from different sources or by increasing their signal precision at a cost. Endogenizing the information acquisition process can be a more realistic approach as agents are allowed to choose whether to acquire information or not depending on its value. Moreover, agents can choose which type of information to attend to.

The literature on endogenous information acquisition has mostly assumed specific functional forms for the distribution of the fundamental and of the signals that the players observe (typically Gaussian). The information acqusition technology is — in this sense — rigid, and thus restrictive. Real-world agents often have access to a multitude of informational sources almost free of any cost. For example, virtually all information relevant to stock exchange movements can be easily accessed on the internet. What is costly is the *processing* of information that these sources convey, as it requires agents' attention — a scarce resource. Capturing the flexibility of indiviudals to process information in any way they want, the *rational inattention* framework Sims (1998, 2003) offers a way to go beyond the rigid Gaussian functional form assumtion.[1]

In rational inattention models, players are not passive recipients of information, nor

[1] Sims (1998, 2003) introduced rational inattention as a possible explanation of sluggish macroeconomic adjustments. Woodford (2008, 2009) has used it in state-dependent pricing settings where producers can revise their prices in continuous time, and they do so depending on market conditions. Rational inattention has also been used to analyze strategic interactions: Maćkowiak and Wiederholt (2009) study how rationally inattentive firms set prices in a monopolistic competition environment, whereas Yang (2015) focuses on investors' binary decisions to invest or not. Maćkowiak, Matějka, and Wiederholt (2018) provide a recent survey of the rational inattention literature.

is their information signal restricted to a particular form (e.g. Gaussian). Instead, each player receives information about the fundamental by designing her own information channel: an endogenous process that narrows down her belief about the fundamental. The more informative this "experiment" is, the more attention it requires and the more costly it is. In particular, the cost can be linear in the expected reduction of Shannon entropy (Shannon 1948) between the player's prior and posterior beliefs.[2] When deciding, each player chooses how to *efficiently* allocate her attention: not only does she trade off the benefit of the extra bit of information to its cost but she also decides to which events this extra bit is going to be allocated. In this sense, information acquisition is *flexible* (Yang 2015, 2019).

Among others, Matějka and McKay (2015) recognise that without the normal prior assumption continuous solutions to rational inattention problems are hard to calculate in decision-theoretic problems, let alone strategic environments. However, in order to explore how sensitive equilibrium results are with respect to the prior distribution, it is important to model these processes without the normality assumption. The present paper is the first to address this sensitivity issue in spite of the apparent difficulty.

To this goal, the paper provides a characterization of continuous-strategy equilibria without assuming a normal prior and establishes some of their properties. It then proceeds to show that tractable equilibria for which the average action is affine in the fundamental exist *only if* the prior is normal. This hints that results derived for normal priors may not extend to other distributions. Inquiring further, it demonstrates that even when prior distributions are very close to the normal, the distributions of the individual action and the population-wide average action can be significantly different from those derived for a Gaussian prior.

Put differently, it is shown that similar problems can have sharply different distributions of optimal choices, much in the spirit of Jung et al. (2019). The source of the difference, though, does not lie in the discreteness of the optimal solution (as in Jung

[2] Csiszár (2008) provides an axiomatization of the Shannon entropy information measure. For a more detailed argumentation about why Shannon entropy is an appropriate measure for information in economic models, see Jung et al. (2019, section 9.2). However, Denti, Marinacci, and Rustichini (2019) point out that costs based on mutual information cannot be derived from costs on actual experiments taken as primitives, since the former vanish as the prior becomes more dogmatic. The present paper abstracts from the details of the process that lead to mutual information costs.

et al. 2019) but is rather an equilibrium effect resulting from the strategic complementarities present in coordination games.

Small departures from the normal prior can yield qualitatively novel results. For example, they can lead to highly skewed distributions of equilibrium objects. So, high skewedness in observed equilibrium quantities (e.g. prices) can potentially be explained as a result of small non-Gaussian "distortions" of the prior, amplified by the coordination motive of market participants. Therefore, the normal prior assumption cannot be made without loss of generality. While using a normal prior can buy the modeler a lot in terms of tractability, it also mutes by construction any potential novel results — especially if higher moments of distributions are of interest as in, for example, finance applications.

Along with Myatt and Wallace (2012), who study economic beauty contests with endogenous information acquisition, and Yang (2015), who introduced flexible information acquisition technology, Denti (2019) is closely related to this study.[3] Denti (2019) studies a more general strategic environment and allows players to design signals that can be correlated even after conditioning on the fundamental, while paying a cost that is increasing in the Blackwell order (Blackwell 1951, 1953). In this sense, his agents use an *unrestricted* information acquisition technology. The information structure used by Denti (2019) is, therefore, richer than the one used herein, and leads to Bayes correlated equilibria (Bergemann and Morris 2016).[4] A detailed discussion of how results in the existing litarature relate those presented here is postponed to Section 7.

The paper is organized as follows: Section 2 sets up the model. Section 3 studies how players best-respond to continuous strategy profiles. Section 4 characterizes equilibria where players use continuous strategies and generalizes the main result to populations with heterogeneous costs. Section 5 takes an exhaustive look into the case of aggregately affine equilibria where the average action of the population is an affine function of the realization of the fundamental. Section 6 introduces a method to study non-normal priors and discusses the implications of non-normality. Section 7 discusses the results of the paper and compares them to the ones in the existing literature.

---

[3] Myatt and Wallace (2012), like Morris and Shin (2002), assume a uniform (improper) prior for the fundamental. Section 7 explains how one should compare the present setting to theirs.

[4] More recently, Hébert and La'O (2020) use a similar technology to study efficiency and non-fundamental volatility under different information costs.

2 The Model

Consider a large population of (ex-ante) identical expected utility-maximizing players, who are indexed by $i \in [0,1]$. Players are incentivized in two different ways: they want to (a) coordinate (take actions close to one another) and (b) take actions close to a value θ. The parameter $\gamma \in [0,1)$ determines how strong the coordination motive is. Each of them obtains utility given by

$$u_i = -(1-\gamma)(a_i - \theta)^2 - \gamma(a_i - \bar{a})^2 \tag{1}$$

where $a_i \in A_i = \mathbb{R}$ is player i's action and $\bar{a} = \int_0^1 a_i \, \mathrm{d}i$ represents the players' (population-wide) average action. It is assumed that players act in a way such that $\bar{a}$ is well-defined. Indeed, in all equilibria presented in the following sections this holds true.[5]

The value θ is unknown to the players. In particular, it is the realization of a random variable $\boldsymbol{\theta} : \Omega \to \Theta$ — the *fundamental* — where (Ω, Σ, P) is an underlying probability space and $\Theta = \mathbb{R}$. The fundamental is distributed according to $P_\theta \in \Delta(\Theta)$ which is commonly known, absolutely continuous, and has full support.[6,7] The probability density function (PDF) of P_θ is denoted by $p(\cdot)$ and is analytic. Moreover, $\boldsymbol{\theta}$ is assumed to have a well-defined mean $\bar{\theta}$ and variance σ^2.[8]

Before choosing her action, each player i gets to privately observe a *message*, which is the realization of her *signal*, a random variable. In particular, player i's signal is a measurable function $\boldsymbol{s}_i : \Omega \to S_i$, where S_i is some rich enough message space.[9] The signal is the only channel through which the player will receive information.

[5]For a discussion on this point see Myatt and Wallace (2012, footnotes 3 and 6). The integration over i should be interpreted as the limit of a weighted average of the actions of the "first n players" as $n \to \infty$. For a more formal exposition and discussion see the appendix of Acemoglu and Jensen (2010). The main implication from assuming an infinite amount of players is that individual players cannot affect the average action of the population and, thus, take it as given.

[6]Throughout the paper, $\Delta(X)$ denotes the space of probability measures over space X and P_x denotes the probability distribution of random variable $\boldsymbol{x}$.

[7]Note that in Morris and Shin (2002) and Myatt and Wallace (2012) players arrive at a common belief through updating a diffuse prior based on a publicly observed signal. This is discussed in Section 7.

[8]Note that $p \in L^2(\mathbb{R})$ (where $\mathbb{R}$ is endowed with the Lebesgue measure) is sufficient for $\bar{\theta}$ and σ^2 to be well-defined. Analyticity of $p(\cdot)$ is imposed in order to satisfy the conditions of Matějka and Sims (2010, proposition 2). Most distributions used by economists in practice satisfy this assumption.

[9]The message space S_i is endowed with a σ-algebra such that $\boldsymbol{s}_i$ is measurable and the pushforward measure.

Importantly, information acquisition is endogenous and each player gets to design her own signal/channel. Signal design takes place in two steps: (a) choosing the signal's support and (b) choosing the signal's distribution. Players can be very flexible when designing their signals: any measurable function will do as long as $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are independent for any $i \neq j$, conditional on the realization θ.[10] The flexibility of signal design allows players to decide not only how much information they want to receive but also about which events they want to get more information (where they want to focus their attention).

Information comes at a cost represented by the function $C(\cdot)$, which is the same across players. Following the standard literature on rational inattention, information costs are assumed to be linear in Shannon mutual information between the signal and the fundamental. Mutual information measures by how much observing one variable reduces one's uncertainty about some other random variable. It is defined as the Kullback-Leibler divergence between the joint and the product distributions of the two variables. Explicitly, mutual information between the fundamental $\boldsymbol{\theta}$ and the signal $\boldsymbol{s}_i$ is given by

$$I(\boldsymbol{\theta}, \boldsymbol{s}_i) = \int_{\Theta, S_i} \log \frac{\mathrm{d}P_{\theta, s_i}}{\mathrm{d}P_{\theta \times s_i}} \, \mathrm{d}P_{\theta, s_i}$$

if $P_{\theta, s_i} \ll P_{\theta \times s_i}$ and $I(\boldsymbol{\theta}, \boldsymbol{s}_i) = +\infty$ otherwise.[11] In the former case, P_{θ, s_i} admits a probability density function and $P_{s_i|\theta}$ can be described by conditional probability density functions $q(\cdot|\theta)$. Then the cost of signal $\boldsymbol{s}_i$ is given by

$$C(\boldsymbol{s}_i) = \mu \cdot I(\boldsymbol{\theta}, \boldsymbol{s}_i) = \mu \left(\int_{\Theta} \int_{S_i} p(\theta) q_i(s_i|\theta) \log \frac{q_i(s_i|\theta)}{Q_i(s_i)} \, \mathrm{d}s_i \, \mathrm{d}\theta \right) \tag{2}$$

where $Q_i(\cdot) = \int_{\Theta} q(\cdot|\theta) p(\theta) \, \mathrm{d}\theta$ is the (marginal) PDF of $\boldsymbol{s}_i$ and $\mu \geq 0$ is the cost per unit of information.[12] The more informative signal $\boldsymbol{s}_i$ is, the higher the cost of information.

[10] Conditional signal independence is a natural assumption. It essentially means that players cannot condition on each other's messages, since these are private. Thus, any correlation between signals should be the outcome of the players' conditioning on the fundamental (i.e. getting information about θ) and not on one another's messages. See also Section 7 for further discussion.

[11] Mutual information is symmetric and non-negatively defined. The derivative appearing in the expression is the Radon-Nikodym derivative between the joint distribution and the product of the marginal distributions of $\boldsymbol{\theta}$ and $\boldsymbol{s}_i$. For a standard textbook treatment of the topic see Cover and Thomas (2006).

[12] Throughout the paper, log denotes the natural logarithm and so the unit of measurement of infor-

Say, for example, that $S_i = \mathbb{R}$ and that $q(\cdot|\theta)$ is very concentrated and changes rapidly with θ, then the information structure is very informative and will come at a high cost. If, in contrast, $q(\cdot|\theta)$ does not change with θ, the information structure bears no information to player i and therefore has zero cost.

Upon receiving her message s_i, player i has to decide upon an action to take. This decision, in general, is a mixed strategy i.e. a probability measure $P_{a_i|s_i} \in \Delta(A_i)$ for each message $s_i \in S_i$.[13]

The timing is as follows:

1. Each player i — independently from and simultaneously with all other players — designs her signal $\boldsymbol{s}_i$ and pays the associated cost $C(\boldsymbol{s}_i)$.

2. The value of θ is realized.

3. Each player receives a message according to her chosen signal and the realization of θ. That is, player i's message is distributed according to $P_{s_i|\theta}$.

4. Each player — independently from and simultaneously with others — takes an action $a_i \in A_i$ contingent on the signal she received.

5. Players receive payoffs according to (1).

Given the above, a player's strategy consists of two parts: the signal part $\boldsymbol{s}_i$ and the action part $\boldsymbol{a}_i$. Let player i's strategy be denoted by $m_i = (\boldsymbol{s}_i, \boldsymbol{a}_i)$. The whole population's strategy profile will be denoted by m and the strategy profile of the population excluding player i by m_{-i}. Given a strategy profile m, the population-wide average action conditional on θ is given by the function $\bar{a} : \mathbb{R} \to \mathbb{R}$ defined through

$$\bar{a}(\theta) \equiv \int_0^1 \left(\int_{S_j} \left(\int_{A_j} a_j \, \mathrm{d}P_{a_j|s_j}(a_j|s_j) \right) \mathrm{d}P_{s_j|\theta}(s_j|\theta) \right) \mathrm{d}j. \tag{3}$$

As mentioned before, it is required that $\bar{a}(\cdot)$ is well-defined for (almost) all θ. A sufficient condition for this is that all $\boldsymbol{a}_i|\theta$ have a variance for (almost) all θ, and

mation is the *nat*. If the logarithms were taken with a base 2, the unit of measurement of information would be the *bit*. Notice that the choice of the unit of measurement does not qualitatively change the results as 1 bit equals $\log 2$ nats. So, the cost parameter μ is given in utils per nat of information.

[13] Formally, player i's action strategy is a random variable $\boldsymbol{a}_i : \Omega \to A_i$ which, conditionally on s_i, is independent from $\boldsymbol{\theta}$, $\boldsymbol{s}_j$, and $\boldsymbol{a}_j$ $(j \neq i)$.

that $\{\text{Var}(\boldsymbol{a}_i|\theta)\}_{i\in[0,1]}$ is bounded. The rest of the analysis focuses on strategy profiles whereby $\bar{a}(\cdot)$ is measurable.

3 Best Responses

Following standard arguments (see Woodford 2008; Yang 2015, for example), in optimal strategies messages used by players should correspond to actions: player i takes action a_i if and only if she has received a uniquely defined message $s_i(a_i)$.[14] Therefore, player i's best response can be summarized by a family of conditional probability measures $P_{a_i|\theta}$ that give the distribution over actions (i.e. mixed strategy) conditional on the realization θ—thus skipping the intermediate step of messages (as each message corresponds to exactly one action, and vice versa). Let $r_i(\cdot|\theta)$ denote the PDF of $P_{a_i|\theta}$.

Now, observe that from player i's point of view, the only way that the other players are affecting her payoff is through the effect of their strategies on the average action $\bar{a}$. Thus, player i is not affected by *the way* that the particular $\bar{a}$ comes about. This means that the object to which she is best-responding is the function $\bar{a}(\cdot)$ which summarizes all of her opponents' strategies (and which she cannot affect as she is "small"). So, the decision problem of player i is the following:

$$\max_{r_i} \; U(r_i, r_{-i}) - \mu I(\boldsymbol{\theta}, \boldsymbol{a}_i)$$

where

$$\begin{aligned} U(r_i, r_{-i}) &= -(1-\gamma)\int_\Theta \int_{A_i} (a_i-\theta)^2 r_i(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \\ &- \gamma \int_\Theta \int_{A_i} (a_i - \bar{a}(\theta))^2 r_i(a_i|\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta \end{aligned}$$

with $\bar{a}(\theta)$ given by

$$\bar{a}(\theta) = \int_0^1 \int_{A_j} a_j r_j(a_j|\theta)\,\mathrm{d}a_j\,\mathrm{d}j.$$

As a first result, it is easy to show that if information is costless, player i has an essentially unique pure-strategy best response to any $\bar{a}(\cdot)$.

[14] A detailed proof is provided in the Online Appendix for completeness.

Proposition 1. *Let (p, γ, μ) be a beauty contest with flexible information acquisition. If $\mu = 0$, then for any $\bar{a}(\cdot)$ player i has an essentially unique best response that assigns probability mass of 1 to the action*

$$b(\theta) = (1 - \gamma)\theta + \gamma \bar{a}(\theta) \tag{4}$$

(p-almost all $\theta \in \Theta$).

Proof. See Appendix A.1.

The function $b : \mathbb{R} \to \mathbb{R}$ defined through equation (4) will be refered to as the *best action function* of the strategy profile. It gives the action that a fully informed individual would take when she best-responds to a profile with average action $\bar{a}(\cdot)$. The focus from now on is on strategy profiles that satisfy some smoothness conditions.

Definition 1 (Smooth, monotone, full-support profile)**.** *A strategy profile r is a* smooth, monotone, full-support profile *if*

1. *the profile's average action function $\bar{a}(\cdot)$ is analytic in its argument, and*
2. *$\bar{a}'(\theta) > -\frac{1-\gamma}{\gamma}$ for all $\theta \in \mathbb{R}$.*

The two requirements ensure that the best action function $b(\cdot)$ is strictly increasing, analytic and bijective. Notice that 2 is not too restrictive since it is satisfied for any weakly increasing $\bar{a}(\cdot)$ but allows for the average action function to be decreasing in the realization of the fundamental (but not too fast), even though such behavior may not make intuitive sense.[15] Since in a smooth, monotone, full-support profile $b(\cdot)$ is bijective, it is also invertible. Let $\theta(\cdot) := b^{-1}(\cdot)$ denote the inverse of $b(\cdot)$ and $g(\cdot)$ denote the PDF of the distribution that the best action follows. The PDF $g(\cdot)$ is, then,

$$g(\cdot) = p\left(\theta(\cdot)\right)\theta'(\cdot) \tag{5}$$

and is analytic. The variance of the best action (the variance of $g(\cdot)$) is denoted by σ_b^2.

[15] In fact, in Proposition 5 it will be shown that such unintuitive behavior does not take place in equilibrium, even though the definition of smooth, monotone, full-support profile does not preclude it.

The analysis presented here and in Section 4 focuses on players using strategies that have densities, termed *continuous straetegies*.[16]

Definition 2 (Continuous strategy). *Strategy r_i of player i is* continuous *if $r_i(\cdot|\theta)$ is absolutely continuous with respect to the Lebesgue measure for (p-almost) all $\theta \in \Theta$.*

The following proposition provides necessary and sufficient conditions for the existence of a continuous best response to a smooth, monotone, full-support profile (see also Matějka and Sims 2010).

Proposition 2. *Let (p, γ, μ) with $\mu > 0$ be a beauty contest with flexible information acquisition. Let also r_{-i} be a smooth, monotone, full-support strategy profile of player i's opponents. Player i has a continuous best response to r_{-i} if and only if*

$$R_i := \mathscr{F}_\xi^{-1}\left[\exp(\mu\pi^2\xi^2)\hat{g}(\xi)\right] \qquad \text{is the PDF of a probability distribution.} \tag{6}$$

This continuous strategy is her unique best response and is given by

$$r_i(a_i|\theta) = R_i(a_i)\frac{b'(\theta)}{p(\theta)}\frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{(a_i - b(\theta))^2}{\mu}\right)$$

where $R_i(a_i)$ is the marginal density of action a_i.

Proof. See Appendix A.2.

Throughout the paper, $\mathscr{F}_x$ and $\mathscr{F}_\xi^{-1}$ denote the Fourier and inverse Fourier transforms defined as

$$\mathscr{F}_x[f(x)](\xi) = \int_{-\infty}^{+\infty} f(x)\exp(-2\pi\imath x\xi)\,\mathrm{d}x \tag{7}$$

$$\text{and } \mathscr{F}_\xi^{-1}[F(\xi)](x) = \int_{-\infty}^{+\infty} F(\xi)\exp(2\pi\imath x\xi)\,\mathrm{d}\xi \tag{8}$$

respectively ($\imath$ is the imaginary unit). Moreover, the shorthand notation with the hat operator $\hat{f}(\xi) := \mathscr{F}_x[f(x)](\xi)$ is used sometimes.

[16] The focus on continuous strategies may seem quite restrictive at first, especially since Jung et al. (2019) find that even in continuous environments optimal strategies might be "discrete," i.e., admit no density. At the end of Section 3 it is argued that in the present setting discontinuous strategies are unlikely to appear in equilibrium when information costs are not too large, so as to allow for enough information acquisition.

In the remainder of the paper, the terms "best response to r_{-i}," "best response to $\bar{a}(\cdot)$," and "best response to $b(\cdot)$" will be used interchangeably as all smooth, monotone, full-support profiles r_{-i} that yield the same $\bar{a}(\cdot)$ will lead to the same best response from player i and for a given γ there is a one-to-one relation between $\bar{a}(\cdot)$ and $b(\cdot)$ (see (4)).

Notice that it can be the case that $\hat{R}_i(\cdot)$ as calculated from (6) is the Fourier transform of a probability distribution containing atoms. If so, then the solution to the optimization problem is still given by Proposition 2 but it is no longer in continuous strategies (see also the comment on discontinuous strategies at the end of this Section).

Condition (6) states that $g(\cdot)$ can be written as the convolution of two distributions, one of them being normal with variance $\mu/2$ and the other one being $R_i(\cdot)$. If this is possible, then R_i is the marginal distribution of player i's action. Put differently, the best action $\boldsymbol{b}$, viewed as a random variable, should be able to be written as the sum of two other independently distributed random variables — $\boldsymbol{a}_i$ and $\boldsymbol{\varepsilon}_i$ — where the latter is noise following $N(0, \mu/2)$. Clearly, for that to be feasible the "resolution" $\sqrt{\mu/2}$ must be small enough.[17]

From the above analysis, the posterior belief of a player that has received message a_i should be normally distributed. This is formally captured in the following Proposition.

Proposition 3. *Let (p, γ, μ) be a beauty contest with flexible information acquisition and r_{-i} be a smooth, monotone, full-support strategy profile of player i's opponents such that condition (6) holds. In player i's best response, her posterior belief about the best action b has a PDF given by*

$$\varrho_i(b|a_i) = \frac{1}{\sqrt{\pi\mu}} \exp\left(-\frac{(a_i - b)^2}{\mu}\right).$$

Proof. See Appendix A.3.

In her best response, the posterior that player i has when she takes action a_i is *independent* of the prior distribution of the fundamental. Moreover, this posterior follows a normal distribution. This is a result of two things: the quadratic-losses objective and the Shannon-entropy-based information costs. Firstly, the quadratic losses form of the objective function gives incentives to take an action as close as possible to $b(\cdot)$ and

[17] A necessary condition is that R_i as calculated from (6) has a positive variance, i.e. that $\sigma_b^2 > \mu/2$. Another, stronger, necessary condition is that $\mathrm{Var}(\boldsymbol{a}_i|b) > 0$ for all b (see Proposition 5).

with the smallest possible variance of deviation from that. Secondly, the entropy-based information costs give incentives to have posteriors with high entropy. Additionally, among the family of distributions with full support on $\mathbb{R}$ and a given mean x_0 and variance σ^2, the Gaussian $N(x_0, \sigma^2)$ is the distribution with the maximum entropy (in this sense, the normal distribution is "informationally efficient"). Therefore, a normally distributed posterior is the "cheapest" one that achieves any given variance level.[18]

In light of Proposition 3, the effect of an increasing cost μ to player i's strategy becomes clear. A higher μ forces the player to have a less accurate posterior on what the best action is (as a more accurate belief would be more costly). Since $r(\cdot|\theta)$ needs to be a PDF, $\int_{-\infty}^{+\infty} r(a_i|\theta)\,\mathrm{d}a_i = 1$ must hold for all θ. As higher μ induces more dispersed $\varrho(\cdot|a_i)$, in order for this condition to be satisfied, the marginal distribution of actions $R_i(\cdot)$ needs to become more concentrated. When the value of μ reaches the point where condition (6) ceases to hold, player i switches to discontinuous strategies.

Proposition 4. *Let $g(\cdot)$ be the distribution of the best action of a smooth, monotone, full-support strategy profile of player i's opponents that satisfies condition* (6). *In player i's best response, her expected action conditional on the best action being b is given by*

$$\alpha(b) := \mathbb{E}(a_i|b) = b + \frac{\mu}{2}\frac{\mathrm{d}}{\mathrm{d}b}\left(\log(g(b))\right). \tag{9}$$

Proof. See Appendix A.4.

The above proposition shows a rationally inattentive player's expected response to a linear-quadratic problem in which her "target" b is distributed according to $g(\cdot)$. The player's expected action is higher than b when $g'(b)$ is positive, whereas it is "lagging" behind it when $g'(b)$ is negative. Moreover, the less likely b is (the lower $g(b)$ is), the higher the error $|\alpha(b) - b|$. Finally, errors are larger with higher information costs μ. So, the average action tends to be more concentrated around peaks in b's distribution with higher costs leading to larger discrepancies.

Proposition 5. *Let $g(\cdot)$ be the distribution of the best action of a smooth, monotone, full-support strategy profile of player i's opponents that satisfies condition* (6).

[18] The Online Appendix has a variant of this result for a broader class of games. Just as in Proposition 3, a player's posterior on b does not depend on the prior but only depends on the payoff function and the information cost. It is Gaussian iff payoffs are given by quadratic losses. See also Jung et al. (2019).

1. *The variance of player i's action in her best response is given by*

$$\mathrm{Var}(\boldsymbol{a}_i) = \mathrm{Var}(\boldsymbol{b}) - \mu/2 \tag{10}$$

2. *and its variance conditional on the best action being b is given by*

$$\mathrm{Var}(\boldsymbol{a}_i|b) = \frac{\mu}{2} + \frac{\mu^2}{4}\frac{\mathrm{d}^2}{\mathrm{d}b^2}(\log(g(b)))\,. \tag{11}$$

3. *Moreover, $\alpha(\cdot)$ and $\theta \mapsto \mathbb{E}(a_i|\theta)$ are increasing functions.*

Proof. See Appendix A.5.

Points 1 and 2 provide necessary conditions that the distribution of the best action to a smooth, monotone, full-support profile has to satisfy. Point 3 is a reassuring result confirming that in a best response the expected action moves in "the right way," i.e., follows the direction in which the best action and the fundamental move.

Results up to now concerned a player's best response to a "best action" distribution $g(\cdot)$: an equilibrium object. The way in which $g(\cdot)$ and other equilibrium objects are shaped by the model's primitives (the prior p, coordination motive γ and cost μ) is addressed in Sections 4–6

Comment: Continuous and Discontinuous Strategies

Jung et al. (2019) find that continuous problems with mutual information costs can have discontinuous solutions. Below it is argued that unless information costs are large, in the present setting discontinuous strategies are unlikely to appear in equilibrium.

Begin with the following observation: In discrete optimal strategies à la Jung et al. (2019) the conditional probability $\pi(a|\theta)$ received by an action a in the (countable) consideration set is a continuous function of θ. In fact, it is analytic since it is driven by the analytic utility function (eq. (1)) and is modulated through the also analytic $p(\cdot)$. Therefore, even in a strategy profile where all players use discrete strategies, the average action $\bar{a}(\cdot)$ and, in consequence, the best action $b(\cdot)$ are analytic functions of θ. Now, as long as the information cost is low enough compared to the best action's distribution (condition (6)), the fact that $p(\cdot)$ and $b(\cdot)$ are analytic is sufficient to ensure that the best response has, indeed, a density (Matějka and Sims 2010, proposition 2). So, when information is not too expensive, continuous equilibria should be more likely.

If condition (6) does not hold (i.e., if μ is high), then the solution to player i's optimization problem is in *discontinuous strategies*. In this case, player i's best response includes putting probability atoms on some subset of the action space that has no limit points instead of following a strategy with full support. The solution to this problem can be identified numerically (as in Jung et al. 2019) and the action distribution follows a multinomial logit rule as in Matějka and McKay (2015). In the extreme case, the support of $R(\cdot)$ is a single point and the player acquires no information (see also Section 5).

4 Equilibrium

Building on the results of Section 3, the class of equilibria in which players use continuous strategies (see Definition 2) is characterized in what follows. Some properties of this equilibrium class are described in Section 4.2, whereas 4.3 generalizes the main characterization result in populations with heterogeneous information costs.

4.1 Smooth, monotone, full-support equilibria

Having established conditions for the existence of a continuous best response to smooth, monotone, full-support strategy profiles (see Proposition 2), attention is now shifted towards equilibria whereby individual strategies are continuous. Such equilibria are defined formally as follows.

Definition 3 (Smooth, monotone, full-support equilibrium)**.** *A strategy profile r is called a* smooth, monotone, full-support equilibrium (SMFE) *if*

1. *Profile r is a smooth, monotone, full-support profile,*

and the strategy r_i of each player $i \in [0,1]$ is

2. *continuous and*
3. *a best response to r.*

As discussed in Section 3, when information costs are low enough, if the candidate equilibrium strategy profile satisfies point 1 of the above definition, then the best response to that would necessarily be continuous (i.e. it would satisfy point 2). Point 3 states that an SMFE is a fixed point. The following proposition provides a characterization of the class of SMFE.

Proposition 6. *Let (p, γ, μ) be a beauty contest with flexible information acquisition. Then the following two statements are equivalent*

(A) $\theta(\cdot)$ *is the inverse of the best action function and* $g(\cdot)$ *is the PDF of the distribution of the best action in an SMFE.*

(B) $\theta : \mathbb{R} \to \mathbb{R}$ *is a strictly increasing bijection,* $\mathscr{F}_{\xi}^{-1}[\exp(\mu\pi^2\xi^2)\hat{g}(\xi)]$ *is a probability distribution,*

$$\theta(b) = b - \frac{\mu\gamma}{2(1-\gamma)}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b))), \quad \textit{and} \tag{12}$$

$$g(b) = p(\theta(b))\theta'(b). \tag{13}$$

Proof. See Appendix A.6.

On the one hand, equation (12) holds independently of the prior and has to do with the way that individuals acquire information. On the other hand, equation (13) forces the distribution $g(\cdot)$ to be generated by the particular best action function (or, to be precise, its inverse) given the fundemental's prior distribution $p(\cdot)$, as seen in equation (5). According to Proposition 6, a distribution $g(\cdot)$ generates a unique $\theta(\cdot)$ through (12). Similarly, a best action function with inverse $\theta(\cdot)$ generates a unique distribution $g(\cdot)$ through (13). If these hold simultaneously, then an equilibrium has been identified.

Corollary 1. *In an SMFE, the best action function satisfies*

$$b(\theta) = \theta + \frac{\mu\gamma}{2(1-\gamma)}\frac{1}{b'(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\log\left(\frac{p(\theta)}{b'(\theta)}\right)\right) \tag{14}$$

and the average action function satisfies

$$\bar{a}(\theta) = \theta + \frac{\mu}{2(1-\gamma)(1+\gamma(\bar{a}'(\theta)-1))}\left(\frac{p'(\theta)}{p(\theta)} - \frac{\gamma\bar{a}''(\theta)}{1+\gamma(\bar{a}'(\theta)-1)}\right). \tag{15}$$

Sections 5 and 6 demonstrate that SMFE exist for certain classes of priors and for information costs μ bounded away from zero. With this in hand, the existence of SMFE for other families of priors and non-vanishing costs can be established through homotopy arguments.

Examining the equations of Corollary 1, it is seen that the equilibrium $b(\theta)$ deviates from θ to the extent that $\mu\gamma/2(1-\gamma)$ is positive. This has two implications. Firstly, absent coordination motives ($\gamma = 0$) the best action function coincides with θ. Secondly, as information costs μ decrease, the deviations of the equilibrium best action and average action from θ are be becoming smaller. In particular, only when $\mu = 0$ does equation (15) give $\bar{a}(\theta) = \theta$ identically. Of course — as seen in Proposition 1 — the best response to a smooth, monotone, full-support profile is not continuous any more. Nevertheless, the unique equilibrium that arises in this case is the one where all players acquire full information and use $r_i(a_i|\theta) = \delta(a_i - \theta)$ almost surely. Clearly, in this equilibrium it is true that $\bar{a}(\theta) = \theta$, which is what equation (15) yields for $\mu = 0$.

The Online Appendix characterizes the SMFE in a broader class of games. The absence of explicit payoff formulas in these more general games, though, limits the usability of the characterization therein.

4.2 Equilibrium properties

Both (14) and (15) are second-order nonlinear differential equations to which a general solution is not possible to find in closed form. However, it is still possible to describe some properties that any SMFE should have. Two sets of results are provided. Firstly the behavior of $b(\cdot)$ and $\bar{a}(\cdot)$ close to $\pm\infty$ as well as their ex-ante expected value are examined. Secondly, a relation between the variance of the best action and the variance of the fundamental σ^2 is established. Together with Propositions 4 and 5, these results relate to Bergemann and Morris (2013)'s agenda to identify moments of equilibrium distributions of variables.

Proposition 7. *Let r be an SMFE with average action function $\bar{a}(\cdot)$ and best action function $b(\cdot)$. Then*

1. $\int_{-\infty}^{+\infty} b(\theta)p(\theta)\,\mathrm{d}\theta = \int_{-\infty}^{+\infty} \bar{a}(\theta)p(\theta)\,\mathrm{d}\theta = \int_{-\infty}^{+\infty} \theta p(\theta)\,\mathrm{d}\theta = \bar{\theta}$
2. $\lim_{\theta\to+\infty} b(\theta) - \theta < 0$ *and* $\lim_{\theta\to+\infty} \bar{a}(\theta) - \theta < 0$

3. $\lim_{\theta\to-\infty} b(\theta)-\theta > 0$ *and* $\lim_{\theta\to-\infty} \bar{a}(\theta)-\theta > 0$

Proof. See Appendix A.7.

Bergemann and Morris (2013) derive the result of point 1 for a normally distributed prior and point out that it should hold for any prior. This result says that the "mean error" that players make is zero. They will miss their target θ most of the time but on average they should be correct. Points 2 and 3 show that — in equilibrium — players are biased towards the center of the distribution and take actions with extreme values (as compared to the ex-ante mean of the distribution) less often.

Proposition 8. *Let (p,γ,μ) be a beauty contest with flexible information acquisition that admits an SMFE and σ_b^2 be the variance of the best action b in that SMFE. Then*

1. $\sigma^2-\sigma_b^2 = \frac{\mu\gamma}{1-\gamma} + \mathrm{Var}(\theta-b)$.
2. $\mu < \frac{2(1-\gamma)}{1+\gamma}\sigma^2$.

Proof. See Appendix A.8.

An immediate consequence of point 1 of Proposition 8 is that in an SMFE the best action is more concentrated than the fundamental ($\sigma^2 \geq \sigma_b^2$). Point 2 gives an upper bound to the value of μ. This upper bound is not tight: what one can say for sure is that if μ exceeds this value, then (p,γ,μ) has no SMFE.

4.3 Heterogeneous costs

The preceding equilibrium analysis assumed that all players face the same information costs. It might well be plausible, though, that in the real world agents face heterogeneous costs. Maintaining the assumption that players are rationally inattentive and facing Shannon entropy costs (i.e. that the costs follow the functional form (2)), players are allowed to have differing unit information cost. In particular, each agent i has a cost $\mu_i \in [\mu_{\min},\mu_{\max}]$ $(0 \leq \mu_{\min} < \mu_{\max} < \infty)$. Let also the distribution of agents' costs be $M \in \Delta([\mu_{\min},\mu_{\max}])$ with expected value $\mathbb{E}(\mu) = \int_{\mu_{\min}}^{\mu_{\max}} \mu \,\mathrm{d}M(\mu) = \bar{\mu}$.

The following proposition looks at the aggregate equilibrium behavior of a heterogeneous population M of players that follow continuous strategies.

Proposition 9. *Let a population with costs distributed according to $M \in \Delta([\mu_{\min}, \mu_{\max}])$ play a beauty contest with prior p and coordination motive γ. Let also g be the distribution of the best action in an equilibrium. If $\mathscr{F}_\xi^{-1}\left[\exp(\mu_{\max}\pi^2\xi^2)\hat{g}(\xi)\right]$ is the PDF of a probability distribution, then all players follow continuous strategies in equilibrium, the inverse of the best action is given by*

$$\theta(b) = b - \frac{\bar{\mu}\gamma}{2(1-\gamma)}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b))),$$

and $g(\cdot) = p(\theta(\cdot))\theta'(\cdot)$. In fact, this equilibrium is an SMFE.

Proof. See Appendix A.9.

As a consequence of Proposition 9, a heterogeneous population M with low costs can be conveniently treated as a population of homogeneous players all endowed with the population-wide average cost $\bar{\mu}$ of M. In particular, the equilibrium average action and best action should satisfy (14) and (15) with $\mu := \bar{\mu}$.

5 Aggregately Affine Equilibria

As mentioned earlier, it is not possible to solve (14) or (15) in the general case. Instead, this section examines smooth, monotone, full-support equilibria of a specific form. In particular, equilibria whereby the average action (and, thus, also the best action) is affine in θ are identified.

Definition 4 (Aggregately affine equilibrium). *An SMFE will be called an* aggregately affine equilibrium (AAE) *if the best action function has the form $b(\theta) = \kappa\theta + d$ for some constants $\kappa > 0$ and $d \in \mathbb{R}$.*

The following proposition gives a necessary and sufficient condition for an AAE to exist.

Proposition 10. *Let (p, γ, μ) with $\mu > 0$ and $\gamma > 0$ be a beauty contest with flexible information acquisition. The following statements are equivalent:*

(A) (p, γ, μ) admits an aggregately affine equilibrium.

(B) p is a normal distribution, and

(i) either $\gamma \leq \frac{1}{2}$ *and* $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$

(ii) or $\gamma > \frac{1}{2}$ *and* $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$.

Proof. See Appendix A.10.

The result of Proposition 10 is stark. It implies that in the presence of strategic motives and information costs nicely tractable equilibria exist only under the assumption of a normal prior, which is prevalent in the literature. Under any other prior distribution, analytically identifying SMFEs becomes a formidable task.

Now, turning to the characterization of the set of equilibria, even though one cannot easily find the number of equilibria, the following proposition shows that there are cases where multiple equilibria exist.

Proposition 11. *Let* (p, γ, μ) *be a beauty contest with flexible information with* p *being a normal distribution and* $\mu > 0$.

1. *If either*

   (a) $\gamma \leq \frac{1}{2}$ *and* $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$ *or*

   (b) $\gamma > \frac{1}{2}$ *and* $\sigma^2 \geq \frac{\mu}{2(1-\gamma)^2}$

   then (p, γ, μ) *admits exactly one AAE with* $\kappa = \frac{1}{2}\left(1 + \sqrt{1 - \frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right)$.

2. *If* $\gamma > \frac{1}{2}$ *and* $\sigma^2 \in \left(\frac{2\mu\gamma}{1-\gamma}, \frac{\mu}{2(1-\gamma)^2}\right)$ *then* (p, γ, μ) *admits exactly two AAE: one with* $\kappa = \frac{1}{2}\left(1 + \sqrt{1 - \frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right)$ *and one with* $\kappa = \frac{1}{2}\left(1 - \sqrt{1 - \frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right)$.

Proof. See Appendix A.11.

Interestingly, under the conditions where two AAE exist, there also exists an equilibrium where players do not obtain any information (which happens when $\sigma_b^2 < \mu/2$). In this sense, three equilibria of the classes considered exist under these parameter configurations. Moreover, if neither the conditions of point 1 nor of point 2 hold, there exists an equilibrium without information acquisition. Notice that in this equilibrium $b'(\theta) = 1 - \gamma$ for all values of θ. So, the first derivative of $b(\cdot)$ is constant *even though*

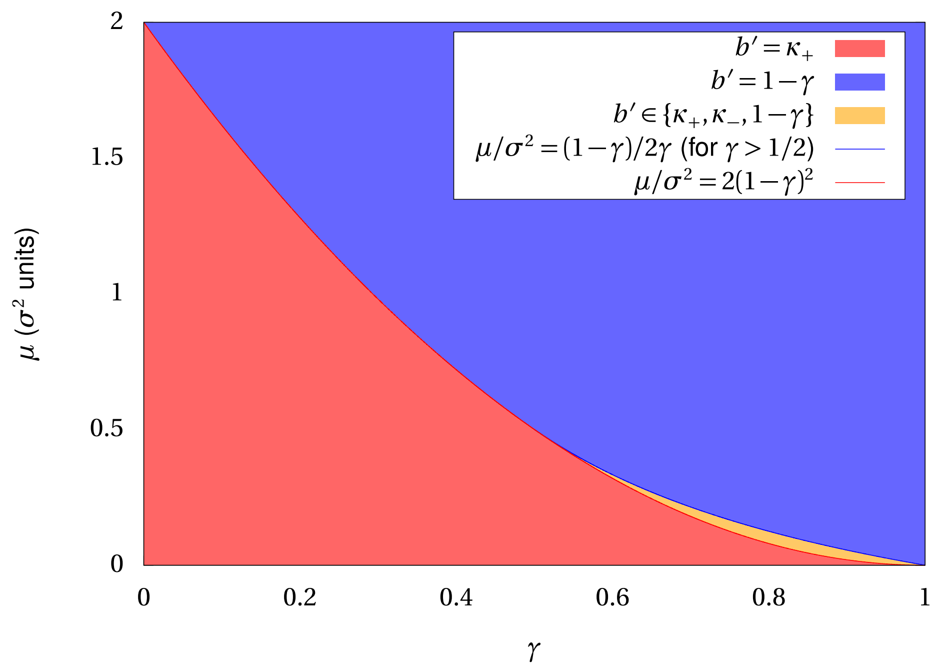


Figure 1: Aggregately affine equilibria and equilibria without information acquisition in different regions of μ and γ when θ is normally distributed with variance σ^2. Red area: one AAE; Orange area: two AAE and no-info equilibrium; Blue area: no-info equilibrium.

this is not an AAE/SMFE, as the strategy that the players use is not continuous. Figure 1 summarizes these results.

The intuition on how multiple equilibria appear under these parameter configurations is the following. Fix the information cost μ and start increasing γ. When γ is small ($\gamma < 1/2$), players are able to coordinate on a unique AAE where they acquire information. As γ increases, less information is acquired — as the importance of getting close to the realized value of θ is decreasing and the motive to coordinate increases. When γ reaches the critical value for which $\mu = 2(1-\gamma)^2\sigma^2$, the coordination motive becomes so strong that an equilibrium where no player acquires any information is established: if no player acquires information, they are sure to perfectly coordinate at $\bar{\theta}$ — thus saving the costs of acquiring information. When costs are high, there is no overlapping region where an AAE and an equilibrium without information acquisition coexist. However, when information costs are lower, there is a region of the coordination parameter that allows for the existence of an AAE and an equilibrium without information acquisition at the same time: coordinating on acquiring and using some information and on not acquiring any information are both equilibria.

As mentioned previously, under the parameter configurations that allow for the co-existence of an AAE and an equilibrium without information acquisition, a second AAE also exists under which each of the players acquires a smaller amount of information compared to the first, stable AAE (κ is smaller). This second AAE is created as an in-between case of the two equilibria described above. Moreover, it is *unstable* in the sense that iterative best responses lead away from this equilibrium. This argument is explained in more detail in Appendix B where stability analysis is conducted.

6 Beyond Normality: Method and Implications

6.1 A method to address non-normal priors

Despite it being challenging to identify equilibria in the usual way (i.e. to find smooth, monotone, full support equilibria for a particular parameter combination (p, γ, μ)), considerable progress can be made through following a "backwards" procedure. In particular, one can postulate some distribution $g(\cdot)$ to be the PDF of the best action b in an SMFE and then, making use of equation (12), calculate (analytically or numerically) the prior distribution of the fundamental through

$$p(\theta) = g(b(\theta))b'(\theta).$$

A major challenge that arises during this process rests in confirming whether condition (6) holds, i.e., whether the calculated distribution of a player's action is, indeed, a distribution. In order to overcome this problem, this section makes use of best action distributions $g(\cdot)$ which are conjugate for the normal distribution.[19]

When g is conjugate for the Gaussian, $g(\cdot)$ and $R(\cdot)$ belong to the same distribution family with $R(\cdot)$ having the same mean as $g(\cdot)$ and a variance reduced by $\mu/2$. In this way, confirming that (6) gives, indeed, the PDF of a probability distribution boils down to making sure that the parameters calculated for $R(\cdot)$ fall within the allowed ranges for the particular distribution family. The method is demonstrated for the skew normal

[19] Recall that from the result of Section 3, $\boldsymbol{b}$ is the sum of two variables, $a_i \sim R$ and $\varepsilon_i \sim N(0, \mu/2)$.

distribution in Section 6.2, while Section 6.3 shows that moving away from the Gaussian prior leads to new insights about how equilibrium actions and economic fundamentals are related.

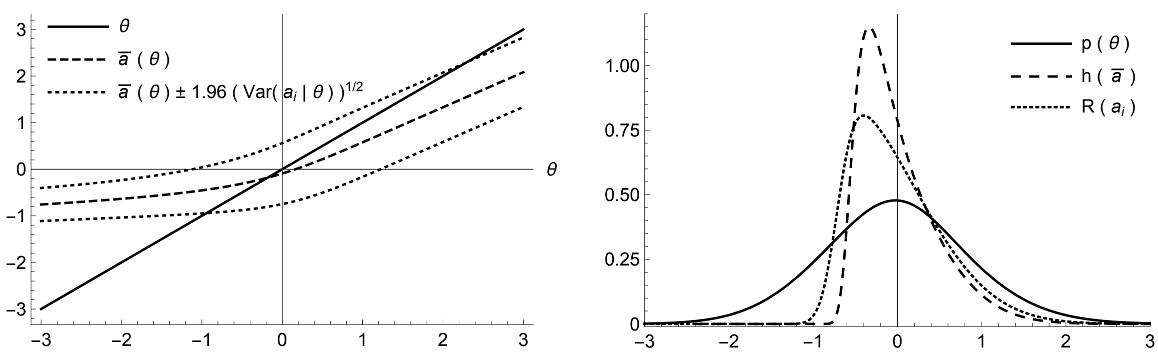
6.2 Application: skew normal $g(\cdot)$

The skew normal distribution $SN(b_0, \omega, \lambda)$ with parameters $b_0 \in \mathbb{R}$, $\omega \in (0, \infty)$, and $\lambda \in \mathbb{R}$ (introduced by O'Hagan and Leonard 1976), is a continuous distribution over $\mathbb{R}$ with PDF

$$f(b; b_0, \omega, \lambda) = \frac{2}{\omega}\phi\left(\frac{b - b_0}{\omega}\right)\Phi\left(\lambda\left(\frac{b - b_0}{\omega}\right)\right)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the PDF and cumulative distribution function (CDF) of the standard normal distribution ($N(0,1)$). The important variable here is λ that adds skewness to the distribution (notice that when $\lambda = 0$, the distribution boils down to a Gaussian).

Using the fact that the skew normal distribution is conjugate for the normal distribution, if $b \sim SN(b_0, \omega, \lambda)$, then $a_i \sim SN\left(b_0,\ \sqrt{\omega^2 - \mu/2},\ \lambda\omega\left(\omega^2 - \mu(1+\lambda^2)/2\right)^{-1/2}\right)$ (see Azzalini 1985). So, as long as information costs are not too large ($\mu < 2\omega^2/(1 + \lambda^2)$), the function $R(\cdot)$ defined through (6) is the PDF of a skew normal probability distribution, and the players have a continuous best response to $g(\cdot)$.



(a) Equilibrium average action and "confidence intervals" as a function of the fundamental

(b) Equilibrium distributions of the fundamental, average action, and the action of a player

Figure 2: Prior and equilibrium objects for an SMFE that has a $g(\cdot)$ which is skew normal with $(\omega, \lambda) = (1, 2)$ and b_0 such that $\mathbb{E}(b) = 0$. $\mu = 0.35$, $\gamma = 0.5$.

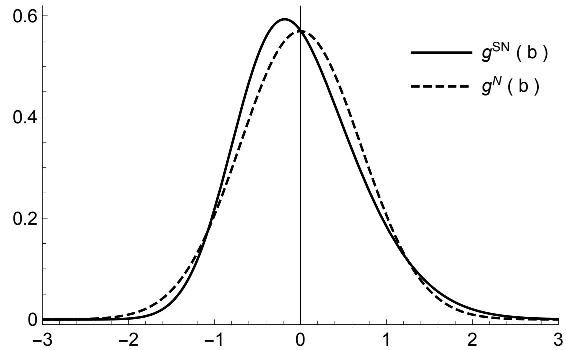
Figure 2 presents the prior along with equilibrium objects for an SMFE in which the best action follows a skew normal distribution. In Figure 2a, it is seen that the population's average action follows the fundamental more closely for θ to the right of $\theta^* : \bar{a}(\theta^*) = \theta^*$ than for θ to the left of θ^*. This is explained by the fact that the distribution is right-tailed and therefore there is more variation to which the players pay attention towards the right end of the distribution. Figure 2b presents the equilibrium densities of θ, $\bar{a}$, and a_i. As expected, the distribution of a_i and — even more so — that of $\bar{a}$ are more concentrated than the fundamental's PDF $p(\cdot)$. Moreover, although the prior distribution $p(\cdot)$ is only slightly (barely perceivably) skewed, individual actions a_i are clearly skewed (Figure 2b), along with the equilibrium average action $\bar{a}$.

6.3 Implications of non-normality

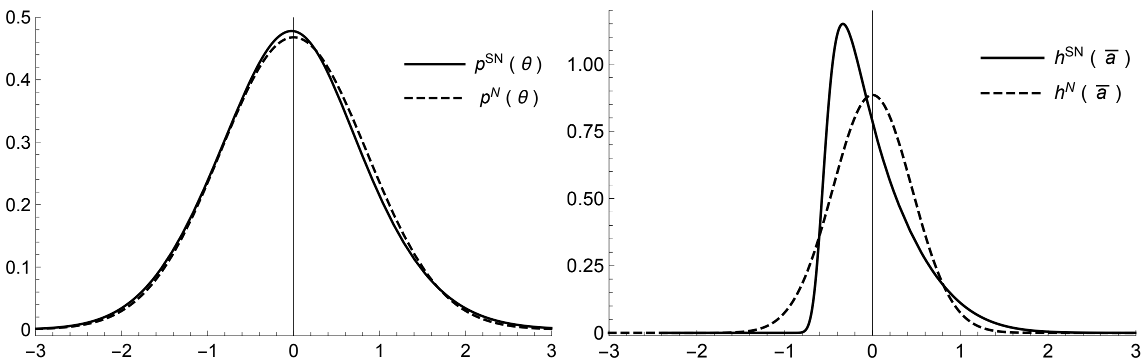
In theoretical research, using normal priors when real-world distributions of economic fundamentals are, in fact, non-normal will obviously lead to predictions that are not entirely accurate. One may wonder, though, to what extent a simplifying assumption of normality produces inaccurate results and, importantly, whether such a simplification leads to the researcher missing out on new insights, especially given the high tractability of models with normal distributions (see Section 5). Using the example of 6.2, this section demonstrates that even when prior distributions are very close to the normal, distributions of equilibrium objects can be significantly and qualitatively different from those derived for the normal.

In order to illustrate these differences, consider two beauty contests both of which have a coordination motive $\gamma = 0.5$ and information cost $\mu = 0.35$. In the first beauty contest there is an SMFE in which the best action is distributed according to $g^{\mathrm{SN}}(\cdot)$, a skew normal distribution with mean 0 and parameters $(\omega, \lambda) = (1, 2)$. Similarly, in the second beauty contest there is an SMFE in which the best action is distributed according to a normal distribution $g^{\mathrm{N}}(\cdot)$ such that $\mathbb{E}(g^{\mathrm{N}}) = \mathbb{E}(g^{\mathrm{SN}}) = 0$ and $\mathrm{Var}(g^{\mathrm{N}}) = \mathrm{Var}(g^{\mathrm{SN}})$. The two distributions can be seen in Figure 3a.

Using the method described in 6.1, one can calculate the prior distribution of the fundamental and the equlibirium distribution of the average action for each of the two beauty contests. These are seen in panels 3b and 3c, respectively. From the diagrams it is clear that although the two priors are barely different, the equilibrium distributions of



(a) Distribution of the equilibrium best action: $g^N(\cdot)$ is Gaussian while $g^{SN}(\cdot)$ is skew normal with positive skewness. The two distributions have the same mean and variance.



(b) Prior distribution of the fundamental

(c) Distribution of the average action

Figure 3: Comparison of distributions of prior and equilibrium objects between SMFE with normal $g(\cdot)$ and skew normal $g(\cdot)$. In both cases $\mu = 0.35$ and $\gamma = 0.5$.

$\bar{a}$ differ significantly. In particular, the distribution of the best action in the SMFE with g^{SN} is highly skewed (the left tail almost disappears), even though the prior that gives rise to it is only slightly away from the Gaussian prior of the second beauty contest.

The preceding analysis shows that highly skewed market obervables do not necessarily reflect highly skewed fundamentals but can be the result of *small* fundamental non-normalities, amplified by strong coordination motives and costly information.

Take, for example, a situation based on the model of Hellwig and Veldkamp (2009): Competing firms choose their (log) price a_i trying to match a demand shock (the fundamental) in the presence of coordination motives (higher average price leads to higher best-response price from an individual firm). Now consider an analyst who wants to study the distribution of the demand shock θ in this market. As fundamentals are unobservable, what the analyst can observe are individual and (more realistically) aggregate data, such as the average price. After multiple observations, the analyst observes a highly skewed distribution of the average price and concludes that the underlying distribution of the fundamental has a similar property and is by no means close to a Gaussian. The example presented above, shows that this reasoning is heavily flawed and that, in fact, the fundamental can be very close to a Gaussian.

Naïvely attempting to draw conclusions about the (unobservable) distribution of the fundamental from observed market behavior may lead to over-estimating its skewness and under-estimating its variance. Consequently, this can exacerbate the issues that statistical inference with skewed distributions brings about (e.g. constructing confidence intervals). These effects become stronger as information costs or coordination motives increase.

7 Discussion

This paper studied large coordination games played by rationally inattentive agents in the presence of fundamental motives. What follows discusses and relates the paper's results to the ones found in other studies.

Equilibrium sensitivity

Jung et al. (2019) make the point that small differences in the prior of decision-theoretic problems of rational inattention can lead to very different behavior of the agent. In particular, the agent may switch from continuous to discontinuous strategies in problems whose priors are very close. In Section 6 it was shown that in *games* with rationally inattentive agents problems with very similar priors can lead to very different equilibrium behavior even if the equilibria in both problems are in continuous strategies. The reason for the high sensitivity of solutions to the prior in Jung et al. (2019) is the information cost being too high for continuous strategies to be optimal in one of the problems, or the distribution having fat tails. In contrast, the driving force behind the similar result of Section 6 lies in the coordination motive working together with the information cost. Even small asymmetries in the prior can lead to players having a preference for one side of the distribution over the other when they coordinate (since they cannot follow the fundamental closely because of information costs). As they make zero error in expectation (Proposition 7), this results in highly skewed distributions of individual and average actions.

Equilibrium multiplicity

Existing literature has pointed out that entropy-related information costs, as the ones used in this paper, can lead to multiple equilibria (e.g. Hellwig, Kohls, and Veldkamp 2012). With flexible information acquisition technology when information is cheap, players obtain more of it and the game gets closer to a complete-information one. Yang (2015) uses such a technology to study a two-player coordination game. The complete-information game has multiple equilibria for a range of realizations of the random variable and this multiplicity is recovered when information costs are low. It is therefore unclear whether the multiplicity of equilibria is present in his model because of the form of the underlying game or because of the information acquisition technology employed.

In Section 5 it was shown that multiple AAE arise in for *intermediate* values of μ (bounded away from zero for any fixed $\gamma > 1/2$). On the one hand, this differs from the result of Yang (2015) who finds multiple equilibria for low values of the information cost — even though both the interaction studied here and the one in his study are coordination games (with the caveat that the statement is only for affine equilibria in the present setting). On the other hand, in agreement with results in Yang (2015), when information costs are small, the equilibrium structure of the complete-information game is recovered: a unique equilibrium in the present case, multiple in Yang's. So, vanishing information costs under flexible information acquisition lead to recovery of the equilibrium structure of the full-information game (*cf.* global games, Carlsson and van Damme 1993). This observation is also consistent with Morris and Yang (2016). In their setting (which has multiple equilibria under full information), continuous choice breaks when it is sufficiently easy to distinguish nearby states, and multiple equilibria appear.[20] Under the rational inattention framework of this paper, nearby and far away states are always equally easy to distinguish. This, in turn, leads to the equilibrium structure of the full-information game being recovered when information costs vanish.

[20] A recent experiment by Goryunov and Rigos (2019) reports that the behavior of laboratory participants is consistent with this intuition.

Moreover, the results of Section 5 confirm that with mutual information costs equilibrium multiplicity can arise, even when attention is restricted to a very particular class of tractable equilibria.

Finally, relating to Section 4.3, populations with heterogeneous costs can give rise to situations in which some individuals use continuous strategies while others, with higher costs, use discrete strategies (or even not acquire information whatsoever). Such cases introduce a new channel for equilibrium multiplicity which can be investigated further.

Linear equilibria

The notion of a linear equilibrium is often encountered in existing literature studying beauty contest-like games (see for example Angeletos and Pavan 2007; Morris and Shin 2002; Myatt and Wallace 2012). In linear equilibria each player takes an action that is a linear combination of the messages she receives from (potentially) different sources. When signal noises and the prior follow normal distributions — which is the common modelling choice in the aforementioned literature — then any linear equilibrium is also aggregately affine in the sense of Section 5. Proposition 10 shows that tractability of equilibria when players are rationally inattentive is heavily dependent on this very assumption. As soon as one parts with the normal prior assumption, one has to accept that deriving equilibria in closed form may be impossible. Importantly, Proposition 6 and the method of Section 6.1 can help identify or approximate equilibria even when the prior is not normal.

Improper priors

In order to compare this paper's results with those in settings with uniform (improper) priors (e.g. Morris and Shin 2002; Myatt and Wallace 2012) one has to treat the proper prior $p(\cdot)$ as an interim belief shared by all players after observing a public signal and before observing their private ones. If, for example, the public signal is distributed around the realization of the fundamental as $N(\theta, \sigma^2)$, then it should lead to a common Gaussian interim belief (which is the "prior" in the present model). Myatt and Wallace (2012) find that under certain conditions two linear equilibria may arise under a mutual information-based cost specification: one with and one without information acquisition.

The analysis conducted in Section 5 shows that there is one more aggregately affine equilibrium with a lower degree of information acquisition which is albeit unstable (Appendix B).

Conditionally correlated signals

Hellwig and Veldkamp (2009) point out that strategic complementarities lead to complementarities in players' information acquisition decisions. Moreover, one can think of situations where a player may want to have information about other players' signal realizations or may even want other players to have information about her own realization (as in Kozlovskaya 2018, for example). The model presented here does not allow for such correlation as the only information players can obtain is about the fundamental and not about others' signal realizations (signals are always conditionally independent). However — as Denti (2019) argues — when all players are "small" and aggregate behavior is all that matters, incentives to learn about others' signal realizations disappear as aggregate behavior becomes a deterministic function of the fundamental.

Appendix

A Omitted Proofs

A.1 Proof of Proposition 1

As $\mu = 0$, player i can obtain full information on the value of θ without paying any costs. So, conditional on the value of θ, her optimization problem becomes

$$\max_{a_i} \quad -(1-\gamma)(a_i - \theta)^2 - \gamma(a_i - \bar{a}(\theta))^2$$

Taking a first order condition, one obtains that the optimal action is given by

$$(1-\gamma)\theta + \gamma\bar{a}(\theta).$$

So, given any $\bar{a}(\cdot)$ and any value of θ, player i has a unique best action given by the expression $b(\theta) = (1-\gamma)\theta + \gamma\bar{a}(\theta)$. Thus, her best response is to assign a probability mass of 1 to that action (conditional on θ). That is, her best response is to use r_i given by $r_i(a_i|\theta) = \delta(a_i - b(\theta))$ with δ being Dirac's delta function (almost all θ).

A.2 Proof of Proposition 2

Consider variations of the strategy of player i. These variations will be of the type $\tilde{r} = r + \varepsilon\eta$ for some $\varepsilon > 0$. These variations should still be feasible. That is, for all θ, it is required that $r(\cdot|\theta) + \varepsilon\eta(\cdot|\theta)$ is a probability distribution over A_i. It is required, thus, that for all θ, $\int_{A_i} r(a_i|\theta) + \varepsilon\eta(a_i|\theta)\,\mathrm{d}a_i = 1$ which leads to the condition that for all θ, $\int_{A_i} \eta(a_i|\theta)\,\mathrm{d}a_i = 0$. It also has to be that $r(a_i|\theta) + \varepsilon\eta(a_i|\theta) \geq 0$ and so $\eta(a_i|\theta) \geq -r(a_i|\theta)/\varepsilon$ for all a_i and θ. From the above equations, the following is calculated:[21]

$$\begin{aligned} U(r_i + \varepsilon\eta, r_{-i}) &= -(1-\gamma)\int_\Theta \int_{A_i} (a_i - \theta)^2 (r_i(a_i|\theta) + \varepsilon\eta(a_i|\theta))p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \\ &- \gamma \int_\Theta \int_{A_i} (a_i - \bar{a}(\theta))^2 (r_i(a_i|\theta) + \varepsilon\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta. \end{aligned} \tag{16}$$

And the derivatives:

$$\begin{aligned} \left.\frac{\mathrm{d}U(r + \varepsilon\eta, r_{-i})}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} &= -(1-\gamma)\int_\Theta \int_{A_i} (a_i - \theta)^2 \eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \qquad (17) \\ &- \gamma \int_\Theta \int_{A_i} (a_i - \bar{a}(\theta))^2 \eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta \end{aligned}$$

$$\begin{aligned} \left.\frac{\mathrm{d}I(r + \varepsilon\eta)}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} &= \int_\Theta \int_{A_i} \log(r(a_i|\theta))\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \qquad (18) \\ &- \int_{A_i} \log(R_i(a_i))H(a_i)\,\mathrm{d}a_i \end{aligned}$$

with $H(a_i) = \int_\Theta \eta(a_i|\theta)p(\theta)\,\mathrm{d}\theta$.

Since the variations considered have to be feasible, player i has to solve the following constrained optimization problem:

$$\max_{r_i \in L^1(\Theta,p)} U(r_i, r_{-i}) - \mu I(r_i)$$

[21]The effect of the other players' strategies is incorporated in $\bar{a}(\theta)$.

$$\text{s.t.} \int_{A_i} r_i(a_i|\theta)\,\mathrm{d}a_i = 1 \qquad \text{for all } \theta \in \Theta.$$

So, the Lagrangian for player i's decision problem will be

$$\mathscr{L}(r_i, k(\theta)) = U(r_i, r_{-i}) - \mu I(\boldsymbol{\theta}, \boldsymbol{a}_i) - \int_{\Theta} k(\theta)\left(\int_{A_i} r(a_i|\theta)\,\mathrm{d}a_i - 1\right)p(\theta)\,\mathrm{d}\theta$$

where $k(\theta)$ is the Lagrange multiplier for the θ-constraint.

Therefore, for any given $\theta \in \Theta$ and all possible perturbations η, an optimal strategy r should satisfy the following first order conditions:

$$\left.\frac{\mathrm{d}\mathscr{L}(r_i + \varepsilon\eta, k(\theta))}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 0 \Rightarrow$$

$$\int_{\Theta}\int_{A_i}\left[u_i(a_i,\theta) - \mu\log\left(\frac{r(a_i|\theta)}{R_i(a_i)}\right) - k(\theta)\right]\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta = 0 \qquad (19)$$

$$\text{and} \int_{A_i} r_i(a_i|\theta)\,\mathrm{d}a_i = 1 \qquad \text{for all } \theta \in \Theta. \qquad (20)$$

Where

$$u_i(a_i,\theta) = -(1-\gamma)(a_i - \theta)^2 - \gamma(a_i - \bar{a}(\theta))^2. \qquad (21)$$

Since condition (19) has to be satisfied *for all* η, it has to be the case that

$$-(1-\gamma)(a_i-\theta)^2 - \gamma(a_i - \bar{a}(\theta))^2 - \mu\left[\log(r_i(a_i|\theta)) - \log(R_i(a_i))\right] = k(\theta) \text{ for all } \theta \in \Theta.$$

So $r(a_i|\theta)$ has to be:

$$r(a_i|\theta) = R_i(a_i)\exp\left(-\frac{k(\theta)}{\mu}\right)\exp\left(\frac{u_i(a_i,\theta)}{\mu}\right) \qquad (22)$$

and (22) can be rewritten as

$$r(a_i|\theta) = R_i(a_i)K(\theta)\exp\left(\frac{u_i(a_i,\theta)}{\mu}\right). \qquad (23)$$

where $K(\theta) = \exp\left(-\frac{k(\theta)}{\mu}\right)$. All that remains to be done is to determine the functions $K(\cdot)$ and $R_i(\cdot)$. Now, from the definition of $R_i(a_i)$:

$$R_i(a_i) = \int_{\theta} r(a_i|\theta)p(\theta)\,\mathrm{d}\theta \Rightarrow \int_{\Theta}\frac{r(a_i|\theta)}{R_i(a_i)}p(\theta)\,\mathrm{d}\theta = 1.$$

After substituting from (23) and (21), simple calculations give

$$\int_{-\infty}^{+\infty} K(\theta)\exp\left(-\frac{(a_i-b(\theta))^2}{\mu}\right)\exp\left(-\frac{\gamma(1-\gamma)}{\mu}(\theta-\bar{a}(\theta))^2\right)p(\theta)\,\mathrm{d}\theta = 1. \qquad (24)$$

In the above, $b(\theta)=(1-\gamma)\theta+\gamma\bar{a}(\theta)$. By assumption (smooth, monotone, full-support strategy profile), b is invertible with b^{-1} being the inverse of b. Because of assumption 2 of Definition 1, $b(\cdot)$ is bijective and strictly increasing. Thus, $\lim_{b\to\infty} b^{-1}(b)=\infty$ and $\lim_{b\to-\infty} b^{-1}(b)=-\infty$. So, with a change of the variable of integration from θ to $b=b(\theta)$, and by defining $G(\cdot)$ as

$$G(b)=\frac{K(b^{-1}(b))\exp\left(-\frac{\gamma(1-\gamma)}{\mu}(b^{-1}(b)-\bar{a}(b^{-1}(b)))^2\right)p(b^{-1}(b))}{(1-\gamma)+\gamma\bar{a}'(b^{-1}(b))} \qquad (25)$$

condition (24) can be rewritten as

$$\int_{-\infty}^{+\infty} G(b)\exp\left(-\frac{1}{\mu}(a_i-b)^2\right)\mathrm{d}b = 1. \qquad (26)$$

Notice that the above condition has to hold for all $a_i\in\mathbb{R}$. This can only happen if $G(b)=1/\sqrt{\pi\mu}$.

Proof. Notice that the left-hand side of equation (26) is the convolution of G and f given by $f(x)=\exp(-x^2/\mu)$. Now, take the Fourier transform on both sides and use the convolution theorem:

$$\begin{aligned}\mathscr{F}_{a_i}[(G*f)(a_i)](\xi) &= \mathscr{F}_{a_i}[1](\xi) \Rightarrow \mathscr{F}_{a_i}[G(a_i)](\xi)\cdot\mathscr{F}_{a_i}[f(a_i)](\xi)=\delta(\xi)\\ \Rightarrow \mathscr{F}_{a_i}[G(a_i)](\xi) &= \frac{1}{\sqrt{\pi\mu}}\exp(\mu\pi^2\xi^2)\delta(\xi)\end{aligned}$$

Where $\delta(\cdot)$ is Dirac's delta function. By taking the inverse Fourier transform on both sides, the statement is proven:

$$\begin{aligned}G(b) &= \mathscr{F}_{\xi}^{-1}\left[\frac{1}{\sqrt{\pi\mu}}\exp(\mu\pi^2\xi^2)\delta(\xi)\right](b)\\ &= \frac{1}{\sqrt{\pi\mu}}\int_{-\infty}^{+\infty}\exp(2\pi\imath\xi x)\exp(\mu\pi^2\xi^2)\delta(\xi)\,\mathrm{d}\xi = \frac{1}{\sqrt{\pi\mu}}.\end{aligned}$$

□

So now $K(\theta)$ can be calculated.

$$K(\theta)=\frac{1+\gamma(\bar{a}'(\theta)-1)}{p(\theta)\sqrt{\pi\mu}}\exp\left(\frac{\gamma(1-\gamma)}{\mu}(\theta-\bar{a}(\theta))^2\right) \qquad (27)$$

Using (27) in (23) yields

$$r(a_i|\theta) = R_i(a_i)\frac{1+\gamma(\bar{a}'(\theta)-1)}{p(\theta)\sqrt{\pi\mu}}\exp\left(-\frac{(a_i-b(\theta))^2}{\mu}\right). \tag{28}$$

The solution has to also satisfy $\int_{-\infty}^{+\infty} r(a_i|\theta)\,\mathrm{d}a_i = 1$ for all θ. Again, changing the variable from θ to $b = b(\theta)$, this condition yields

$$\int_{-\infty}^{+\infty} R_i(a_i)\exp\left(-\frac{(b-a_i)^2}{\mu}\right)\mathrm{d}a_i = \sqrt{\pi\mu}p(b^{-1}(b))\left(b^{-1}\right)'(b). \tag{29}$$

Notice that the left-hand side of equation (29) is the convolution of R_i and f. Now, take the Fourier transform on both sides and use the convolution theorem

$$\mathscr{F}_{a_i}[R_i(a_i)](\xi)\cdot\mathscr{F}_b[f(b)](\xi) = \sqrt{\pi\mu}\cdot\mathscr{F}_b[p(b^{-1}(b))\left(b^{-1}\right)'(b)](\xi) \Rightarrow$$

$$\mathscr{F}_{a_i}[R_i(a_i)](\xi) = \exp(\mu\pi^2\xi^2)\cdot\mathscr{F}_b[p(b^{-1}(b))\left(b^{-1}\right)'(b)](\xi) \Rightarrow \tag{30}$$

$$R_i(a_i) = \mathscr{F}_\xi^{-1}[\exp(\mu\pi^2\xi^2)\cdot\mathscr{F}_b[p(b^{-1}(b))\left(b^{-1}\right)'(b)](\xi)](a_i) \tag{31}$$

If the expression above is the PDF of a probability distribution, then the solution is calculated by equation (23) which — after noticing that $g(b) = p(b^{-1}(b))\left(b^{-1}\right)'(b)$ is the PDF of the best action $b = b(\theta)$ — becomes

$$r_i(a_i|\theta) = R_i(a_i)\frac{b'(\theta)}{p(\theta)\sqrt{\pi\mu}}\exp\left(-\frac{(a_i-b(\theta))^2}{\mu}\right) \tag{32}$$

with

$$R_i(a_i) = \mathscr{F}_\xi^{-1}[\exp(\mu\pi^2\xi^2)\cdot\mathscr{F}_b[g(b)](\xi)](a_i). \tag{33}$$

This solution is *unique*. The analyticity of $p(\cdot)$ and $b(\cdot)$ (and, therefore, of $g(\cdot)$) ensures that the solution to the player's decision problem is actually in continuous strategies rather than in strategies that put positive probability mass on a countable set of actions i.e. discrete or strategies with both a discrete and a continuous part (see Matějka and Sims 2010, Proposition 2).

Now, for the "only if" part, if $R(\cdot)$ defined through (33) was not the PDF of a probability distribution, player i would not have a continuous best reply to r_{-i}. Because if she did, the marginal of her action would need to be defined by equation (33). □

A.3 Proof of Proposition 3

From Bayes's rule, one gets:

$$\varrho_i(b|a_i) = \frac{\tau_i(a_i|b)g(b)}{R(a_i)} \tag{34}$$

where $\tau_i(\cdot|b)$ is the PDF of player i's action a_i conditional on the best action being b. As $b(\cdot)$ is bijective with inverse $b^{-1}(\cdot)$, one can derive τ_i from the result of Proposition 2 with a change of variable:

$$\begin{aligned}\tau_i(a_i|b) &= \frac{R_i(a_i)}{p(b^{-1}(b))\,(b^{-1})'(b)}\frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{(a_i-b)^2}{\mu}\right)\\ &= R_i(a_i)\frac{1}{g(b)}\frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{(a_i-b)^2}{\mu}\right)\end{aligned}$$

and comparing with (34), one obtains

$$\varrho_i(b|a_i) = \frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{(a_i-b)^2}{\mu}\right).$$

□

A.4 Proof of Proposition 4

Denote by $\tau(a|b)$ the probability density of action a conditional on the best action being b in player i's best response. From Bayes's rule

$$\tau(a|b) = \frac{\varrho(b|a)R(a)}{g(b)}.$$

Using the property of the Fourier transform (see equation (38)), the expected action of player i, conditional on b, is

$$\alpha(b) = -\frac{1}{2\pi\imath}\left(\mathscr{F}_a[t(a|b)]\right)'(0) = -\frac{1}{g(b)2\pi\imath}\left(\mathscr{F}_a[\varrho(b|a)R(a)]\right)'(0)$$

and using the convolution theorem as well as the properties of the Fourier transform,

$$\alpha(b) = -\frac{1}{g(b)2\pi\imath}\left(\mathscr{F}_a[\varrho(b|a)] * (\mathscr{F}_a[R(a)])'\right)(0). \tag{35}$$

Now

$$\mathscr{F}_a[\varrho(b|a)](x) = \mathscr{F}_a\left[\frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{(a-b)^2}{\mu}\right)\right](x)$$

$$= \frac{1}{\sqrt{\pi\mu}} \exp(-2\pi\imath bx) \mathscr{F}_a \left[\exp\left(-\frac{a^2}{\mu} \right) \right](x)$$

$$= \exp(-2\pi\imath bx) \exp\left(-\mu\pi^2 x^2\right) \equiv \psi(x) \tag{36}$$

And

$$(\mathscr{F}_a[R(a)])'(x) = \frac{\mathrm{d}}{\mathrm{d}\xi} \left(\exp(\mu\pi^2\xi^2) \cdot \mathscr{F}_{\tilde{b}}[g\left(\tilde{b}\right)](\xi) \right)\Big|_{\xi=x}$$

$$= \underbrace{2\mu\pi^2 x \exp\left(\mu\pi^2 x^2\right) \mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](x)}_{\zeta_1(x)} + \underbrace{\exp\left(\mu\pi^2 x^2\right) \left(\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right]\right)'(x)}_{\zeta_2(x)} \tag{37}$$

So,

$$(\psi * \zeta_1)(0) = \int_{-\infty}^{+\infty} \zeta_1(y)\psi(-y)\,\mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} 2\mu\pi^2 y \exp\left(\mu\pi^2 y^2\right) \mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](y) \exp(2\pi\imath by) \exp\left(-\mu\pi^2 y^2\right) \mathrm{d}y$$

$$= 2\mu\pi^2 \int_{-\infty}^{+\infty} \exp(2\pi\imath by) y \mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](y)\,\mathrm{d}y = 2\mu\pi^2 \mathscr{F}^{-1}\left[y\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](y)\right](b) = 2\mu\pi^2 \frac{1}{2\pi\imath} g'(b)$$

and

$$(\psi * \zeta_2)(0) = \int_{-\infty}^{+\infty} \zeta_2(y)\psi(-y)\,\mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} \exp\left(\mu\pi^2 y^2\right) \left(\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right]\right)'(y) \exp(2\pi\imath by) \exp\left(-\mu\pi^2 y^2\right) \mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} \exp(2\pi\imath by) \left(\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right]\right)'(y)\,\mathrm{d}y = \frac{2\pi}{\imath} b g(b)$$

Bringing everything together

$$\alpha(b) = -\frac{1}{g(b) 2\pi\imath} \left((\psi * \zeta_1)(0) + (\psi * \zeta_2)(0)\right)$$

and, finally,

$$\alpha(b) = b + \frac{\mu}{2} \frac{g'(b)}{g(b)}. \qquad \square$$

A.5 Proof of Proposition 5

It follows from the definition of the Fourier transform that for any integrable function f, $\mathscr{F}_x[f(x)](0) = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$. So, as g is a PDF, $\mathscr{F}_x[g(x)](0) = 1$. Moreover, the mean of a random variable $\boldsymbol{x}$ with PDF $p_{\boldsymbol{x}}$ is given by

$$\mathbb{E}(\boldsymbol{x}) = \frac{1}{-2\pi\imath}(\mathscr{F}_x[p_{\boldsymbol{x}}(x)])'(0) \tag{38}$$

and its variance given by

$$\mathrm{Var}(\boldsymbol{x}) = \sigma_x^2 = \left(\frac{1}{-2\pi\imath}\right)^2 (\mathscr{F}_x[p_{\boldsymbol{x}}(x)])''(0) - (\mathbb{E}(\boldsymbol{x}))^2. \tag{39}$$

Point 1:
So, taking the first derivative on both sides of equation (33) at $\xi = 0$ and multiplying by $(-2\pi\imath)^{-1}$ results in

$$\mathbb{E}(\boldsymbol{a}_i) = \mathbb{E}(\boldsymbol{b}) \tag{40}$$

and taking the second derivative on both sides of equation (33) at $\xi = 0$, multiplying by $(-2\pi\imath)^{-2}$ and taking into account that $\mathbb{E}(\boldsymbol{a}_i) = \mathbb{E}(\boldsymbol{b})$ results in

$$\mathrm{Var}(\boldsymbol{a}_i) = -\frac{\mu}{2} + \sigma_b^2. \tag{41}$$

Point 2:
Following the same process for $\tau(\cdot|b)$, yields

$$\mathrm{Var}(\boldsymbol{a}_i|b) = \left(\frac{1}{-2\pi\imath}\right)^2 (\mathscr{F}_a[\tau(a|b)])''(0) - (\mathbb{E}(\boldsymbol{a}_i|b))^2. \tag{42}$$

Now

$$(\mathscr{F}_a[\tau(a|b)])''(x) = \left(\mathscr{F}_a\left[\frac{\varrho(b|a)R(a)}{g(b)}\right]\right)''(x) = \frac{1}{g(b)}(\mathscr{F}_a[\varrho(b|a)R(a)])''(x)$$

and

$$(\mathscr{F}_a[\varrho(b|a)R(a)])''(x) = \left(\mathscr{F}_a[\varrho(b|a)] * (\mathscr{F}_a[R(a)])''\right)(x). \tag{43}$$

Taking (37) and calculating the derivative, one gets

$$(\mathscr{F}_a[R(a)])''(x) = \hat{R}''(x) =$$
$$\underbrace{4\mu^2\pi^4x^2\exp\left(\mu\pi^2x^2\right)\hat{g}(x)}_{\zeta_3(x)} + \underbrace{4\mu\pi^2x\exp\left(\mu\pi^2x^2\right)\hat{g}'(x)}_{\zeta_4(x)}$$

$$+\underbrace{2\mu\pi^2\exp\left(\mu\pi^2x^2\right)\hat{g}(x)}_{\zeta_5(x)}+\underbrace{\exp\left(\mu\pi^2x^2\right)\hat{g}''(x)}_{\zeta_6(x)}.$$

Moreover, from (36)

$$\mathscr{F}_a[\varrho(b|a)](x)=\exp(-2\pi\imath bx)\exp\left(-\mu\pi^2x^2\right)\equiv\psi(x)$$

and

$$\begin{aligned}
(\psi*\zeta_3)(0) &= \int_{-\infty}^{+\infty}\zeta_3(y)\psi(-y)\,\mathrm{d}y=(2\mu\pi^2)^2\mathscr{F}_y^{-1}\left[y^2\hat{g}(y)\right](b)=-\mu^2\pi^2g''(b)\\
(\psi*\zeta_4)(0) &= \int_{-\infty}^{+\infty}\zeta_4(y)\psi(-y)\,\mathrm{d}y=4\mu\pi^2\mathscr{F}_y^{-1}[y\hat{g}'(y)](b)=-4\mu\pi^2\left(g(b)+bg'(b)\right)\\
(\psi*\zeta_5)(0) &= \int_{-\infty}^{+\infty}\zeta_5(y)\psi(-y)\,\mathrm{d}y=2\mu\pi^2g(b)\\
(\psi*\zeta_6)(0) &= \int_{-\infty}^{+\infty}\zeta_6(y)\psi(-y)\,\mathrm{d}y=\mathscr{F}_y^{-1}[\hat{g}''(y)](b)=-4\pi^2b^2g(b)
\end{aligned}$$

Substituting the above together with $\mathbb{E}(\boldsymbol{a}_i|b)=b+\frac{\mu}{2}\frac{g'(b)}{g(b)}$ into (42) yields the result:

$$\mathrm{Var}(\boldsymbol{a}_i|b)=\frac{\mu}{2}+\frac{\mu^2}{4}\frac{\mathrm{d}^2}{\mathrm{d}b^2}\log(g(b))$$

Point 3:

From equation (9), one gets:

$$\alpha'(b)=1+\frac{\mu}{2}\frac{\mathrm{d}^2}{\mathrm{d}b^2}\log(g(b))$$

and, using the result of point 2,

$$\mathrm{Var}(\boldsymbol{a}_i|b)=\frac{\mu}{2}\alpha'(b).$$

Now, since $\tau(\cdot|b)$ is a probability distribution, its conditional variance should be non-negative and, since g is analytic, well-defined (finite). So, since $\mathrm{Var}(\boldsymbol{a}_i|b)\geq 0$, the above equation leads to $\alpha'(b)\geq 0$. As $b'(\theta)>0$ for all θ, $\mathbb{E}(a_i|\theta)$ is an increasing function of θ. □

A.6 Proof of Proposition 6

Start with the following Lemma.

Lemma 1. *Consider a beauty contest with flexible information acquisition. Then all SMFE are essentially symmetric i.e. in equilibrium all players use strategies that are equal to the same strategy $\tilde{r}$ almost everywhere.*

Proof. As there is a continuum of players, any single player i cannot influence the average action taken by the population for any value of θ. This means that all players face the same decision problem. Recall that each player has a unique best reply (up to deviations of measure zero, see Proposition 2) to a smooth, monotone, full-support profile. Thus, in equilibrium, the strategies that the players are using should be equal to the same strategy $\tilde{r}$ almost everywhere. □

"$A \Rightarrow B$"

In light of Lemma 1, since all players have essentially the same best response to the equilibrium profile, the average action of the population conditional on b is given by

$$\alpha(b) = b + \frac{\mu}{2}\frac{g'(b)}{g(b)}.$$

In equilibrium, the best action b should be the one that is generated by aggregating the best responses of the players, i.e.,

$$b = \gamma\alpha(b) + (1-\gamma)\theta(b)$$

and, therefore, in equilibrium

$$\theta(b) = b - \frac{\gamma\mu}{2(1-\gamma)}\frac{g'(b)}{g(b)}.$$

Moreover, $g(\cdot)$ should be the distribution that is generated by $\theta(\cdot)$, i.e., (see eq. (5))

$$g(b) = p(\theta(b))\theta'(b).$$

"$B \Rightarrow A$"

Firstly, if $\theta(\cdot)$ is the inverse of the best action function, then the best action's distribution has the PDF $g(b) = p(\theta(b))\theta'(b)$. Since $g(\cdot)$ has a variance larger than $\mu/2$, the unique best response to it is continuous (see Proposition 2).

The fact that $\theta(\cdot)$ and $g(\cdot)$ satisfy (12) says that the profile where all players best respond to $\theta(\cdot)$ (equivalently, $b(\cdot)$) gives rise to $\theta(\cdot)$ as the inverse of the best action function i.e. that it is an SMFE. $\square$

A.7 Proof of Proposition 7

Point 1:

From the definition of b it is clear that $\mathbb{E}(\bar{\boldsymbol{a}}) = \mathbb{E}(\boldsymbol{b}) \Leftrightarrow \mathbb{E}(\boldsymbol{b}) = \bar{\theta}$. So, all that needs to be shown is that $\mathbb{E}(\bar{\boldsymbol{a}}) = \mathbb{E}(\boldsymbol{b})$.

Let r_i be the best response to $b(\cdot)$ and begin from the left-hand side of the above equation:

$$\mathbb{E}(\bar{\boldsymbol{a}}) = \int_{-\infty}^{+\infty} \bar{a}(\theta)p(\theta)\,\mathrm{d}\theta = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} a_i r_i(a_i|\theta)\,\mathrm{d}a_i p(\theta)\,\mathrm{d}\theta =$$
$$\int_{-\infty}^{+\infty} a_i \int_{-\infty}^{+\infty} r_i(a_i|\theta)p(\theta)\,\mathrm{d}\theta\,\mathrm{d}a_i = \int_{-\infty}^{+\infty} R(a_i)a_i\,\mathrm{d}a_i = \mathbb{E}(\boldsymbol{a}_i)$$

From the proof of Section A.5, since r_i is a best response to $b(\cdot)$ one gets that $\mathbb{E}(\boldsymbol{a}_i) = \mathbb{E}(\boldsymbol{b})$ so $\mathbb{E}(\bar{\boldsymbol{a}}) = \mathbb{E}(\boldsymbol{b})$. $\square$

Points 2 and 3:

It is first shown that $\lim_{\theta\to+\infty} p(\theta)/b'(\theta) = 0$. Begin by integrating condition (14).

$$\int_{-\infty}^{+\infty} b(\theta)p(\theta)\,\mathrm{d}\theta = \int_{-\infty}^{+\infty} \theta p(\theta)\,\mathrm{d}\theta + \frac{\mu\gamma}{2(1-\gamma)}\int_{-\infty}^{+\infty} \frac{1}{b'(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\log\left(\frac{p(\theta)}{b'(\theta)}\right)\right)p(\theta)\,\mathrm{d}\theta$$

The above expression is well-defined in a smooth, monotone, full-support profile. From the proof of point 1, $\int_{-\infty}^{+\infty} b(\theta)p(\theta)\,\mathrm{d}\theta = \int_{-\infty}^{+\infty} \theta p(\theta)\,\mathrm{d}\theta$. So:

$$\int_{-\infty}^{+\infty} \frac{p(\theta)}{b'(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\log\left(\frac{p(\theta)}{b'(\theta)}\right)\right)\mathrm{d}\theta = \int_{-\infty}^{+\infty} \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{p(\theta)}{b'(\theta)}\right)\mathrm{d}\theta = \left[\frac{p(\theta)}{b'(\theta)}\right]_{\theta=-\infty}^{\theta=+\infty} = 0.$$

So, $\lim_{\theta\to+\infty} \frac{p(\theta)}{b'(\theta)} = -\lim_{\theta\to-\infty} \frac{p(\theta)}{b'(\theta)}$ and the two limits exist. As p is a PDF, it has to be that $\lim_{\theta\to+\infty} p(\theta) = \lim_{\theta\to-\infty} p(\theta) = 0$. Now focus on $\lim_{\theta\to+\infty} p(\theta)/b'(\theta)$. There are three possible cases:

(i) $\lim_{\theta\to+\infty} p(\theta)/b'(\theta) = +\infty$.
As $\lim_{\theta\to+\infty} p(\theta) = 0$, it has to be that $\lim_{\theta\to+\infty} b'(\theta) = 0$. But then, there exists a θ' such that $b(\theta) < \theta$ for all $\theta > \theta'$. So, from equation (12), it has to be that

$p(\theta)/b'(\theta)$ is decreasing for all $\theta > \theta'$. This contradicts $\lim_{\theta\to+\infty} p(\theta)/b'(\theta) = +\infty$.

(ii) $\lim_{\theta\to+\infty} p(\theta)/b'(\theta) = l > 0$.
In this case, there exists a θ'' such that $p(\theta)/b'(\theta) \geq l/2$ for all $\theta \geq \theta''$. Since $b' > 0$, b is strictly increasing and thus $\lim_{\theta\to+\infty} b(\theta)$ is well-defined (possibly infinite). So, for $\theta \geq \theta''$ it has to be that $b'(\theta) \leq (2/l)p(\theta)$ and integrating this gives $\int_{\theta''}^{\theta} b'(x)\,\mathrm{d}x \leq (2/l)\int_{\theta''}^{\theta} p(x)\,\mathrm{d}x \leq 2/l$ and so $b(\theta) - b(\theta'') \leq 2/l$. So, $\lim_{\theta\to\infty} b(\theta) \leq 2/l + b(\theta'') < +\infty$. This contradicts that $b(\cdot)$ is bijective.

(iii) $\lim_{\theta\to+\infty} p(\theta)/b'(\theta) = 0$.
Since the other two cases lead to contradictions, it has to be that this is the case.

A similar argument can be made for the case where $\theta \to -\infty$.

So, $\lim_{\theta\to+\infty} p(\theta)/b'(\theta) = \lim_{\theta\to-\infty} p(\theta)/b'(\theta) = 0$.

By solving condition (14) for p one gets

$$p(\theta) = \frac{p(\theta')}{b'(\theta')} b'(\theta) \exp\left(\frac{2(1-\gamma)}{\mu\gamma} \int_{\theta'}^{\theta} b'(t)(b(t)-t)\,\mathrm{d}t \right) \tag{44}$$

for *any* $\theta' \in \mathbb{R}$. And so

$$\frac{p(\theta)}{b'(\theta)} = \frac{p(\theta')}{b'(\theta')} \exp\left(\frac{2(1-\gamma)}{\mu\gamma} \int_{\theta'}^{\theta} b'(t)(b(t)-t)\,\mathrm{d}t \right).$$

Now, taking the limit for $\theta \to +\infty$:

$$\lim_{\theta\to+\infty} \frac{p(\theta)}{b'(\theta)} = \frac{p(\theta')}{b'(\theta')} \exp\left(\frac{2(1-\gamma)}{\mu\gamma} \int_{\theta'}^{+\infty} b'(t)(b(t)-t)\,\mathrm{d}t \right).$$

As $\lim_{\theta\to+\infty} p(\theta)/b'(\theta) = 0$ and $p(\theta')/b'(\theta') > 0$ for any θ', it has to be that

$$\int_{\theta'}^{+\infty} b'(t)(b(t)-t)\,\mathrm{d}t = -\infty$$

for all $\theta' \in \mathbb{R}$. Clearly, as $b'(\theta) > 0$ for all θ this can happen only if $\lim_{\theta\to+\infty} b(\theta) - \theta < 0$. The same arguments for $\bar{a}$ can be given if one takes into account the definition of $b(\theta) = (1-\gamma)\theta + \gamma\bar{a}(\theta)$. A similar argument can be given for $\theta \to -\infty$. □

A.8 Proof of Proposition 8

Point 1:

From player i's point of view, and given that she knows the function $b(\cdot)$, there are two random variables: θ and a_i. One can define more random variables, namely $y = \mathbb{E}(a_i|\theta)$ which is the (equilibrium) average action given θ and $x = (1-\gamma)\theta + \gamma y$, which is the best action given θ. Using the variance decomposition formula for a_i, one obtains

$$\mathrm{Var}(a_i) = \mathbb{E}(\mathrm{Var}(a_i|\theta)) + \mathrm{Var}(\mathbb{E}(a_i|\theta)) = \mathbb{E}(\mathrm{Var}(a_i|\theta)) + \mathrm{Var}(y)$$

Using this and from equation (41) (in the proof of Proposition 2), one gets

$$\mathrm{Var}(x) = \frac{\mu}{2} + \mathrm{Var}(a_i) = \frac{\mu}{2} + \mathbb{E}(\mathrm{Var}(a_i|\theta)) + \mathrm{Var}(y). \tag{45}$$

As $y = x/\gamma + (1-\gamma)\theta/\gamma$,

$$\mathrm{Var}(y) = \left(\frac{1}{\gamma}\right)^2 \mathrm{Var}(x) + \left(\frac{1-\gamma}{\gamma}\right)^2 \mathrm{Var}(\theta) - \frac{2(1-\gamma)}{\gamma^2} \mathrm{Cov}(x,\theta). \tag{46}$$

Substituting (46) into equation (45) and after calculations, one gets

$$\gamma(\mathrm{Var}(\theta) - \mathrm{Var}(x)) = \frac{\mu\gamma^2}{2(1-\gamma)} + \frac{\gamma^2}{1-\gamma}\mathbb{E}(\mathrm{Var}(a_i|\theta)) + \mathrm{Var}(\theta) + \mathrm{Var}(x) - 2\,\mathrm{Cov}(x,\theta)).$$

Now, notice that

$$\mathrm{Var}(\theta) + \mathrm{Var}(x) - 2\,\mathrm{Cov}(x,\theta) = \mathrm{Var}(\theta - x)$$

and thus

$$\sigma^2 - \sigma_b^2 = \frac{\mu\gamma}{2(1-\gamma)} + \frac{\gamma}{1-\gamma}\mathbb{E}(\mathrm{Var}(a_i|\theta)) + \frac{1}{\gamma}\mathrm{Var}(\theta - b) \tag{47}$$

where $\sigma^2 = \mathrm{Var}(\theta)$ and $\sigma_b^2 = \mathrm{Var}(x)$.

Now, from the result of Proposition 5:

$$\mathbb{E}(\mathrm{Var}(a_i|\theta)) = \frac{\mu}{2} + \frac{\mu^2}{4}\int_{-\infty}^{+\infty} \frac{\mathrm{d}^2}{\mathrm{d}b^2}(\log(g(b)))g(b)\,\mathrm{d}b =$$

$$\frac{\mu}{2} + \frac{\mu^2}{4}\left\{\left[\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b)))g(b)\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b)))g'(b)\,\mathrm{d}b\right\} \Rightarrow$$

$$\mathbb{E}(\mathrm{Var}(a_i|\theta)) = \frac{\mu}{2} + \frac{\mu^2}{4}\int_{-\infty}^{+\infty}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b)))g'(b)\,\mathrm{d}b \tag{48}$$

Moreover, from (9) and the fact that $\mathbb{E}(\bar{\boldsymbol{a}}) = \mathbb{E}(\boldsymbol{b})$:

$$\text{Var}(\bar{a} - b) = \frac{\mu^2}{4} \int_{-\infty}^{+\infty} \frac{\mathrm{d}}{\mathrm{d}b} (\log(g(b))) g'(b) \,\mathrm{d}b$$

Using this into (48) and substituting into (47):

$$\sigma^2 - \sigma_b^2 = \frac{\mu\gamma}{2(1-\gamma)} + \frac{\gamma}{1-\gamma} \left(\frac{\mu}{2} + \text{Var}(\bar{a} - b) \right) + \frac{1}{\gamma} \text{Var}(\theta - b)$$

Finally, as $(b - \bar{a}) = \frac{1-\gamma}{\gamma}\theta - b$, one gets that $\text{Var}(b - \bar{a}) = \frac{(1-\gamma)^2}{\gamma^2} \text{Var}(\theta - b)$. So

$$\sigma^2 - \sigma_b^2 = \frac{\mu\gamma}{1-\gamma} + \text{Var}(\theta - b)$$

Point 2:

As in an SMFE r_i is the best response to a smooth, monotone, full-support strategy profile that induces a best action function $b(\cdot)$, it has to be that $\text{Var}(b) > \mu/2$. Using this along with the result of Proposition 8, point 1, one gets that

$$\frac{\mu}{2} < \sigma^2 - \frac{\mu\gamma}{1-\gamma} - \text{Var}(\theta - b) \Rightarrow$$
$$\mu < \frac{2(1-\gamma)}{1+\gamma}\sigma^2 - \frac{2(1-\gamma)}{1+\gamma}\text{Var}(\theta - b) < \frac{2(1-\gamma)}{1+\gamma}\sigma^2 \qquad \square$$

A.9 Proof of Proposition 9

By Bochner's theorem (Bochner 1933; Rudin 1962, p.19), since $R_{\max} \equiv \mathscr{F}_{\xi}^{-1}\left[\exp(\mu_{\max}\pi^2\xi^2)\hat{g}(\xi)\right]$ is the PDF of a probability distribution, $\hat{R}_{\max}$, given by $\hat{R}_{\max}(\xi) = \exp(\mu_{\max}\pi^2\xi^2)\hat{g}(\xi)$, is a positive definite function. Begin by the following observation:

$$\exp(\mu\pi^2\xi^2)\hat{g}(\xi) = \exp(-(\mu_{\max} - \mu)\pi^2\xi^2)\exp(\mu_{\max}\pi^2\xi^2)\hat{g}(\xi) = \exp(-(\mu_{\max} - \mu)\pi^2\xi^2)\hat{R}_{\max}.$$

So, $\exp(\mu\pi^2\xi^2)\hat{g}(\xi)$ is a positive definite function as the product of two positive definite functions (notice that $\exp(-(\mu_{\max} - \mu)\pi^2\xi^2)$ is the Fourier transform of the normal distribution $N(0, (\mu_{max} - \mu)/2)$ and, thus, positive definite).

Therefore, according to Proposition 2, all players' best responses are in continuous strategies. Moreover, the expected action of a player with information cost μ_i conditional on the best action being b is given by:

$$\alpha_i(b, \mu_i) = b + \frac{\mu_i}{2}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b))).$$

So, the average action conditional on b is given by

$$\alpha(b) = \int_{\mu_{\min}}^{\mu_{\max}} \alpha_i(b,\mu_i)\,\mathrm{d}M(\mu_i) = \int_{\mu_{\min}}^{\mu_{\max}} b + \frac{\mu_i}{2}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b)))\,\mathrm{d}M(\mu_i)$$
$$= b + \frac{\bar{\mu}}{2}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b)))$$

The final part follows from an argument identical to the one used in Proposition 6. □

A.10 Proof of Proposition 10

"$A \Rightarrow B$"

In an AAE the best action function is given by $b(\theta) = \kappa\theta + d$. So, $b'(\theta) = \kappa$ and $b''(\theta) = 0$ for all θ. Moreover, an AAE is an SMFE, so $b(\cdot)$ should satisfy (14). From equation (14) one obtains:

$$\kappa\theta + d = \theta + \frac{\mu\gamma}{2(1-\gamma)}\frac{1}{\kappa}\frac{\mathrm{d}}{\mathrm{d}\theta}\log p(\theta).$$

And thus,

$$\log p(\theta) = \int \frac{2(1-\gamma)\kappa}{\mu\gamma}((\kappa-1)\theta + d)\,\mathrm{d}\theta + C$$

where $C \in \mathbb{R}$ is an integrating constant. It will have to be chosen so that the condition $\int_{-\infty}^{+\infty} p(\theta)\,\mathrm{d}\theta = 1$ is satisfied. From the previous equation:

$$\log p(\theta) = \frac{(1-\gamma)\kappa}{\mu\gamma}((\kappa-1)\theta^2 + 2d\theta) + C.$$

Completing the square in the brackets and taking the exponential of both sides one obtains:

$$p(\theta) = \exp(C')\exp\left(\frac{(1-\gamma)\kappa(\kappa-1)}{\mu\gamma}\left(\theta - \frac{d}{1-\kappa}\right)^2\right)$$

for some other constant C'. Now, for $\int_{-\infty}^{+\infty} p(\theta)\,\mathrm{d}\theta = 1$ to be satisfied, it has to be that $\kappa \in (0,1)$, otherwise the resulting p will not be integrable. It is clear that — for an appropriate selection of C' — the previous expression is a normal distribution with a mean $\theta_0 = d/(1-\kappa)$ and variance

$$\sigma^2 = \frac{\mu\gamma}{2(1-\gamma)\kappa(1-\kappa)}.$$

More than that, since in an AAE it has to be that $\sigma_b^2 > \mu/2$, one gets that $\kappa^2\sigma^2 > \mu/2$ i.e. that $\sigma^2 > \mu/2\kappa^2$. Using this together with the above equation, one gets that $\kappa > 1-\gamma$. So, a lower bound for the value of σ^2 is given by the solution to the problem

$$\min_{\kappa\in(0,1)} \frac{\mu\gamma}{2(1-\gamma)\kappa(1-\kappa)}$$
$$\text{s.t. } \kappa \geq 1-\gamma$$

The solution is $\kappa = 1/2$ when $\gamma > 1/2$ and $\kappa = 1-\gamma$ when $\gamma \leq 1/2$ yielding the lower bounds of the variance to be

$$\sigma^2 > \frac{\mu}{2(1-\gamma)^2} \quad \text{when } \gamma \leq \frac{1}{2} \quad \text{and } \sigma^2 > \frac{2\mu\gamma}{1-\gamma} \quad \text{when } \gamma > \frac{1}{2}.$$

"$B \Rightarrow A$"

According to Proposition 6 if $b(\theta) = \kappa\theta + d$ satisfies (14) and condition (6), then $b(\cdot)$ is the best action function of an SMFE; and since it is affine, it is also the best action function of an AAE. All that needs to be shown is that such $\kappa > 0$ and $d \in \mathbb{R}$ exist.

The fundamental is distributed according to

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta-\bar{\theta})^2}{2\sigma^2}\right).$$

So, $\frac{p'(\theta)}{p(\theta)} = -\frac{\theta-\bar{\theta}}{\sigma^2}$. This, along with $b'(\theta) = \kappa$ and $b''(\theta) = 0$ make equation (12), read:

$$\kappa\theta + d = \theta - \frac{\mu\gamma}{2(1-\gamma)\kappa}\frac{\theta-\bar{\theta}}{\sigma^2}.$$

Solving for κ and d, one obtains two solutions:

$$\kappa_+ = \frac{1}{2}\left(1+\sqrt{1-\frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right) \qquad d_+ = \frac{\mu\gamma}{2(1-\gamma)\sigma^2\kappa_+}\bar{\theta}$$

and

$$\kappa_- = \frac{1}{2}\left(1-\sqrt{1-\frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right) \qquad d_- = \frac{\mu\gamma}{2(1-\gamma)\sigma^2\kappa_-}\bar{\theta}.$$

For either of κ_+ or κ_- to be positive reals, it is needed that $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$.

The second requirement for $b(\theta) = \kappa\theta + d$ to qualify for an AAE best action function is that $\mathrm{Var}(b) > \mu/2$ i.e. that $\kappa^2\sigma^2 > \mu/2$. This condition for the solution κ_+, d_+ implies that either $\sigma^2 > \frac{\mu}{1-\gamma}$ or $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$. Together with $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$, the restrictions require that

- either $\gamma \leq 1/2$ and $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$
- or $\gamma > 1/2$ and $\sigma^2 > \frac{2\gamma\mu}{1-\gamma}$.

These are exactly the conditions assumed in statement B. So, the solution with slope κ_+ is always the best action function of an AAE. □

A.11 Proof of proposition 11

Already from the proof of section A.10, it is known that the solution with slope κ_+ is the best action function of an AAE when either $\gamma \leq 1/2$ and $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$; or $\gamma > 1/2$ and $\sigma^2 > \frac{2\gamma\mu}{1-\gamma}$. All that remains to be shown is that the solution with slope κ_- is an AAE iff $\gamma > 1/2$ and $\sigma^2 \in \left(\frac{2\mu\gamma}{1-\gamma}, \frac{\mu}{2(1-\gamma)^2}\right)$. Again, the requirements are that: $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$ and that $\kappa_-^2\sigma^2 > \mu/2$. After substituting κ_-, the resulting system of inequalities is

$$\sigma^2 > \frac{2\mu\gamma}{1-\gamma} \qquad \sigma^2 > \frac{\mu}{1-\gamma} \qquad \sigma^2 < \frac{\mu}{2(1-\gamma)^2}$$

which coincide only for $\gamma > 1/2$ and $\sigma^2 \in \left(\frac{2\mu\gamma}{1-\gamma}, \frac{\mu}{2(1-\gamma)^2}\right)$. □

B AAE stability

In the cases where multiple equilibria exist, it is important to determine which of them are in some sense "stable." The approach taken here is based on recursive best responses and follows from the following observation.

Say the population of players follows a strategy profile under which the best action function is affine and its average is $\bar{\theta}$ (i.e. satisfies point 1 of Proposition 7), which makes it an AAE candidate. Then the best action function has the form

$$b(\theta) = \kappa\theta + (1-\kappa)\bar{\theta}$$

for some $\kappa > 0$. Now assume that all players are best-responding to $b(\cdot)$ and find the best action function of the resulting profile. If $\mathrm{Var}(b) < \mu/2$ — which happens for $\kappa^2 < \mu/2\sigma^2$ — then the best response is to acquire no information and the resulting slope of the best action function is $\kappa = 1-\gamma$. If, on the other hand $\kappa^2 > \mu/2\sigma^2$, then from the proof of Proposition 6 — applied for the case of a normal prior — one can see that the new best action function is given by

$$\tilde{b}(\theta) = \left(1 - \gamma\left(\frac{\mu}{2\kappa\sigma^2} + (1-\kappa)\right)\right)\theta + \left(\gamma\left(\frac{\mu}{2\kappa\sigma^2} + (1-\kappa)\right)\right)\bar{\theta}.$$

Which is also affine with a slope of $\kappa' = 1 - \gamma\left(\frac{\mu}{2\kappa\sigma^2} + (1-\kappa)\right)$ and has an average of $\bar{\theta}$.

So, the best response to an aggregately affine profile with slope κ_n has a best action function whose slope is given by the mapping

$$\kappa_{n+1} = \begin{cases} 1-\gamma & \text{if } \kappa_n^2 \leq \frac{\mu}{2\sigma^2} \\ 1 - \gamma\left(\frac{\mu}{2\kappa_n\sigma^2} + (1-\kappa_n)\right) & \text{otherwise} \end{cases} \tag{49}$$

and the slope of any AAE should be a fixed point of (49).

Thinking about iterative best responses and conducting standard stability analysis, one can see that when multiple AAE exist (see point 2 of Proposition 11) the fixed point at $\kappa_- = \frac{1}{2}\left(1 - \sqrt{1 - \frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right)$ is unstable whereas the one at $\kappa_+ = \frac{1}{2}\left(1 + \sqrt{1 - \frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right)$ is stable. These results are summarized in Figure 4.

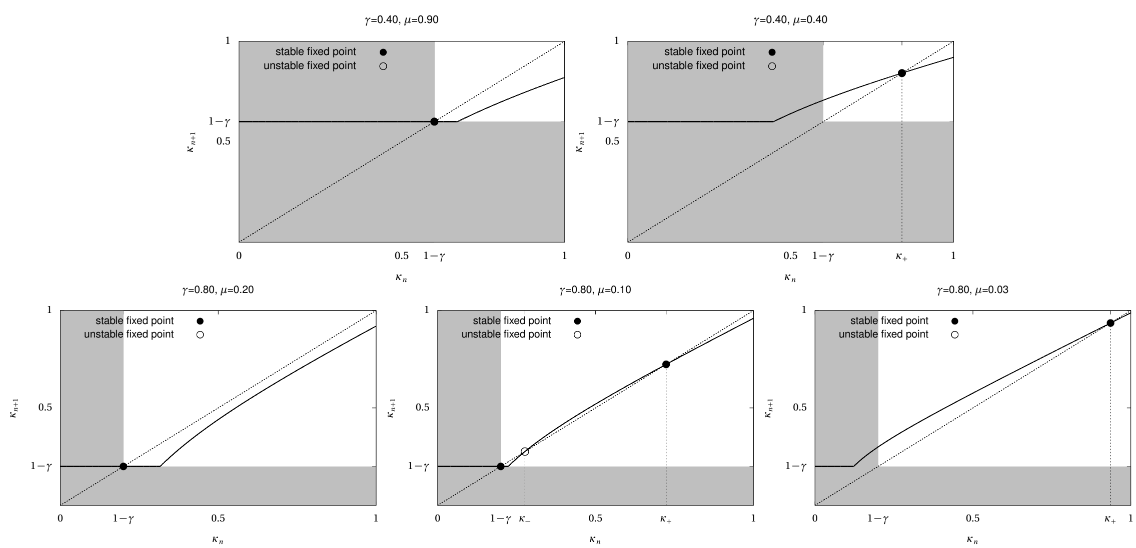


Figure 4: Stability analysis for the slope κ of the equilibrium best action function in an AAE. In the upper row $\gamma < 1/2$ whereas in the lower row $\gamma > 1/2$. Information costs μ are decreasing from left to right. The shaded region depicts areas that imply an average action function with a negative slope.

References

Acemoglu, D. and M. K. Jensen (2010). "Robust comparative statics in large static games". In: *Decision and Control (CDC), 2010 49th IEEE Conference on*. IEEE, pp. 3133–3139.

Allen, F., S. Morris, and H. S. Shin (2006). "Beauty Contests and Iterated Expectations in Asset Markets". *Review of Financial Studies* 19 (3), pp. 719–752.

Angeletos, G.-M. and A. Pavan (2004). "Transparency of Information and Coordination in Economies with Investment Complementarities". *American Economic Review* 94 (2), pp. 91–98.

Angeletos, G.-M. and A. Pavan (2007). "Efficient Use of Information and Social Value of Information". *Econometrica* 75 (4), pp. 1103–1142.

Azzalini, A. (1985). "A Class of Distributions Which Includes the Normal Ones". *Scandinavian Journal of Statistics* 12 (2), pp. 171–178.

Bergemann, D. and S. Morris (2013). "Robust Predictions in Games With Incomplete Information". *Econometrica* 81 (4), pp. 1251–1308.

Bergemann, D. and S. Morris (2016). "Bayes correlated equilibrium and the comparison of information structures in games". *Theoretical Economics* 11 (2), pp. 487–522.

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

Blackwell, D. (1953). "Equivalent Comparisons of Experiments". *The Annals of Mathematical Statistics* 24 (2), pp. 265–272.

Bochner, S. (1933). "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse". *Mathematische Annalen* 108 (1), pp. 378–410.

Carlsson, H. and E. van Damme (1993). "Global games and equilibrium selection". *Econometrica* 61, pp. 989–989.

Cover, T. M. and J. A. Thomas (2006). *Elements of information theory*. 2nd Edition. John Wiley & Sons.

Csiszár, I. (2008). "Axiomatic characterizations of information measures". *Entropy* 10 (3), pp. 261–273.

Denti, T. (2019). "Unrestricted Information Acquisition". *mimeo*.

Denti, T., M. Marinacci, and A. Rustichini (2019). "Experimental Cost of Information". *IGIER Working Paper 657*.

Dewan, T. and D. P. Myatt (2008). "The qualities of leadership: Direction, communication, and obfuscation". *American Political Science Review* 102 (03), pp. 351–368.

Goryunov, M. and A. Rigos (2019). "Discontinuous and Continuous Stochastic Choice and Coordination in the Lab". *mimeo*.

Hébert, B. M. and J. La'O (2020). *Information Acquisition, Efficiency, and Non-Fundamental Volatility*. Working Paper 26771. National Bureau of Economic Research.

Hellwig, C., S. Kohls, and L. Veldkamp (2012). "Information Choice Technologies". *American Economic Review* 102 (3), pp. 35–40.

Hellwig, C. and L. Veldkamp (2009). "Knowing What Others Know: Coordination Motives in Information Acquisition". *The Review of Economic Studies* 76 (1), pp. 223–251.

Jung, J., J. H. J. Kim, F. Matějka, and C. A. Sims (2019). "Discrete Actions in Information-Constrained Decision Problems". *The Review of Economic Studies* 86 (6), pp. 2643–2667.

Kozlovskaya, M. (2018). "Industrial espionage in duopoly games". *Available at SSRN 3190093*.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2018). "Rational Inattention: A Disciplined Behavioral Model". *mimeo*.

Maćkowiak, B. and M. Wiederholt (2009). "Optimal sticky prices under rational inattention". *American Economic Review* 99 (3), pp. 769–803.

Matějka, F. and A. McKay (2015). "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model". *American Economic Review* 105 (1), pp. 272–98.

Matějka, F. and C. A. Sims (2010). "Discrete actions in information-constrained tracking problems". *Princeton University manuscript*.

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

Morris, S. and M. Yang (2016). "Coordination and Continuous Choice". *Available at SSRN 2889861*.

Myatt, D. P. and C. Wallace (2012). “Endogenous Information Acquisition in Coordination Games”. *The Review of Economic Studies* 79 (1), pp. 340–374.

Myatt, D. P. and C. Wallace (2015). “Cournot competition and the social value of information”. *Journal of Economic Theory* 158, Part B. Symposium on Information, Coordination, and Market Frictions, pp. 466–506.

O’Hagan, A. and T. Leonard (1976). “Bayes estimation subject to uncertainty about parameter constraints”. *Biometrika* 63 (1), pp. 201–203.

Rudin, W. (1962). *Fourier analysis on groups*. Vol. 121967. Wiley Online Library.

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

Sims, C. A. (1998). “Stickiness”. In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 49. Elsevier, pp. 317–356.

Sims, C. A. (2003). “Implications of rational inattention”. *Journal of monetary Economics* 50 (3), pp. 665–690.

Woodford, M. (2008). “Inattention as a source of randomized discrete adjustment”. *unpublished, Columbia University*.

Woodford, M. (2009). “Information-constrained state-dependent pricing”. *Journal of Monetary Economics* 56, S100–S124.

Yang, M. (2015). “Coordination with flexible information acquisition”. *Journal of Economic Theory* 158, Part B, pp. 721–738. Symposium on Information, Coordination, and Market Frictions.

Yang, M. (2019). “Optimality of Debt under Flexible Information Acquisition”. *The Review of Economic Studies* 87 (1), pp. 487–536.