Flexible Information Acquisition in Large Coordination Games

Alexandros Rigos*

July 2019

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 determine the distribution of the signal that they receive and arbitrarily correlate it with the fundamental, paying costs linear in Shannon mutual information. Without assuming a normal prior for the fundamental, the class of equilibria in continuous strategies is characterized. Populations with heterogeneous costs exhibit the same aggregate behavior as homogeneous populations with the same average cost. Equilibria where the population-wide average action is an affine function of the fundamental exist only when the fundamental is normally distributed. Finally, a novel method allows to study non-normal priors, leading to new insights. For example, the distribution of the equilibrium action exhibits an amplified skewness compared to the distribution of the fundamental.

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@nek.lu.se. 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. Thanks go also to Jörgen Weibull, Heinrich Nax, Erik Mohlin, Ming Yang, Christoph Wolf, Mogens Fosgerau, Petra Thiemann, Jens Gudmundsson and André Stenzel 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 Pavvan 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 2015a). As one would expect, solving for equilibrium in a complex environment like that can be challenging without imposing simplifying assumptions. Despite this difficulty, the present study characterizes equilibria in (absolutely) continuous strategies by providing an easy-to-verify condition without imposing a normal prior (as is often done in existing studies). Furthermore, it shows that equilibria in which the population-wide average action follows the fundamental linearly exist only if the fundamental is normally distributed. It also demonstrates how strong coordination motives can lead to multiple equilibria. Finally, using numerical methods, it shows that departing from the normal distribution can lead to new insights into how markets aggregate information.

As in Morris and Shin (2002), these interactions are modelled through a beauty contest game. The economic fundamental that is relevant for players’ payoffs (henceforth the fundamental) is modelled as a random variable following 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. This modelling choice allows one to directly study the effects of coordination motives on equilibrium outcome(s). In this setting, information about the fundamental can be very valuable to the players as it can help them make accurate guesses and also serve as a coordination device.

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 is clearly 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.

Despite providing a more realistic approach, 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 acquisition technology is — in this sense — rigid, and may not reflect real-world conditions. On the one hand, real-world agents 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. It is therefore natural to assume that agents are fully flexible to determine not only the amount of information they want to receive or which information sources to “listen” to, but also on which specific events they want to focus their attention. On the other hand, processing the information that these sources convey is costly as it requires their attention — a scarce resource.

This situation is exactly captured by the notion of rational inattention, pioneered by Sims (1998, 2003).¹ Players are not passive receivers of information, nor are they

¹ Sims (1998, 2003) introduced rational inattention as a possible explanation of sluggish macroe-
bounded to receive information in some particular form (e.g. normally distributed). Instead, each player can endogenously acquire information about the fundamental by designing her own information channel: an experiment that narrows down her belief about the fundamental. The more informative the 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 the posterior beliefs. 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 2015a,b).

Keeping the tractable combination of linear-quadratic losses and Shannon entropy costs, the present paper departs from the normal prior assumption. It argues that equilibria are in continuous strategies when information costs are low (cf. Jung et al. 2016) and proceeds to characterize them. Moreover, it shows that the aggregate behavior of a population with heterogeneous (but low) information costs is the same as that of a heterogeneous population in which all players have cost equal to the original population’s mean information cost.

Along with Myatt and Wallace (2012), who study economic beauty contests with endogenous information acquisition, and Yang (2015a), who introduced flexible information acquisition technology, Denti (2018) is closely related to this study. Denti (2018) studies a more general game-theoretic 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 economic 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 (2015a) focuses on investors’ binary decisions to invest or not. Maćkowiak, Matějka, and Wiederholt (2018) provide a recent survey of the rational inattention literature.

---

2 Csiszár (2008) provides an axiomatic extension 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. (2016, section III).

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.
sense, his agents use an unrestricted information acquisition technology. The information structure used by Denti (2018) is, therefore, richer than the one used herein, and leads to Bayes correlated equilibria (Bergemann and Morris 2016).

In strategic settings which include the beauty contest — but restricted to the Gaussian case — Bergemann and Morris (2013) aim to put restrictions on the moments of equilibrium distributions of actions and the fundamental. Results in this direction are reported in Sections 3 and 4. A more detailed discussion of how results in the existing literature 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. It shows that equilibria of this form exist if and only if the fundamental is normally distributed and has large enough variance. Section 6 introduces a method to study distributions other than the normal and provides an application. Section 7 discusses the results of the paper and compares them to the ones in the existing literature.

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 $\theta$. The parameter $\gamma \in [0, 1)$ determines how strong the coordination motive is. Each of them obtains utility given by

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

(1)

where $a_i \in A_i = \mathbb{R}$ is player $i$’s action and $\bar{a} = \int_0^1 a_i \, di$ 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.\(^4\)

\(^4\)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
The value $\theta$ is unknown to the players. In particular, it is the realization of a random variable $\theta : \Omega \to \Theta$ — the *fundamental* — where $(\Omega, \Sigma, P)$ 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. The probability density function (PDF) of $P_\theta$ is denoted by $p(\cdot)$ and is analytic. Moreover, $\theta$ is assumed to have a well-defined mean $\bar{\theta}$ and variance $\sigma^2$.

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 $s_i : \Omega \to S_i$, where $S_i$ is some rich enough message space. The signal is the only channel through which the player will receive information.

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 $s_i$ and $s_j$ are independent for any $i \neq j$, conditional on the realization $\theta$. 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 $n \to \infty$. 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.

Throughout the paper, $\Delta(X)$ denotes the space of probability measures over space $X$ and $P_x$ denotes the probability distribution of random variable $x$.

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.

Note that $p \in L^2(\mathbb{R})$ (where $\mathbb{R}$ is endowed with the Lebesgue measure) is sufficient for $\bar{\theta}$ and $\sigma^2$ to be well-defined.

The message space $S_i$ is endowed with a $\sigma$-algebra such that $s_i$ is measurable and the pushforward measure.

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 $\theta$) and not on one another’s messages. See also Section 7 for further discussion.
costs are assumed to be linear in the chosen channel’s capacity as measured by Shannon’s mutual information. 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 $\theta$ and the signal $s_i$ is given by

$$ I(\theta, s_i) = \int_{\Theta, S_i} \log \frac{dp_{\theta,s_i}}{dp_{\theta} \times ds_i} dP_\theta s_i $$

if $P_{\theta,s_i} \ll P_\theta \times s_i$ and $I(\theta, s_i) = +\infty$ otherwise.  

In the former case, $P_{\theta,s_i}$ 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 $s_i$ is given by

$$ C(s_i) = \mu \cdot I(\theta, s_i) = \mu \left( \int_{\Theta} \int_{S_i} p(\theta) q_i(s_i|\theta) \log \frac{q_i(s_i|\theta)}{Q_i(s_i)} ds_i d\theta \right) $$

where $Q_i(\cdot) = \int_\Theta q(\cdot|\theta)p(\theta) d\theta$ is the (marginal) PDF of $s_i$ and $\mu \geq 0$ is the cost per unit of information.  

The more informative signal $s_i$ is, the higher the cost of information. Say, for example, that $S_i = \mathbb{R}$ and that $q(\cdot|\theta)$ is very concentrated and changes rapidly with $\theta$, then the information structure is very informative and will come at a high cost. If, in contrast, $q(\cdot|\theta)$ does not change with $\theta$, 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 $s_i \in S_i$.

The timing is as follows:

---

10 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 $\theta$ and $s_i$. For a standard textbook treatment of the topic see Cover and Thomas (2006).

11 Throughout the paper, log denotes the natural logarithm and so the unit of measurement of information 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 $\mu$ is given in utils per nat of information.

12 Formally, player $i$’s action strategy is a random variable $a_i : \Omega \rightarrow A_i$ which, conditionally on $s_i$, is independent from $\theta$, $s_j$, and $a_j$ ($j \neq i$).
1. Each player \(i\) — independently from and simultaneously with all other players — designs her signal \(s_i\) and pays the associated cost \(C(s_i)\).

2. The value of \(\theta\) is realized.

3. Each player receives a message according to her chosen signal and the realization of \(\theta\). 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 \(s_i\) and the action part \(a_i\). Let player \(i\)'s strategy be denoted by \(m_i = (s_i, a_i)\). 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 \(\theta\) is given by the function \(\bar{a}: \mathbb{R} \rightarrow \mathbb{R}\) defined through

\[
\bar{a}(\theta) \equiv \int_{0}^{1} \left( \int_{S_j} \left( \int_{A_j} a_j dP_{a_j|s_j}(a_j) dP_{s_j|\theta}(s_j) \right) dA_j \right) dj. \tag{3}
\]

As mentioned before, it is required that \(\bar{a}(\cdot)\) is well-defined for (almost) all \(\theta\). A sufficient condition for this is that all \(a_i|\theta\) have a variance for (almost) all \(\theta\), and that \(\{\text{Var}(a_i|\theta)\}_{i \in [0,1]}\) 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 2015a, 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 \(s_i(a_i)\). 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

\[\text{A detailed proof is provided in the Online Appendix for completeness.}\]
the realization $\theta$—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:

$$\max_{r_i} U(r_i, r_{-i}) - \mu I(\theta, a_i)$$

where

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

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

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

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 $\bar{a}(\cdot)$.

**Proposition 1.** Let $(p, \gamma, \mu)$ 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

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

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

**Proof.** See Appendix B.1.

The function $b : \mathbb{R} \to \mathbb{R}$ defined through Equation (4) will be referred 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 its best action function, defined by (4), is:

1. strictly increasing,
2. bijective, and
3. analytic in its argument.

The requirement that $b(\cdot)$ is bijective — which, along with monotonicity, implies that $\lim_{\theta \to +\infty} b(\theta) = +\infty$ and $\lim_{\theta \to -\infty} b(\theta) = -\infty$ — is not too restrictive since it is satisfied for any weakly increasing and any decreasing $\bar{a}(\cdot)$, as long as $\bar{a}'(\theta) > -(1-\gamma)/\gamma$. It 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.\(^{14}\)

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 given by

$$g(\cdot) = p(\theta(\cdot)) \theta'(\cdot)$$

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

**Continuous strategies**

The analysis presented here and in Section 4 focuses on equilibria whereby players use strategies that have densities.

**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 focus on continuous strategies may seem quite restrictive at first. Jung et al. (2016) find that even in continuous environments optimal strategies might be “discrete,” i.e. admit no density. So, one could think that players in the present setting

---

\(^{14}\)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.
could find such strategies optimal. Below it is argued that in the present setting discontinuous strategies are unlikely to appear in equilibrium when information costs are low.

Begin with the following observation: In discrete optimal strategies à la Jung et al. (2016) the conditional probability $\pi(a|\theta)$ received by an action $a$ in the (countable) consideration set is a continuous function of $\theta$. In fact, it is analytic since it is driven by the analytic utility function (eq. (1)) and is modulated through the also analytic $p$. 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 $\theta$. Now, as long as the information cost is low enough compared to the best action’s distribution (condition (6) below), 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 costs are relatively low, it should be expected that continuous equilibria are more likely.

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 $(p, \gamma, \mu)$ 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

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

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 B.2.

Throughout the paper, $F_x$ and $F_x^{-1}$ denote the Fourier and inverse Fourier transforms defined as

$$F_x[f(x)](\xi) = \int_{-\infty}^{+\infty} f(x) \exp(-2\pi i x \xi) \, dx$$

10
\[ F^{-1}(F(\xi))(x) = \int_{-\infty}^{+\infty} F(\xi) \exp(2\pi i x \xi) \, d\xi \] (8)

respectively \((i\) is the imaginary unit). Moreover, the shorthand notation with the hat operator \(\hat{f}(\xi) := F_x[f(x)](\xi)\) is used sometimes.

In the remainder of the paper, the terms “best response to \(r_{-i}\),” “best response to \(\tilde{a}(\cdot)\),” and “best response to \(b(\cdot)\)” will be used interchangeably as all smooth, monotone, full-support profiles \(r_{-i}\) that yield the same \(\tilde{a}(\cdot)\) will lead to the same best response from player \(i\) and for a given \(\gamma\) there is a one-to-one relation between \(\tilde{a}(\cdot)\) 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 \(b\), viewed as a random variable, should be able to be written as the sum of two other independently distributed random variables — \(a_i\) and \(\epsilon_i\) — 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.\(^{15}\)

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 \((p, \gamma, \mu)\) 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

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

**Proof.** See Appendix B.3.

Observe that, 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

\(^{15}\)A necessary condition is that \(R_i\) as calculated from (6) has a positive variance, i.e. \(\sigma_b^2 > \mu/2\). Another, stronger, necessary condition is that Var\((a_i|b) > 0\) for all \(b\) (see Proposition 5).
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 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 \( \sigma^2 \), 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.

In light of Proposition 3, the effect of an increasing cost \( \mu \) to player \( i \)'s strategy becomes clear. A higher \( \mu \) 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, \( \int_{-\infty}^{+\infty} r(a_i|\theta)da_i = 1 \) must hold for all \( \theta \). As higher \( \mu \) 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 \( \mu \) 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

\[
\alpha(b) := \mathbb{E}(a_i|b) = b + \frac{\mu}{2} \frac{d}{db} \left( \log(g(b)) \right).
\]

**Proof.** See Appendix B.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 \( \mu \). 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).

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

\[
\text{Var}(a_i) = \text{Var}(b) - \mu/2
\]

(10)

2. and its variance conditional on the best action being \( b \) given by

\[
\text{Var}(a_i|b) = \frac{\mu}{2} + \frac{\mu^2}{4} \frac{d^2}{db^2} (\log(g(b))).
\]

(11)

3. Moreover, her expected action conditional on \( b \) is increasing, i.e., \( \alpha'(b) \geq 0 \) and \( \bar{a}'(\theta) \geq 0 \).

Proof. See Appendix B.5.

Points 1 and 2 provide necessary conditions that the distribution of the best action has to satisfy in an SMFE. 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 fundamental moves.

Comment: Discontinuous Strategies

If condition (6) does not hold (i.e., if \( \mu \) 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. 2016) 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,
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 \((p, \gamma, \mu)\) 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 strictly increasing, \( \mathcal{F}_\xi^{-1}[\exp(\mu \pi^2 \xi^2) \hat{g}(\xi)] \) is a probability distribution,

\[
\theta(b) = b - \frac{\mu\gamma}{2(1-\gamma)} \int \frac{d}{db} \log(g(b))) \mathrm{d}b, \quad \text{and} \quad (12) \\
g(b) = p(\theta(b))\theta'(b). \quad (13)
\]
Proof. See Appendix B.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 fundamental’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.

Examining equation (12) it is seen that the equilibrium best action function $b(\cdot)$ deviates from $\theta$ to the extent that $\mu \gamma / 2(1 - \gamma)$ is positive. This has two implications. Firstly, when $\gamma = 0$ (i.e. when there is no coordination motive) the best action function coincides with $\theta$. This should be expected as in the absence of strategic motives, the best a player can do is to guess the true value of the fundamental. Secondly, as the information cost $\mu$ decreases, the deviations of the best action function from $\theta$ should be becoming smaller. In particular, when $\mu = 0$ equation (12) gives $b(\theta) = \theta$. 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 $b(\theta) = \theta$, which is accurately described by equation (12) (setting $\mu = 0$).

4.2 Equilibrium properties

Equations (12) and (13) together imply

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

which is a second-order nonlinear differential equation 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 $b(\cdot)$ close to $\pm \infty$ as well as its ex-ante expected value is examined. Secondly, a relation between the variance of the best action and the variance of the fundamental $\sigma^2$ 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. \[ \int_{-\infty}^{+\infty} b(\theta) p(\theta) d\theta = \int_{-\infty}^{+\infty} \bar{a}(\theta) p(\theta) d\theta = \int_{-\infty}^{+\infty} \theta p(\theta) d\theta = \bar{\theta} \]
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 B.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 \( \theta \) 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 \((p, \gamma, \mu)\) be a beauty contest with flexible information acquisition that admits an SMFE and \( \sigma_{b}^2 \) be the variance of the best action \( b \) in that SMFE. Then

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

**Proof.** See Appendix B.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 \( \mu \). This upper bound is not tight: what one can say for sure is that if \( \mu \) exceeds this value, then \((p, \gamma, \mu)\) 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 $\mu_i \in [\mu_{\text{min}}, \mu_{\text{max}}]$ ($0 \leq \mu_{\text{min}} < \mu_{\text{max}} < \infty$). Let also the distribution of agents’ costs be $M \in \Delta([\mu_{\text{min}}, \mu_{\text{max}}])$ with expected value $\mathbb{E}(\mu) = \int_{\mu_{\text{min}}}^{\mu_{\text{max}}} \mu \, dM(\mu) = \bar{\mu}$.

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

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

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

**Proof.** See Appendix B.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 $M$.

5 Aggregately Affine Equilibria

As mentioned earlier, it is not possible to solve (14) in the general case. Instead, this section examines smooth, monotone, full-support equilibria of a specific form. In particular, equilibria whereby the average action function is affine in $\theta$ 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 \((p, \gamma, \mu)\) with \(\mu > 0\) and \(\gamma > 0\) be a beauty contest with flexible information acquisition. The following statements are equivalent:

(A) \((p, \gamma, \mu)\) 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 B.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 \((p, \gamma, \mu)\) 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 \((p, \gamma, \mu)\) admits exactly one AAE.

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 \((p, \gamma, \mu)\) admits exactly two AAE.

**Proof.** See Appendix B.11.

Interestingly, under the conditions where two AAE exist, there also exists an equilibrium where players do not obtain any information (which happens when \(\sigma_b^2 < \mu/2\)).
Figure 1: Aggregately affine equilibria and equilibria without information acquisition in different regions of $\mu$ and $\gamma$ when $\theta$ is normally distributed with variance $\sigma^2$. Red area: one AAE; Orange area: two AAE and no-info equilibrium; Blue area: no-info equilibrium.

In this sense, three equilibria of the classes considered exist under these parameter configurations. Moreover, if neither the conditions of point 1 nor point 2 hold, there exists an equilibrium without information acquisition. Notice that in this equilibrium $b'(\theta) = 1 - \gamma$ for all values of $\theta$. So, the first derivative of $b(\cdot)$ is constant even though 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 $\mu$ and start increasing $\gamma$. When $\gamma$ is small ($\gamma < 1/2$), players are able to coordinate on a unique AAE where they acquire information. As $\gamma$ increases, less information is acquired — as the importance of getting close to the realized value of $\theta$ is decreasing and the motive to coordinate increases. When $\gamma$ 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 \( \hat{\theta} \) — 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 coexistence 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 (\( \kappa \) is smaller). This second AAE is created as an in-between case of the two equilibria described above. More than that, it is unstable in the sense that iterative best responses lead away from this equilibrium. This argument is explained in more detail in Appendix A where stability analysis is conducted.

### 6 Beyond the Normal Distribution

Despite it being challenging to identify equilibria in the usual way (i.e. to find SMFE for a particular parameter combination \( (p, \gamma, \mu) \)), 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 is the difficulty of confirming whether condition (6) holds. In order to overcome this problem, this section makes use of priors \( g \) which are conjugate for the normal distribution.\(^{16}\) Therefore, \( g \) and \( R \) belong to the same distribution family. In this way, confirming that (6) gives, indeed, the PDF of a probability distribution boils down to making sure that the parameters calculated for \( R \) fall within the allowed ranges for the particular distribution family. The method is demonstrated for the skew normal distribution and shows that moving away from

\(^{16}\)Recall that from the result of Section 3, \( b \) is the sum of two variables, \( a_i \sim R \) and \( \epsilon_i \sim N(0, \mu/2) \).
the Gaussian prior leads to new insights about how equilibrium actions and economic fundamentals are related.

**Application: Skew Normal** \( g(\cdot) \)

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

\[
\phi(b; b_0, \sigma_b, \lambda) = \frac{2}{\sigma_b} \phi\left( \frac{b - b_0}{\sigma_b} \right) \Phi\left( \lambda \frac{b - b_0}{\sigma_b} \right)
\]

where \( \phi(\cdot) \) and \( \Phi(\cdot) \) are the PDF and cumulative distribution function (CDF) of the standard normal distribution \( N(0, 1) \). The important variable here is \( \lambda \) 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 \( b \sim SN(b_0, \sigma_b, \lambda) \), then \( a_i \sim SN\left(b_0, \sqrt{\sigma_b^2 - \mu/2}, \frac{\lambda \sigma_b (\sigma_b^2 - (1 + \lambda^2)/2)^{-1/2}}{2} \right) \) (see Azzalini 1985). So, as long as information costs are low \( (\mu < 2\sigma_b^2/(1 + \lambda^2)) \), 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 \( g(\cdot) \). If \( \mu \) is larger than the above threshold, then the best response of the players will be in discontinuous strategies.

Trying to connect the theoretical model with potential “observable” variables, it is interesting to see how good the population is as a whole (as summarized by their average action \( \bar{a} \)) at being close to the fundamental \( \theta \). To this end, Figure 2 presents the prior along with equilibrium objects for an SMFE in which the best action follows a skew normal distribution.

The bottom panel of Figure 2 presents the equilibrium densities of \( \theta, \bar{a}, \) 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 \( p \). Moreover, although the prior distribution \( p \) is only slightly (barely perceivably) skewed, individual actions \( a_i \) are clearly skewed (right panel), along with the equilibrium average action function \( \bar{a}(\cdot) \). On the top panel, it is seen that the population’s average action follows the fundamental more closely for \( \theta \) to the right of \( \theta^* : \bar{a}(\theta^*) = \theta^* \) than for \( \theta \) to the left of \( \theta^* \). This is explained by the fact
Figure 2: Prior and equilibrium objects for an SMFE with skew normal $g(\cdot)$. The position parameter $b_0$ is such that $\mathbb{E}(b) = 0$. 
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.

The analysis presented shows that the distribution of observable behavior (i.e. $a_i$ and $\bar{a}$) can have significantly different features than the fundamental distribution when individuals are rationally inattentive. In particular, attempting to learn about the (un-observable) 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 and/or the coordination motive increase.

7 Discussion

This paper studied large coordination games played by rationally inattentive agents in the presence of fundamental motives. The set of equilibria in continuous strategies was characterized (Section 4). The analysis conducted in Section 5 found that aggregately affine equilibria exist only when the prior on the fundamental is normal, and that multiple equilibria can occur when coordination motives are strong. Moreover, Section 6 introduced a method that helps investigate and visualize the relationship between economic fundamentals and the behavior of market participants. What follows discusses and relates these results to the ones found in other studies.

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 full-information one. Yang (2015a) uses such a technology to study a two-player coordination game. The full-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 $\mu$ (bounded away from zero for any fixed $\gamma > 1/2$). On the one hand, this differs from the result of Yang (2015a) 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 (2015a), when information costs are small, the equilibrium structure of the full-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. 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. Moreover, reported results confirm that with entropy information costs equilibrium multiplicity can arise, even when attention is restricted to a very particular class of tractable equilibria.\footnote{Note that Denti (2018) finds a unique equilibrium of this class but this is because he restricts himself to “cheap enough” information ($\mu < \frac{1-\gamma}{2\gamma} \sigma^2$ in terms of this paper).}

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 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 \( p \) as an interim belief shared by all players after observing a public signal and before observing their private one. 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). In this sense, Myatt and Wallace (2012) find conditions under which two linear equilibria may arise under a mutual information-based cost specification: one with and one without information acquisition. The analysis conducted here shows that there is one more aggregately affine equilibrium with a lower degree of information acquisition which is albeit unstable.

**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 (2018) 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 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 \( \text{Var}(b) < \mu/2 \) — 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 \( \kappa_n \) has a best action function whose slope is given by the mapping

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

and the slope of any AAE should be a fixed point of (16).

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 3.
Figure 3: Stability analysis for the slope $\kappa$ 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 $\mu$ are decreasing from left to right. The shaded region depicts areas that imply an average action function with a negative slope.

B Omitted Proofs

B.1 Proof of Proposition 1

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

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

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 $\theta$, 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 $\theta$). That is, her best response is to use $r_i$ given by $r_i(a_i|\theta) = \delta(a_i - b(\theta))$ with $\delta$ being Dirac’s delta function (almost all $\theta$).
B.2 Proof of Proposition 2

Consider variations of the strategy of player $i$. These variations will be of the type $\tilde{r} = r + \epsilon \eta$ for some $\epsilon > 0$. These variations should still be feasible. That is, for all $\theta$, it is required that $r(\cdot|\theta) + \epsilon \eta(\cdot|\theta)$ is a probability distribution over $A_i$. It is required, thus, that for all $\theta$, $\int_{A_i} r(a_i|\theta) + \epsilon \eta(a_i|\theta) da_i = 1$ which leads to the condition that for all $\theta$, $\int_{A_i} \eta(a_i|\theta) da_i = 0$. It also has to be that $r(a_i|\theta) + \epsilon \eta(a_i|\theta) \geq 0$ and so $\eta(a_i|\theta) \geq -r(a_i|\theta)/\epsilon$ for all $a_i$ and $\theta$. From the above equations, the following is calculated:\textsuperscript{18}

$$U(r_i + \epsilon \eta, r_{-i}) = -(1-\gamma) \int_{\Theta} \int_{A_i} (a_i - \theta)^2 (r_i(a_i|\theta) + \epsilon \eta(a_i|\theta)) p(\theta) da_i d\theta -$$

$$- \gamma \int_{\Theta} \int_{A_i} (a_i - \bar{a}(\theta))^2 (r_i(a_i|\theta) + \epsilon \eta(a_i|\theta)) p(\theta) da_i d\theta.$$ \hspace{1cm} (17)

And the derivatives:

$$\frac{dU(r + \epsilon \eta, r_{-i})}{d\epsilon} \bigg|_{\epsilon=0} = -(1-\gamma) \int_{\Theta} \int_{A_i} (a_i - \theta)^2 \eta(a_i|\theta) p(\theta) da_i d\theta -$$

$$- \gamma \int_{\Theta} \int_{A_i} (a_i - \bar{a}(\theta))^2 \eta(a_i|\theta) p(\theta) da_i d\theta.$$ \hspace{1cm} (18)

$$\frac{dI(r + \epsilon \eta)}{d\epsilon} \bigg|_{\epsilon=0} = \int_{\Theta} \int_{A_i} \log(r(a_i|\theta)) \eta(a_i|\theta) p(\theta) da_i d\theta -$$

$$- \int_{A_i} \log(R_i(a_i)) H(a_i) da_i.$$ \hspace{1cm} (19)

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

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

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

$$\text{s.t. } \int_{A_i} r_i(a_i|\theta) da_i = 1 \quad \text{for all } \theta \in \Theta.$$  

\textsuperscript{18}The effect of the other players' strategies is incorporated in $\bar{a}(\theta)$.
So, the Lagrangian for player \( i \)'s decision problem will be

\[
\mathcal{L}(r_i, k(\theta)) = U(r_i, r_{-i}) - \mu I(\theta, a_i) - \int_{\Theta} k(\theta) \left( \int_{A_i} r(a_i|\theta) \, da_i - 1 \right) p(\theta) \, d\theta
\]

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

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

\[
\left. \frac{d\mathcal{L}(r_i + \epsilon \eta, k(\theta))}{d\epsilon} \right|_{\epsilon=0} = 0 \Rightarrow
\]

\[
\int_{\Theta} \int_{A_i} \left[ u_i(a_i, \theta) - \mu \log \left( \frac{r(a_i|\theta)}{R_i(a_i)} \right) - k(\theta) \right] \eta(a_i|\theta) p(\theta) \, da_i \, d\theta = 0 \quad (20)
\]

and

\[
\int_{A_i} r_i(a_i|\theta) \, da_i = 1 \quad \text{for all } \theta \in \Theta. \quad (21)
\]

Where

\[
u_i(a_i, \theta) = -(1 - \gamma)(a_i - \theta)^2 - \gamma(a_i - \bar{a}(\theta))^2. \quad (22)\]

Since condition (20) has to be satisfied for all \( \eta \), it has to be the case that

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

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

\[
r(a_i|\theta) = R_i(a_i) \exp \left( -\frac{k(\theta)}{\mu} \right) \exp \left( \frac{u_i(a_i, \theta)}{\mu} \right) \quad (23)
\]

and (23) can be rewritten as

\[
r(a_i|\theta) = R_i(a_i) K(\theta) \exp \left( \frac{u_i(a_i, \theta)}{\mu} \right). \quad (24)
\]

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 \( R_i(a_i) \):

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

After substituting from (24) and (22), simple calculations give

\[
\int_{-\infty}^{+\infty} K(\theta) \exp \left( -\frac{(a_i - b(\theta))^2}{\mu} \right) \exp \left( -\frac{\gamma(1 - \gamma)}{\mu} (\theta - \bar{a}(\theta))^2 \right) p(\theta) \, d\theta = 1. \quad (25)
\]
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 assumptions 1 and 2 of Definition 1, $\lim_{b \to \infty} b^{-1}(b) = \infty$ and $\lim_{b \to -\infty} b^{-1}(b) = -\infty$. So, with a change of the variable of integration from $\theta$ to $b = b(\theta)$, and by defining $G(\cdot)$ as

$$G(b) = \frac{K(b^{-1}(b)) \exp\left(-\frac{\gamma(1-\gamma)}{\mu}(b^{-1}(b) - \bar{a}(b^{-1}(b)))^2\right) p(b^{-1}(b))}{(1-\gamma) + \gamma \bar{a}'(b^{-1}(b))} \tag{26}$$

condition (25) can be rewritten as

$$\int_{-\infty}^{+\infty} G(b) \exp\left(-\frac{1}{\mu}(a_i - b)^2\right) \, db = 1. \tag{27}$$

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 (27) 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:

$$\mathcal{F}_{a_i}[(G * f)(a_i)](\xi) = \mathcal{F}_{a_i}[G(a_i)](\xi) \cdot \mathcal{F}_{a_i}[f(a_i)](\xi) = \delta(\xi)$$

$$\Rightarrow \mathcal{F}_{a_i}[G(a_i)](\xi) \cdot \frac{1}{\sqrt{\pi\mu}} \exp(\mu \pi^2 \xi^2) \delta(\xi)$$

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

$$G(b) = \mathcal{F}_{\xi}^{-1}\left[\frac{1}{\sqrt{\pi\mu}} \exp(\mu \pi^2 \xi^2) \delta(\xi)\right](b)$$

$$= \frac{1}{\sqrt{\pi\mu}} \int_{-\infty}^{+\infty} \exp(2\pi i \xi x) \exp(\mu \pi^2 \xi^2) \delta(\xi) \, d\xi = \frac{1}{\sqrt{\pi\mu}}.$$

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

$$K(\theta) = \frac{1 + \gamma \bar{a}'(\theta) - 1}{p(\theta) \sqrt{\pi\mu}} \exp\left(\frac{\gamma(1-\gamma)}{\mu}(\theta - \bar{a}(\theta))^2\right) \tag{28}$$

Using (28) in (24) yields

$$r(a_i|\theta) = R_i(a_i) \frac{1 + \gamma \bar{a}'(\theta) - 1}{p(\theta) \sqrt{\pi\mu}} \exp\left(-\frac{(a_i - b(\theta))^2}{\mu}\right). \tag{29}$$
The solution has to also satisfy \( \int_{-\infty}^{+\infty} r(a_i|\theta) \, da_i = 1 \) for all \( \theta \). Again, changing the variable from \( \theta \) to \( b = b(\theta) \), this condition yields

\[
\int_{-\infty}^{+\infty} R_i(a_i) \exp \left( -\frac{(b - a_i)^2}{\mu} \right) \, da_i = \sqrt{\pi \mu} p(b^{-1}(b)) (b^{-1})'(b). \tag{30}
\]

Notice that the left-hand side of equation (30) is the convolution of \( R_i \) and \( f \). Now, take the Fourier transform on both sides and use the convolution theorem

\[
\mathcal{F}_{a_i}[R_i(a_i)](\xi) \cdot \mathcal{F}_{b}[f(b)](\xi) = \sqrt{\pi \mu} \cdot \mathcal{F}_{b}[p(b^{-1}(b)) (b^{-1})'(b)](\xi) \Rightarrow \mathcal{F}_{a_i}[R_i(a_i)](\xi) = \mathcal{F}_{\xi}^{-1}[\exp(\mu \pi^2 \xi^2) \cdot \mathcal{F}_{b}[p(b^{-1}(b)) (b^{-1})'(b)](\xi)](a_i) \tag{31}
\]

If the expression above is the PDF of a probability distribution, then the solution is calculated by equation (24) which — after noticing that \( g(b) = p(b^{-1}(b)) (b^{-1})'(b) \) is the PDF of the best action \( b = b(\theta) \) — becomes

So, the expression for \( r \) given by equation (24) is well-defined. Thus, the best reply is given by the following formula:

\[
r_i(a_i|\theta) = R_i(a_i) \frac{b'(\theta)}{p(\theta) \sqrt{\pi \mu}} \exp \left( -\frac{(a_i - b(\theta))^2}{\mu} \right) \tag{33}
\]

with

\[
R_i(a_i) = \mathcal{F}_{\xi}^{-1}[\exp(\mu \pi^2 \xi^2) \cdot \mathcal{F}_{b}[p(b^{-1}(b)) (b^{-1})'(b)](\xi)](a_i). \tag{34}
\]

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 (34) 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 (34). \( \Box \)

**B.3 Proof of Proposition 3**

From Bayes’s rule, one gets:

\[
\varrho_i(b|a_i) = \frac{\tau_i(a_i|b) g(b)}{R(a_i)} \tag{35}
\]
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 \( \tau_i \) from the result of Proposition 2 with a change of variable:

\[
\tau_i(a_i|b) = R_i(a_i) \frac{1}{g(b)} \frac{1}{\sqrt{\pi \mu}} \exp \left( -\frac{(a_i - b)^2}{\mu} \right)
\]

and comparing with (35), one obtains

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

\[\Box\]

### B.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 (39)), the expected action of player \( i \), conditional on \( b \), is

\[
\alpha(b) = -\frac{1}{2\pi i} (\mathcal{F}_a[t(a|b)])'(0) = -\frac{1}{g(b)2\pi i} (\mathcal{F}_a[\varrho(b|a)R(a)])'(0)
\]

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

\[
\alpha(b) = -\frac{1}{g(b)2\pi i} (\mathcal{F}_a[\varrho(b|a)] * (\mathcal{F}_a[R(a)])')'(0).
\]

Now

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

\[
= \frac{1}{\sqrt{\pi \mu}} \exp(-2\pi ibx) \mathcal{F}_a \left[ \exp \left( -\frac{a^2}{\mu} \right) \right](x)
\]

\[
= \exp(-2\pi ibx) \exp(-\mu \pi^2 x^2) \equiv \psi(x)
\]

(37)
And

\[ (\mathcal{F}_b[R(a)])' (x) = \frac{d}{d\xi} \left( \exp(\mu \pi^2 \xi^2) \cdot \mathcal{F}_b[g \left( \frac{\xi}{b} \right)](\xi) \right)_{\xi = x} \]

\[ = 2\mu \pi^2 x \exp(\mu \pi^2 x^2) \mathcal{F}_b \left[ g \left( \frac{\xi}{b} \right) \right] (x) + \exp(\mu \pi^2 x^2) (\mathcal{F}_b \left[ g \left( \frac{\xi}{b} \right) \right])' (x) \]  

(38)

So,

\[ (\psi * \zeta_1)(0) = \int_{-\infty}^{+\infty} \zeta_1(y) \psi(-y) \, dy \]

\[ = \int_{-\infty}^{+\infty} 2\mu \pi^2 y \exp(\mu \pi^2 y^2) \mathcal{F}_b \left[ g \left( \frac{\xi}{b} \right) \right] (y) \exp(2\pi ib y) \exp(-\mu \pi^2 y^2) \, dy \]

\[ = 2\mu \pi^2 \int_{-\infty}^{+\infty} \exp(2\pi ib y) \mathcal{F}_b \left[ g \left( \frac{\xi}{b} \right) \right] (y) \, dy = 2\mu \pi^2 \mathcal{F}^{-1} \left[ y \mathcal{F}_b \left[ g \left( \frac{\xi}{b} \right) \right] (y) \right] (b) = \frac{1}{2\pi i} g'(b) \]

and

\[ (\psi * \zeta_2)(0) = \int_{-\infty}^{+\infty} \zeta_2(y) \psi(-y) \, dy \]

\[ = \int_{-\infty}^{+\infty} \exp(\mu \pi^2 y^2) (\mathcal{F}_b \left[ g \left( \frac{\xi}{b} \right) \right])' (y) \exp(2\pi ib y) \exp(-\mu \pi^2 y^2) \, dy \]

\[ = \int_{-\infty}^{+\infty} \exp(2\pi ib y) (\mathcal{F}_b \left[ g \left( \frac{\xi}{b} \right) \right])' (y) \, dy = \frac{2\pi}{i} b g(b) \]

Bringing everything together

\[ \alpha(b) = -\frac{1}{g(b)2\pi i} ((\psi * \zeta_1)(0) + (\psi * \zeta_2)(0)) \]

and, finally,

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

\[ \square \]

B.5 Proof of Proposition 5

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

\[ \mathbb{E}(x) = \frac{1}{-2\pi i} (\mathcal{F}_x[p_x(x)])'(0) \]  

(39)
and its variance given by

\[ \text{Var}(x) = \sigma_x^2 = \left( \frac{1}{-2\pi i} \right)^2 (\mathcal{F}_x[p_x(x)])''(0) - (\mathbb{E}(x))^2. \] (40)

**Point 1:**
So, taking the first derivative on both sides of equation (34) at \( \xi = 0 \) and multiplying by \((-2\pi i)^{-1}\) results in

\[ \mathbb{E}(a_i) = \mathbb{E}(b) \] (41)

and taking the second derivative on both sides of equation (34) at \( \xi = 0 \), multiplying by \((-2\pi i)^{-2}\) and taking into account that \( \mathbb{E}(a_i) = \mathbb{E}(b) \) results in

\[ \text{Var}(a_i) = -\frac{\mu}{2} + \sigma_b^2. \] (42)

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

\[ \text{Var}(a_i|b) = \left( \frac{1}{-2\pi i} \right)^2 (\mathcal{F}_a[\tau(a|b)])''(0) - (\mathbb{E}(a_i|b))^2. \] (43)

Now

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

and

\[ (\mathcal{F}_a[g(b|R(a))]')''(x) = (\mathcal{F}_a[g(b|R)])' * (\mathcal{F}_a[R(a)])''(x). \] (44)

Taking (38) and calculating the derivative, one gets

\[ (\mathcal{F}_a[R(a)])''(x) = \hat{R}''(x) = \frac{4\mu^2 \pi^4 x^2 \exp(\mu \pi^2 x^2) g(x) + 4\mu \pi^2 x \exp(\mu \pi^2 x^2) \hat{g}'(x)}{\zeta_5(x)} + \frac{2\mu \pi^2 \exp(\mu \pi^2 x^2) \hat{g}(x) + \exp(\mu \pi^2 x^2) \hat{g}''(x)}{\zeta_6(x)}. \]

Moreover, from (37)

\[ \mathcal{F}_a[g(b|R)](x) = \exp(-2\pi i bx) \exp(-\mu \pi^2 x^2) \equiv \psi(x) \]
and

\[(\psi \ast \zeta_3)(0) = \int_{-\infty}^{+\infty} \zeta_3(y)\psi(-y)\,dy = (2\mu \pi^2)^2 \mathcal{F}^{-1}[y^2 \hat{g}(y)](b) = -\mu^2 \pi^2 g''(b)\]

\[(\psi \ast \zeta_4)(0) = \int_{-\infty}^{+\infty} \zeta_4(y)\psi(-y)\,dy = 4\mu \pi^2 \mathcal{F}^{-1}[y \hat{g}'(y)](b) = -4\mu \pi^2 (g(b) + b g'(b))\]

\[(\psi \ast \zeta_5)(0) = \int_{-\infty}^{+\infty} \zeta_5(y)\psi(-y)\,dy = 2\mu \pi^2 g(b)\]

\[(\psi \ast \zeta_6)(0) = \int_{-\infty}^{+\infty} \zeta_6(y)\psi(-y)\,dy = \mathcal{F}^{-1}[\hat{g}''(y)](b) = -4\pi^2 b^2 g(b)\]

Substituting the above together with \(\mathbb{E}(a_i|b) = b + \frac{\mu g'(b)}{2g(b)}\) into (43) yields the result:

\[
\text{Var}(a_i|b) = \frac{\mu}{2} + \frac{\mu^2}{4} \frac{d^2}{db^2} \log(g(b))
\]

**Point 3:**

From equation (9), one gets:

\[
\alpha'(b) = 1 + \frac{\mu}{2} \frac{d^2}{db^2} \log(g(b))
\]

and, using the result of point 2,

\[
\text{Var}(a_i|b) = \frac{\mu}{2} \alpha'(b).
\]

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 \(\text{Var}(a_i|b) \geq 0\), the above equation leads to \(\alpha'(b) \geq 0\). \(\Box\)

**B.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 \(\theta\). 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

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

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.

**B.7 Proof of Proposition 7**

**Point 1:**

From the definition of \( b \) it is clear that \( \mathbb{E}(\bar{a}) = \mathbb{E}(b) \Leftrightarrow \mathbb{E}(b) = \bar{\theta} \). So, all that needs to be shown is that \( \mathbb{E}(\bar{a}) = \mathbb{E}(b) \).
Let $r_i$ be the best response to $b(\cdot)$ and begin from the left-hand side of the above equation:

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

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

Points 2 and 3:

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

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

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

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

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 $\theta'$ 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 $\theta''$ 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 $\int_{\theta''}^{\theta} b'(x)\,dx \leq (2/l)\int_{\theta''}^{\theta} p(x)\,dx \leq 2/l$ and so $b(\theta) - b(\theta'') \leq 2/l$. So, $\lim_{\theta \to +\infty} b(\theta) \leq 2/l + b(\theta'') < +\infty$. This contradicts condition 2 of definition 1.
\[
\lim_{\theta \to +\infty} p(\theta)/b'(\theta) = 0.
\]
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
\[
p(\theta) = \frac{p(\theta')}{b'(\theta')} b'(\theta) \exp \left( \frac{2(1-\gamma)}{\mu \gamma} \int_{\theta'}^{\theta} b'(t)(b(t)-t) \, dt \right)
\]
for any \( \theta' \in \mathbb{R} \). And so
\[
\frac{p(\theta)}{b'(\theta)} = \frac{p(\theta')}{b'(\theta')} \exp \left( \frac{2(1-\gamma)}{\mu \gamma} \int_{\theta'}^{\theta} b'(t)(b(t)-t) \, dt \right).
\]
Now, taking the limit for \( \theta \to +\infty \):
\[
\lim_{\theta \to +\infty} \frac{p(\theta)}{b'(\theta)} = \frac{p(\theta')}{b'(\theta')} \exp \left( \frac{2(1-\gamma)}{\mu \gamma} \int_{\theta'}^{+\infty} b'(t)(b(t)-t) \, dt \right).
\]
As \( \lim_{\theta \to +\infty} p(\theta)/b'(\theta) = 0 \) and \( p(\theta')/b'(\theta') > 0 \) for any \( \theta' \), it has to be that
\[
\int_{\theta'}^{+\infty} b'(t)(b(t)-t) \, dt = -\infty
\]
for all \( \theta' \in \mathbb{R} \). Clearly, as \( b'(\theta) > 0 \) for all \( \theta \) 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 \( \theta \to -\infty \). \( \blacksquare \)

### B.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: \( \theta \) and \( a_i \). One can define more random variables, namely \( y = \mathbb{E}(a_i|\theta) \) which is the (equilibrium) average action given \( \theta \) and \( x = (1-\gamma)\theta + \gamma y \), which is the best action given \( \theta \). Using the variance decomposition formula for \( a_i \), one obtains
\[
\text{Var}(a_i) = \mathbb{E}(\text{Var}(a_i|\theta)) + \text{Var}(\mathbb{E}(a_i|\theta)) = \mathbb{E}(\text{Var}(a_i|\theta)) + \text{Var}(y)
\]

Using this and from equation (42) (in the proof of Proposition 2), one gets

$$\text{Var}(x) = \frac{\mu}{2} + \text{Var}(a_i) = \frac{\mu}{2} + \mathbb{E}(\text{Var}(a_i|\theta)) + \text{Var}(y).$$ \hspace{1cm} (46)

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

$$\text{Var}(y) = \left(\frac{1}{\gamma}\right)^2 \text{Var}(x) + \left(\frac{1 - \gamma}{\gamma}\right)^2 \text{Var}(\theta) - \frac{2(1 - \gamma)}{\gamma^2} \text{Cov}(x, \theta).$$ \hspace{1cm} (47)

Substituting (47) into equation (46) and after calculations, one gets

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

Now, notice that

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

and thus

$$\sigma^2 - \sigma_b^2 = \frac{\mu\gamma}{2(1 - \gamma)} + \frac{\gamma}{1 - \gamma} \mathbb{E}(\text{Var}(a_i|\theta)) + \frac{1}{\gamma} \text{Var}(\theta - b)$$ \hspace{1cm} (48)

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

Now, from the result of Proposition 5:

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

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

$$\mathbb{E}(\text{Var}(a_i|\theta)) = \frac{\mu}{2} + \frac{\mu^2}{4} \int_{-\infty}^{+\infty} \frac{d}{db} (\log(g(b)))g'(b) \, db$$ \hspace{1cm} (49)

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

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

Using this into (49) and substituting into (48):

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

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

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

B.9 Proof of Proposition 9

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

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

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 $\exp(-(\mu_{\text{max}} - \mu) \pi^2 \xi^2)$ is the Fourier transform of the normal distribution $N(0, (\mu_{\text{max}} - \mu)/2)$ 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 $\mu_i$ conditional on the best action being $b$ is given by:

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

So, the average action conditional on $b$ is given by

$$\alpha(b) = \int_{\mu_{\text{min}}}^{\mu_{\text{max}}} \alpha_i(b, \mu_i) dM(\mu_i) = \int_{\mu_{\text{min}}}^{\mu_{\text{max}}} b + \frac{\mu_i}{2} \frac{d}{db} (\log(g(b))) dM(\mu_i)$$

$$= b + \hat{\mu} \frac{d}{db} (\log(g(b)))$$

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

B.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 \( \theta \). Moreover, an AAE is an SMFE, so \( b(\cdot) \) should satisfy (14). From equation (14) one obtains:

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

And thus,

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

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

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

for some other constant \( C' \). Now, for \( \int_{-\infty}^{+\infty} p(\theta) d\theta = 1 \) 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 \( \sigma_0^2 > \mu/2 \), 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 \( \sigma^2 \) is given by the solution to the problem

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

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

\[
\sigma^2 > \frac{\mu}{2(1 - \gamma)^2} \quad \text{when} \quad \gamma \leq \frac{1}{2} \quad \text{and} \quad \sigma^2 > \frac{2\mu \gamma}{1 - \gamma} \quad \text{when} \quad \gamma > \frac{1}{2}.
\]
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 \( \kappa \) and \( d \), one obtains two solutions:
\[
\kappa_+ = \frac{1}{2} \left( 1 + \sqrt{1 - \frac{2\mu \gamma}{(1-\gamma)\sigma^2}} \right) \quad d_+ = \frac{\mu \gamma}{2(1-\gamma)\sigma^2 \kappa_+} \frac{\theta - \bar{\theta}}{\sigma^2}
\]
and
\[
\kappa_- = \frac{1}{2} \left( 1 - \sqrt{1 - \frac{2\mu \gamma}{(1-\gamma)\sigma^2}} \right) \quad d_- = \frac{\mu \gamma}{2(1-\gamma)\sigma^2 \kappa_-} \frac{\theta - \bar{\theta}}{\sigma^2}.
\]

For either of \( \kappa_+ \) or \( \kappa_- \) 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 \( \text{Var}(b) = \mu/2 \) i.e. that \( \kappa^2 \sigma^2 > \mu/2 \). This condition for the solution \( \kappa_+, d_+ \) implies that either \( \sigma^2 > \frac{\mu}{1-\gamma} \) or \( \sigma^2 > \frac{2\mu \gamma}{(1-\gamma)^2} \). 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 \( \sigma^2 > \frac{2\mu \gamma}{1-\gamma} \).

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

\[\square\]

**B.11 Proof of proposition 11**

Already from the proof of section B.10, it is known that the solution with slope \( \kappa_+ \) 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 $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$. All that remains to be shown is that the solution with slope $\kappa_-$ 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 $\kappa_-$, the resulting system of inequalities is

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

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

\[\Box\]

**References**


Online Appendix: Determining the support of the signal and the message-to-action distribution

In order to determine the optimal signal support that a player $i$ will use, let $m_{-i}$ be any strategy profile that player $i$’s opponents are using. Let $s_i$ be the message that player $i$ obtained after having chosen channel $s_i$. Given $s_i$, player $i$ forms a posterior belief about $\theta$ and from this belief and $m_{-i}$ (by pushing forward), a posterior belief about the value of $\bar{a}$.\footnote{Formally it should be $\bar{a}_{-i}$ (i.e. the average action of all players excluding player $i$) but as the contribution of a single player to the average action of a continuum of players is zero, $\bar{a}_{-i} = \bar{a}$.} From these beliefs, player $i$ forms expectations $\theta^i(s_i; s_i)$ and $\bar{a}^i(s_i; s_i, m_{-i})$ on the respective variables.

Lemma 2. In a best response of player $i$, almost all messages $s_i \in S_i$ have the following property: there exists a unique action $a^i \in A_i$ such that $\Pr(a^i|s_i) = 1$.

Proof. For simplicity of exposition it is assumed that $P_{s_i|\theta}$ is described by a PDF $q_i(\cdot|\theta)$ and $P_{a_i|s_i}$ is described by a PDF $\lambda_i(\cdot|s_i)$. The proof is essentially the same in the more general case.

Let $p^i(\cdot|s_i; s_i)$ denote the PDF of the posterior belief that player $i$ has about the fundamental (conditional on $i$ receiving message $s_i$ while using channel $s_i$). It is calculated by Bayes’s rule:

$$p^i(\theta|s_i; s_i) = \frac{q_i(s_i|\theta)p(\theta)}{\int_\Theta q_i(s_i|\theta)p(\theta)d\theta}.$$  

Given player $j$’s strategy and upon receiving message $s_i$, player $i$ forms a posterior belief about $j$’s message applying Bayes’s rule once more. This is given by:

$$q^i_j(s_j|s_i; s_i, s_j) = \int_\Theta q_j(s_j|\theta)p^i(\theta|s_i; s_i)d\theta.$$  

So, player $i$’s posterior belief about player $j$’s action is

$$\lambda^i_j(a_j|s_i; s_i, m_j) = \int_{S_j} \lambda_j(a_j|s_j) q^i_j(s_j|s_i; s_i, m_j) ds_j$$  

and player $i$’s expectation of player $j$’s action is

$$a^i_j(s_i; s_i, m_j) = \int_{A_j} a_j \lambda^i_j(a_j|s_i; s_i, m_j) da_j.$$
Therefore, player $i$’s expectation of the average action of her opponents is

$$\bar{a}_i(s_i; s_{-i}) = \bar{a}(s_i; s_{-i}) = \int_0^1 \int_{A_{-i}} a_j \lambda_j(a_j; s_i, m_j) da_j dj$$

Notice that this expectation is equal to $i$’s expectation over the population-wide average action $\bar{a}$ as player $i$’s action cannot affect the mean action in a continuum population. Finally, player $i$’s expectation of the value of the fundamental is

$$\bar{\theta}(s_i; s_i) = \int_{\Theta} \theta p_i(\theta|s_i, s_i) d\theta.$$ 

Any costs player $i$ has spent on acquiring information are sunk at the time she observes $s_i$. So, her expected utility at that point is calculated by

$$\mathbb{E}_i(u_i|s_i; s_i, m_{-i}) = -(1-\gamma) \int_{\Theta} (a_i - \theta)^2 p_i(\theta|s_i, m_{-i}) d\theta - \int_{\Theta} \gamma (a_i - \bar{a}(\theta))^2 p_i(\theta|s_i, m_{-i}) d\theta.$$ 

Given that player $i$ maximizes expected utility, any “best” action $b_i$ that receives positive density in a best response of player $i$ has to satisfy the following first order condition.

$$b_i(s_i; s_i, m_{-i}) = (1-\gamma) \int_{\Theta} \theta p_i(\theta|s_i, s_i) d\theta + \gamma \int_{\Theta} \bar{a}(\theta) p_i(\theta|s_i, s_i) d\theta.$$ 

As long as the integrals appearing in the right-hand side of the above equation are well-defined, there is a unique value of $b_i$ that satisfies the above condition. Therefore, a best response should put all probability to that action, i.e., $\Pr(a^i|s_i) = 1$ with $a^i$ given by the above equation. \hfill \qed

Moreover, there should be a unique message that maps to each action.

**Lemma 3.** In a best response of player $i$, almost all actions $a_i \in A_i$ have the following property: there exists a unique message $s^{a_i} \in S_i$ such that $\Pr(a^i|s^{a_i}) = 1$.

**Proof.** Consider two strategies: $m_i = (s_i, a_i)$ under which each action has a unique message that maps to it (through $a_i$) and $m'_i = (s'_i, a'_i)$ under which a set of actions of positive measure (under the measure induced on $A_i$ by $m'_i$) have multiple messages that
map to them. For each action $a_i \in A_i$, denote by $S'(a_i)$ the set of messages that map to $a_i$ under $a'_i$ i.e. $S'(a_i) = \{s'_i \in S'_i : \Pr(a_i = a_i | s'_i) = 1\}$ and by $s^{a_i}$ the (unique) message that maps to $a_i$ under $a_i$. Note that $S'(a_i)$ should be nonempty for almost all $a_i$ as a result of Lemma 2. Let also $q(s^{a_i} | \theta) = \sum_{s' \in S'(a_i)} q(s' | \theta)$. It is clear that the expected value of $-(1 - \gamma)(a_i - \theta)^2 - \gamma(a_i - \bar{a})^2$ from the two strategies will be the same as they induce the same probability distribution on $A_i$ for the same values of $\theta$. It is also true that $I(\theta, a'_i) > I(\theta, a_i)$ due to the convexity of mutual information in $q$ (see Fozunbal, McLaughlin, and Schafer 2005). Therefore, $m'_i$ is more expensive to player $i$ than $m_i$ and thus not an optimal choice.

From Lemmas 2 and 3, the action part of the strategy $a_i$ can be summarized by a bijection from $S_i$ to $A_i$ such that $a_i$ gives probability one to a unique action $a_i$ for each message $s_i \in S_i$, and for each action $a_i \in \mathbb{R}$ there exists a unique message for which $\Pr(a_i | s_i) = 1$. Thus, the message space should have the same cardinality as the action space. This should happen even if some of these messages are never used. Of course, if any of the messages is not to be used, this would immediately mean that the corresponding action would never be used by player $i$. So, player $i$’s message space can be reduced to be a space equinumerous with $A_i = \mathbb{R}$. Since signals are important only as far as they prescribe probabilities over actions, the exact choice of the message space will not change players’ actions as long as it has the same cardinality as $\mathbb{R}$, and $a_i$ can be described by a bijection, as explained above.