No.43

NOVIEMBRE DE 2019

Documentos CEDE

ISSN 1657-7191 Edición electrónica.

The Less I Know The Better? A Model of
Rational Attention and Experimentation

Juan-Camilo Chaves

CEDE
CENTRO DE ESTUDIOS SOBRE DESARROLLO ECONÓMICO

**Universidad de
los Andes**
Facultad de Economía

Serie Documentos Cede, 2019-43
ISSN 1657-7191 Edición electrónica.
Noviembre de 2019

© 2019, Universidad de los Andes, Facultad de Economía,
CEDE. Calle 19A No. 1 – 37 Este, Bloque W.
Bogotá, D. C., Colombia Teléfonos: 3394949- 3394999,
extensiones 2400, 2049, 3233
infocede@uniandes.edu.co
<http://economia.uniandes.edu.co>

Impreso en Colombia – Printed in Colombia

La serie de Documentos de Trabajo CEDE se circula con propósitos de discusión y divulgación. Los artículos no han sido evaluados por pares ni sujetos a ningún tipo de evaluación formal por parte del equipo de trabajo del CEDE.

El contenido de la presente publicación se encuentra protegido por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por tanto su utilización, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso, digital o en cualquier formato conocido o por conocer, se encuentran prohibidos, y sólo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito del autor o titular. Las limitaciones y excepciones al Derecho de Autor, sólo serán aplicables en la medida en que se den dentro de los denominados Usos Honrados (Fair use), estén previa y expresamente establecidas, no causen un grave e injustificado perjuicio a los intereses legítimos del autor o titular, y no atenten contra la normal explotación de la obra.

Universidad de los Andes | Vigilada Mineducación
Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

The Less I Know The Better? A Model of Rational Attention and Experimentation^{*}

Juan-Camilo Chaves[†]

Abstract

This paper explores the binary choice of an agent who must allocate her attention to noisy information obtained from costly experimentation to improve her knowledge in a context of uncertainty. We model a two period-lived agent who must decide whether to invest or not in each period. Initial experimentation (investing) enables access to noisy information about the state of nature, which can be refined by paying attention. We find that there is no case for rational inattention. The agent internalizes the fact that information is costly and thus cannot be wasted. Hence, when information is available, attention is a rational action. Moreover, initial experimentation is driven by an endogenous cutoff strategy: the prior beliefs and the cost of attention determine the extent to which information is valuable. Finally, an important finding is that limited attention has no effect on the initial experimentation amount because it never worsens the quality of information. Experimentation is understood as paying for information and therefore the agent pays the lowest price regardless of the quality of information (perfectly informative or noisy).

Keywords: Rational inattention, experimentation, learning, uncertainty.

JEL Codes: C72, C73, D83, D91

^{*}I am grateful to my advisor Paula Jaramillo for her patience and support in this process. I also thank Lina Marcela Ramírez, Álvaro José Pinzón, and Édgar Sánchez-Cuevas for their invaluable contributions to my research and for their delightful company. I am also thankful to Tomás Rodríguez for his guidance on theoretical issues. Finally, thanks to my family for encouraging me to always do my best.

[†]Department of Economics, Universidad de los Andes. Email: jc.chaves348@uniandes.edu.co.

¿Entre Menos Sé Mejor? Un Modelo de Atención Racional y Experimentación^{*}

Juan-Camilo Chaves^{**}

Resumen

Este trabajo explora la decisión binaria de un agente que debe asignar su atención a información ruidosa que se obtiene de una experimentación costosa en un contexto de incertidumbre. Nosotros modelamos un agente en dos periodos que debe decidir si invertir o no en cada periodo. La experimentación (invertir) inicial permite el acceso a información ruidosa sobre el estado de la naturaleza, la cual puede ser refinada prestando atención. Encontramos que no es el caso de inatención racional. El agente internaliza el hecho de que la información es costosa y, por ende, no puede ser desperdiciada. Además, la experimentación inicial está guiada por un umbral determinado endógenamente: las creencias iniciales y el costo de la atención determinan qué tan valiosa es la información. Finalmente, un resultado importante es que la atención limitada no tiene efecto en la cantidad de experimentación inicial porque la calidad de la información siempre es la misma. La experimentación es entendida como pagar por información y, como consecuencia, el agente paga el precio más bajo por esta sin importar la calidad de la información (perfectamente informativa o ruidosa).

Palabras Clave: Inatención racional, experimentación, aprendizaje, incertidumbre.

Códigos JEL: C72, C73, D83, D91

^{*}Estoy agradecido con mi asesora Paula Jaramillo por su paciencia y su apoyo en este proceso. También le agradezco a Lina Ramírez, Álvaro José Pinzón y a Édgar Sánchez-Cuevas por sus valiosas contribuciones a mi investigación y por su agradable . Le doy las gracias a Tomás Rodríguez por su guía en cuestiones teóricas. Finalmente, le doy gracias a mi familia por animarme a siempre dar lo mejor de mí.

^{**}Departamento de Economía, Universidad de los Andes. Email: jc.chaves348@uniandes.edu.co.

I Introduction

Understanding the cognitive limits of agents has been central in economic analysis for its implications on decision making. Classic economic models assume rational agents which are capable of solving decision problems with several variables. However, psychology studies of bounded rationality have shown that people have computational constraints ([Miller, 1956](#)). In particular, people face limited attention and cannot take into account all the dimensions that they desire ([Kahneman, 1973](#)). Instead, they face a trade-off over which variables to focus their attention on. One of these variables is information.

The key role of information in economics is that it is useful inasmuch as it leads to optimal decision making. In contexts of uncertainty or information asymmetries, acquiring information often entails better decisions.¹ However, information is not always available and even if it is, the benefits are not completely internalized. People pay for available information, yet they cannot process it as a whole. Nowadays Internet is at our disposal almost for free and we choose to consider only a part of it in our decisions ([De Los Santos et al., 2012](#)). When buying a good in a market we would desire to know and understand all the properties to choose, for example, the one with the highest quality among a set of alternatives. However, time constraints and other barriers force us to pay attention only to a subset of these properties ([Goeree, 2008](#); [Wilson, 2008](#)). In other words, before making a decision, people may have the opportunity to study some information, though in some cases it is too costly to inquire to the point where decision making is free from uncertainty.

Thus attention, like any other scarce resource, must be allocated efficiently. People must confront the decision of which pieces of information to avoid and which to attend to. Moreover, people must choose how much attention to pay to each piece. We can have a preference relation over the properties of the good and allocate our attention according to this ordering ([Manzini and Mariotti, 2014](#)).

In this paper, we study the binary choice of an agent who must allocate her attention to

¹In special cases of uncertainty people occasionally demand non-instrumental information ([Eliaz and Schotter, 2010](#)).

the information obtained from experimentation in a context of uncertainty.² We model an extensive-form game of one player in which an agent acquires information in two ways.³ First, experimentation enables access to noisy information. Second, the precision of information depends on the attention paid. This setting allows us to explore some features in experimentation decision.

To fix ideas we can think about a two period-lived farmer that decides whether to invest in a new crop in the context of unknown climate conditions. In particular, the agent maximizes its utility function by choosing the level of investment that deviates the least from the state of the world θ , which is understood as the investment target under specific climate conditions. Thus, she must choose between a safe alternative, whose payoff is known ex-ante, and a risky alternative, whose payoff depends on the state of nature. We set the safe alternative (the no experimentation action) as a default action by centering the distribution of the state of the world in zero. If she experiments (i.e. she chooses the risky choice), then she receives a noisy signal related to the true state. However, the agent can choose the level of noise by paying attention. The more attention she pays, the more precise the signal is. However, attention is costly. Costs are modeled using Shannon Entropy ([Sims, 2003](#)). This cost is represented as the reduction in uncertainty due to the knowledge of new information, which is measured as an information cost. For a typical farmer, this attention cost stands for the cost of monitoring the rainfall, the cost of keeping a detailed log book, and the effort cost of being alert for some eventuality. We model attention as a decreasing function in the variance of the noise following [Wiederholt et al. \(2010\)](#). For instance, the higher the attention, the lower the noise. Moreover, this setting prevents full attention, which is consistent with psychology studies ([Kahneman, 1973](#); [Styles, 2006](#)). This modeling allows us to narrow the action set of attention, and thus have a more intuitive notion of what attention represents. For this reason, attention lies in the interval $[0, 1]$, whose limits imply zero and full attention, respectively. Afterwards, the agent faces again the decision of whether to invest or not.

We show that when information is available, there is no case for rational inattention. This

²In fact, our model considers a continuous decision set. However, rationally inattentive agents can narrow their decision problem to a discrete choice set ([Matějka, 2015](#)).

³The word *acquisition* may be ambiguous. Actually, experimentation is associated with acquisition and attention to absorption. However, in the literature, both account for acquisition costs.

scenario occurs when the agent acquires information and wastes it (i.e. does not pay attention to it). One explanation is that as experimentation is needed to access to an exogenously determined signal (the source of information), attending to some degree is an optimal strategy since the agent internalizes that she cannot waste information that she has already paid for. For instance, initial investment entails risk-prone behavior as the no attention decision is equivalent to the default action (a safe alternative).

Moreover, we find an endogenously determined threshold in which uncertainty is the main driver for initial experimentation. The agent only considers experimenting if the variance of the state of the world is sufficiently high. The reason is that experimentation implies learning. The higher the uncertainty, the more she can learn.⁴ Hence, experimentation is valuable to the extent that it allows learning. This result is consistent with the essential feature of rational inattention literature that choice depends on prior beliefs (Caplin et al., 2013).

Finally, noisy information has no effect on initial experimentation. In other words, we find that conditional on experimenting in a scenario with limited attention (i.e. noisy information), and with no limited attention (i.e. perfectly informative information), the agent initially experiments the same amount. Furthermore, she experiments the least she can, which is the level attained by just investing the fixed cost of experimentation. This finding is consistent with our second result. Experimentation is equivalent to paying for information since the information obtained is the same for any experimentation amount. Therefore, the agent pays the lowest price for experimentation regardless of how she uses the acquired information.

We contribute to the rational inattention literature by providing a simple model where information is not available. Rational inattention models have focused on how to model attention and the costs of attention considering scenarios where information is always at disposal for agents, but cognitive limitations block the complete absorption of information. We set aside these concerns and focus on the necessary conditions for attention choice when the agent must pay before to obtain noisy information, and the impact of limited attention on the experimentation decision.

⁴In words of Miller (1956): “When we have a large variance, we are very ignorant about what is going to happen. If we are very ignorant, then when we make the observation it gives us a lot of information.”

2 Related Literature

The literature on attention has two main strands. Attention can be treated either as a psychological bias or as a cognitive limitation. In the former, people's decisions are driven by hedonic motivations. People may be attentive to positive information while remaining inattentive to adverse or negative information (even if that information is useful). [Karlsson et al. \(2009\)](#) propose a model in which investors are inattentive to their financial portfolios when news about the stock market is unfavorable. They referred to this behavior as an "ostrich effect". This attitude may help investors preventing them from selling under panic when markets are down. [Golman et al. \(2017\)](#) model a situation where attention is driven by surprising information since people receive anticipated utility from the dimension related to this new information. News about future illness focuses our attention on health issues through feelings of anxiety, while news about sales on plane tickets drives our attention to leisure aspects through excitement ([Tasoff and Madarasz, 2009](#)). Finally, [Prat-Ortega and de la Rocha \(2018\)](#) show how people are attentive to the news that supports their previous hypotheses leading to confirmation bias. Our model does not account for hedonic motivations. We use a quadratic utility function (an approximation around its maximum). This maximum is reached at zero, which coincides with the expected value of the state of the world. Hence positive and negative deviations are weighted the same.

In the latter, people's decisions are guided by scarcity. Cognitive psychologists have pointed out that attention is limited and must be efficiently employed to enhance information processing ([Simon, 1971](#); [Shiffrin and Schneider, 1977](#)). Economics research on this insight is called rational inattention literature.⁵ The key assumption in these models is that available information is not completely absorbed by the agents. Thus people must choose how to allocate their attention to this information. [Szeidl and Koszegi \(2012\)](#) propose a model in which people focus their attention disproportionately on certain features of their choice set. This effect increases when two options differ the most and disappears when the choice set is balanced. For instance, we may choose one place for vacations rather than other guided mainly by climate instead of

⁵This literature has applications on a variety of fields such as macroeconomics ([Coibion and Gorodnichenko, 2012](#); [Paciello and Wiederholt, 2013](#); [Matějka, 2015](#)), finance ([Mondria and Quintana-Domeque, 2012](#); [Kacperczyk et al., 2016](#)), labor ([Bartoš et al., 2016](#)), and political economy ([Matějka and Tabellini, 2015](#)).

other determinants (Schkade and Kahneman, 1998). Gabaix (2014) develops a “sparse max” operator as an approach to model a decision problem with limited attention. People form a sparse idea of the world and hence consider only first-order variables. Sparsity means assigning optimally zeros and non-zero weights to some components of the vector of variables to represent the attention allocated to each variable. Manzini and Mariotti (2014) show that rational inattention causes the formation of consideration sets according to a preference relation over the set of available options. In other words, attention leads to dismiss a subset of available options and build a new set of choices. Reis (2006) models attention as a binary choice of whether or not become fully informed to answer *when* to attend. In his model, consumers occasionally update their information and calculate their consumption plans, while they remain inattentive between updating dates. Finally, Caplin and Dean (2015) claim that observed mistakes in people’s decisions are explained by a trade-off between the costs of information acquisition and expected benefits of learning rather than by irrational behavior. They model attention as an information cost to characterize all patterns of stochastic choice.

Our work lies in this last literature. Rational inattention is mainly explained by the costs of attending. Seminal work by Sims (2003) establishes the idea of measuring attention as a cost. This cost is measured with the information theory notion of Shannon Entropy.⁶ Informally speaking, entropy is a measure of uncertainty of a random variable. Attention is measured as the reduction in entropy (uncertainty) due to the knowledge of new information (in our model the noisy signal). This is called the mutual information. The reasons for using this approach of Shannon Entropy are that it is tractable (easy to compute), its qualitative properties are reasonable (more precision at higher costs), and that it is justified in information theory (Mackowiak et al., 2018). However, our results are robust if we use reduction in variance instead of a reduction in entropy.⁷

Finally, our work is related to bandit models of information searching. The basic intuitive idea behind these models is explained in the following passage:

⁶See Appendix B for detailed definitions.

⁷This is straightforward since the entropy of a normally distributed variable (Appendix B) is an increasing function in the variance of the state of the world σ_θ^2 .

In the multi-arm bandit problem, the gambler has to decide which arm of K different slot machines to play in a sequence of trials so as to maximize his reward. This classical problem has received much attention because of the simple model it provides of the trade-off between exploration (trying out each arm to find the best one) and exploitation (playing the arm believed to give the best payoff). Each choice of an arm results in an immediate random payoff, but the process determining these payoffs evolves during the play of the bandit (Bergemann and Välimäki, 2008).

Our model can be thought of as a one-armed bandit model in the sense that experimentation is the only source for learning. However, in standard bandit models, the agents learn the probability distribution of each arm directly from observed payoffs. Instead, in our model learning is determined by exogenous signals, which agents can only access through experimentation. One of our results indicates that experimentation implies exploitation for a restricted set of parameters. The trade-off between experimentation and exploitation is given by a cutoff strategy.

3 Theoretical Framework

3.1 Model Without Limited Attention

A two period-lived peasant is interested in investing in a new crop. We model this agent as two incarnations and so we denote the self-playing in period 1 as Player 1 and the self in period 2 as Player 2⁸. Her decision y (i.e. the amount of crop to harvest) depends on the state of nature $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$, which depicts the current climate conditions. This distribution coincides with the agent's prior knowledge. In particular, θ can be thought of as the target investment under

⁸The reasons why we model a game with one player (even if the definition of a game involves the interaction of more than one player whose payoffs depend on the actions of all the players) rather than an usual economic inter-temporal problem of one agent are two. First, we capture the idea of mistakes so the agent playing in period one can differ from the player in period two. This idea is not captured in the inter-temporal problem. Second, WBPE concept restrict the actions of players outside the equilibrium path and thus we can learn from the decisions of the agents when they make mistakes.

specific climate conditions. For each period $i \in \{1, 2\}$ the peasant's immediate payoff is given by

$$\mathcal{U}_i(\theta, y_i) = \mathbb{E}_\theta[-(\theta - y_i)^2] - c\mathbb{1}[y_i > 0]. \quad (\text{I})$$

This objective function can be understood as an approximation of the chosen investment y around its maximum θ less a fixed cost of investment $c > 0$, for example, an input cost (fertilizers, crops or services). Hence, $y \in \{0\} \cup [c, \infty)$. If Player 1 decides to invest (i.e. $y_1 > 0$), then Player 2 receives a perfectly informative signal about the true state of nature. Otherwise, she observes no signal. Then she decides whether to invest or not. Thus, the decision problem for each Player is given by:

$$\begin{aligned} \max_{y_2} \mathcal{V}_2(\theta, y_2) &= \max_{y_2} \mathcal{U}_2(\theta, y_2) \\ \max_{y_1} \mathcal{V}_1(\theta, y_1) &= \max_{y_1} \mathcal{U}_1(\theta, y_1) + \delta \mathcal{V}_2(\theta, y_2) \end{aligned}$$

Where $\delta \in (0, 1)$ is the inter-temporal discount rate. Notice that $\mathcal{V}_i(\cdot)$ entails $y_i = 0$ as a safe choice.⁹

The extensive-form game is depicted in Figure 1. Notice that there are two types of information sets for Player 2. Depending on Player 1's choice, Player 2 either knows the true θ due to the knowledge of the signal or has the same information as Player 1's, which is illustrated by the dashed circle.

Definition 1 (Information Sets). *Let h_1 be the unique information set of Player 1, h_2^0 be the information set of Player 2 when Player 1 chooses $y_1 = 0$ in h_1 , and $(h_2^{y_1})_{y_1 \in \mathbb{R}_{>0}}$ be a vector of information sets of Player 2 when Player 1 chooses $y_1 > 0$ in h_1 .*

Proposition 1. *The model without limited attention has a WPBE equilibrium for each prior θ*

⁹For all $i \in \{1, 2\}$ and for any $\theta, \mathcal{U}_i(\theta, y_i = 0) = -\sigma_\theta^2$. In other words, since the distribution of θ is known ex ante by the agent (i.e. she knows σ_θ^2), the immediate payoff obtained from no experimentation is known ex ante. Therefore, it is independent from the state of the world.

Figure 1: Model - No Limited Attention

given by the following behavioral strategy and partial system of beliefs:¹⁰

$$b = \begin{cases} b(h_1) = ((1)x_1, (0)(y_1)_{y_1 \in \{0\} \cup [c, \infty) \setminus \{x_1\}}) \\ b(h_2^{y_1}) = ((1)x_2^1, (0)(y_2)_{y_2 \in \{0\} \cup [c, \infty) \setminus \{x_2^1\}}), \\ b(h_2^0) = ((1)x_2^2, (0)(y_2)_{y_2 \in \{0\} \cup [c, \infty) \setminus \{x_2^2\}}) \end{cases} \quad \forall y_1 \in \{0\} \cup [c, \infty) \quad (2)$$

$$\mu_b = \begin{cases} \mu(h_1) = (P_b(x | h_1))_{x \in h_1}, & P_b(x | h_1) \sim \mathcal{N}(0, \sigma_\theta^2) \\ \mu(h_2^{y_1}) = ((1)), & \forall y_1 \in \{0\} \cup [c, \infty) \\ \mu(h_2^0) = (P_b(x | h_2^0))_{x \in h_2^0}, & P_b(x | h_2^0) \sim \mathcal{N}(0, \sigma_\theta^2) \end{cases} \quad (3)$$

¹⁰In fact, h_2^0 has probability zero of being reached. However, if we perturb the behavioral strategy following the Sequential Equilibrium concept, we find that $P_b(x | h_2^0) \sim \mathcal{N}(0, \sigma_\theta^2)$, which is intuitive given that Player 2 should have the same information of θ as Player 1 when experimentation does not occur.

Where $x_1 = c$, $x_2^1 = \theta \mathbb{1}[\theta \geq c]$, and $x_2^2 = \mathbb{E}[\theta]$ and a set of parameters σ_θ^2 , δ , and c such that $\sigma_\theta^2 \geq c \left[\frac{c+1}{\delta \mathbb{P}(\theta < c)} + 1 \right]$.

Proof. Appendix A. □

The proposition shows some trivial and non-trivial results. First, when Player 2 has full information (i.e. when Player 1 experiments) experimentation occurs if the true state is suitable for investing ($\theta \geq c$); otherwise, she chooses the safe action: $y_2 = \mathbb{E}[\theta] = 0$. Moreover, when Player 2 has the same information as Player 1 about the true state, she abstains from experiment since objective 1 is maximized by choosing the current belief about θ , i.e., her default choice: $y_2 = \mathbb{E}[\theta] = 0$. This is consistent with the fact that default actions matter as opting out is costly in expected utility terms (Carroll et al., 2009). On the other hand, the level of uncertainty determines the extent to which Player 1 experiments. Player 1 considers only to experiment if uncertainty is above of an endogenously determined threshold $\sigma_\theta^2 \geq c \left[\frac{c+1}{\delta \mathbb{P}(\theta < c)} + 1 \right]$ because Player 1 discounts the immediate payoff of Player 2 obtained from receiving perfect information. Hence, uncertainty entails Player 1 to acquire information for Player 2. However, variance and discount rate are negatively related. The higher the discount rate, the lower the variance needed for experimentation. This reflects a possible substitution between uncertainty and impatience. As δ tends to zero, Player 1 would need an infinite variance to invest, which is not likely to occur given θ 's distribution.¹¹ Instead, she behaves as Player 2, i.e., she solely cares about her immediate payoff and therefore, she takes the safe choice. In other words, the decision problem for each Player is given by Equation 1 and thus is maximized by choosing the default $\mathbb{E}[\theta] = 0$. Furthermore, as stated above, Player 1 invests inasmuch as she cares about Player 2. For instance, Player 1 invests the minimum, i.e., the cost of investing c regardless of the penalty from deviating from θ . Player 1 acknowledges that initial experimentation action is, in fact, paying for information. Thus, she pays the lowest price. This result explains the positive relationship between uncertainty and experimentation costs. The higher the cost, the

¹¹If $\sigma_\theta^2 \rightarrow \infty$, then $\theta \sim Uniform$, letting all the states equally probable. This scenario is the case with the highest uncertainty measured in entropy. Therefore, any source of information would have learning marginally infinite.

higher the variance needed for experimentation. We conduct simulations to exhibit these relationships in Figure 2.¹² Finally, for σ_θ^2 below the threshold, neither Player invests, because uncertainty is so low that Player 1 takes the risk of avoiding costly information and choose her belief to maximize the immediate payoff.

Figure 2: Model Without Limited Attention - Simulations

3.2 The Model

Now consider the following scenario. If Player 1 decides to invest (i.e. $y_1^* > 0$), then Player 2 receives a noisy signal about the true state of nature. The signal is defined as $S = \theta + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The precision of the signal depends on her choice of attention. Attention $\mathcal{A}$ is measured as a decreasing function in σ_ϵ^2 to capture the distortion limits of the noise following [Wiederholt et al. \(2010\)](#). In particular, $\mathcal{A}$ is defined by

$$\mathcal{A} := \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}.$$

¹²The threshold function depends on the cumulative distribution of θ . Thus, the implicit derivatives $\frac{\partial \sigma_\theta^2}{\partial \delta}$ and $\frac{\partial \sigma_\theta^2}{\partial c}$ cannot be calculated. However, we conduct numerical simulations to show the relationships.

Note that as $\sigma_\epsilon^2 \rightarrow 0$ (i.e. no noise) $\mathcal{A} \rightarrow 1$ and $\sigma_\epsilon^2 \rightarrow \infty$ (i.e. full noise) implies $\mathcal{A} \rightarrow 0$. Thus, the more attention she pays, the more precise the signal she receives, which removes uncertainty in decision making. However, attention is costly. The costs are defined by $\lambda\kappa(\mathcal{A})$, where $\lambda > 0$ is a unit cost of information, and $\kappa(\mathcal{A})$ is the quantity of information that Player 2 pays attention to.¹³ This amount is measured as the reduction in the uncertainty of θ due to the knowledge of S , which is expressed by differential entropy.

$$\kappa(\mathcal{A}) = \begin{cases} \mathcal{H}(\theta) - \mathcal{H}(\theta|S) & \text{if } y_1 > 0 \\ 0 & \text{if } y_1 = 0 \end{cases}$$

Where $\mathcal{H}(\theta) - \mathcal{H}(\theta | S) = \frac{1}{2} \ln\left(\frac{1}{1-\mathcal{A}}\right)$. The chosen κ is zero when Player 1 does not invest, since information is only available when experimentation occurs. Thus, the decision problem for each Player is now given by:

$$\begin{aligned} \max_{y_2} \mathcal{V}_2(\theta, y_2) &= \max_{y_2} \mathcal{U}_2(\theta, y_2) - \lambda\kappa \\ \max_{y_1} \mathcal{V}_1(\theta, y_1) &= \max_{y_1} \mathcal{U}_1(\theta, y_1) + \delta \mathcal{V}_2(\theta, y_2) \end{aligned}$$

Where $\delta \in (0, 1)$ is the inter-temporal discount rate.

The extensive-form game is depicted in Figure 3. Notice that there is no case for full information in contrast with the Benchmark model. Even if Player 2 decides to pay full attention to reduce all uncertainty, it would be unfeasible given the costs of attending. Thus, after the decision on attention, there is still imperfect information. However, *a priori* information is different from *a posteriori*'s. This is exhibited in the size of the dashed circles. Even though attention does not remove completely uncertainty, it does improve knowledge about θ .¹⁴ Hence, the size of the dashed circle after paying attention is smaller. Therefore, paying attention helps Player 2 to choose her investment better.

¹³In fact, $\kappa(\mathcal{A})$ is an injective function. Thus, choosing κ is equivalent to choosing $\mathcal{A}$. So we define the space of actions in $h_2^{y_1}$ by $[0, 1]$, which is the space of $\mathcal{A}$.

¹⁴As $\mathcal{A}$ tends to zero, σ_ϵ^2 tends to infinity. Hence, ϵ would be a random variable with probability distribution function of zero (a uniformly continuous random variable over $\mathbb{R}$) and the noise would not have any distortion in the information of θ . Thus, signal information would be at least the same as θ 's. Likewise, mutual information (the reduction in entropy) is always non-negative (Cover and Thomas, 2012).

Figure 3: Model - Limited Attention

Definition 2 (Information Sets). Let h_1 be the unique information set of Player 1, $h_2^{y_1=0}$ be the information set of Player 2 when Player 1 chooses $y_1 = 0$ in h_1 , $h_2^{y_1>0}$ be the information set of Player 2 when Player 1 chooses $y_1 > 0$ in h_1 , $h_2^{A=0}$ be the information set of Player 2 when Player 2 chooses $A = 0$ in $h_2^{y_1>0}$, and $h_2^{A>0}$ be the information sets of Player 2 when Player 2 chooses $A \in (0, 1]$ in $h_2^{y_1>0}$.

Proposition 2. The model has a WPBE equilibrium for each prior θ given by the following

behavioral strategy and partial system of beliefs:¹⁵

$$b = \begin{cases} b(h_1) = \left((1)x_1, (0)(y_1)_{y_1 \in \{0\} \cup [c, \infty) \setminus \{x_1\}} \right) \\ b(h_2^{y_1 > 0}) = \left((1)x_2^1, (0)(\mathcal{A})_{\mathcal{A} \in [0, 1] \setminus \{x_2^1\}} \right) \\ b(h_2^{y_1 = 0}) = \left((1)x_2^2, (0)(\mathcal{A})_{\mathcal{A} \in [0, 1] \setminus \{x_2^2\}} \right) \\ b(h_2^{A > 0}) = \left((1)x_2^3, (0)(y_2)_{y_2 \in \{0\} \cup [c, \infty) \setminus \{x_2^3\}} \right) \\ b(h_2^{A = 0}) = \left((1)x_2^4, (0)(y_2)_{y_2 \in \{0\} \cup [c, \infty) \setminus \{x_2^4\}} \right) \end{cases} \quad (4)$$

Where $x_1 = c$, $x_2^1 = \max \left\{ 0, 1 - \frac{\lambda}{2\sigma_\theta^2 \mathbb{P}(\mathbb{E}[\theta | S] \geq c)} \right\}$, $x_2^2 = \mathbb{E}[\theta] = 0$, $x_2^3 = \mathbb{E}[\theta | S] \mathbb{1}[\mathbb{E}[\theta | S] \geq c]$, $x_2^4 = \mathbb{E}[\theta] = 0$, and a set of parameters λ , c , δ , and σ_θ^2 such that $\sigma_\theta^2 \geq f(\sigma_\theta^2, c, \delta, \lambda, \mathbb{P}(\mathbb{E}[\theta | S] \geq c))$

$$\mu_b = \begin{cases} \mu(h_1) = \left(P_b(x | h_1) \right)_{x \in h_1}, & P_b(x | h_1) \sim \mathcal{N}(0, \sigma_\theta^2) \\ \mu(h_2^{y_1 > 0}) = \left(P_b(x | h_2^{y_1 > 0}) \right)_{x \in h_2^{y_1 > 0}}, & P_b(x | h_2^{y_1 > 0}) \sim \mathcal{N}(0, \sigma_\theta^2) \\ \mu(h_2^{y_1 = 0}) = \left(P_b(x | h_2^{y_1 = 0}) \right)_{x \in h_2^{y_1 = 0}}, & P_b(x | h_2^{y_1 = 0}) \sim \mathcal{N}(0, \sigma_\theta^2) \\ \mu(h_2^{A > 0}) = \left(P_b(x | h_2^{A > 0}) \right)_{x \in h_2^{A > 0}}, & P_b(x | h_2^{A > 0}) \sim \mathcal{N}(0, \frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2}) \\ \mu(h_2^{A = 0}) = \left(P_b(x | h_2^{A = 0}) \right)_{x \in h_2^{A = 0}}, & P_b(x | h_2^{A = 0}) \sim \mathcal{N}(0, \sigma_\theta^2) \end{cases} \quad (5)$$

Proof. Appendix C. □

The solution has some striking features. First, when information is available there is no case for rational inattention: $\mathcal{A} = 1 - \frac{\lambda}{2\sigma_\theta^2 \mathbb{P}(\mathbb{E}[\theta | S] \geq c)} > 0$. This could be explained by the fact that the agent cannot waste information that she already has paid for. For instance, if experimentation is seen as a synonym of paying for information, then it is rational to use the information once she pays for it. In contrast, if the information would be received at zero cost ($\lambda = 0$), for sufficiently low uncertainty, she would ignore information as precision is not necessary for determining her strategy. Likewise, if the costs of attention are sufficiently high, the agent would abstain from paying attention as the expectation of payoffs in

¹⁵Analogously as in the Model Without Limited Attention, if we perturb the pure behavioral strategies, $P_b(x | h_2^{y_1 = 0}) \sim \mathcal{N}(0, \sigma_\theta^2)$ and $P_b(x | h_2^{A = 0}) \sim \mathcal{N}(0, \sigma_\theta^2)$.

bad states is lower than this cost. Second, as in the Benchmark model, initial experimentation is driven by a cutoff strategy depending on uncertainty. This is reflected by the threshold $\sigma_\theta^2 \geq f(\sigma_\theta^2, c, \delta, \lambda, \mathbb{P}(\mathbb{E}[\theta \mid S] \geq c))$. The agent only considers experimenting if the variance of the state of the world is sufficiently high. This result is consistent with the feature of rational inattention models that attention is mainly driven by beliefs. However, in contrast with discrete choice models where the agents allocate their attention according to their beliefs about the set of available options, beliefs are seen here as the variance of the distribution since the set of options in our model is unitary: $\{\theta\}$. In such models, attention is decreasing in uncertainty (i.e. when the agents have the same beliefs for all options) because any option would be *a priori* rational to attend. In our case, the value of information is increasing in the level of uncertainty and thus the agent chooses to pay attention to remove this uncertainty. The reason is that the higher the uncertainty, the higher the learning. Furthermore, as proved in the Benchmark Model, the relationship between the uncertainty and discount rate is negative, while uncertainty and the cost of experimentation are positively related. Moreover, the addition of limited attention to the Model Without Limited Attention does not change the decision about the initial experimentation. In both scenarios, as stated above, experimentation is considered as paying for information. Thus, the agent is willing to pay the lowest price to acquire this information, which is the cost of investing regardless of the expected quality of information nor the level of attention paid to it. We conduct simulations to exhibit these relationships in Figure 5 and Figure 6 of Appendix D. The figure shows that the inclusion of limited attention to the initial setting does not modify the relations exhibited in the Model Without Limited Attention. However, Figure 4 shows a difference in the parameters needed to guarantee the initial experimentation equilibrium. Given a fixed level of variance of the state of the world, the $f(\sigma_\theta^2, c, \delta, \lambda, \mathbb{P}(\mathbb{E}[\theta \mid S] \geq c))$ threshold is closer to the level of variance than $c \left[\frac{c+1}{\delta \mathbb{P}(\theta < c)} + 1 \right]$, whenever the level of variance is greater than the thresholds. This means that every level of cost of investment and any discount factor that supports the initial experimentation equilibrium in the Model With Limited Attention, then it supports the same equilibrium in the Model Without Limited Attention. That is, the addition of limited attention to our setting dismisses some levels of parameters that support the initial experimentation equilibrium meaning that it is “harder” to reach the equilibrium in the model.

Figure 4: Contrast Between Models - Simulations

4 Concluding Remarks

We introduce the necessary conditions for initial experimentation and attention when information is available. These conditions display the main feature of rational inattention models where attention is guided by the prior beliefs and the costs of attention. We use the Shannon entropy to model the information flow attended by the agent. This modeling allows us to describe an intuitive notion of attention measured as the reduction in uncertainty of the state of the world. The addition of the experimentation setting enriches this feature by including the discount factor as a determinant for the cutoff strategy. Moreover, we find that when information is available, there is no case for rational inattention. This finding is explained by the fact that information acquired in a costly fashion cannot be wasted even if attention is costly as well. Finally, initial experimentation is independent of the quality of the information acquired conditional on certain parameters. In other words, the limited attention does not affect the initial investing since the noise never worsens the quality of information. Experimentation is understood as paying for information and therefore the agent pays the lowest price regardless of the quality of information (perfectly informative or noisy).

The results show that our setting can be enriched to exploit the relationship between experimentation and attention. The fact that our model does not exhibit rational inattention may be explained because experimentation is binary. A dynamic multi-armed bandit model with armed-dependent experimentation costs could reflect rational inattention if we model attention as a scarce inter-temporal resource. That is, the level of attention may decrease in time as a result of fatigue or use. Thus, the agents may refine the knowledge of the arms with relatively favorable payoffs or beliefs and may select which periods to rationally attend. A different approach is to model the quality of information as an increasing function in the level of experimentation. Therefore, attention may not be needed when the quality of received information is sufficient to support the strategy of the agent, while the initial level of experimentation could vary from the no limited attention setting to the one with limited attention.

A Model Without Limited Attention Solution

First, we check the consistency of beliefs:

1. In h_1 , $P_b(h_1) = 1$ since every state of the world is played with positive probability.

Hence, for every $x \in h_1$, $P_b(x)$ coincides with the prior θ . Thus,

$$P_b(x | h_1) = \frac{P_b(x)}{P_b(h_1)} \sim \mathcal{N}(0, \sigma_\theta^2).$$

2. For every $y_1 \in \mathbb{R}_{>0}$, in $h_2^{y_1}$ uncertainty disappears due to the knowledge of $S = \theta$.

Thus,

$$P_b(x | h_2^{y_1}) = 1.$$

3. In h_2^0 , $P_b(h_2^0) = P_b(y_1 = 0) = 0$. Hence, for every node $x \in h_2^0$, $P_b(x) = 0$. Thus,

$$P_b(x | h_2^0) = \frac{P_b(x)}{P_b(h_2^0)} = p \in [0, 1].$$

Beliefs, calculated this way, are consistent with Bayes' rule.

Second, we verify that b is sequentially rational under μ_b .

1. In h_2^0 , $\mathcal{V}_2(\theta, y_2)$ reaches its maximum when $y_2 = \mathbb{E}[\theta] = 0$.¹⁶ Hence,

$$\mathcal{V}_2(\theta, y_2) = \mathbb{E}[-(\theta - \mathbb{E}[\theta])^2] = -\sigma_\theta^2. \quad (6)$$

Therefore, b is rational in h_2^0 given μ_b .

2. For every $y_1 > 0$, in $h_2^{y_1}$ Player 2 knows the true state of the world. So she maximizes

$\mathcal{V}_2(\theta, y_2)$ choosing $y_2 = \theta \mathbb{1}[\theta \geq c]$. Thus,

$$\mathcal{V}_2(\theta, y_2) = \begin{cases} -c & \text{if } \theta \geq c \\ -\theta^2 & \text{if } \theta < c \end{cases} \quad (7)$$

Hence, b is rational in $h_2^{y_1}$ for every $y_1 > 0$ given μ_b .

¹⁶For every $x, x' \in h_2^0$ with $x \neq x'$, $\mathcal{U}_2(b | x) = \mathcal{U}_2(b | x')$ for every $b \in \mathcal{B}$. Therefore, $P_b(x | h_2^0)$ is unnecessary for determining rationality of b , in other words, the expectation of $\mathcal{U}_2(b)$ is equal to $\mathcal{U}_2(b)$ in any arbitrary node.

3. In h_1 , Player 1 has “two” choices: $y_1 > 0$ or $y_1 = \mathbb{E}[\theta] = 0$.

(a) If $y_1 > 0$, $\mathcal{U}_1(\theta, y_1 > 0) = \mathbb{E}[-(\theta - y_1)^2] = -\sigma_\theta^2 - y_1^2 - c$. Then

$$\mathcal{V}_1(\theta, y_1 > 0) = \begin{cases} -\sigma_\theta^2 - y_1^2 - c - \delta c & \text{if } \theta \geq c \\ -\sigma_\theta^2 - y_1^2 - c - \delta \sigma_\theta^2 & \text{if } \theta < c \end{cases} \quad (8)$$

since θ^2 is unknown for Player 1 when $\theta \leq 0$.

(b) If $y_1 = 0$, $\mathcal{U}_1(\theta, y_1 = 0) = \mathbb{E}[-(\theta - 0)^2] = -\sigma_\theta^2$. Then

$$\mathcal{V}_1(\theta, y_1 = 0) = -\sigma_\theta^2 - \delta \sigma_\theta^2 = -\sigma_\theta^2(1 + \delta) \quad (9)$$

It is rational for Player 1 to play $y_1 > 0$ if

$$\begin{aligned} \mathbb{P}(\theta \geq c)\mathcal{V}_1(\theta \geq c, y_1 > 0) + \mathbb{P}(\theta < c)\mathcal{V}_1(\theta < c, y_1 > 0) &\geq \mathcal{V}_1(\theta, y_1 = 0) \\ \Leftrightarrow y_1 &\leq \sqrt{\delta \sigma_\theta^2 \mathbb{P}(\theta < c) - c(1 + \delta - \delta \mathbb{P}(\theta \geq c))} \end{aligned} \quad (10)$$

where

$$\mathbb{P}(\theta \geq c) = \int_c^\infty \frac{1}{\sqrt{2\pi\sigma_\theta^2}} e^{-\frac{\theta^2}{2\sigma_\theta^2}} d\theta.$$

Recall that $y_1 > 0$ implies $y_1 \geq c$. Thus, Player 1's optimal strategy is given by:

$$y_1 = c \quad (11)$$

if and only if $\sigma_\theta^2 \geq c \left[\frac{c+1}{\delta \mathbb{P}(\theta < c)} + 1 \right]$.¹⁷

Therefore, (b, μ_b) is sequentially rational. Thus, (b, μ_b) is a WPBE.

¹⁷This also ensures that $\delta \sigma_\theta^2 \mathbb{P}(\theta < c) \geq c(1 + \delta - \delta \mathbb{P}(\theta \geq c))$.

B Shannon Cost

Definition 3 (Differential Entropy). *The differential entropy $\mathcal{H}(X)$ of a continuous random variable X with a density $f(x)$ is defined as*

$$\mathcal{H}(X) = \mathbb{E}[-\log(f(x))] = - \int_S f(x) \log(f(x)) dx,$$

where S is the support set of the random variable X .¹⁸

Definition 4 (Jointly Normal). *Two random variables X and Y are said to be jointly normal if they can be expressed in the form*

$$X = aU + bV, \quad Y = cU + dV$$

where U and V are independent random variables and a, b, c , and d are scalars.

Definition 5 (Conditional Differential Entropy). *If X, Y have a joint density function $f(x, y)$, the conditional differential entropy $\mathcal{H}(X | Y)$ is defined as*

$$\mathcal{H}(X | Y) = - \int f(x, y) \log[f(x | y)] dx dy.$$

Definition 6 (Mutual Information). *The mutual information $\mathcal{I}_{X,Y}$ between two random variables X and Y is defined as*

$$\mathcal{I}_{X,Y} = \mathcal{H}(X) - \mathcal{H}(X | Y).$$

- i. Find $\mathcal{H}(\theta)$. Recall that $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$

$$\begin{aligned} \mathcal{H}(\theta) &= \mathbb{E}[-\ln(f(\theta))] = - \int_S f(\theta) \ln(f(\theta)) d\theta \\ &= \int_S f(\theta) \ln \left[\frac{1}{\sqrt{2\pi\sigma_\theta^2}} e^{-\frac{\theta^2}{2\sigma_\theta^2}} \right] d\theta = - \int_S f(\theta) \ln \left[\frac{1}{\sqrt{2\pi\sigma_\theta^2}} - \frac{\theta^2}{2\sigma_\theta^2} \right] d\theta \\ &= \int_S f(\theta) \ln \left[\sqrt{2\pi\sigma_\theta^2} \right] d\theta + \int_S f(\theta) \frac{\theta^2}{2\sigma_\theta^2} d\theta \\ &= \frac{1}{2} \ln(2\pi\sigma_\theta^2) \int_S f(\theta) d\theta + \frac{1}{2\sigma_\theta^2} \int_S f(\theta) \theta^2 d\theta \\ &= \frac{1}{2} \ln(2\pi\sigma_\theta^2) + \frac{1}{2\sigma_\theta^2} \mathbb{E}[\theta^2] = \frac{1}{2} \ln(2\pi\sigma_\theta^2) + \frac{\sigma_\theta^2}{2\sigma_\theta^2} \\ &= \frac{1}{2} \ln(2\pi\sigma_\theta^2) + \frac{1}{2} \ln(e) = \frac{1}{2} \ln(2\pi e \sigma_\theta^2) \end{aligned} \tag{12}$$

where S is the support of θ .

¹⁸See [Cover and Thomas \(2012\)](#).

2. Show that S and θ are jointly normal.

Let $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Now define $S = 1\theta + 1\epsilon$ and $\theta = 1\theta + 0\epsilon$. Since θ and ϵ are normally distributed random independent variables, S and θ are jointly normal.

3. Calculate $\mathbb{E}[\theta]$, $\mathbb{E}[S]$, $\text{Var}[\theta]$, $\text{Var}[S]$, $\text{Cov}[\theta, S]$, $\rho_{\theta, S}$, $\mathbb{E}[\theta | S]$, and $\text{Var}[\theta | S]$.

$$\mathbb{E}[\theta] = 0$$

$$\mathbb{E}[S] = \mathbb{E}[\theta + \epsilon] = \mathbb{E}[\theta] + \mathbb{E}[\epsilon] = 0$$

$$\text{Var}[\theta] = \sigma_\theta^2$$

$$\text{Var}[S] = \mathbb{E}[S^2] - \mathbb{E}[S]^2 = \mathbb{E}[(\theta + \epsilon)^2] = \sigma_\theta^2 + \sigma_\epsilon^2 = \sigma_S^2$$

$$\text{Cov}[\theta, S] = \mathbb{E}[\theta S] - \mathbb{E}[\theta]\mathbb{E}[S] = \mathbb{E}[\theta(\theta + \epsilon)] = \mathbb{E}[\theta^2] = \sigma_\theta^2$$

$$\rho_{\theta, S} = \frac{\text{Cov}[\theta, S]}{\sigma_\theta \sigma_S} = \frac{\sigma_\theta^2}{\sigma_\theta \sqrt{\sigma_\theta^2 + \sigma_\epsilon^2}} = \frac{\sigma_\theta}{\sqrt{\sigma_\theta^2 + \sigma_\epsilon^2}}$$

$$\mathbb{E}[\theta | S] = \rho_{\theta, S} \frac{\sigma_\theta}{\sigma_S} S = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2} (\theta + \epsilon)$$

$$\text{Var}[\theta | S] = (1 - \rho_{\theta, S}^2) \sigma_\theta^2 = \left[1 - \left(\frac{\sigma_\theta}{\sqrt{\sigma_\theta^2 + \sigma_\epsilon^2}} \right)^2 \right] \sigma_\theta^2 = \frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2} = \sigma_{\theta|S}^2$$

where the conditional expectation and the conditional variance come from the fact that S and θ are jointly normal.

4. Find $\mathcal{H}(\theta | S)$.

$$\begin{aligned} \mathcal{H}(\theta | S) &= \frac{1}{2} \ln(2\pi e \sigma_{\theta|S}^2) \\ &= \frac{1}{2} \ln \left(2\pi e \frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2} \right) \end{aligned} \tag{I3}$$

5. Find $\mathcal{I}_{\theta, S}$.

$$\begin{aligned} \mathcal{I}_{\theta, S} &= \mathcal{H}(\theta) - \mathcal{H}(\theta | S) \\ &= \frac{1}{2} \ln(2\pi e \sigma_\theta^2) - \frac{1}{2} \ln \left(2\pi e \frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2} \right) \\ &= \frac{1}{2} \ln \left(\frac{\sigma_\theta^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2} \right) \end{aligned} \tag{I4}$$

C Model With Limited Attention Solution

First, we check the consistency of beliefs:

1. In h_1 , $P_b(h_1) = 1$ since every state of the world is played with positive probability.

Hence, for every $x \in h_1$, $P_b(x)$ coincides with the prior θ . Thus,

$$P_b(x | h_1) = \frac{P_b(x)}{P_b(h_1)} \sim \mathcal{N}(0, \sigma_\theta^2).$$

2. In $h_2^{y_1 > 0}$, $P_b(h_2^{y_1 > 0}) = 1$ since for every θ , $y_1 > 0$ is played with positive probability.

Hence, for every $x \in h_2^{y_1 > 0}$, $P_b(x)$ coincides with the prior θ . Thus,

$$P_b(x | h_2^{y_1 > 0}) \sim \mathcal{N}(0, \sigma_\theta^2).$$

3. In $h_2^{y_1 = 0}$, $P_b(h_2^{y_1 = 0}) = P_b(y_1 = 0) = 0$. Hence, for every node $x \in h_2^{y_1 = 0}$, $P_b(x) = 0$.

Thus,

$$P_b(x | h_2^{y_1 = 0}) = \frac{P_b(x)}{P_b(h_2^{y_1 = 0})} = p \in [0, 1].$$

4. In $h_2^{A=0}$, $P_b(h_2^{A=0}) = P_b(y_1 = 0) = 0$. Hence, for every node $x \in h_2^{A=0}$, $P_b(x) = 0$.

Thus,

$$P_b(x | h_2^{A=0}) = \frac{P_b(x)}{P_b(h_2^{A=0})} = p \in [0, 1].$$

5. In h_2^A uncertainty is reduced due to the knowledge of $S = \theta + \epsilon$. Thus,

$$P_b(x | h_2^A) \sim \mathcal{N}\left(0, \frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2}\right).$$

Beliefs, calculated this way, are consistent with Bayes' rule.

Second, we verify that b is sequentially rational under μ_b .

1. In $h_2^{A=0}$, $\mathcal{V}_2(\theta, y_2)$ reaches its maximum when $y_2 = \mathbb{E}[\theta] = 0$ ¹⁹. Hence,

$$\mathcal{V}_2(\theta, y_2) = \mathbb{E}[-(\theta - \mathbb{E}[\theta])^2] = -\sigma_\theta^2. \quad (15)$$

Therefore, b is rational in $h_2^{A=0}$ given μ_b .

¹⁹Since for every $x, x' \in h_2^0$ with $x \neq x'$, $\mathcal{U}_2(b | x) = \mathcal{U}_2(b | x')$ for every $b \in \mathcal{B}$. This allows to fix the distribution over the nodes in $h_2^{A=0}$ and $h_2^{y_1=0}$ as $\mathcal{N}(0, \sigma_\theta^2)$, which is intuitive given that Player 2 should have the same information of θ as Player 1 when experimentation does not occur.

2. For every $\mathcal{A} \in (0, 1]$, in $h_2^{\mathcal{A}>0}$ Player 2 has more information of θ due to the knowledge of S . So she maximizes $\mathcal{V}_2(\theta, y_2)$ by choosing $y_2 = \mathbb{E}[\theta \mid S] \mathbb{1}[\mathbb{E}[\theta \mid S] \geq c]$. Thus, if $\mathbb{E}[\theta \mid S] \geq c$,

$$\begin{aligned}\mathcal{V}_2(\theta, y_2) &= \mathbb{E}_\theta \left[\mathbb{E}_S [-(\theta - \mathbb{E}[\theta \mid S])^2] \right] - \lambda \kappa - c \\ &= \mathbb{E}_\theta \left[\mathbb{E}_S \left[-\left(\theta - \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2} S \right)^2 \right] \right] - \frac{\lambda}{2} \ln \left(\frac{\sigma_\theta^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2} \right) - c \\ &= -\frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2} - \frac{\lambda}{2} \ln \left(\frac{\sigma_\theta^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2} \right) - c\end{aligned}\quad (16)$$

If $\mathbb{E}[\theta \mid S] < c$, then $\mathcal{V}_2(\theta, y_2) = -\sigma_\theta^2 - \frac{\lambda}{2} \ln \left(\frac{\sigma_\theta^2 + \sigma_\epsilon^2}{\sigma_\theta^2} \right)$.

Hence, b is rational in $h_2^{\mathcal{A}>0}$ for every $\mathcal{A} \in (0, 1]$ given μ_b .

3. In $h_2^{y_1=0}$, $\mathcal{V}_2(\theta, y_2)$ reaches its maximum when $y_2 = \mathbb{E}[\theta] = 0$. Hence, $\mathcal{V}_2(\theta, y_2) = \mathbb{E}[-(\theta - \mathbb{E}[\theta])^2] = -\sigma_\theta^2$.

Therefore b is rational in $h_2^{y_1=0}$ given μ_b .

4. For every $y_1 > 0$, in $h_2^{y_1>0}$ Player 2 wishes to

$$\max_{\sigma_\epsilon^2 \geq 0} \mathcal{V}_2(\theta, y_2) = \begin{cases} -\frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2} - \frac{\lambda}{2} \ln \left(\frac{\sigma_\theta^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2} \right) - c \mathbb{1}[y_2 > 0] & \text{if } \mathbb{E}[\theta \mid S] \geq c \\ -\sigma_\theta^2 - \frac{\lambda}{2} \ln \left(\frac{\sigma_\theta^2 + \sigma_\epsilon^2}{\sigma_\theta^2} \right) - c \mathbb{1}[y_2 > 0] & \text{if } \mathbb{E}[\theta \mid S] < c \end{cases}$$

Recall that $\mathcal{A} := \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}$ is the level of attention chosen by Player 2. Rewriting $\mathcal{V}_2(\theta, y_2)$ in terms of $\mathcal{A}$, Player 2 wishes to

$$\max_{\mathcal{A} \in [0, 1]} \mathcal{V}_2(\theta, y_2) = \begin{cases} -(1 - \mathcal{A})\sigma_\theta^2 - \frac{\lambda}{2} \ln \left(\frac{1}{1 - \mathcal{A}} \right) - c \mathbb{1}[y_2 > 0] & \text{if } \mathbb{E}[\theta \mid S] \geq c \\ -\sigma_\theta^2 - \frac{\lambda}{2} \ln \left(\frac{1}{1 - \mathcal{A}} \right) - c \mathbb{1}[y_2 > 0] & \text{if } \mathbb{E}[\theta \mid S] < c \end{cases}$$

Let $\mathbb{P} := \mathbb{P}(\mathbb{E}[\theta \mid S] \geq c)$. Therefore, optimal attention is given by $\mathcal{A} = \max \left\{ 0, 1 - \frac{\lambda}{2\sigma_\theta^2 \mathbb{P}} \right\}$. Thus,

$$\mathcal{V}_2(\theta, y_2) = \begin{cases} -\frac{\lambda}{2} \ln \left(\frac{2\sigma_\theta^2 e \mathbb{P}}{\lambda} \right) - c \mathbb{P} - (1 - \mathbb{P})\sigma_\theta^2 & \text{if } 1 > \frac{\lambda}{2\sigma_\theta^2 \mathbb{P}} \\ -\mathbb{P}c - (1 - \mathbb{P})\sigma_\theta^2 & \text{if } 1 \leq \frac{\lambda}{2\sigma_\theta^2 \mathbb{P}} \end{cases}\quad (17)$$

Finally, notice that $\mathcal{A} = 1$ if and only if $\lambda = 0$, but $\mathcal{V}_2(\theta, y_2)$ is undetermined for $\lambda = 0$. However, using L'Hôpital, $\lambda \rightarrow 0$ implies $\mathcal{V}_2(\theta, y_2) \rightarrow -c$, which depicts the full information setting.²⁰

Hence, b is rational in $h_2^{y_1 > 0}$ given μ_b .

5. In h_1 , Player I has “two” choices: $y_1 > 0$ or $y_1 = \mathbb{E}[\theta] = 0$.

(a) If $y_1 > 0$, $\mathcal{U}_1(\theta, y_1 > 0) = \mathbb{E}[-(\theta - y_1)^2] - c = -\sigma_\theta^2 - y_1^2 - c$. Then

$$\mathcal{V}_1(\theta, y_1 > 0) = \begin{cases} -\sigma_\theta^2 - y_1^2 - c - \delta \left[\frac{\lambda}{2} \ln \left(\frac{2\mathbb{P}\sigma_\theta^2 e}{\lambda} \right) + \mathbb{P}c + (1 - \mathbb{P})\sigma_\theta^2 \right] & \text{if } \mathcal{A} > 0 \\ -\sigma_\theta^2 - y_1^2 - c - \delta \left[\mathbb{P}c + (1 - \mathbb{P})\sigma_\theta^2 \right] & \text{if } \mathcal{A} \leq 0 \end{cases}$$

(b) If $y_1 = 0$, $\mathcal{U}_1(\theta, y_1 = 0) = \mathbb{E}[-(\theta - 0)^2] = -\sigma_\theta^2$. Then

$$\mathcal{V}_1(\theta, y_1 = 0) = -\sigma_\theta^2 - \delta\sigma_\theta^2 = -\sigma_\theta^2(1 + \delta)$$

It is rational for Player I to play $y_1 > 0$ if

$$\begin{aligned} \mathcal{V}_1(\theta, y_1 > 0) &\geq \mathcal{V}_1(\theta, y_1 = 0) \\ \Leftrightarrow y_1 &\leq \sqrt{\delta \left[\sigma_\theta^2 - \frac{\lambda}{2} \ln \left(\frac{2\mathbb{P}\sigma_\theta^2 e}{\lambda} \right) \right] - c(1 + \delta\mathbb{P})} \end{aligned} \quad (18)$$

Thus, Player I's optimal strategy is given by:

$$y_1 = c \quad (19)$$

if and only if $\sigma_\theta^2 \geq f(\sigma_\theta^2, c, \delta, \lambda, \mathbb{P}(\mathbb{E}[\theta | S] \geq c))$, where

$$f(\sigma_\theta^2, c, \delta, \lambda, \mathbb{P}(\mathbb{E}[\theta | S] \geq c)) = \frac{c^2 + c(1 + \delta\mathbb{P})}{\delta\mathbb{P}} + \frac{\delta}{2\mathbb{P}} \ln \left(\frac{2\mathbb{P}\sigma_\theta^2 e}{\lambda} \right)$$

Therefore, (b, μ_b) is sequentially rational. Thus, (b, μ_b) is a WPBE.

²⁰See Secondary Results D.

D Secondary Results Model

1. $\lim_{\lambda \rightarrow 0} \mathcal{V}_2(\theta, y_2) = 0$: Let $\mathcal{A} > 0$, i.e., $\mathcal{A} = 1 - \frac{\lambda}{2\sigma_\theta^2\mathbb{P}}$, then

$$\begin{aligned} \lim_{\lambda \rightarrow 0} \mathcal{V}_2(\theta, y_2) &= \lim_{\lambda \rightarrow 0} -\frac{\lambda}{2} \ln\left(\frac{2\mathbb{P}\sigma_\theta^2 e}{\lambda}\right) - c = \lim_{\lambda \rightarrow 0} -\frac{\ln\left(\frac{2\mathbb{P}\sigma_\theta^2 e}{\lambda}\right)}{\frac{2}{\lambda}} - c \\ &= \lim_{\lambda \rightarrow 0} \frac{\frac{\lambda}{2\mathbb{P}\sigma_\theta^2 e} \frac{2\mathbb{P}\sigma_\theta^2 e}{\lambda^2}}{-\frac{2}{\lambda^2}} - c = \lim_{\lambda \rightarrow 0} -\frac{\lambda}{2} - c = -c \end{aligned}$$

2. Simulations Cost of Investment and Discount Factor:

Figure 5: Model With Limited Attention - Simulations I

3. Simulation Cost of Attention:

Figure 6: Model With Limited Attention - Simulations II

References

- Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–75.
- Bergemann, D. and Välimäki, J. (2008). Bandit problems. *The New Palgrave Dictionary of Economics: Volume 1–8*, pages 336–340.
- Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203.
- Caplin, A., Dean, M., and Leahy, J. (forthcoming). Rational inattention, optimal consideration sets and stochastic choice. *The Review of Economic Studies*.
- Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., and Metrick, A. (2009). Optimal Defaults and Active Decisions*. *The Quarterly Journal of Economics*, 124(4):1639–1674.
- Coibion, O. and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1):116–159.
- Cover, T. M. and Thomas, J. A. (2012). *Elements of Information Theory*. John Wiley & Sons.
- De Los Santos, B., Hortaçsu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80.
- Eliasz, K. and Schotter, A. (2010). Paying for confidence: An experimental study of the demand for non-instrumental information. *Games and Economic Behavior*, 70(2):304–324.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.
- Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, 76(5):1017–1074.
- Golman, R., Hagmann, D., and Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, 55(1):96–135.
- Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds’ attention allocation. *Econometrica*, 84(2):571–626.
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall series in experimental psychology. Prentice-Hall.

- Karlsson, N., Loewenstein, G., and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, 38(2):95–115.
- Mackowiak, B. A., Matejka, F., Wiederholt, M., et al. (2018). Survey: Rational inattention, a disciplined behavioral model. Technical report, CEPR Discussion Papers.
- Manzini, P. and Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176.
- Matějka, F. (2015). Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, 83(3):1125–1155.
- Matějka, F. and Tabellini, G. (2015). Electoral competition with rationally inattentive voters. *CERGE-EI Working Paper Series*.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81.
- Mondria, J. and Quintana-Domeque, C. (2012). Financial contagion and attention allocation. *The Economic Journal*, 123(568):429–454.
- Paciello, L. and Wiederholt, M. (2013). Exogenous information, endogenous information, and optimal monetary policy. *Review of Economic Studies*, 81(1):356–388.
- Prat-Ortega, G. and de la Rocha, J. (2018). Selective attention: A plausible mechanism underlying confirmation bias. *Current Biology*, 28(19):1151–1154.
- Reis, R. (2006). Inattentive consumers. *Journal of Monetary Economics*, 53(8):1761–1800.
- Schkade, D. A. and Kahneman, D. (1998). Does living in california make people happy? a focusing illusion in judgments of life satisfaction. *Psychological Science*, 9(5):340–346.
- Shiffrin, R. M. and Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2):127.
- Simon, H. A. (1971). Designing Organizations for an Information-Rich World. *Computers, Communications, and the Public Interest*, pages 37–72.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.
- Styles, E. (2006). *The psychology of attention*. Psychology Press.
- Szeidl, A. and Koszegi, B. (2012). A Model of Focusing in Economic Choice. *The Quarterly Journal of Economics*, 128(1):53–104.

- Tasoff, J. and Madarasz, K. (2009). A model of attention and anticipation. *Available at SSRN 2629404*.
- Wiederholt, M. et al. (2010). Rational inattention. *The New Palgrave Dictionary of Economics (Online Edition ed.)*.
- Wilson, C. J. (2008). Consideration sets and political choices: A heterogeneous model of vote choice and sub-national party strength. *Political Behavior*, 30(2):161–183.