CERGE
Center for Economics Research and Graduate Education
Charles University Prague

Essays on the implications of bounded rationality to choice

Sergei Mikhalishchev

Dissertation

Prague, May 2022

Sergei Mikhalishchev

Essays on the implications of bounded rationality to choice

Dissertation

Prague, May 2022

Dissertation Committee

Filip Matějka (CERGE-EI; chair)

Jan Zápal (CERGE-EI)

Avner Shaked (University of Bonn)

Referees

Chen Lian (University of California, Berkeley)

Peter Zorn (Ludwig Maximilian University of Munich)

Table of Contents

Abstract

In the first chapter, we introduce a new role of quotas, e.g., labor market quotas: the attentional role. We study the effects of quota implementation on the attention allocation strategy of a rationally inattentive (RI) manager. We find that quotas induce attention: a RI manager who is forced to fulfill a quota, unlike an unrestricted RI manager, never rejects minority candidates without acquiring information about them. We also demonstrate that, in our model, quotas are behaviorally equivalent to subsidies. We further analyze different goals that the social planner can achieve by implementing quotas. First, quotas can eliminate statistical discrimination, i.e., make chances of being hired independent from group identity. Second, when the hiring manager has inaccurate beliefs about the distribution of candidates' productivities, the social planner can make the manager behave as if she has correct beliefs. Finally, we show how our results can be used to set a quota level that increases the expected value of the chosen candidates.

In the second chapter, we study the information choice of exchange-traded funds (ETF) investors, and its impact on the price efficiency of underlying stocks. First, we show that the learning of stock-specific information happens at the ETF level. Further, our results suggest that ETF investors respond endogenously to changes in the fundamental value of underlying stocks, in line with the rational inattention theory. Second, we provide evidence that ETFs facilitate propagation of idiosyncratic shocks across its constituents.

In the third chapter, I study how an optimal menu chosen by a social planner depends on whether agents receive imperfect signals about her true taste (imperfect self-knowledge) or the properties of available alternatives (imperfect information). Under imperfect self-knowledge, it is not optimal to offer fewer alternatives than the number of different tastes present in the population, unless noise is infinite (agents have no clue about their true preferences). As noise increases, the social planner would offer menu items that are closer together (more similar), in the limit only offering one choice matching the mean preference in the population. However, under imperfect information, as noise increases, the social planner prefers to restrict the number of alternatives. Whether he makes them more or less similar is non-linear in noise.

Abstrakt

abstrakt

Acknowledgments

I am extremely thankful to my supervisor, Filip Matějka, for his patient guidance, support, and constructive criticism. I am also grateful to Jan Zápal, who was always available to discuss my research and encouraged me to aspire more. I would also like to express my gratitude to Avner Shaked and Ole Jann, who were supportive and helpful throughout.

I am indebted to many other inspiring people, especially to Daniel Martin for valuable discussions and help with organizing my research stay at Northwestern University. I am thankful to my co-authors, Andrei Matveenko and Mariia Kosar. I would like to thank all other academics who I met for numerous discussions of different parts of this thesis, as well as participants in various seminars, conferences, and workshops, in which I had opportunities to present my papers.

Furthermore, I also want to thank Andrea Downing, Grayson Krueger, and Deborah Novakova, who provided academic writing support and performed the language editing of this thesis. I am also grateful to the CERGE-EI community - faculty members, library, and staff for their assistance. I would also like to express my special thanks to Elizaveta Zasukhina for human support.

Lastly, and most importantly, I am thankful to my family, who accepted my choice of academic study, though they were uncertain about what I have been doing all these years. I especially thankful to my partner, Ekaterina Travova, who has made this journey feel less isolating.

Financial support from the Charles University project GA UK project No. 36119 and the ERC-2015-STG (H2020) project no. 678081 is gratefully acknowledged.

All errors remaining in this text are my responsibility.

Czech Republic, Prague Sergei
May 2022

Introduction

In information-rich world, where every day we face a choice, it is often difficult to make the best decision possible. The unifying theme of all three chapters of this dissertation centers on the idea that people make mistakes and that there are different ways to provide assistance and/or affect others. In the first two chapters, we apply the theory of rational inattention (RI), a disciplined model of how persons whose time and energy is limited choose what information to acquire. The key assumption is that there is generally more information available than a decision maker can pay attention to, though she can choose what information to focus on. In the first chapter, Andrei Matveenko and I investigate the influence the influence of a new affirmative action policy on the behavior of a rationally inattentive HR manager. In the second chapter, Maria Kosar and I study the information choices made by exchange-traded funds (ETF) investors, and its impact on the price efficiency of underlying stocks. the third chapter focuses on an environment in which people makes mistakes because they have imperfect self-knowledge or imperfect information. I study how optimal menus chosen by a social planner depend on the source of mistakes.

In the first chapter, we focus on labor market quotas, which have become a heavily-used governmental policy instrument in recent years. For example, in 2006, all publicly listed companies in Norway were required to increase female representation on their boards of directors to 40 percent. While there is a large body of literature that studies the effect of quota implementation on market outcomes, there is a lack of research that focuses on individual decision-making when an agent is forced to fulfill a quota. We address this gap in the research and introduces a new role of quotas: the attentional role. We find that quotas induce attention: RI manager who is forced to fulfill a quota, unlike an unrestricted RI manager, never rejects minority candidates, without acquiring information about them. We also demonstrate that, in our model, quotas are behaviorally equivalent to subsidies. In addition, we analyze different goals that the social planner can achieve by implementing quotas. First, quotas can help to eliminate statistical discrimination, i.e., make the chances of being hired independent of group identity. Second, when the hiring manager has inaccurate beliefs about the distribution of candidates' productivities, the social planner can spur the manager behave as if she has correct beliefs. Finally, we show how our results can be used to set a quota level that increases the expected value of the candidates chosen.

In the second chapter, we consider exchange-traded funds (ETFs), which have gained popularity among investors over the past decades, and have rapidly grown in terms of assets under management and trading volume. These instruments have attracted attention from both scholars and practitioners due to important asset pricing implications for their underlying securities. However, there is still a question whether ETFs can facilitate stock-specific price discovery, and what the net effect it has for the ETF underlying bundle. In this chapter we investigate this question. First, we show that the learning of stock-specific fundamental information can happen at the ETF level. More interestingly, our results suggest that ETF investors endogenously respond to changes in the fundamental value of underlying stocks, in line with the rational inattention theory. Second, we provide evidence that this pattern of learning affects the ETFs underlying bundles, leading to abnormal idiosyncratic volatility (AIV) co-movements across underlying stocks.

In the third chapter, I focus on situations when people choose from a discrete menu, for example, when choosing an insurance plan, school for our children, or pension fund. People often make mistakes in these important decisions, for two potential reasons. First, we misperceive the true properties of alternatives, i.e., we have imperfect information. Second, we misperceive our own tastes, i.e., we have imperfect self-knowledge. I study how choosing an optimal menu by a social planner depends on the source of agent's mistakes. Under imperfect self-knowledge, it is not optimal to offer fewer alternatives than the number of different tastes present in the population unless noise is infinite (meaning that agents have no clue about their true preferences). As noise increases, the social planner would offer menu items that are closer together (more similar), in the limit only offering one choice matching the mean preference in the population. However, under imperfect information, as noise increases, the social planner wants to restrict the number of alternatives. Whether he makes them more or less similar is non-linear in noise.

Chapter 1

Attentional Role of Quota Implementation

Co-authored with Andrei Matveenko (University of Copenhagen).

1.1 Introduction

Labor market quotas have become a heavily-used governmental policy instrument in recent years. For example, in 2006 all publicly listed companies in Norway were required to increase female representation on their boards of directors to 40 percent. Following Norway's lead, the European Union and several countries worldwide have passed similar reforms (Bertrand et al. 2019). While there is a large body of literature that studies the effect of quota implementation on market outcomes, there is a lack of research that focuses on individual decision-making when an agent is forced to fulfill a quota. This paper addresses the latter gap in the research and introduces a new role of quotas: the attentional role.

We consider the following setup. A human resources (HR) manager in a large firm is a rationally inattentive (RI) decision-maker. Each day she encounters a group of candidates. First, she sees a candidate's ethnicity and gender (or other observable characteristics), which forms her prior beliefs about the candidate's qualities and potential future productivity levels. Subsequently, the manager can acquire additional information about candidates – she can read resumes, ask questions, conduct tests, and use other learning strategies. The key feature of the process is that the information acquisition is flexible and endogenous – the manager does not have a fixed guide on how to learn about a potential worker's future productivity. At the same time, she has cognitive (and/or time) limitations. We model these limitations as costly information acquisition. Therefore, the manager faces a trade-off between acquiring more precise information about candidates and the cost of this information. After acquiring optimal information, the manager hires the candidate with the highest expected value for the firm.

We follow the setup introduced by Matějka and McKay (2015), in which the agent is uncertain about the values of available options. These values are modeled as an unknown draw from the known distribution. The agent has an opportunity to receive additional information about the realization of the draw in the manner that is optimal given the costs, which we model using the rational inattention framework introduced by Sims (2003). We think of the labor market candidates as available options and candidates' productivity levels as the values of options. Matějka and McKay (2015) show that the choice of a RI agent is typically stochastic and is characterized by the vectors of conditional and unconditional choice probabilities. In this paper, we explore the effect of quota implementation on the behavior of a RI manager. We model a quota as a constraint on the unconditional choice probability of choosing a candidate from a particular group. Due to the law of large numbers, such a limitation on unconditional choice probability is essentially a limitation on the share of workers from a particular group in the overall composition of workers in the firm.

We analyze the behavior of the RI manager when quotas restrict her choice, and compare it with an unrestricted case and the situation in which a social planner subsidizes the manager's choice of certain alternatives. We find that the choice probabilities of the manager in the constrained problem have the form of a generalized multinomial logit as in Matějka and McKay (2015) with an additional state independent component. In a choice among N candidates with the realized values $v(i|\omega)$ for $i \in \{1, ..., N\}$, our modified logit formula implies that the probability of choosing a candidate from group i is:

$$\mathcal{P}(i|\omega) = \frac{q_i e^{(v(i|\omega)-\varphi_i)/\lambda}}{\sum_{j=1}^{N} q_j e^{(v(j|\omega)-\varphi_j)/\lambda}},$$

where λ is the marginal cost of information, the q_i terms are quotas, and φ_i are state independent components. The form of choice probabilities shows that the manager behaves as if the value of the candidate is lower by φ_i. That is, the φ_i component induces an additive utility shifter in the decision-maker's preferences. Therefore, if a choice of a particular alternative is subsidized by $-\varphi_i$ then such a subsidy has exactly the same effect as the quota, which is the result we show in Section 1.3.3.

These adjustments to the logit model lead to the following change in the manager's behavior. If the choice problem is nontrivial, the RI manager who is forced to fulfill a quota always acquires information about candidates (Proposition 2). This feature is absent in the unconstrained problem, in which the manager has prior beliefs for which she decides not to acquire any additional information.

Further, we analyze the implications of choosing different vectors of quotas by the social planner. In Section 1.4 we prove that the social planner using quotas can induce choice probabilities which coincide with those of the unrestricted RI manager with any prior. This result has two immediate implications. Firstly, the social planner can always find a quota that makes the probability of a certain candidate being hired independent from the group identity, i.e. induces the fair (meritocratic) choice. Secondly, when the manager has inaccurate prior beliefs about the distribution of candidates' productivities, the social planner using quotas can induce the manager to behave as if she has correct beliefs.

Finally, using an example with candidates from two distinct social groups, we show a scenario when the social planner maximizes the expected value of the chosen candidates without taking into account the information costs. In general, the social planner benefits by forcing the manager to fulfill quotas, but for some priors the social planner prefers not to impose a quota.

Although we primarily focus on the effect of a quota in the labor market, the results of our analysis can be applied to studying individual behavior in other areas, e.g. a quota on the proportion of safe assets that must be in the portfolio of a financial manager or a quota on the number of orders a taxi driver can reject when searching for a client using peer-to-peer ride sharing applications (such as Uber, Lyft, or Yandex). We briefly discuss these applications in Section 1.6.

In the next section we review the related literature. Section 1.3 states the formal model of the manager's behavior with quotas and subsidies. Section 1.4 studies how the social planner can induce fair choice and correct wrong beliefs. Section 1.5 demonstrates the implications of the model using the binary example and discusses the value maximizing quota.

1.2 Literature

Our work contributes to the research on affirmative action and labor market discrimination. Affirmative action is "...any measure, beyond simple termination of a discriminatory practice, adopted to correct or compensate for past or present discrimination or to prevent discrimination from recurring in the future" (U. S. Commission on Civil Rights, 1977 p. 2). One of the most hotly debated types of affirmative action is the implementation of quotas. Coate and Loury (1993), in their famous paper, analyze a model of job assignment and show that quotas may lead to equilibria with persistent discrimination, due to feedback effects between expected job assignments and incentives to invest in human capital. Moro and Norman (2003) study the same problem in the general equilibrium setting and confirm the possibility that quotas can hurt the intended beneficiaries. These articles examine how affirmative action influences the behavior of the target group, and then its interaction with the behavior of the firm. In contrast, our study aims to investigate the individual decision-making process under quotas and consequences for policy design. A review of early studies on affirmative action can be found in Fang and Moro (2010).

To date there is mixed empirical evidence of the effect of quotas on the quality of workers and firm's revenue. For example, Ahern and Dittmar (2012) show that firm value declined with a law mandating 40% representation of each gender on the board of public limited liability companies in Norway. They also show that the average age and experience of the new female directors were significantly lower than that of the existing male directors and argue that this change led to a deterioration in operating performance. At the same time, Eckbo, Nygaard, and Thorburn (2019), using econometric adjustments and a larger data set, argue that the effect of implementation of quotas on both the value of firms and on the quality of directors was insignificant. Bertrand et al. (2019), exploiting the same intervention, document that a quota resulted in significant improvement of the average observable qualifications of the women appointed to the boards and a decrease in the gender gap in earnings within boards. Besley et al. (2017), using Swedish data on the performance of politicians, show that a gender quota on the ballot increased the competence of male politicians. Ibanez and Riener (2018) use data from three field experiments in Colombia and show that the gains from attracting female applicants far outweigh the losses from deterring male applicants. Our paper proposes a mechanism that can possibly explain why the evidence on consequences of quotas is mixed.

Our study fits into the rational inattention literature, which originated in studies by Sims (2003). As a benchmark, we use the modified multinomial logit model of (Matějka and McKay 2015), in which agents choose between alternatives without precise information about their values, but with an opportunity to study the options for some cost[1]. We analyze this model with an additional constraint on the unconditional probabilities of the choice of a certain alternative. Lindbeck and Weibull (2020) analyze investment decisions with delegation to a RI agent. They find that optimal contracts for an agent include a high reward for good investments and punishment for bad investments. Lipnowski, Mathevet, and Wei (2020) study a model where a principal provides information to a RI agent, but does not internalize the agent's cost of processing information. They show that if there are more than two alternatives, a principal can improve material benefits from the choice by manipulating information. We analyze a similar principal-agent problem in Section 1.5.2, but with a different mechanism. We show that, for a set of parameters, a principal can force the manager to acquire more information by defining the level of quotas, and thereby increase the expected value of the manager's choice.

Fosgerau, Sethi, and Weibull (2020) characterize equilibrium with a RI employer and candidates who choose how much effort to invest before being screened and use it to examine categorical inequality, including statistical discrimination, prejudice, and social capital. Acharya and Wee (2020) consider a search and matching model with RI firms and find that during recessions firms become more selective. The results of our paper are complementary to these studies and can be used to investigate how affirmative action influences equilibrium outcomes.

Bartoš et al. (2016), in a field experiment, show that HR managers and landlords allocate their attention to job and rental applicants in line with the rational inattention theory. For example, a non-European name or recent unemployment induces the HR manager to read a job application and CV in less detail, which negatively affects the probability of the applicant being invited for a job interview. The results of our study predict the attention allocation of decision-makers, such as HR managers, in the presence of quotas, i.e. whether they would blindly choose the quoted option or whether a quota would lead to greater information acquisition about the target group. Thus, the results of this study can provide a starting point for the empirical investigation of the effect of a quota on attention allocation. A detailed review of the RI literature can be found in

[1]See also Caplin and Dean (2015) for an alternative method for characterizing solutions in a similar environment.

Maćkowiak, Matějka, and Wiederholt (2021).

Our study also relates to the discussion on whether directly administering an activity is better than fixing transfer prices and relying on utility maximization to achieve the same results in a decentralized fashion (Weitzman 1974). We contribute to the discussion on this issue by comparing the behaviour of a manager operating under quotas and a manager whose choice is subsidized by a social planner.

1.3 The model

This section begins with a benchmark model – we describe the standard RI problem as in Matějka and McKay (2015) and Caplin, Dean, and Leahy (2019) and its implications. Then, we state our problem with quotas and discuss the properties of the solution. Finally, we analyze the RI problem with subsidies.

The HR manager faces N candidates and wants to select the candidate with the highest value for the employer. We refer to a candidate i as a candidate from a category i. There are finitely many states of the world Ω, with $\omega \in \Omega$ denoting a generic state. The values of the candidates differ from state to state, $v(i|\omega) \in \mathbb{R}$ is the value of the candidate from category $i \in \{1, ..., N\}$ in the state $\omega \in \Omega$. The decision-maker is uncertain about the realization of the state of the world. However, she knows a distribution of possible states of the world – this prior knowledge is described by a distribution $\mu \in \Delta(\Omega)$, where $\Delta(\Omega)$ denotes the set of all probability distributions over Ω, we assume that $\mu(\omega) > 0 \ \forall \omega \in \Omega$. She can refine her knowledge by processing costly information about the realization. Information processing results in a stochastic (possibly not purely stochastic) choice and, at the optimum, the decision problem can be treated as a problem of choosing conditional choice probabilities rather than the choice of information structure (Corollary 1 in Matějka and McKay 2015). We denote the conditional probability of a candidate i being selected when the realized state is ω as $\mathcal{P}(i|\omega)$.

1.3.1 Standard RI problem

The standard RI manager's problem is formalized as follows.

Standard (unconstrained) RI problem. The manager's problem is to find a vector function of conditional choice probabilities $\mathcal{P}^U = \{\mathcal{P}^U(i|\omega)\}_{i,\omega}, \ \ i \in \{1, ..., N\}, \ \ \omega \in \Omega$, (the superscript "U" stands for "unrestricted") that maximizes expected payoff less the information cost:

$$\max_{\{\mathcal{P}^U(i|\omega)\}_{i=1}^N} \left\{ \sum_{i=1}^{N} \sum_{\omega \in \Omega} v(i|\omega)\mathcal{P}^U(i|\omega)\mu(\omega) - \lambda\kappa(\mathcal{P}^U) \right\}$$

subject to

$$\forall i \in \{1, ..., N\}, \quad \forall \omega \in \Omega : \quad \mathcal{P}^U(i|\omega) \geq 0, \tag{1.1}$$

$$\forall \omega \in \Omega : \quad \sum_{i=1}^{N} \mathcal{P}^U(i|\omega) = 1, \tag{1.2}$$

where unconditional choice probabilities are

$$\mathcal{P}^U(i) = \sum_{\omega \in \Omega} \mathcal{P}^U(i|\omega)\mu(\omega), \quad i \in \{1, ..., N\}. \tag{1.3}$$

The cost of information is $\lambda\kappa(\mathcal{P}^U)$, where $\lambda \in (0, +\infty)$ is a given unit cost of information and κ is the amount of information that the manager processes, which is measured by the expected reduction in the entropy (Shannon 1948, Cover and Thomas 2012):

$$\kappa(\mathcal{P}^U) = -\sum_{i=1}^{N} \mathcal{P}^U(i) \log \mathcal{P}^U(i) + \sum_{i=1}^{N} \sum_{\omega \in \Omega} \mathcal{P}^U(i|\omega) \log \mathcal{P}^U(i|\omega)\mu(\omega). \tag{1.4}$$

The entropy shape of the information cost is common in the literature on rational inattention. Its use has been justified both axiomatically and through links to optimal coding in information theory (see Sims (2003), Matějka and McKay (2015), and Denti, Marinacci, and Montrucchio (2019) for discussions).

Matějka and McKay (2015) show that, at the optimum, the conditional probabilities of choosing a candidate $i \in \{1, ..., N\}$ follow the generalized logit form.

Theorem 1 (Matějka and McKay 2015)**.** *Conditional on the realized state of the world $\omega \in \Omega$ the optimal choice probabilities satisfy:*

$$\mathcal{P}^U(i|\omega) = \frac{\mathcal{P}^U(i)e^{v(i|\omega)/\lambda}}{\sum_{j=1}^{N} \mathcal{P}^U(j)e^{v(j|\omega)/\lambda}}.$$

The shape of conditional choice probabilities is similar to a multinomial logit, but is weighted with the coefficients $\mathcal{P}^U(i)$, which are endogenous to the decision problem and represent the probability of selecting a candidate i before the manager starts processing any information. These adjustments to the logit model reflect the fact that some candidates may, a priori, look better than others.

An important property of the solution is that parameters of the model may exist for which the manager decides not to acquire any information, and instead makes her decision based solely on her prior knowledge. In this situation, she simply chooses the candidate with the highest a priori expected value.[2] In terms of the labor market, this means that some categories of workers may not be given any attention and are consequently not hired. As we show in the next section, the social planner can force the manager to receive at least some information about the candidates – this can be achieved via the use of quotas.

1.3.2 Quotas

We consider a departure from the standard RI problem when the manager is not completely free in her choice. Instead, some authority limits her choice in that, for all categories, the share of the candidates hired from a category i should be equal to $q_i \in (0,1)$.[3]

RI problem with quotas. The manager's problem is to find a vector function of conditional choice probabilities $\mathcal{P} = \{\mathcal{P}(i|\omega)\}_{i,\omega}, \quad i \in \{1,...,N\}, \quad \omega \in \Omega$, that maximizes expected payoff less the information cost:

$$\max_{\{\mathcal{P}(i|\omega)\}_{i=1}^{N}} \left\{ \sum_{i=1}^{N} \sum_{\omega\in\Omega} v(i|\omega)\mathcal{P}(i|\omega)\mu(\omega) - \lambda\kappa(\mathcal{P}) \right\}$$

subject to (1.1)-(1.4) and

$$\forall i \in \{1,...,N\}: \quad \mathcal{P}(i) = \sum_{\omega\in\Omega} \mathcal{P}(i|\omega)\mu(\omega) = q_i, \quad q_i \in (0,1), \tag{1.5}$$

where $\mathbf{q} = (q_1, ..., q_N)^T$ is the vector of quotas and

$$\sum_{i=1}^{N} q_i = 1.$$

[2]See Caplin and Martin (2017), who analyze a discrete choice problem of a RI agent with costly information acquisition, and show that if there is a high quality default option the manager chooses zero attentional effort.

[3]We focus on the case with binding quotas for all alternatives, since (a) if the quotas have a form of weak inequality, then, if it is not binding, the solution is the same as the solution to the unconstrained problem; and (b) the case with quotas only for some categories is considered in Appendix 1.D where we show that results are similar. In addition, we restrict the quota vector to be interior because if some of its components are 0, the problem boils down to the situation in which the choice is limited and the solution coincides with the solution to the standard RI problem with restricted menu. When the quota equals 1 for a particular category, the solution is trivial – the manager does not acquire any information.

Choice probabilities at the optimum follow:

$$\mathcal{P}(i|\omega) = \frac{q_i e^{(v(i|\omega)-\varphi_i)/\lambda}}{\sum_{j=1}^{N} q_j e^{(v(j|\omega)-\varphi_j)/\lambda}}. \tag{1.6}$$

This result is formalized in the following proposition:

Proposition 1. *Choice probabilities that are the solution to the RI manager problem with quotas are of a generalized logit form: logit choice probabilities with an additive state-independent component. Conditional choice probabilities* $\{\mathcal{P}(i|\omega)\}_{i,\omega}$, $i \in \{1, ..., N\}$, $\omega \in \Omega$, *are the solution to the RI problem with quotas if they satisfy (1.5) and (1.6). Moreover, the solution to the RI problem with quotas is unique.*

Proof. See Appendix 1.A.1. □

The terms φ_i are the Lagrange multipliers on the constraints on unconditional choice probabilities. In choice probabilities they play the role of utility shifters, that is, the manager's behavior follows the logit rule, but with utilities which are changed by some value that depends on the marginal cost of information, prior beliefs and the value of a quota. In Section 1.3.3 we relate $\boldsymbol{\varphi}$ to subsidies which are required to be paid to the manager when she hires workers from a certain category.

Proposition 1 states that the solutions to the standard RI problem and the RI problem with quotas have a similar form. However, there is a crucial difference in the information acquisition strategies of a RI manager with and without quotas. We express this in the following proposition:

Proposition 2. *If* $v(i|\omega) - v(j|\omega)$ *is not constant across states* $\forall i, j \in \{1, ..., N\}$ *and* $i \neq j$, *then the RI manager with quotas always acquires information.*

Proof. See Appendix 1.A.2. □

Proposition 2 states that the employer will never blindly choose a candidate from a certain group based only on her prior beliefs, but will acquire some information. The intuition for this result comes from the fact that a quota forces the manager to choose not only between candidates from different categories, but also between candidates from the same category, which brings incentives for information acquisition. According to the condition in Proposition 2, there is a beneficial deviation from the state-independent probabilities. Therefore, the manager can benefit from acquiring at least some information in order to improve her choice.

The assumption that the cost of information is proportional to the expected reduction in entropy is not crucial for Proposition 2 to hold. In particular, in Appendix 1.A.2 we show that it holds for any posterior separable attention cost function (Caplin, Dean, and Leahy 2021), with an extra assumption that the cost of marginal change in choice probabilities is zero at the point of initial uncertainty. Also, let us note that in the case which seems to be important in many applications, namely, when there is a fixed cost to acquire any positive amount of information $\underline{k} > 0$, Proposition 2 does not hold. For example, the manager chooses not to acquire any information even when restricted by quotas if $v(i|\omega) - v(j|\omega) < \underline{k}$ for all $i, j \in \{1, ..., N\}$, $\omega \in \Omega$.

1.3.3 Subsidies

We are interested in understanding how a manager's attention strategy depends on the particular form of affirmative action chosen by the government. One form of affirmative action, and an alternative to quotas, is an employment subsidy policy, under which a firm receives subsidies if it employs certain categories of workers (for survey see, e.g. Card, Kluve, and Weber 2017). In this situation, the government does not have access to the values of specific candidates for a firm and introduces the same subsidy for all possible realizations of the values of candidates hired from a particular category.

Such a policy in our setting changes the values of candidates from $v(i|\omega)$ to $v(i|\omega) + S_i, \quad \forall \omega \in \Omega$, where S_i is a subsidy for choosing a candidate from category i. Let us begin the analysis with the definition of the RI problem with subsidies. If the government introduces an employment subsidy policy, then the manager solves the following problem:

RI problem with subsidies. The manager's problem is to find a vector function of conditional choice probabilities $\mathcal{P}^S = \{\mathcal{P}^S(i|\omega)\}_{i,\omega}, \quad i \in \{1, ..., N\}, \quad \omega \in \Omega$, (the superscript "S" stands for "subsidy") that maximizes expected payoff less the information cost:

$$\max_{\{\mathcal{P}^S(i|\omega)\}_{i=1}^N} \left\{ \sum_{i=1}^N \sum_{\omega \in \Omega} (v(i|\omega) + S_i) \mathcal{P}^S(i|\omega) \mu(\omega) - \lambda \kappa(\mathcal{P}^S) \right\},$$

subject to (1.1)-(1.4) and where S_i is a subsidy for choosing a candidate i.

In this case, the solution to the manager's problem follows the standard modified generalized multinomial logit formula, but with the changed value of the candidate i by

S_i:

$$\mathcal{P}^S(i|\omega) = \frac{\mathcal{P}^S(i) e^{(v(i|\omega)+S_i)/\lambda}}{\sum_{j=1}^N \mathcal{P}^S(j) e^{(v(j|\omega)+S_j)/\lambda}}. \tag{1.7}$$

And unconditional choice probabilities are

$$\mathcal{P}^S(i) = \sum_{\omega\in\Omega} \mathcal{P}^S(i|\omega)\mu(\omega).$$

Equation (1.7) provides intuition about the nature of the additive component $\boldsymbol{\varphi}$ from the solution to the RI problem with quotas. This component can be interpreted as the government subsidies that are needed to be added to the values of candidates in order to induce the RI manager to choose them with the required unconditional probabilities.

In the following proposition we establish formally that for any quotas $\mathbf{q}$ there exists a subsidies vector $\mathbf{S}$, such that the optimal behavior of a manager who is facing quotas $\mathbf{q}$ is also optimal when she is facing subsidies $\mathbf{S}$.

Proposition 3. *For any vector of quotas* $\mathbf{q}$ *there exists a vector of subsidies* $\mathbf{S}$, *such that* $\mathcal{P}^S(i) \equiv q_i$, $i \in \{1, ..., N\}$, *are optimal unconditional choice probabilities in the RI manager problem with subsidies* $\mathbf{S}$. *Moreover, optimal conditional choice probabilities in the RI manager problem with quotas* $\mathbf{q}$ *are also optimal in the problem with subsidies* $\mathbf{S}$.

Proof. See Appendix 1.A.3. □

Proposition 3 establishes the existence of subsidies such that for a given vector of quotas there is a common optimum across the problems with subsidies and quotas. However, the problem with subsidies might have several solutions and the desired quota might not be achieved if the social planner chooses a subsidy policy. At the same time, such a situation is possible only in extreme cases with a very specific value co-movement pattern. Proposition 4 states the condition that is sufficient for behavioral equivalence of quotas and subsidies[4].

Proposition 4. *If for any vector* $\{\zeta_i\}_{i=1}^N$ *the values* $\{e^{(v(i|\cdot)+\zeta_i)/\lambda}\}_{i=1}^N$ *are affinely independent in* $\mathbb{R}^\Omega$, *i.e. there does not exist a set* $\{a_j\}_{j\neq i}$ *such that* $\sum_{j\neq i} a_j = 1$ *and*

$$\forall\omega\in\Omega: \quad e^{(v(i|\omega)+\zeta_i)/\lambda} = \sum_{j\neq i} a_j e^{(v(j|\omega)+\zeta_j)/\lambda}, \tag{1.8}$$

[4] In Appendix 1.C we provide a solution for a binary example with subsidies. We show that while the behavior of the manager under quotas and subsidies is the same, the utility of the manager is different.

then quotas and subsidies are behaviorally equivalent, i.e. for any $\mathbf{q}$ *there exists* $\mathbf{S}$*, such that, in the problem with subsidies* $\mathbf{S}$*,* $\mathbf{q}$ *is the unique vector of optimal unconditional choice probabilities and optimal conditional choice probabilities in problems with quotas and subsidies coincide.*

Proof. See Appendix 1.A.4. □

Let us note, that condition (1.8) is a stronger version of the affine independence condition (see Matějka and McKay 2015, Caplin and Dean 2013). For example, in a binary choice case affine independence requires that the values of the candidates are not the same in all states of the world, while condition (1.8) requires that the values do not differ by a state-independent constant in all states of the world. In addition, the results of Propositions 3 and 4 are much more sensitive to the functional form of information cost than the results of Proposition 2.

1.4 Effect of quotas on conditional choice probabilities

So far we have analyzed the manager's problem for a given level of quota. In this section, we analyze the features of conditional probabilities which the social planner can induce by quotas. We start by stating a technical lemma which we apply later in the section.

Lemma 1. *For any vector of weighting coefficients* $\boldsymbol{\beta} \in [0,1]^N$*, such that* $\sum_{i=1}^N \beta_i = 1$*, there exists a vector of quotas* $\mathbf{q} \in [0,1]^N$*, such that* $\sum_{i=1}^N q_i = 1$*, which induces the following choice probabilities as a solution to the RI problem with quotas* $\mathbf{q}$

$$\forall i \in \{1, ..., N\}, \quad \omega \in \Omega: \quad \mathcal{P}(i|\omega) = \frac{\beta_i e^{v(i|\omega)/\lambda}}{\sum_{j=1}^N \beta_j e^{v(j|\omega)/\lambda}}. \tag{1.9}$$

Proof. See Appendix 1.A.5. □

Lemma 1 establishes an important feature of quotas: for any distribution of states of the world the social planner can implement a vector of quotas such that the manager's behavior replicates the behavior of the unconstrained RI manager with any prior beliefs. Now we consider two situations in which implementing such a quota is optimal.

1.4.1 Fair quota

The original goal of affirmative action was, and for many policies still is, to ensure that candidates are "treated [fairly] during employment, without regard to their race, creed, color, or national origin" (John F. Kennedy 1961). In terms of our model this goal is to ensure that the probability of a certain worker being hired does not depend on group identity and depends only on the job-relevant characteristics of candidates. In this section, we show that there exists a quota that achieves this goal.

Formally, the solution to the standard RI maximization problem is

$$\forall i \in \{1, ..., N\}, \quad \omega \in \Omega : \quad \mathcal{P}^U(i|\omega) = \frac{\mathcal{P}^U(i) e^{v(i|\omega)/\lambda}}{\sum_{j=1}^{N} \mathcal{P}^U(j) e^{v(j|\omega)/\lambda}}, \tag{1.10}$$

where $\mathcal{P}^U(i)$ corresponds to the bias towards hiring a candidate from group i. Such choice probabilities may result in a situation in which $v(i|\omega) = v(j|\omega)$ for some $i \neq j$, $\omega \in \Omega$, but $\mathcal{P}^U(i|\omega) > \mathcal{P}^U(j|\omega)$ just because a candidate i comes from a group that seems a priori better.

The goal of the social planner is to make choice probabilities depend on the realized values only, i.e. eliminate $\mathcal{P}^U(i)$ on the right-hand side in equation (1.10). The affirmative action policy of such a social planner results in equal chances of being hired for candidates with the same productivity.

One policy that aspires to achieve this goal is a blind or de-identified approach to reviewing candidates (Goldin and Rouse 2000). In practice, however, it is not always possible to fully hide the group identifier, since it can be approximated by other information, which can lead to lower quality of the choice (see, for example, Ray and Sethi (2010) and Antonovics and Backes (2014) on the adverse effects of color-blind affirmative action). Moreover, such policies can have unexpected results and lower the representation of the discriminated group if there exists positive discrimination (Hiscox et al. 2017).

Proposition 5 states that the social planner can choose such a quota, which we call *fair,* that makes conditional choice probabilities dependent on candidates' job-relevant characteristics only. Therefore, the resulting choice outcomes with a fair quota are as if there is a blind or de-identified policy. At the same time, by using a fair quota the social planner can eliminate the bias, while keeping the information about the group identity.

Proposition 5. *There exists a vector of quotas* $\mathbf{q}$ *and a function* $p : \mathbb{R} \times \mathbb{R}^{N-1} \to (0, 1)$ *such that* $p(v, \mathbf{v})$ *is increasing in* v *and optimal conditional choice probabilities in the RI* *manager problem with quotas* $\mathbf{q}$ *do not depend on the candidate's identity* $i \in \{1, ..., N\}$ *but only on the candidates' values:*

$$\forall i \in \{1, ..., N\}, \quad \forall \omega \in \Omega: \quad \mathcal{P}(i|\omega) = p(v(i|\omega), \mathbf{v}(\omega)),$$

where $\mathbf{v}(\omega) = \{v(j|\omega)\}_{j \neq i}$.

Proof. According to Lemma 1, there is a vector of quotas which induces the following choice probabilities

$$\forall i \in \{1, ..., N\}, \quad \omega \in \Omega: \quad \mathcal{P}(i|\omega) = \frac{e^{v(i|\omega)/\lambda}}{\sum_{j=1}^{N} e^{v(j|\omega)/\lambda}}. \tag{1.11}$$

Choice probabilities (1.11) depend on the candidates' values only. □

1.4.2 Inaccurate priors

Usually the literature highlights two types of discrimination: taste-based discrimination (Becker 1957) and statistical discrimination (Phelps 1972). Without any additional assumptions, the rational inattention theory allows us to model the latter type, which is typically assumed to be driven by limited information about the particular candidate's characteristics and correct beliefs about the group distributions of the relevant outcome. At the same time, people often have inaccurate beliefs about the performance distributions of particular groups (Bohren et al. 2020). Such misperceptions can not only increase discrimination, but are also harmful for the manager, since she would acquire information in a sub-optimal way and choose candidates with lower expected productivity. In this section, we show that the social planner, who knows the correct distribution of productivities within and between different groups, can implement a quota that induces the manager to behave as if she also has correct beliefs.

In terms of our model, misperception means that the manager has a wrong prior: she believes that the distribution of possible states of the world is $\mu^W \in \Delta(\Omega)$, (the superscript "W" stands for "wrong") while in reality this distribution is $\mu \in \Delta(\Omega)$. Such an incorrect belief, if the choice of the manager is unconstrained, results in conditional choice probabilities $\mathcal{P}^W(i|\omega)$. The optimal choice probabilities of a manager with a correct belief are $\mathcal{P}^U(i|\omega)$.

The social planner introduces a vector of quotas $\mathbf{q}$. The quota enters the manager's maximization problem as a binding constraint. The manager chooses the information acquisition strategy given her beliefs and quota, and commits to a choice rule which, in the manager's opinion, will result in the required unconditional choice probabilities. However, since the manager has incorrect beliefs, such an information acquisition strategy results in choice probabilities which are different from the intended, and the constraint is not actually satisfied. That is, the vector of quotas corresponds not to actual choice probabilities but to the manager's belief about them. At the same time, the social planner, who knows the true distribution and wrong belief of the manager, can easily verify whether the hiring strategy has been adjusted or not. Over time, the manager, of course, realizes that the choice probabilities do not satisfy the constraint and updates her belief. We focus on the short term effect of the quota and do not model the process of belief updating.

We consider the following possible goal of the social planner – to "de-bias" the hiring manager. That is, the social planner wants to make conditional choice probabilities appear as if the manager has correct beliefs. More formally, to make $\mathcal{P}(i|\omega) = \mathcal{P}^U(i|\omega)$ for all $i \in \{1, ..., N\}$, $\omega \in \Omega$. Proposition 6 states that the social planner can choose a quota that achieves this goal. It is important to note that a quota is not simply $q_i = \mathcal{P}^U(i), \ \ \forall i \in \{1, ..., N\}$, but is adjusted to the prior of the manager.

Proposition 6. *For any misperceived distribution of possible states of the world* $\mu^W \in (0,1)$ *there exists such* $\mathbf{q}$ *which induces*

$$\forall i \in \{1, ..., N\}, \ \ \forall \omega \in \Omega : \quad \mathcal{P}(i|\omega) = \frac{\mathcal{P}^U(i) e^{v(i|\omega)/\lambda}}{\sum_{j=1}^N \mathcal{P}^U(j) e^{v(j|\omega)/\lambda}} = \mathcal{P}^U(i|\omega),$$

where $\{\mathcal{P}^U(i|\omega)\}_{i,\omega}$ *are conditional choice probabilities of a manager with the correct prior belief.*

1.5 Example with two groups of candidates

In this section, in order to illustrate the logic of the model, we consider a simple example with two groups of candidates. We then use this setup and analyze the situation in which the social planner wants to maximize the expected value of the chosen candidates without taking into account the information costs of the manager.

1.5.1 Setup

The manager chooses between two candidates from two different social groups, $i \in \{1,2\}$. There are two states of the world, $\omega \in \{1,2\}$. For simplicity, the type 1 candidate is the safe choice that always has the constant value $v(1) = v(1|1) = v(1|2) = C$. The type 2 candidate is the risky choice, which can take values $v(2|1) = 0$ with the probability b and $v(2|2) = 1$ with the probability $1-b$. In a labor market context, we can assume that the share b of workers from a category 2 has low productivity, while the share $1-b$ has high productivity. These probabilities (or shares) are prior of the manager and she does not know what the realization of the state of the world is. The manager has an opportunity to acquire some costly information about the realization.

The manager's choice is restricted in that, on average, the share q of hired candidates should be type 2 (risky) and share $1-q$ of chosen candidates should be type 1 (safe). In terms of our model the manager has restrictions on unconditional choice probabilities.

To solve the problem we must find conditional probabilities $\mathcal{P}(i|\omega)$. We show in Appendix 1.B that the solution is

$$\mathcal{P}(2|0) = \frac{-b-q+(b+q-1)e^{\frac{1}{\lambda}} + \sqrt{(b+q-(b+q-1)e^{\frac{1}{\lambda}})^2 + 4q(be^{\frac{1}{\lambda}}-b)}}{2(be^{\frac{1}{\lambda}}-b)},$$

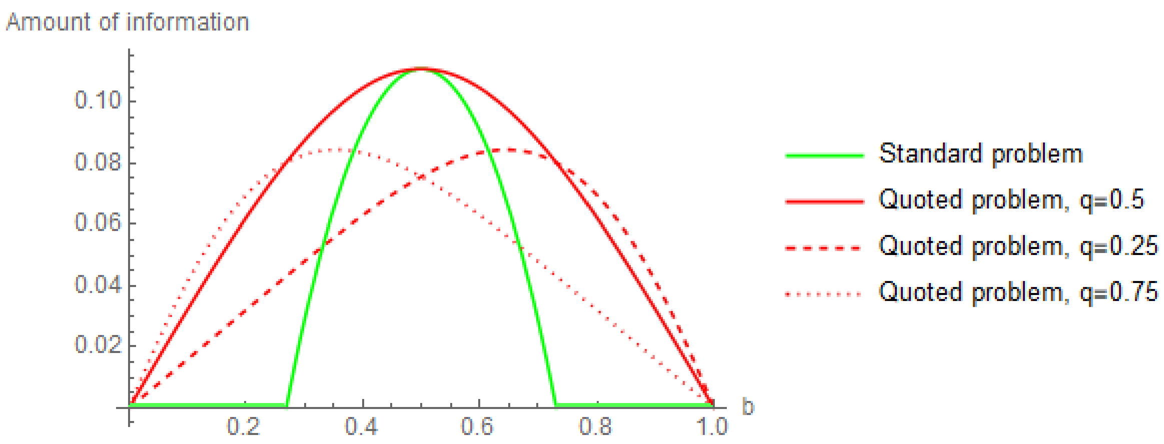
$$\mathcal{P}(2|1) = \frac{q-b\mathcal{P}(2|0)}{1-b}.$$

It is worth noticing that in the formulas above conditional choice probabilities $\mathcal{P}(2|0)$ and $\mathcal{P}(2|1)$ do not depend on the value of the safe candidate C, unlike in the unconstrained RI problem. There are two intuitive explanations for this. First, it follows from Proposition 3 that the required unconditional choice probabilities can be achieved by subsidies. In this example it is enough to subsidize (or put a fine on) candidates from the safe group, making its value for the firm equal to some $\tilde{C}(q)$, which, of course, does not depend on the initial C. The second intuitive reason is as follows. Whereas an unrestricted manager is interested in the relative payoffs of hiring candidates from different groups in the same state, a quota-bounded manager compares the relative payoffs of hiring candidates from the same group in different states.

For a given set of parameters, Figure 1.1 shows the expected reduction in entropy as a function of b. In the standard RI problem, when b is close to 0 or 1 the manager decides not to process information and selects one of the candidates with certainty. However, when the manager is forced to fulfill the quota, she always acquires information, and hence there are no non-learning areas. For example, when b is close to 1, she is forced to choose a risky candidate with positive probability, and it is profitable to acquire information in order to choose the risky candidate with high value rather than pick a random candidate from risky group.

At the same time, the quota-bounded manager can acquire less information than in the standard RI problem (Figure 1.1). Accordingly, the effect of the quota on the amount of acquired information is ambiguous.

Figure 1.1: Amount of information as a function of b and $\lambda = 0.5$, $C = 0.5$. The green curve is for the standard RI problem and the red curves are for the quoted RI problem: the solid curve is for $q = 0.5$, the dotted curve for $q = 0.75$ and the dashed curve for $q = 0.25$.

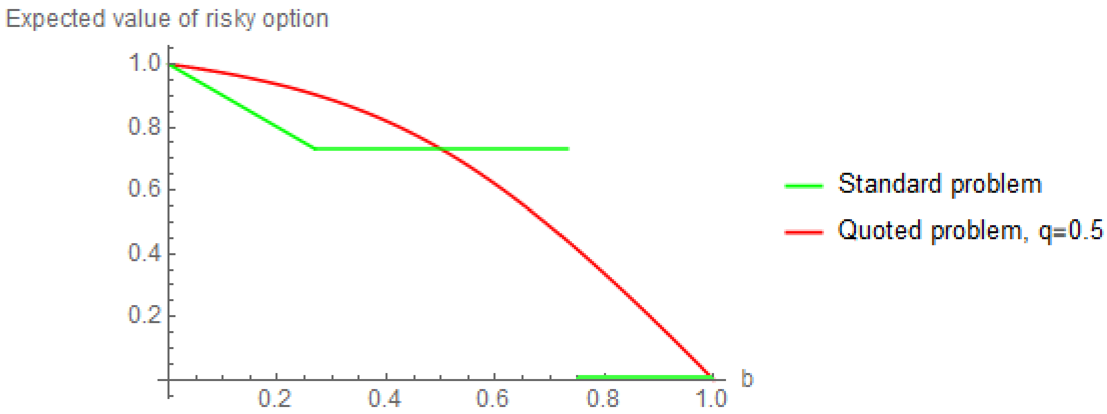


We now explore how the quota affects the expected value of the chosen candidates. In terms of the labor market this question can be restated in the following way: does a quota necessarily mean that the expected value (or productivity) of the hired workers will fall? The definition of the expected value of the chosen risky candidate can be found below.

Definition 1. The expected value of the chosen risky candidate is $\frac{(1-b)\mathcal{P}(2|1)}{\mathcal{P}(2)}$. This is the ratio of the probability of the chosen risky candidate being of high value to the probability of choosing any risky candidate.

Figure 1.2 illustrates that the expected value of the chosen risky candidate is higher (lower) when the quota on it is smaller (larger) than the unconditional probability of choosing it in the standard RI problem.

Figure 1.2: The expected value of the risky candidate conditional on being chosen as a function of b and $\lambda = 0.5$, $q = 0.5$, $C = 0.5$. The green curve is for the standard RI problem and the red curve is for the quoted RI problem.



1.5.2 Value maximizing quota

We follow Lipnowski, Mathevet, and Wei (2020) and consider the following problem. The social planner maximizes the expected value of the chosen candidate and does not take into account the cost of information. This is reasonable, for instance, if positive production externalities exist. However, while the manager is making hiring decisions she may not take these externalities into account. Thus, the maximization problem of the manager and the social planner (organization, industry as a whole, or government) may differ. However, in contrast to Lipnowski, Mathevet, and Wei (2020), in our case the social planner chooses the optimal quotas $\mathbf{q}$ and not the information structure available to the manager. The maximization problem is

$$\max_{\mathbf{q}} \left\{ \sum_{i=1}^{N} \sum_{\omega \in \Omega} v(i|\omega) \mathcal{P}(i|\omega, \mathbf{q}) \mu(\omega) \right\},$$

where $\mathcal{P}(i|\omega, \mathbf{q})$ is a solution to the RI problem with the quota $\mathbf{q}$.

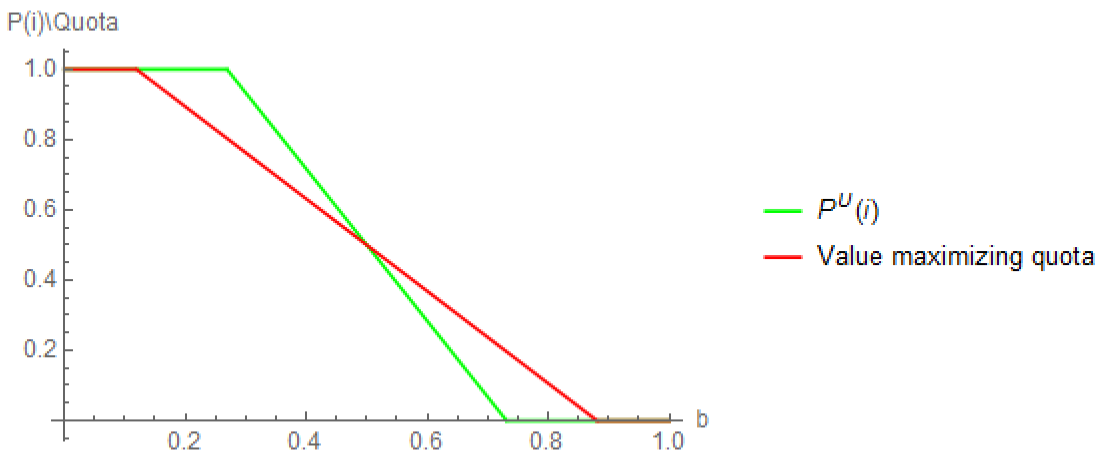
We refer to the solution of this problem as a *Value maximizing quota.* In contrast to information filtering proposed by Lipnowski, Mathevet, and Wei (2020), a quota allows the social planner to increase the expected value of the chosen candidate even when the state is binary.

Figure 1.3 illustrates the solution to the social planner's problem as a function of b, and $\lambda = 0.5$, $C = 0.5$. When the social planner maximizes the expected value of chosen alternatives, there are still non-learning areas, but they are smaller than in the standard

RI problem. The reason for the presence of the non-learning areas is as follows. Consider the situation in which b is small, i.e. the probability of the risky candidate being good is high and the manager chooses him with certainty without acquiring any information. When the non-trivial quota is implemented, the manager will acquire some information in order to find out whether the risky candidate is good or bad, but the improvement in the expected value of chosen risky candidates would not compensate for the loss of good risky candidates that were mistakenly rejected. Therefore, the social planner prefers not to constrain the manager, or, in other words, he prefers to implement the quota that would force the manager to always choose a risky candidate – the same action that the manager would take without any constraints. Similar logic applies when the probability of the bad state is high.

Outside these non-learning areas, the social planner induces the manager to acquire more information than in the standard RI problem (Figure 1.5), and hence increases the expected value of the chosen candidate (Figure 1.4). Thus, in this example, it is optimal to establish a quota that is higher (lower) than the unconditional probability in the standard RI problem when the state is more likely to be bad (good).[5]

Figure 1.3: Optimal quota as a function of b and $\lambda = 0.5$, $C = 0.5$. The green curve is for the standard RI problem and the red curve is for the quoted RI problem.



[5]In our working paper we provide an example when the social planner has limited information about the distribution of alternatives' values. In this situation, setting a quota may decrease the expected value of the chosen candidates.

Figure 1.4: The expected value of the chosen candidates as a function of b and $\lambda = 0.5$, $C = 0.5$. The green curve is for the standard RI problem and the red curve is for the quoted RI problem.

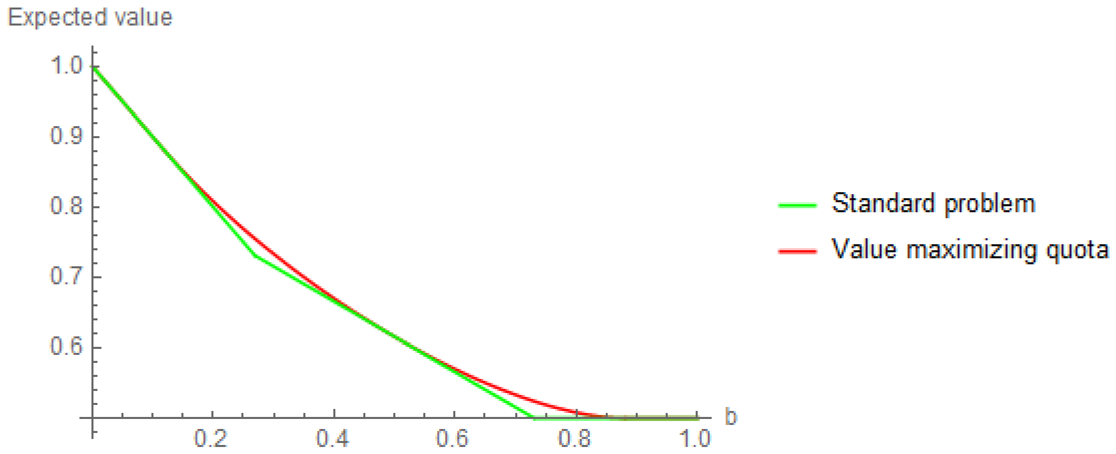
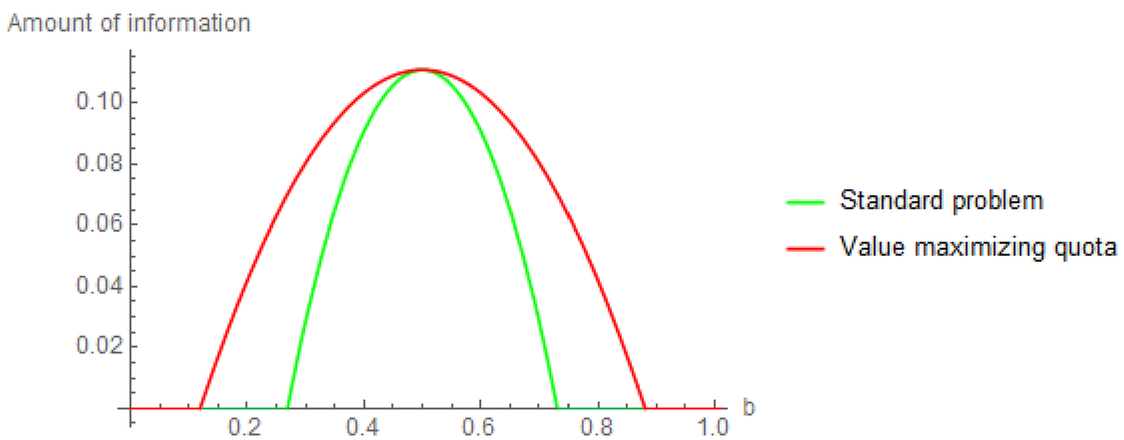


Figure 1.5: Amount of information as a function of b and $\lambda = 0.5$, $C = 0.5$. The green curve is for the standard RI problem and the red curve is for the quoted RI problem.



1.6 Conclusion

In this paper, we study the optimal behavior of a RI manager who is forced to fulfill quotas when making a choice from a discrete menu. Whilst throughout the paper we have used labor market settings to illustrate how quotas can influence the manager's attention allocation, the proposed model can also be used to analyze the effect of quotas in other areas. For example, the model in Section 1.5.1 can be applied to analyze financial interventions. Consider a situation in which the agent manages a financial portfolio and chooses between risky and safe assets. The financial regulator wants to increase the proportion of safe assets in the agent's portfolio. Thus, the regulator imposes a quota. This intervention can lead to higher screening efforts by the agent, and hence to the more profitable and diversified portfolio.

The model can also be applied to analyze the effect of quotas that are nowadays used in many mobile applications. For example, in many peer-to-peer ride sharing applications, the driver does not know some details of the order before accepting it. In addition, she faces a quota on the number of orders that she can reject. The primary goal of such quotas is to ensure that drivers accept a sufficient number of orders that may not be as profitable for her as some other orders. This restriction forces the driver to calculate the benefits and costs of accepting an order based on the distance, road condition, traffic congestion, etc. At the same time, such a policy can force the driver to switch to a competing platform. Our model can be used to find the optimal quota that will be beneficial for the platform and not too restrictive for the drivers.

1.A Main proofs

1.A.1 Proposition 1

We use the Karush-Kuhn-Tucker theorem in order to find optimal choice probabilities. The Lagrangian of the manager's problem can be written as

$$\sum_{i=1}^{N}\sum_{\omega\in\Omega} v(i|\omega)\mathcal{P}(i|\omega)\mu(\omega) - \lambda\left(-\sum_{i=1}^{N}\mathcal{P}(i)\log\mathcal{P}(i) + \sum_{i=1}^{N}\sum_{\omega\in\Omega}\mathcal{P}(i|\omega)\log\mathcal{P}(i|\omega)\mu(\omega)\right)$$
$$+\sum_{\omega\in\Omega}\xi_i(\omega)\mathcal{P}(i|\omega)\mu(\omega) - \sum_{\omega\in\Omega}\psi(\omega)(\sum_{i=1}^{N}\mathcal{P}(i|\omega)-1)\mu(\omega) - \sum_{i=1}^{N}\varphi_i(\sum_{\omega\in\Omega}\mathcal{P}(i|\omega)\mu(\omega)-q_i),$$

where $\psi(\omega)$, $\xi_i(\omega)$ and $\varphi_i \in \mathbb{R}_+$ are Lagrange multipliers.

The first order condition with respect to $\mathcal{P}(i|\omega)$ is

$$v(i|\omega) + \xi_i(\omega) - \psi(\omega) + \lambda(\log\mathcal{P}(i) - \log\mathcal{P}(i|\omega)) - \varphi_i = 0. \tag{1.12}$$

Let us first show that $\mathcal{P}(i|\omega) > 0$ for all $i \in \{1, ..., N\}$, $\omega \in \Omega$. Suppose to the contrary that $P(i|\omega) = 0$. Then the term $-\lambda \log P(i|\omega)$ goes to infinity. The only terms which can balance it in order to make the equation (1.12) hold are $\psi(\omega)$ and φ_i. Thus, either $\psi(\omega)$ or φ_i goes to infinity. $P(i|\omega)$ cannot be zero for all $\omega \in \Omega$. That is, there exists a state of the world ω' in which $P(i|\omega') > 0$. In this state of the world $\xi_i(\omega') = 0$ and the first order condition is

$$v(i|\omega') + \lambda\log\mathcal{P}(i) - \lambda\log\mathcal{P}(i|\omega') - \psi(\omega') - \varphi_i = 0.$$

Therefore φ_i cannot go to infinity because there is no other term which can balance it. Thus, $\psi(\omega)$ goes to infinity. $P(i|\omega)$ cannot be zero for all $i \in \{1, ..., N\}$. That is, there exists option j such that $P(j|\omega) > 0$. The first order condition for this option is

$$v(j|\omega) + \lambda\log\mathcal{P}(j) - \lambda\log\mathcal{P}(j|\omega) - \psi(\omega) - \varphi_j = 0.$$

The last first order condition cannot hold since there is nothing to balance minus infinity of $-\psi(\omega)$. We have arrived to a contradiction, therefore, $\mathcal{P}(i|\omega) > 0$ for all $i \in \{1, ..., N\}$, $\omega \in \Omega$. Hence we have $\xi_i(\omega) = 0$ and the first order condition (1.12) can be rearranged to:

$$\mathcal{P}(i|\omega) = \mathcal{P}(i)e^{(v(i|\omega)-\psi(\omega)-\varphi_i)/\lambda}. \tag{1.13}$$

Plugging (1.13) into (1.2), we obtain:

$$e^{\psi(\omega)/\lambda} = \sum_{i=1}^{N} \mathcal{P}(i)e^{(v(i|\omega)-\varphi_i)/\lambda},$$

which we again use in (1.13) and find:

$$\mathcal{P}(i|\omega) = \frac{\mathcal{P}(i)e^{(v(i|\omega)-\varphi_i)/\lambda}}{\sum_{j=1}^{N} \mathcal{P}(j)e^{(v(j|\omega)-\varphi_j)/\lambda}}.$$

Finally, using (1.5) we obtain:

$$\mathcal{P}(i|\omega) = \frac{q_i e^{(v(i|\omega)-\varphi_i)/\lambda}}{\sum_{j=1}^{N} q_j e^{(v(j|\omega)-\varphi_j)/\lambda}}. \tag{1.14}$$

The uniqueness of the optimal conditional choice probabilities and the sufficiency follow from the fact that the maximand in the RI problem with quotas is a strictly concave function, since the payoffs are linear in choice probabilities and the cost is strictly convex (Theorem 2.7.2 in Cover and Thomas 2012).

1.A.2 Proposition 2

First, we relax the condition on the form of the cost function and prove a more general result. In particular, let $\kappa(\mathcal{P})$ be any posterior separable attention cost function (Definition 2 in Caplin, Dean, and Leahy 2021). The properties of posterior-separable cost functions guarantee that decision problems with such attention costs can be solved with the method of Lagrange multipliers. Proposition 7 provides conditions under which the RI manager with a posterior-separable cost function always acquires information.

Proposition 7. *If (i) $v(i|\omega) - v(j|\omega)$ is not constant across states $\forall i, j \in \{1, ..., N\}$ and $i \neq j$, and (ii) $\left.\frac{\partial \kappa(\mathcal{P})}{\partial \mathcal{P}(i|\omega)}\right|_{\mathcal{P}(i|\omega)\equiv\mathcal{P}(i)} = 0$ for all $i \in \{1, ..., N\}$, $\omega \in \Omega$, when $\mathcal{P}(i) > 0$, then the RI manager with quotas always acquires information.*

Proof. Let us assume the opposite. If the manager does not acquire information, then conditional choice probabilities are state-independent and coincide with quotas at the optimum $\mathcal{P}(i|\omega) \equiv \mathcal{P}_i \equiv q_i$ for all $i \in \{1, ..., N\}$, $\omega \in \Omega$. We use the Karush-Kuhn-Tucker theorem in order to analyze the optimal choice probabilities. The Lagrangian of the manager's problem can be written as

$$\sum_{i=1}^{N}\sum_{\omega\in\Omega} v(i|\omega)\mathcal{P}(i|\omega)\mu(\omega) - \lambda\kappa(\mathcal{P})+$$

$$+\sum_{\omega\in\Omega}\xi_i(\omega)\mathcal{P}(i|\omega)\mu(\omega) - \sum_{\omega\in\Omega}\psi(\omega)(\sum_{i=1}^{N}\mathcal{P}(i|\omega)-1)\mu(\omega) - \sum_{i=1}^{N}\varphi_i(\sum_{\omega\in\Omega}\mathcal{P}(i|\omega)\mu(\omega)-q_i),$$

where $\psi(\omega), \xi_i(\omega)$ and $\varphi_i \in \mathbb{R}_+$ are Lagrange multipliers.

The first order condition with respect to $\mathcal{P}(i|\omega)$ is

$$v(i|\omega) + \xi_i(\omega) - \psi(\omega) - \lambda\frac{\partial\kappa(\mathcal{P})}{\partial\mathcal{P}(i|\omega)} - \varphi_i = 0. \tag{1.15}$$

Let us show that the choice probabilities which coincide with the quotas cannot satisfy equation (1.15). Since $q_i > 0$, $\xi_i(\omega) = 0$, and, according to condition (ii), $\frac{\partial\kappa(\mathcal{P})}{\partial\mathcal{P}(i|\omega)}\Big|_{\mathcal{P}=q} = 0$ for all $i \in \{1, ..., N\}$, $\omega \in \Omega$. This leaves

$$v(i|\omega) - \psi(\omega) - \varphi_i = 0.$$

The same holds for all other options. Therefore,

$$v(i|\omega) - \psi(\omega) - \varphi_i = v(j|\omega) - \psi(\omega) - \varphi_j.$$

Or

$$v(i|\omega) - v(j|\omega) = \varphi_i - \varphi_j.$$

This equation cannot hold for all realizations of ω. This is because the LHS of the above equation is state-dependent, while the RHS is state-independent, which contradicts condition (i). $\square$

The expected reduction in Shannon's entropy cost function is a posterior-separable cost function which satisfies the condition (ii) in Proposition 7. Therefore, the case in Proposition 2 is a special case of Proposition 7.

1.A.3 Proposition 3

For a given vector of quotas $\mathbf{q}$ there exists a vector of Lagrange multipliers $\boldsymbol{\varphi}$ which corresponds to the optimal choice probabilities for the problem with quotas. Then, for a given vector of quotas, we can construct a vector of subsidies $\mathbf{S} = -\boldsymbol{\varphi}$. In a problem with such a vector of subsidies, the conditional choice probabilities

$$\forall i \in \{1, ..., N\}: \quad \mathcal{P}^S(i|\omega) = \frac{q_i e^{(v(i|\omega)+S_i)/\lambda}}{\sum_{j=1}^N q_j e^{(v(j|\omega)+S_j)/\lambda}}$$

are optimal, since, by construction

$$\forall i \in \{1, ..., N\}: \quad q_i = \sum_{\omega \in \Omega} \mathcal{P}^S(i|\omega)\mu(\omega),$$

hence, the conditions of Proposition 1 of Caplin, Dean, and Leahy (2019) are satisfied, and $\mathcal{P}^S$ is a vector function of optimal conditional choice probabilities for the problem with subsidies.

1.A.4 Proposition 4

The condition (1.8) secures that the solution to the problem with subsidies is unique and there does not exist another optimum due to Lemma S2 from the online Appendix to Matějka and McKay (2015). The rest follows from Proposition 3.

1.A.5 Lemma 1

We start the proof by stating a lemma, which appears to be useful.

Lemma 2. *Vector* $\boldsymbol{\varphi}$*, which appears in the solution for the RI problem with subsidy* $\mathbf{q}$*, is continuous in* $\mathbf{q}$.

Proof. The idea of the proof is the following. We write an equation which provides an implicit function between $\mathbf{q}$ and $\boldsymbol{\varphi}$ and show that the conditions of implicit function theorem are satisfied, so $\boldsymbol{\varphi}$ is a continuously differentiable function of $\mathbf{q}$, and thus, continuous.

Without loss of generality we normalize $\sum_{i=1}^N \varphi_i$ to some constant c and assume that $\lambda = 1$. Let us consider the following function:

$$F(\mathbf{q}, \boldsymbol{\varphi}) = \mathbf{q} - \mathbb{E}\left[\frac{\mathbf{q} \circ e^{\mathbf{v}-\boldsymbol{\varphi}}}{\mathbf{q}^T e^{\mathbf{v}-\boldsymbol{\varphi}}}\right] + \mathbf{i}^T \boldsymbol{\varphi} \mathbf{i} - c\mathbf{i}.$$

Here "$\circ$" denotes element-wise multiplication, $\mathbf{q}^T$ is the transpose of vector $\mathbf{q}$, $e^{\mathbf{v}-\boldsymbol{\varphi}}$ is a vector with elements $e^{v_i-\varphi_i}$, $\mathbf{i}$ is a N-dimensional vector with all elements equal to 1, $\mathbf{q}^T e^{\mathbf{v}-\boldsymbol{\varphi}}$ is a scalar product. This function, if we consider the equality $F(\mathbf{q},\boldsymbol{\varphi}) = 0$, provides an implicit function between $\mathbf{q}$ and $\boldsymbol{\varphi}$. Equality $\mathbf{q} - \mathbb{E}\left[(\mathbf{q}\circ e^{\mathbf{v}-\boldsymbol{\varphi}})/(\mathbf{q}^T e^{\mathbf{v}-\boldsymbol{\varphi}})\right] = 0$ is derived by plugging equation (1.6) into equation (1.5) and $\mathbf{i}^T\boldsymbol{\varphi}\mathbf{i} - c\mathbf{i} = 0$ is the normalization condition $\sum_{i=1}^N \varphi_i - c = 0$.

If we show that $\nabla_{\boldsymbol{\varphi}} F$ is invertible, we can use the implicit function theorem, and thus prove that $\mathbf{q} \to \boldsymbol{\varphi}$ is a continuously differentiable correspondence. It is easy to show that

$$\nabla_{\boldsymbol{\varphi}} F = \mathbb{E}\left[diag(Q) - QQ^T\right] + ii^T,$$

where $Q = \frac{\mathbf{q}\circ e^{\mathbf{v}-\boldsymbol{\varphi}}}{\mathbf{q}^T e^{\mathbf{v}-\boldsymbol{\varphi}}}$. $\mathbb{E}\left[diag(Q) - QQ^T\right]$ has rank $N-1$ and is positive semi-definite (PSD)[6]. Then the matrix

$$G = \mathbb{E}\left[diag(Q) - QQ^T\right] + ii^T$$

is PSD. That is so since

$$\mathbf{z}^T G\mathbf{z} = \underbrace{\mathbb{E}\left[\mathbf{z}^T(diag(Q) - QQ^T)\mathbf{z}\right]}_{\geq 0} + \underbrace{(\mathbf{z}^T i)^2}_{\geq 0} \geq 0.$$

A PSD matrix has full rank if and only if it is positive definite (PD). That is, we would like to show that $\mathbf{z}^T G\mathbf{z} \neq 0$. Let us prove it by contradiction. Let us assume the opposite: $\mathbf{z}^T A\mathbf{z} = 0$. Then $\mathbf{z}^T(diag(Q) - QQ^T)\mathbf{z} = 0$. Since the dimensionality of the kernel of $diag(Q) - QQ^T$ is one, the solution which has the form $\mathbf{z} = \alpha i$, $\alpha \neq 0$ is the only possible one. But in this case $(\mathbf{z}^T i)^2 = \alpha^2(i^T i)^2 = \alpha^2 N^2 > 0$, since $\alpha \neq 0$.

Therefore, the matrix G is PD, and thus it has full rank. We can apply an implicit function theorem, and thus $\boldsymbol{\varphi}$ is continuous in $\mathbf{q}$. □

Let us now continue the proof. The idea of the proof is the following. First, we consider a mapping $A : \mathbf{q} \to (\boldsymbol{\beta}\circ e^{\boldsymbol{\varphi}})/(\boldsymbol{\beta}^T e^{\boldsymbol{\varphi}})$, where $\boldsymbol{\varphi}$ is a vector from solution (1.6) and show that such mapping has a fixed point. Second, we show that the equation which determines the fixed point coincides with a condition which is equivalent to equation (1.9), thus proving the existence of quotas $\mathbf{q}$ which induce desired choice probabilities. Throughout the proof, without loss of generality, we assume that $\lambda = 1$.

[6]See, for example, Fosgerau and Nielsen 2021.

Let us consider the following mapping $A : [0,1]^N \to [0,1]^N$:

$$A : \mathbf{q} \to \frac{\boldsymbol{\beta} \circ e^{\boldsymbol{\varphi}}}{\boldsymbol{\beta}^T e^{\boldsymbol{\varphi}}},$$

where φ is a vector of Lagrange multipliers from equation (1.6).

Let us show that mapping A has a fixed point. According to Brouwer's fixed point theorem, there exists a fixed point of a continuous mapping of a compact convex set to itself.[7] The conditions of the theorem are satisfied, since the unit simplex in N-dimensional Euclidean space is compact and convex, $(\boldsymbol{\beta} \circ e^{\boldsymbol{\varphi}})/(\boldsymbol{\beta}^T e^{\boldsymbol{\varphi}})$ clearly belongs to the unit simplex for any $\omega \in \Omega$, mapping A is continuous in $\mathbf{q}$, since $\boldsymbol{\varphi}$ is continuous in $\mathbf{q}$ (Lemma 2), and the vector of generalized logit choice probabilities is continuous in $\boldsymbol{\varphi}$.

Let us show that the equation which determines the fixed point for mapping A and the equation which determines the quota from Proposition 1 are equivalent. The fixed point for mapping A is

$$A(\mathbf{q}) = \mathbf{q} = \frac{\boldsymbol{\beta} \circ e^{\boldsymbol{\varphi}}}{\boldsymbol{\beta}^T e^{\boldsymbol{\varphi}}},$$

which can be rewritten as

$$\log q_i = \log \beta_i + \varphi_i - \log(\sum_{j=1}^{N} \beta_j e^{\varphi_j}), \quad \forall i \in \{1, ..., N\}. \tag{1.16}$$

In order to satisfy equation (1.9), the vector $\mathbf{q}$ should be chosen to satisfy

$$\forall i \in \{1, ..., N\} : \quad \mathcal{P}(i|\omega) = \frac{q_i e^{v(i|\omega) - \varphi_i}}{\sum_j q_j e^{v(j|\omega) - \varphi_j}} = \frac{\beta_i e^{v(i|\omega)}}{\sum_j \beta_j e^{v(j|\omega)}}, \quad \forall \omega \in \Omega.$$

The latter equation does not change if we multiply the nominator and denominator of the LHS by $\sum_{j=1}^{N} \beta_j e^{\varphi_j}$. Therefore,

$$\forall i \in \{1, ..., N\} : \quad \frac{q_i e^{v(i|\omega) - \varphi_i} \sum_{j=1}^{N} \beta_j e^{\varphi_j}}{\sum_j q_j e^{v(j|\omega) - \varphi_j} \sum_{j=1}^{N} \beta_j e^{\varphi_j}} = \frac{\beta_i e^{v(i|\omega)}}{\sum_j \beta_j e^{v(j|\omega)}}, \quad \forall \omega \in \Omega.$$

One of the possibilities of satisfaction of the last equation is

$$\forall i \in \{1, ..., N\} : \quad q_i e^{v(i|\omega) - \varphi_i} \sum_{j=1}^{N} \beta_j e^{\varphi_j} = \beta_i e^{v(i|\omega)}, \quad \forall \omega \in \Omega,$$

[7] It is important to notice that Brouwer's fixed point theorem does not require the mapping to be surjective.

which implies

$$\forall i \in \{1, ..., N\}: \quad \log q_i = \log \beta_i + \varphi_i - \log(\sum_{j=1}^{N} \beta_j e^{\varphi_j}). \tag{1.17}$$

The latter equation coincides with equation (1.16) and, as we showed earlier, equation (1.16) has a solution. Thus, there exists a quota $\mathbf{q}$ that induces choice probabilities (1.9).

1.B Solution for the example with two candidates

In order to find conditional probabilities $\mathcal{P}(i|\omega) = \frac{q_i e^{(v(i|\omega)-\varphi_i)/\lambda}}{\sum_{j=1}^{N} q_j e^{(v(j|\omega)-\varphi_j)/\lambda}}$ we must find φ_i. According to Proposition 3 we can find a vector of subsidies that induces the same behavior, and hence set $S_1 = \varphi_1 = C$. Therefore, the only parameter that we need to find is φ_2. Probabilities must satisfy the equation 1.5:

$$q = \frac{qe^{(-\varphi_2)/\lambda}}{qe^{(-\varphi_2)/\lambda} + (1-q)} b + (1-b) \frac{qe^{(1-\varphi_2)/\lambda}}{qe^{(1-\varphi_2)/\lambda} + (1-q)}.$$

Solving this equation for φ_2 and plugging it into $\mathcal{P}(2|0)$ yields two solutions:

$$\mathcal{P}(2|0) \in \{\frac{-b - q + (b+q-1)e^{\frac{1}{\lambda}} + \sqrt{(b+q-(g+q-1)e^{\frac{1}{\lambda}})^2 + 4q(be^{\frac{1}{\lambda}} - b)}}{2(be^{\frac{1}{\lambda}} - b)}, \frac{-b - q + (b+q-1)e^{\frac{1}{\lambda}} - \sqrt{(b+q-(b+q-1)e^{\frac{1}{\lambda}})^2 + 4q(be^{\frac{1}{\lambda}} - b)}}{2(be^{\frac{1}{\lambda}} - b)}\}.$$

The solution to the manager's problem should be positive. Only the first root is positive. This is so since the denominator $2(be^{\frac{1}{\lambda}} - b)$ is positive. For the root to be positive, the nominator should be positive. The second root is negative since $4q(be^{\frac{1}{\lambda}} - b)$ is positive, so the square root is larger than the term in front of the square root. For a similar reason, the first root is positive.

That is, the solution to the manager's problem is

$$\mathcal{P}(2|0) = \frac{-b - q + (b+q-1)e^{\frac{1}{\lambda}} + \sqrt{(b+q-(b+q-1)e^{\frac{1}{\lambda}})^2 + 4q(be^{\frac{1}{\lambda}} - b)}}{2(be^{\frac{1}{\lambda}} - b)}$$

and

$$\mathcal{P}(2|1) = \frac{q - b\mathcal{P}(2|0)}{1-b}.$$

1.C Details of the solution for the example with two candidates: subsidies

We use the example from Section 1.5.1, but now the social planner sets up a subsidy for the risky candidate: the manager receives extra payment of S if she chooses the risky candidate. In this case, the solution has the standard modified multinomial logit form but with the value of the risky candidate increased by S. Namely,

$$\mathcal{P}(2|0) = \frac{\mathcal{P}(2)e^{S/\lambda}}{\mathcal{P}(2)e^{S/\lambda} + \mathcal{P}(1)e^{C/\lambda}}$$

$$\mathcal{P}(2|1) = \frac{\mathcal{P}(2)e^{(1+S)/\lambda}}{\mathcal{P}(2)e^{(1+S)/\lambda} + \mathcal{P}(1)e^{C/\lambda}}.$$

In order to compare the manager's behavior under both policies we need to find a level of subsidies for which the risky candidate would be chosen by the manager with the required probability q:

$$(1-b)\mathcal{P}(2|1) + b\mathcal{P}(2|0) = q.$$

The unconditional probabilities in the case of the manager's problem with subsidies are as follows:

$$\mathcal{P}(2) = \max\{0, \min\{1, \frac{-e^{C/\lambda}(-e^{(1+S)/\lambda} + e^{C/\lambda} - be^{S/\lambda} + be^{(1+S)/\lambda})}{(e^{(1+S)/\lambda} - e^{C/\lambda})(-e^{S/\lambda} + e^{C/\lambda})}\}\}$$

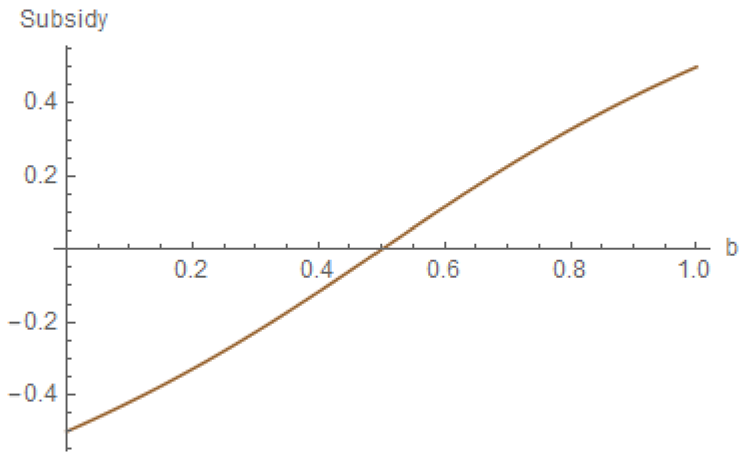
$$\mathcal{P}(1) = 1 - \mathcal{P}(2).$$

Figure 1.6 shows the optimal subsidy that is necessary in order to equalize the unconditional probability of choosing the risky candidate to 0.5 as a function of b.

We see that for small b the social planner sets a financial penalty for choosing the risky candidate. That is because the risky candidate is likely to be productive and the manager would prefer to choose it more often than in half of the cases. In contrast, if b is high, the social planner supports the choice of the risky candidate by establishing a positive subsidy.

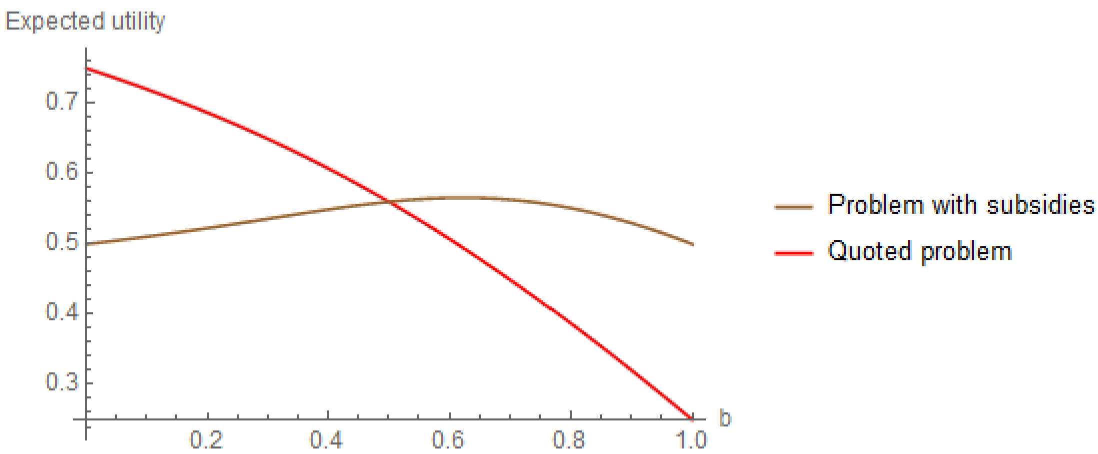
For high b the utility of the firm in the case of subsidies is higher than in the case

Figure 1.6: Optimal subsidy as a function of b and $\lambda = 0.5$, $q = 0.5$, $C = 0.5$.



of quotas (Figure 1.7). Therefore, one can speculate that it is impossible to extract all subsidies from the firm afterwards and hence it is more beneficial for firms to lobby for subsidies rather than quotas.

Figure 1.7: Utility of the manager as a function of b and $\lambda = 0.5$, $q = 0.5$, $C = 0.5$. The red curve is for the quoted RI problem and the brown curve is for the RI problem with subsidies.



1.D Details of the solution for the model with non-binding quotas

Let us assume that there are N alternatives and there is only one restriction on unconditional probabilities: $\mathcal{P}(1) = q$. Accordingly, this constraint implies that $\sum_{j=2}^{N} \mathcal{P}(j) = 1 - q$. Then the Lagrangian of the manager's problem described in Section 1.3.2 is as follows:

$$\sum_{i=1}^{N}\sum_{\omega\in\Omega} v(i|\omega)\mathcal{P}(i|\omega)\mu(\omega) - \lambda\left(\sum_{i=1}^{N}\mathcal{P}(i)\log\mathcal{P}(i) + \sum_{i=1}^{N}\sum_{\omega\in\Omega}\mathcal{P}^U(i|\omega)\log\mathcal{P}^U(i|\omega)\mu(\omega)\right)$$
$$-\sum_{\omega\in\Omega}\psi(\omega)(\sum_{i=1}^{N}\mathcal{P}(i|\omega)-1)\mu(\omega)-\varphi_1(\sum_{\omega\in\Omega}\mathcal{P}(1|\omega)\mu(\omega)-q)-\varphi_2(\sum_{j=2}^{N}\sum_{\omega\in\Omega}\mathcal{P}(j|\omega)\mu(\omega)-1+q),$$

where $\psi(\omega)$, φ_1, and φ_2 are Lagrange multipliers. The first order condition with respect to $\mathcal{P}(1|\omega)$ is:

$$v(1|\omega) - \psi(\omega) + \lambda(\log\mathcal{P}(1) - \log\mathcal{P}(1|\omega)) - \varphi_1 = 0,$$

and with respect to $\mathcal{P}(j|\omega)$ is:

$$v(j|\omega) - \psi(\omega) + \lambda(\log\mathcal{P}(j) - \log\mathcal{P}(j|\omega)) - \varphi_2 = 0.$$

Following the same procedure described in Section 1.3.2 this can be rearranged to:

$$\mathcal{P}(1|\omega) = \frac{qe^{(v(1|\omega)-\varphi_1)/\lambda}}{\sum_{j=2}^{N}\mathcal{P}(j)e^{(v(j|\omega)-\varphi_2)/\lambda} + qe^{(v(1|\omega)-\varphi_1)/\lambda}},$$

and

$$\mathcal{P}(j|\omega) = \frac{\mathcal{P}(j)e^{(v(j|\omega)-\varphi_2)/\lambda}}{\sum_{j=2}^{N}\mathcal{P}(j)e^{(v(j|\omega)-\varphi_2)/\lambda} + qe^{(v(1|\omega)-\varphi_1)/\lambda}}.$$

Therefore, the solution to the problem will be similar to that described in Section 1.3.2. The only difference is that now, for all alternatives for which the quota is not binding and for which $\mathcal{P}(j) > 0$, the additive state-independent component φ_2 is the same. This logic extends to any situation in which not all quotas are binding.

Chapter 2

Inattentive Price Discovery in Exchange-Traded Funds

Co-authored with Mariia Kosar (CERGE-EI).

2.1 Introduction

Exchange-traded funds (ETFs) have gained popularity among investors over the past decades, and have rapidly grown in terms of assets under management and trading volume. These instruments have attracted the attention of both scholars and practitioners due to the important asset pricing implications for their underlying securities. The most well-documented concern about ETFs is their disposition to noise and factor trading that, combined with the continuous arbitrage mechanism, may lead to propagation of noise to the underlying assets (Bhattacharya and O'Hara 2020). However, there is still a question regarding whether ETFs can facilitate stock-specific price discovery, and if yes what net effect it has for the ETF's underlying bundle.

In this paper we investigate this question. First, we show that the learning of stock-specific fundamental information can occur at the ETF level. Moreover, our results suggest that ETF investors endogenously respond to changes in the fundamental value of underlying stocks, in line with the rational inattention theory[1]. Second, we provide

[1] A recent review of the rational inattention literature can be found in Maćkowiak, Matějka, and Wiederholt (2021).

evidence that this pattern of learning affects ETF's underlying bundles, leading to propagation of idiosyncratic shocks across underlying stocks.

We proceed in two steps. Firstly, in order to demonstrate that the information acquisition can occur at the ETF level, we measure the response of ETF intraday prices to earnings surprises. We use earnings surprises as a measure of stock-specific information released at the time of announcement. We focus on capitalization-based ETFs that are traded on U.S. exchanges and have international exposure. We then select only earnings announcements that occur when underlying exchange is closed, and the U.S. exchange is open. By design, this ensures that price discovery, if any, occurs at the ETF level. Moreover, to make ensure that the responses we measure refer to the specified earnings announcements, we select only announcements that were not surrounded by other announcements. Our results suggest that stock specific price discovery can occur at the ETF level. In addition, the earnings response coefficients are statistically significant only for announcements made by firms with large weights in their corresponding ETFs. Furthermore, we differentiate between non-busy days, when there is relatively low news pressure from the U.S. market in terms of macro and stock-specific announcements, and busy information days.

Secondly, we conduct an empirical analysis of the spillover patterns from ETFs to the stocks in their underlying bundles. Specifically, we compute the abnormal idiosyncratic volatility (AIV) of ETFs and their constituents around earnings announcements (Yang, Zhang, and Zhang 2020). The AIV measures to which extent the idiosyncratic volatility on announcement days is abnormal compared to the aggregate idiosyncratic volatility over a given period. Then, we estimate the relationship between the AIV of constituent stocks and their corresponding ETFs when the underlying markets re-open. Our results suggest that there is a significantly positive relationship between the AIV of constituent stocks and their corresponding ETFs, which is significant only around earnings releases of stocks with large weights in ETFs. This allows us to conclude that learning at the ETF level affects underlying bundles, leading to abnormal co-movement in volatilities across underlying stocks.

Finally, we show that the ETF AIV risk is priced in a sample of all ETF constituents. The abnormal stock returns are loaded on the ETF AIV, which results in positive and significant regression coefficients of future returns on the ETF AIV over a relatively long time horizon (10 days). The relationship is reversed, which implies that the reaction of returns to the ETF AIV was not fundamental.

Literature. This study contributes to several strands of literature. Firstly, the results in this paper relate to the literature on the impact of financial innovation on the efficiency of financial markets (Basak and Pavlova 2013, Appel, Gormley, and Keim 2016). There is a growing academic literature on the effects of ETFs on the asset pricing of their constituents. Many researchers treat ETFs mostly as venues for noise or factor trading, and thus focus on propagation of non-fundamental and factor shocks from ETFs to underlying markets (Wang and Xu 2019, Filippou, Gozluklu, and Rozental 2019, Ben-David, Franzoni, and Moussawi 2018, Shim 2018, Huang, O'Hara, and Zhong 2021, Israeli, Lee, and Sridharan 2017, Glosten, Nallareddy, and Zou 2016, Levy and Lieberman 2019). Two prominent reasons for such concerns are best summarized by Ben-David, Franzoni, and Moussawi (2018) and Shim (2018). Ben-David, Franzoni, and Moussawi (2018) argue that ETF investors are dominated by noise traders, who propagate non-fundamental shocks to prices of underlying assets, amplifying non-fundamental volatility. Shim (2018) takes a different approach,arguing that ETF markets are populated with informed traders who are, however, factor-informed. He shows that, if factor price discovery occurs in ETFs, rather than stocks, underlying securities tend to misreact to factor information. Both approaches ascribe the key role in shock propagation from ETFs to underlying securities to ETF arbitrage mechanism. However, some studies have reached a conclusion that, due to benefits that such instruments bring to the market (i.e., low cost, high liquidity, and hedging opportunities), ETFs can encourage informed trading and information transfers around fundamental news releases, and thus improve the pricing efficiency of their underlying stocks (Ciura 2016, Huang, O'Hara, and Zhong 2021, Bhojraj, Mohanram, and Zhang 2020, Ernst 2021). For example, Bhojraj, Mohanram, and Zhang (2020) focus only on top-weighted stocks and show that ETF mechanic bundle trades help to transfer sector and market-wide information contained in company earnings announcements into the stock prices of its peers, reducing their post-earnings announcement drift and thus contributing to their price efficiency. This is consistent with Savor and Wilson (2016), who show that investors learn both factor and asset-specific components from earnings announcements.

Relative to these studies, we focus on the role of ETFs in transferring an asset's value-specific information to other assets[2]. We use identification strategy, which allows us to

[2]Bhattacharya and O'Hara (2018) theoretically show that ETFs may have a detrimental influence on information propagation from one stock to another, since they can also transfer value-irrelevant firm-specific shocks to their peers, which may lead to market instability and increased synchronicity between stock prices.

study how exactly ETF investors acquire information about their constituents, and to evaluate the net effect of price discovery on the ETF level for underlying bundles.

Secondly, this paper is closely related to literature that links asset price responses to investor inattention. While, there are many empirical studies that document this phenomenon (for example, Hirshleifer, Lim, and Teoh 2009, DellaVigna and Pollet 2009, Fedyk 2021), there is still a lack of empirical literature that studies endogenous investor attention and shows how investors actually behave[3]. Chuprinin, Gorbenko, and Kang (2019) show that firm size is a major determinant of the degree of investor research into a specific stock around fundamental news releases. Li (2021) shows that the efficiency of price reaction to a particular type of risk depends on the value-relevance of that risk. Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) demonstrate that mutual fund managers optimally track information about aggregate shocks in recessions and idiosyncratic shocks in booms. Recent studies by Hirshleifer and Sheng (2021) and Huang, Huang, and Lin (2019) investigate how stock investors allocate attention between systematic and idiosyncratic information. We complement this literature by focusing on endogenous investor attention[4]. However, we focus on ETF which, for example, in contrast to mutual funds, have a fixed weighting scheme that allows us to isolate the effect of news releases on changes in the price of ETF, so that we can obtain a clear measure of attention using intraday data[5].

Finally, this project contributes to the strand of literature on the importance of foreign investments into local financial markets (Figlio and Blonigen 2000, Levy and Lieberman 2019, Filippou, Gozluklu, and Rozental 2019). Specifically, we construct a diverse sample of ETFs that focus on various country and sector indexes. From this diverse sample, we are able to establish the impact of U.S. - traded ETFs on local stocks in their underlying bundles.

The rest of the paper is organized as follows. In Section 2.2 we set up a basic theoretical framework of investor's behavior when she faces information constraint. Empirical research design and data are outlined in Sections 2.3 and 2.4. Section 2.5 discusses the

[3]There are numerous theoretical papers that use endogenous inattention to understand co-movements or sluggishness of prices, for example Coibion and Gorodnichenko (2012), Mackowiak and Wiederholt (2009), Veldkamp (2006)

[4]See also Ben-David et al. (2021) who document competition for attention in the ETF space by creating specialized ETFs.

[5]Ernst (2021) also studies ETF and presents empirical evidence that simultaneous trades of ETFs with their announcing constituent stocks increase on earnings announcement days, and more so for stocks with high weights in ETFs.

results. Finally, Section 2.6 concludes.

2.2 Theoretical framework

We model the investor's behavior following the literature on rational inattention, which originated in studies by Sims (2003). For tractability, we consider a one period two-dimensional tracking problem with quadratic loss[6]. The investor wants to track changes in the value of the ETF: $\Delta V = \sum_i w_i \Delta V_i$, where ΔV_i are changes in the liquidation value of stock $i \in 1, 2$ that enters the ETF with weight $w_i > 0$. However she can process only a finite amount of information. We model the limited ability to process information as a constraint on uncertainty reduction, where uncertainty is measured by entropy (Shannon 1948, Cover and Thomas 2012). The problem is formalized as follows.

Standard (unconstrained) RI problem. The investor's problem is to choose the joint distribution of the decision variable ΔV with the exogenous uncertainty ΔV_i, $i \in \{1, 2\}$ so as to maximize:

$$\max_{\Delta V} \mathbb{E}[-(w_1 \Delta V_1 + w_2 \Delta V_2 - \Delta V)^2],$$

where priors are

$$\forall i \in \{1, 2\} : \quad \Delta V_i \sim N(0, \sigma_i^2).$$

The investor can obtain independent signals about the individual liquidation value of stock i:

$$\forall i \in \{1, 2\} : \quad s_i = \Delta V_i + e_i,$$

where the noise of signals is normally distributed, $e_i \sim N(0, \sigma_{e_i}^2)$. The variance of the signals, $\sigma_{e_i}^2$, is subject to investors choice.

The investor has a capacity constraint in the choice of signal[7]

$$\sum_i \underbrace{\frac{1}{2} \log(\frac{\sigma_i^2}{\sigma_{i|s_i}^2})}_{k_i} \leq k, \tag{2.1}$$

[6]We show in Appendix 2.A.2 that results are qualitatively the same for the multi-dimensional tracking problem. Also see Veldkamp (2006) for more general treatment of the problem.

[7]This can be motivated by investors having just 168 hours a week. An alternative way to model the behavior is to assume information-processing costs, such that investors may be able to expand their attention whenever needed. Therefore, investors attention to the specific asset will not depend on information that is not directly relevant. Our empirical results (see Section 2.5.1) could be interpreted as supporting both models. Hence, we remain agnostic on this question, and additional tests are needed to separate these two models. See Azrieli (2021) for a discussion of the difference between model approaches.

where $\sigma^2_{i|s_i}$ is a conditional variance of changes in the value of individual stock i, k is the bound on the investor's capacity to process information, and k_i is the investor's attention to value-relevant information of the stock i.

In addition, the investor faces the no-forgetting constraint, i.e., condition that she can not increase prior uncertainty about changes in the stock's value:

$$\sigma^2_{i|s_i} \leq \sigma^2_i. \tag{2.2}$$

Because priors and noises are normal, $\sigma^2_{i|s_i}$ is a monotone function of $\sigma^2_{e_i}$: $\sigma^2_{i|s_i} = \frac{\sigma^2_{e_i}\sigma^2_i}{\sigma^2_{e_i}+\sigma^2_i}$. In Appendix 2.A.1 we show that the problem of the investor reduces to the choice of $\sigma^2_{i|s_i}$.

The solution to the problem is formalized in the following lemma:

Lemma 1. The optimal investor's choice of conditional variances of changes in values of individual stocks and attention to value-relevant information of stocks are:

$$\sigma^2_{i|s_i} = \min\{\sigma^2_i, \frac{w_{\neg i}}{w_i}\sqrt{e^{-2k}\sigma^2_i\sigma^2_{\neg i}}\}$$

$$k_i = \max\{0, \frac{1}{2}\log(\frac{w_i\sqrt{\sigma^2_i}}{w_{\neg i}\sqrt{e^{-2k}\sigma^2_{\neg i}}})\}. \tag{2.3}$$

Proof. See Appendix 2.A.1. □

Following Lemma (1) and taking derivatives of equation (2.3) with respect to stock weights, the variance of changes in a stock's value, and an investor's capacity to process information yields the following results:

Corollary 1 (Testable implications)**.** *An investor's attention to a stock's value-relevant information is higher for*

1.1. stocks with higher relative weights in the ETF;

1.2. stocks with higher volatility of changes in the value;

1.3. investors with higher information capacity.

According to Corollary (1.1) the ETF response should be higher for stocks with higher weight in the ETF, controlling for other potential factors. Corollary (1.2) states that, if the volatility of changes is high, which in terms of our empirical exercise means high earnings surprises, then the response of the ETF price will be more efficient. Corollary (1.3) indicates that, if investors have lower information capacity, then the ETF price efficiency with respect to stock information decreases. We test this by comparing the ETF price response in busy days and in days with low numbers of informational announcements.

2.3 Empirical research design

2.3.1 ETF-level analysis

Identifying the response to announcements. The most challenging task in our empirical exercise is to identify the response to the earnings announcement shock on ETF level. The first challenge is to isolate the ETF price response to a specific constituent stock earnings announcement. An average ETF contains dozens of stocks which can make concurrent information releases. To attribute the ETF price response to a specific earnings release, it is necessary to ensure that no other constituent in that ETF makes a competing announcement within a chosen time window. To mitigate this problem, we consider only announcements that are not surrounded by competing earnings releases[8] in the same ETF within a [-1 working day, +1 working day] non-announcement window[9].

The second challenge is to attribute the ETF price response to the price discovery on ETF level. Because of the continuous arbitrage process that occurs between ETFs and their underlying bundles, it can be hard to identify where the price discovery occurs, in the ETF or in it's underlying bundle. To mitigate this issue, we consider only ETFs with asynchronous trading hours with their underlying bundles. Those are ETFs that are traded on U.S. exchanges, but have exposure to international markets. For this sample of ETFs, we are able to observe their price responses when the underlying markets are temporarily closed, but the companies on the underlying markets continue to release earnings announcements. Further, to ensure that the ETFs and their underlying markets do not interact during announcement windows, we require at least 6 hours time lapse from an announcement to the next underlying market's opening. This approach allows us to identify ETFs as a source of price discovery, since the arbitrage mechanism is temporarily switched off.

We include fund fixed effect to capture the differences in fund characteristics, mainly the size and liquidity, which can significantly affect the speed and magnitude of fund

[8]Although information releases are not limited to earnings announcements, usually other information releases are done together with the earnings releases as a part of quarterly/yearly disclosure.

[9]The choice of a non-announcement window is motivated by sample size considerations, as well as by the empirical literature that usually employs a 3-day window for calculating the announcement price response to earnings announcement on stock level.

price response around information releases. We also include stock fixed effect to account for differences in stock-specific characteristics that can affect the investors information choice, such as market capitalization. Finally, day fixed effect is included to capture the overall market differences common for all stocks and funds, for example, market volatility and information quantity released during a particular day.

Empirical specification. To test if the investor's attention to stock's value-relevant information is higher for stocks with higher relative weights in the ETF (Corollary 1.1) and stocks with higher volatility of changes in the value (Corollary 1.2), we measure the response of ETF prices to earnings surprises by computing earnings response coefficients (ERC) over different time horizons for different ETF weight quaniles. ERC present price elasticity with respect to information contained in earnings surprise (Blankespoor, deHaan, and Marinovic 2020), and are obtained from regressing window returns around earnings announcements on earnings surprise. The main empirical model of interest is:

$$ret_{i,j,[0,\tau]} = \alpha SUR_{i,j,t} + \beta I_{W_{i,j} \in q} + \gamma SUR_{i,j,t} * I_{W_{i,j} \in q} + \delta_i + \delta_t + \delta_j + \epsilon_{i,j,[0,\tau]}, \tag{2.4}$$

where $ret_{i,j,[0,\tau]}$ is the return over announcement window $[0, \tau]$; $I_{W_{i,j} \in q}$ is the indicator function that takes the value of 1 if the weight of stock j in the ETF i is in the q^{th} quartile of ETF weights distribution; $SUR_{i,j,t}$ is the earnings surprise. $\delta_i, \delta_t, \delta_j$ are ETF, stock, and day fixed effects.

Savor and Wilson (2016) show that earnings announcements are signals of the future growth prospects of the firm, and use them as firm-level information events. We follow Hirshleifer, Lim, and Teoh (2009) and define the earnings surprise of announcement j in the ETF i on day t as:

$$SUR_{i,j,t} = \frac{Earnings_{i,j,t} - 1/K \sum_{k=1}^{K} Earnings_{i,k,t}}{P_{i,j,t}}, \tag{2.5}$$

where $P_{i,j,t}$ is a closing price of stock j in the ETF i on day t, and a mean forecast of earnings of all K analysts for announcement j in the last quarter prior to announcement j is $1/K \sum_{k=1}^{K} Earnings_{i,k,t}$.

We calculate the return $ret_{i,j,[0,t]}$ over announcement window $[0, \tau]$ as:

$$ret_{i,j,[0,\tau]} = log(P_{i,j,\tau}) - log(P_{i,j,0}), \tag{2.6}$$

where $P_{i,j,\tau}$ is price of the ETF i at τ hours past the announcement of stock j, where τ

spans from 1 to 6 hours, at 1 hour intervals. $P_{i,j,0}$ is the last trading price of the ETF i before earnings announcement j.

To ensure that we correctly measure the weight percentile of each announcing stock j in the ETF i, we compute the respective weight percentiles in the full sample of each ETF i constituents on day t.

To investigate whether the investor's attention to stock's value-relevant information is higher for investors with higher information capacity (Corollary 1.3), we study the ERCs on the busy vs. normal days on the U.S. stock market. The empirical model of interest in this respect is the following:

$$\begin{gathered} ret_{i,j,[0,\tau]} = \alpha SUR_{i,j,t} + +\beta SUR_{i,j,t} * I_{W_{i,j} \in q} + \gamma SUR_{i,j,t} * BUSY_t + \\ \theta SUR_{i,j,t} * I_{W_{i,j} \in q} * BUSY_t + Controls_{i,j} + \delta_i + \delta_t + \delta_j + \epsilon_{i,j,[0,\tau]}. \end{gathered} \tag{2.7}$$

where busy day indicator variable $BUSY_t$ is defined as:

$$BUSY_t = 1 \; if \; News \; Score_t > Q_{0.5} \; or \; N > Q_{0.5}. \tag{2.8}$$

In the above formula, $News \; Score_t$ is the macroeconomic news score of each trading day, and is computed following the methodology of (Xu, Yin, and Zhao 2018):

$$News \; Score_t = \frac{1}{N} \sum_{1}^{N} Score_{j,t},$$

where $Score_{j,t} = \frac{ESS_{j,t} - 50}{50}$ is the normalized Event Sentiment Score ($ESS_{j,t}$) for event j on day t on U.S. market. The Event Sentiment Score indicates the extent to which an event can influence a market price. N is the total number of news events on U.S. market on day t.

2.3.2 Stock-level analysis

Empirical specification. In this section, we introduce an empirical model to test whether learning patterns at the ETF level spill over to their underlying portfolios through instant arbitrage between ETFs and their constituents after underlying markets re-open following announcements. We adopt the approach of Yang, Zhang, and Zhang (2020), who introduce the abnormal idiosyncratic volatility (AIV) as a measure of information risk associated with earnings announcements. The AIV measures the extent to which the idiosyncratic volatility on announcement days is abnormal compared to the aggregate idiosyncratic volatility over a given period. We consider quarterly earnings announcements and, hence, use a quarter period. To measure the AIV of constituent stocks, for each unique ETF i within our sample of fund-announcement data, we collect data on all constituent stocks during our sample period (2016-2017). For each of these stocks, we use data on Fama-French factors, and estimate the idiosyncratic returns with a 3 factor Fama-French model using daily data:

$$ret_{j,t} = \alpha_j + \beta_j^{MKT} MKT_t + \beta_j^{SMB} SMB_t + \beta_j^{HML} HML_t + \epsilon_{j,t}, \tag{2.9}$$

where $ret_{j,t}$ are close-to-close returns stock j from day $t-1$ to t; MKT is the value-weighted market portfolio excess return over the risk-free rate; SMB is the size factor; and HML is the value factor; and $\epsilon_{j,t}$ is the abnormal idiosyncratic return.

Next, for each stock j that entered fund i during the announcement day t we compute the idiosyncratic volatility of a stock within a quarter for the announcement days (IV^{AD}), which are the trading session before an announcement that occurred during off-exchange hours, and the next two trading sessions after the announcement; and for non-announcement days (IV^{NAD}) as the log of the standard deviations of the residual from equation (2.9) during these days, assuming that there are 63 trading days in a quarter. More specifically, we define:

$$IV_{j,i,t}^{AD} = ln\sqrt{\frac{63 * \sum_{t \in AD} \epsilon_{j,i,t}^2}{(n_{AD} - 1)}},$$

$$IV_{j,i,t}^{NAD} = ln\sqrt{\frac{63 * \sum_{t \in NAD} \epsilon_{j,i,t}^2}{(n_{NAD} - 1)}},$$

where n_{AD} and n_{NAD} are the number of days in the pre- and non-announcement periods, respectively. We compute the AIV around announcement day t as the difference in log idiosyncratic volatility:

$$AIV_{j,i,t} = IV_{j,i,t}^{AD} - IV_{j,i,t}^{NAD}.$$

Similarly, we estimate the equation (2.9) for returns of ETF i from day $t-1$ to t:

$$ret_{i,t} = \alpha_i + \beta_i^{MKT} MKT_t + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \epsilon_{i,t},$$

where $ret_{i,t}$ are close-to-close returns of fund i from day $t-1$ to t; and $\epsilon_{i,t}$ is the abnormal idiosyncratic return. We compute the idiosyncratic volatility of ETF i for announcement days (IVF^{AD}), which are the trading day on the U.S. market before an announcement of stock j occurred, the trading day on the U.S. market on which the announcement of stock j occurred, and the next day after the announcement; and for non-announcement days (IVF^{NAD}) in a given quarter as:

$$IVF^{AD}_{j,i,t} = ln\sqrt{63 * \frac{\sum_{t\in AD} \epsilon^2_{i,t}}{(n_{AD}-1)}},$$

$$IVF^{NAD}_{j,i,t} = ln\sqrt{63 * \frac{\sum_{t\in NAD} \epsilon^2_{i,t}}{(n_{NAD}-1)}}.$$

Then, the AIV of ETF i around announcement day t is:

$$AIV_{i,t} = IVF^{AD}_{j,i,t} - IVF^{NAD}_{j,i,t}.$$

We measure the relationship between the AIV of ETFs around the announcement and the AIV of its constituent non-announcing stocks during the next open trading session after the announcement. The empirical model is:

$$AIV_{j,i,t} = \alpha_i + \alpha_j + \alpha_t + \sum_q \gamma_q AIV_{i,t} I_{W_{i,k}\in q} + Controls_{j,t} + \eta_{j,i,t}, \quad (2.10)$$

where $AIV_{j,i,t}$ is the abnormal idiosyncratic volatility of stock j in ETF i on announcement day t; $AIV_{i,t}$ is the abnormal idiosyncratic volatility of ETF i on announcement day t on the U.S. market; $I_{W_{i,k}}$ is the indicator for announcing stock k in ETF i on announcement day t being in the q^{th} quartile of the ETF weights distribution. Following Ben-David, Franzoni, and Moussawi (2018) and Yang, Zhang, and Zhang (2020), $Controls_{j,t}$ include the inverse of price of stock j on day t, $\frac{1}{P_{j,t}}$, the log of market capitalization of stock j on day t, $log(Mkt\ Cap_{j,t})$, the log of Amihud illiquidity measure of stock j on day t, $log(Amihud_{j,t})$, and the lagged returns ($ret_{j,[-1]}$, $ret_{j,[-3,-2]}$, $ret_{j,[-6,-4]}$).

Mechanisms. Bhattacharya and O'Hara (2018) theoretically show that shocks from ETFs are transmitted at a higher degree to the stocks with higher weights in ETFs. To test this empirically, we add the weight of stock j in the ETF i on day t, $W_{i,j,t}$ to our empirical specification, and estimate the following model:

$$AIV_{j,i,t} = \alpha_i + \alpha_j + \alpha_t + \sum_q \gamma_q AIV_{i,t} I_{W_{i,k} \in q} + \sum_q \beta_q log(W_{i,j,t}) I_{W_{i,k} \in q} + \sum_q \beta_q AIV_{i,t} log(W_{i,j,t}) I_{W_{i,k} \in q} + Controls_{j,t} + \eta_{j,i,t}. \quad (2.11)$$

Following Ben-David, Franzoni, and Moussawi (2018) and Shim (2018), we also test whether the arbitrage trades that occurs between ETFs and their underlying bundles can explain the correlations between stocks and ETFs. The model of interest is as follows:

$$AIV_{j,i,t} = \alpha_i + \alpha_j + \alpha_t + \sum_q \gamma_q AIV_{i,t} I_{W_{i,j} \in q} + \sum_q \beta_q \Delta_{i,j,t} I_{W_{i,k} \in q} + \sum_q \beta_q AIV_{i,t} \Delta_{i,j,t} I_{W_{i,k} \in q} + Controls_{j,t} + \eta_{j,i,t}. \quad (2.12)$$

We compute the intensity of arbitrage, $\Delta_{i,j,t}$, as the normalized change in the number of the total shares of stock j held by each ETF i on day t:

$$\Delta_{i,j,t} = \frac{Shares_{j,i,t} - Shares_{j,i,t-1}}{Shares_{j,t}},$$

where $Shares_{j,i,t}$ is the number of shares of stock j held by ETF i on day t; $Shares_{j,t}$ is the total number of shares of stock j on day t.

Further evidence. Finally, we test whether the AIV of ETF is priced. We follow Eugene and French (1992) and estimate the following regression:

$$aret_{j,[t,t+m]} = a + b * AIV_{j,i,t} + \sum_q \gamma_q AIV_{j,i,t} I_{W_{i,j} \in q} + Controls_{j,t} + \epsilon_{j,t}, \quad (2.13)$$

where $aret_{j,[t,t+m]}$ is stock j's cumulative abnormal return, which is the sum of abnormal daily returns, $\epsilon_{j,t}$, from the announcement on day t to day $t+m$; $Controls_{j,t}$ are the same as in previous regressions.

2.4 Data

2.4.1 ETF-level data

Data on daily ETF constituents and their weights in each ETF comes from the ETFDB database. We start with an initial ETF sample that includes all U.S. - traded capitalization-based ETFs with international exposure that active during 2016-2017. We obtain the respective ETF tickers from etf.com. We exclude all sector ETFs from our initial sample, and keep only ETFs with country and regional exposure. Within each ETF, we split all constituent stocks into percentiles by their corresponding weight in the ETF.

To construct a measure of surprise earnings, we collect data on quarterly earnings announcements from I/B/E/S for each ETF constituent. Specifically, we retrieve the following variables from I/B/E/S: the date and time of each announcement, official tickers of the announcing stocks, announced earnings per share (EPS) and the analyst forecasts of EPS for each announcement. The I/B/E/S and ETFDB are matched based on constituent CUSIP.

We obtain daily prices of each announcing ETF constituent from Compustat Daily International. We use this data to compute the earnings surprises. Compustat and I/B/E/S data are matched based on a 6-digit CUSIP obtained from the 8-digit CUSIP in I/B/E/S and from SEDOL in Compustat.

Data on the off-exchange hours of the underlying ETF markets and the opening hours of the U.S. exchanges comes from tradinghours.com. Moreover, we require that there is [-1 day, +1 day] non-announcement window around each announcement. We also ensure that there is at least 6 hours after the announcement prior to the underlying market opening. This procedure leaves us with 842 unique fund-announcement observations.

Data on high-frequency intra-day ETF prices comes from the Trades and Quotes database. We use intra-day trades data to find all trades made during each announcement day. TAQ trades include information on the date and exact time of a trade (up to a millisecond), and data on the prices and sizes of trading orders. We sample the trades data at 5 second frequency. We keep the last price in each 5 second interval, and sum up all trades made during the respective interval to compute the trading volume. Finally, we use price adjustment factors from the Compustat Quarterly database to account for stock splits.

We use the full Dow Jones Edition of the RavenPack News Analytics database to compute the news score of each trading day. Out of all macroeconomic news related to topics of business and economics, we select those with the highest relevance (Event Novelty Score = 100).

Summary statistics appears in Table 2.5 in Appendix 2.B.

2.4.2 Stock-level data

We use data on 842 unique fund-announcement observations from in Section 2.3.2. For each announcement, we identify all ETFs that hold the announcing stock, and all constituents of such ETFs at the time of announcement. Next, we use the Compustat International daily data on prices and shares outstanding of all identified ETF constituents during period of 2016-2017.

We use the ETFDB data on weights and number of shares held by ETFs to compute the intensity of arbitrage, $\Delta_{i,j,t}$, and the weight of each stock in each specific ETF.

Summary statistics appears in Table 2.6 in Appendix 2.B.

2.5 Results

2.5.1 ETF-level results

Table 2.1 shows the estimation results of the empirical model in Equation 2.4: Panel A is for time windows before the announcement, and Panel B is for time windows after the announcement. The response in returns occurs mostly before the information release, which is in accordance with the literature on stock market information processing (Kim and Verrecchia 1997, Bamber, Barron, and Stevens 2011, Back, Crotty, and Li 2018, Yang, Zhang, and Zhang 2020). As estimates in Panel A suggest, the response of ETF returns to announcements is strongest for stocks in the top percentiles of ETF weight distribution. Specifically, the coefficients of the interaction terms $SUR * 1_{Weight>Q_{0.75}}$ become significant and positive 4 hours before the announcement[10], and the coefficients for third and second quartile become positive and significant 2 hours before the announcement. At the same time, there is no significant effect of the announcement on stocks with weight in the first quartile of the distribution. This means that the higher the weight of the stock the earlier and more efficiently traders would react to information about it.

[10] The coefficient 0.126 means that one unit increase in earnings surprise leads on average to 13% increase in log ETF return ($e^{0.126} = 1.13$).

These results provide a strong evidence in support of Corollaries 1.1 and 1.2: investors rationally adjust their attention in response to earnings announcements. These findings cannot be explained by liquidity and transaction costs, because we control for time, fund, and stock specific factors. Moreover, they are not consistent with the salience explanation that investors' attention is drawn to those earnings surprises which are most different relative to the average (Bordalo, Gennaioli, and Shleifer 2013).

We evaluate the model in Equation 2.7 to test Corollary 1.3: traders with less cognitive capacity acquire less information. The results are presented in Table 2.2: Panel A is for time windows before the announcement, and Panel B is for time windows after the announcement. We find the similar pattern in return responses: stocks response before the information release, and stocks with higher weights in ETFs start response to the announcement earlier than others. Notably, the results suggest that there is no difference in responses to announcements occurred on busy and non-busy days. Hence, the results of this exercise cannot be strongly interpreted in favor of Corollary 1.3. However, while these results hold for almost all time windows, one hour before the announcement the coefficients for three highest quartiles on busy days become positive and significant. Before that, the coefficients for these variables, while being insignificant, had negative sign. One of possible explanations for this is that on busy days traders react more sluggishly to announcements and they catch up with that information later. These results are consistent with Corollary 1.3, as well as with the distraction effect theory (Hirshleifer, Lim, and Teoh 2009): the arrival of extraneous news causes prices to react sluggishly to relevant news about a firm. However, we cannot distinguish between purely rational and behavioral explanations and, therefore, further research is needed.

Table 2.1: Earnings response coefficients of ETFs around earnings announcements

	Dependent variable: window return					
	Panel A. Before the announcement					
	-6h	-5h	-4h	-3h	-2h	-1h
SUR	0.108	0.071	-0.059	0.012	-0.039	0.023
	(0.139)	(0.132)	(0.100)	(0.087)	(0.088)	(0.038)
$1_{Weight\in[Q_{0.25},Q_{0.5}]}$	0.0002	0.0002	0.001	0.0003	0.001*	-0.0002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0002)
$1_{Weight\in[Q_{0.5},Q_{0.75}]}$	0.0001	0.0001	0.0003	0.0001	0.001	0.0002
	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0005)	(0.0003)
$1_{Weight>Q_{0.75}}$	-0.001	-0.001	0.0003	0.0001	0.001	0.00003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0003)
$SUR * 1_{Weight\in[Q_{0.25},Q_{0.5}]}$	-0.085	-0.047	0.006	0.017	0.153**	0.085**
	(0.077)	(0.075)	(0.064)	(0.057)	(0.072)	(0.042)
$SUR * 1_{Weight\in[Q_{0.5},Q_{0.75}]}$	-0.076	-0.042	-0.009	0.022	0.139**	0.059
	(0.066)	(0.068)	(0.059)	(0.055)	(0.069)	(0.036)
$SUR * 1_{Weight>Q_{0.75}}$	0.036	0.049	0.126**	0.134**	0.174**	0.045
	(0.094)	(0.094)	(0.061)	(0.054)	(0.082)	(0.037)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	913	917	917	917	921	923
R^2	0.639	0.603	0.843	0.808	0.764	0.799
	Panel B. After the announcement					
	+1h	+2h	+3h	+4h	+5h	+6h
SUR	0.028	-0.002	-0.087**	-0.041	-0.159	-0.091
	(0.026)	(0.042)	(0.043)	(0.039)	(0.228)	(0.407)
$1_{Weight\in[Q_{0.25},Q_{0.5}]}$	-0.0001	-0.0002	-0.001	-0.001**	-0.001	-0.004*
	(0.0002)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.002)
$1_{Weight\in[Q_{0.5},Q_{0.75}]}$	0.0001	0.0003	-0.0001	-0.0001	0.0004	-0.003
	(0.0003)	(0.0002)	(0.0004)	(0.0005)	(0.001)	(0.003)
$1_{Weight>Q_{0.75}}$	-0.0004	0.0001	-0.0004	-0.0004	0.0005	-0.004
	(0.0004)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.003)
$SUR * 1_{Weight\in[Q_{0.25},Q_{0.5}]}$	0.011	0.040	0.023	0.042	0.050	-0.017
	(0.012)	(0.037)	(0.036)	(0.042)	(0.051)	(0.080)
$SUR * 1_{Weight\in[Q_{0.5},Q_{0.75}]}$	0.007	0.037	0.015	0.025	0.037	-0.025
	(0.012)	(0.029)	(0.029)	(0.033)	(0.041)	(0.074)
$SUR * 1_{Weight>Q_{0.75}}$	-0.041	0.019	0.047	0.078**	0.066*	-0.207
	(0.038)	(0.034)	(0.032)	(0.033)	(0.038)	(0.185)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	923	923	923	923	923	923
R^2	0.753	0.787	0.627	0.718	0.745	0.702

*p<0.1; **p<0.05; ***p<0.01

Note: This table presents estimates of regression of ETF returns over specified announcement window (-6,...,+6 hours around the announcement) on a measure of earnings surprise of a stock within corresponding ETF, SUR. $I_{W/\in q}$ is the indicator function that takes the value of 1 if the weight of stock in the ETF is in the q^{th} quartile of ETF weights distribution. Standard errors are clustered at fund level and reported in parentheses. We use fund, stock, and day fixed effects. The description of variables is in Section 2.3.1. The sample period is 2016-2017.

Table 2.2: Earnings response coefficients of ETFs around earnings announcements - normal vs. busy days

	Dependent variable: window return					
	Panel A. Before the announcement					
	-6h	-5h	-4h	-3h	-2h	-1h
SUR	0.147	0.102	-0.114	0.008	-0.022	0.065*
	(0.138)	(0.134)	(0.082)	(0.064)	(0.126)	(0.035)
$SUR * 1_{Weight \in [Q_{0.25}, Q_{0.5}]}$	-0.192*	-0.103	-0.037	-0.139	0.116	0.006
	(0.110)	(0.101)	(0.082)	(0.094)	(0.240)	(0.030)
$SUR * 1_{Weight \in [Q_{0.5}, Q_{0.75}]}$	0.102*	0.084	0.081	0.075	0.186*	-0.021
	(0.052)	(0.072)	(0.067)	(0.063)	(0.108)	(0.020)
$SUR * 1_{Weight > Q_{0.75}}$	0.050	0.038	0.149**	0.117**	0.194	-0.032
	(0.117)	(0.119)	(0.062)	(0.054)	(0.133)	(0.021)
$SUR * BUSY$	0.015	0.075	0.873*	0.527	-0.146	-0.117
	(0.631)	(0.587)	(0.486)	(0.459)	(0.214)	(0.115)
$SUR * 1_{Weight \in [Q_{0.25}, Q_{0.5}]} * BUSY$	0.040	0.019	-0.009	0.134	-0.004	0.128*
	(0.137)	(0.135)	(0.107)	(0.115)	(0.227)	(0.067)
$SUR * 1_{Weight \in [Q_{0.5}, Q_{0.75}]} * BUSY$	-0.250**	-0.167	-0.146	-0.079	-0.086	0.130**
	(0.107)	(0.120)	(0.092)	(0.099)	(0.098)	(0.057)
$SUR * 1_{Weight > Q_{0.75}} * BUSY$	0.114	0.165	-0.076	0.006	-0.093	0.151**
	(0.284)	(0.277)	(0.164)	(0.159)	(0.145)	(0.072)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	902	906	906	906	910	912
R^2	0.638	0.598	0.841	0.802	0.758	0.770
	Panel B. After the announcement					
	+1h	+2h	+3h	+4h	+5h	+6h
SUR	0.054*	0.011	-0.036	0.017	-0.110	0.030
	(0.029)	(0.045)	(0.034)	(0.031)	(0.230)	(0.464)
$SUR * 1_{Weight \in [Q_{0.25}, Q_{0.5}]}$	−0.090**	-0.038	-0.045	-0.033	-0.037	-0.242
	(0.036)	(0.035)	(0.041)	(0.051)	(0.083)	(0.249)
$SUR * 1_{Weight \in [Q_{0.5}, Q_{0.75}]}$	0.0004	0.039***	0.008	0.003	-0.003	-0.065
	(0.016)	(0.014)	(0.020)	(0.038)	(0.041)	(0.159)
$SUR * 1_{Weight > Q_{0.75}}$	-0.058	-0.021	-0.022	-0.021	-0.075	-0.368
	(0.052)	(0.036)	(0.050)	(0.063)	(0.069)	(0.239)
$SUR * BUSY$	-0.084	0.059	−0.748***	−0.644*	-0.371	-0.358
	(0.089)	(0.607)	(0.280)	(0.362)	(0.377)	(0.679)
$SUR * 1_{Weight \in [Q_{0.25}, Q_{0.5}]} * BUSY$	0.107***	0.079	0.073	0.094	0.114	0.317
	(0.039)	(0.059)	(0.060)	(0.112)	(0.132)	(0.291)
$SUR * 1_{Weight \in [Q_{0.5}, Q_{0.75}]} * BUSY$	0.011	-0.002	0.008	0.039	0.063	0.127
	(0.024)	(0.048)	(0.042)	(0.099)	(0.098)	(0.215)
$SUR * 1_{Weight > Q_{0.75}} * BUSY$	0.0004	0.101	0.221	0.314	0.408*	0.757**
	(0.062)	(0.073)	(0.222)	(0.232)	(0.245)	(0.340)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912	912
R^2	0.732	0.776	0.622	0.701	0.720	0.741

*p<0.1; **p<0.05; ***p<0.01

Note: This table presents estimates of regression of ETF returns over specified announcement window (-6,...,+6 hours around the announcement) on a measure of earnings surprise of a stock within corresponding ETF, SUR. Other variables are: the indicator function that takes the value of 1 if the weight of stock in the ETF is in the q^{th} quartile of ETF weights distribution, $I_{W/\in q}$; dummy variable that takes the value 1 when the average news score or the number of relevant events on the U.S. market that day is larger than median, $BUSY$. Standard errors are clustered at fund level and reported in parentheses. We use fund, stock, and day fixed effects. The description of variables is in Section 2.3.1. The sample period is 2016-2017.

2.5.2 Stock-level results

Table 2.3 shows the results of the estimation of the empirical models (2.10)-(2.12). They suggest that the AIV of non-announcing constituent stocks increases with the AIV of their corresponding ETFs. Moreover, the effect is significant only around announcements for stocks that are above the 75^{th} percentile of ETF weight distribution, which is consistent with the results above. This result holds across all three models and suggests that ETFs could be a source of increased idiosyncratic stock-level volatility that is transferred to the underlying stocks.

We do not find any evidence that the weights of constituents stocks influence the relationship between the AIV of stocks and ETFs (column 2), nor do arbitrage trades of Authorized Participants (column 3). It contradicts the findings of Ben-David, Franzoni, and Moussawi (2017), who show that ETF-level shocks are translated to underlying stocks at larger magnitudes if the stock has greater weight in the ETF. However, Bhattacharya and O'Hara (2018) outline a possible no-arbitrage mechanism - direct learning from ETFs prices by stock investors that potentially could be observed in our setting too.

Additionally, the results of the estimation of the model (2.13) in Table 2.4 show that the AIV of ETFs is priced in the higher quartiles of ETF weights distribution. That is, the abnormal stock returns are loaded on the ETF's AIV, which is visible from the positive and significant coefficients on the AIV of stocks with weights primarily in fourth quartile. The relationship is then reversed after 10 days, implying that abnormal returns overreact to the ETF AIV, and the overreaction is subsequently corrected.

Table 2.3: The effect of the AIV of ETFs on the AIV of non-announcing stocks

	Dependent variable: AIV of non − announcing stock		
	(1)	(2)	(3)
AIV	0.002	-0.0004	-0.003
	(0.004)	(0.004)	(0.004)
$AIV*1_{W\in[Q_{0.25},Q_{0.5}]}$	-0.003	0.0002	0.001
	(0.004)	(0.004)	(0.004)
$AIV*1_{W\in[Q_{0.5},Q_{0.75}]}$	0.005	0.006*	0.004
	(0.004)	(0.004)	(0.004)
$AIV*1_{W>Q_{0.75}}$	0.020***	0.022***	0.021***
	(0.004)	(0.004)	(0.005)
$AIV*W$		0.007**	
		(0.003)	
$AIV*W*1_{W\in[Q_{0.25},Q_{0.5}]}$		-0.015**	
		(0.006)	
$AIV*W*1_{W\in[Q_{0.5},Q_{0.75}]}$		-0.005	
		(0.004)	
$AIV*W*1_{W>Q_{0.75}}$		-0.008**	
		(0.004)	
$AIV*\Delta$			0.588
			(0.778)
$AIV*\Delta*1_{W\in[Q_{0.25},Q_{0.5}]}$			-2.324
			(1.594)
$AIV*\Delta*1_{W\in[Q_{0.5},Q_{0.75}]}$			-2.053***
			(0.792)
$AIV*\Delta*1_{W>Q_{0.75}}$			-0.888
			(0.839)
Controls	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Observations	174,789	259,235	434,024
R^2	0.220	0.236	0.204

*p<0.1; **p<0.05; ***p<0.01

Note: This table presents estimates of regression of the non-announcing stock abnormal idiosyncratic volatility on the abnormal idiosyncratic volatility of ETF during the next open trading session after the announcement (AIV). Other variables are: the indicator for announcing stock in ETF on announcement day being in the q^{th} quartile of ETF weights distribution, I_W; the weight of non-announcing stock in ETF, W; intensity of arbitrage, Δ. $Controls$ include the inverse of price of non-announcing stock, $\frac{1}{P}$, the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$, the log of Amihud illiquidity measure of non-announcing stock, $log(Amihud)$, and the lagged returns of non-announcing stock ($ret_{[-1]}$, $ret_{[-3,-2]}$, $ret_{[-6,-4]}$). We use fund, stock, and day fixed effects. Standard errors are clustered at stock level and reported in parentheses. The description of variables is in Section 2.3.2. The sample period is 2016-2017.

Table 2.4: The effect of the AIV of non-announcing stock on abnormal stock returns

	Dependent variable: Fama-French Adjusted Cumulative Returns				
	$aret_{[t,t+1]}$	$aret_{[t,t+5]}$	$aret_{[t,t+10]}$	$aret_{[t,t+20]}$	$aret_{[t,t+30]}$
AIV	-0.0003**	-0.0003	-0.001***	-0.003***	-0.006***
	(0.0002)	(0.0002)	(0.0003)	(0.001)	(0.001)
$AIV*1_{W\in[Q_{0.25},Q_{0.5}]}$	0.0002	0.0002	0.001***	0.012***	0.015***
	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.002)
$AIV*1_{W\in[Q_{0.5},Q_{0.75}]}$	0.0003**	0.0001	0.001	-0.002*	0.001
	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.001)
$AIV*1_{W>Q_{0.75}}$	0.001***	0.001***	0.003***	0.0003	-0.002
	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.002)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	466,440	466,421	445,708	423,597	412,269
R^2	0.108	0.110	0.194	0.115	0.104

*p<0.1; **p<0.05; ***p<0.01

Note: This table presents estimates of regression of the non-announcing stock cumulative abnormal return for a given period on the abnormal idiosyncratic volatility of non-announcing stock within a given ETF (AIV). I_W is the indicator for announcing stock in ETF on announcement day being in the q^{th} quartile of ETF weights distribution. $Controls$ include the inverse of price of non-announcing stock, $\frac{1}{P}$, the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$, the log of Amihud illiquidity measure of non-announcing stock, $log(Amihud)$, and the lagged returns of non-announcing stock ($ret_{[-1]}$, $ret_{[-3,-2]}$, $ret_{[-6,-4]}$). We use fund, stock, and day fixed effects. Standard errors are clustered at stock level and reported in parentheses. The description of variables is in Section 2.3.2. The sample period is 2016-2017.

2.6 Conclusion

In this paper, we show that ETFs can be venues for stock-specific price discovery, and that their learning patterns of stock-specific information are consistent with rational inattention theory. Further, we show that these learning patterns are transferred to underlying bundles of ETFs, leading to increased price co-movements in constituent stocks around information releases. Therefore, stock specific shocks in the ETF can affect underlying market prices, even when such information is irrelevant for a particular underlying asset and, hence, it could lead to greater volatility overall. These results suggest that even rational behavior of constrained individuals combined with the design of the new financial instruments could be a potential weakness for the system, and should be taken into account when thinking about future regulations.

We highlight several directions for future investigation. First, while this paper provides empirical evidence suggesting that ETF prices reflect stock-specific information, it is not entirely clear why investors would trade ETFs instead of stocks around stock-specific news releases. One possible explanation is that ETFs are simply more liquid (Ernst 2021). Future work could analyze the liquidity of ETFs around earnings announcements of their constituents to shed light on this question.

Second, we do not find evidence that arbitrage explains information transfer from ETFs to underlying bundles. Therefore, it could be interesting to explore alternative mechanisms of information transfer, which could explain the learning process fully.

Finally, we find evidence of rational endogenous information acquisition. At the same time, the result, that there is a higher response of window returns to the earnings surprise on non-busy days, are inconclusive and also consistent with behavioral inattention theories (Hirshleifer, Lim, and Teoh 2009). Moreover, there is a question of whether investors face information costs or constraints (Azrieli 2021). Exploring and distinguishing different forces behind these results could be a fruitful direction for future research.

2.A Proofs

2.A.1 Proof of Lemma 1

We start by solving the maximization problem for given exogenous signals s_1 and s_2:

$$\max_{\Delta V} \mathbb{E}[-(w_1 \Delta V_1 + w_2 \Delta V_2 - \Delta V)^2 | s_1, s_2]. \tag{2.14}$$

The first order condition is:

$$\Delta V^* = \mathbb{E}[w_1 \Delta V_1 + w_2 \Delta V_2 | s_1, s_2].$$

Then we plug optimal ΔV^* into equation (2.14) and obtain:

$$\mathbb{E}[(w_1 \Delta V_1 + w_2 \Delta V_2 - \mathbb{E}[w_1 \Delta V_1 + w_2 \Delta V_2 | s_1, s_2])^2 | s_1, s_2]$$

$$= -w_1^2 Var[\Delta V_1 | s_1] - w_2^2 Var[\Delta V_2 | s_2]$$

$$= -w_1^2 \sigma_{1|s_1}^2 - w_2^2 \sigma_{1|s_2}^2.$$

Therefore, now we can reformulate the maximization problem in terms of conditional variances of changes in the values of individual stocks:

$$\max_{\sigma_{1|s_1}^2, \sigma_{1|s_2}^2} -w_1^2 \sigma_{1|s_1}^2 - w_2^2 \sigma_{1|s_2}^2, \tag{2.15}$$

subject to (2.1) and (2.2).

From the constraint (2.1) we obtain $\sigma_{1|s_1}^2 = e^{-2k \frac{\sigma_1^2 \sigma_2^2}{\sigma_{2|s_2}^2}}$ and substitute it to the maximization function (2.15):

$$\max_{\sigma_{2|s_2}^2} -w_1^2 \sigma_{2|s_2}^2 - w_2^2 e^{-2k} \frac{\sigma_1^2 \sigma_2^2}{\sigma_{2|s_2}^2}.$$

The first order conditions yields:

$$\sigma_{1|s_1}^2 = \frac{w_2}{w_1} \sqrt{e^{-2k} \sigma_1^2 \sigma_2^2}$$

$$\sigma_{2|s_2}^2 = \frac{w_1}{w_2} \sqrt{e^{-2k} \sigma_1^2 \sigma_2^2}.$$

Then we apply the non-forgetting constraint (2.2) and obtain Lemma 1.

2.A.2 The multi-dimensional rational inattention problem

Above, we considere the two-dimensional problem. The only difference now is that an ETF consists of $N \in \mathbb{R}$ independent stocks with weights w_i, $i \in 1,...,N$. Following the same steps as in Appendix 2.A.1, it is easy to show that the solution to this problem is:

$$\forall i \in 1,...,N: \quad \sigma^2_{i|s_i} = \sqrt[N]{\frac{\prod_{j=1}^{N} w_2^2}{w_i^4} e^{-2k} \prod_{j=1}^{N} \sigma_i^2}.$$

Therefore, the comparative statics results are similar to the two-dimensional problem, and hence the latter could be considered without loss of generality.

2.B Summary statistics

Table 2.5: Summary statistics for ETF-level analysis

Statistic	N	Mean	St. Dev.	Min	Q_{25}	Q_{75}	Max
$ret_{i,j,[-6,0]}$	913	0.001	0.010	−0.030	−0.002	0.004	0.182
$ret_{i,j,[-5,0]}$	917	0.001	0.009	−0.043	−0.002	0.003	0.182
$ret_{i,j,[-4,0]}$	917	0.001	0.006	−0.043	−0.002	0.003	0.028
$ret_{i,j,[-3,0]}$	917	0.0005	0.005	−0.026	−0.001	0.002	0.023
$ret_{i,j,[-2,0]}$	921	0.0003	0.004	−0.029	−0.001	0.001	0.022
$ret_{i,j,[-1,0]}$	923	0.00005	0.002	−0.025	−0.0004	0.001	0.012
$ret_{i,j,[0,1]}$	923	0.00001	0.002	−0.022	−0.0004	0.001	0.021
$ret_{i,j,[0,2]}$	923	0.00002	0.003	−0.025	−0.001	0.001	0.027
$ret_{i,j,[0,3]}$	923	−0.0001	0.005	−0.100	−0.001	0.001	0.030
$ret_{i,j,[0,4]}$	923	−0.00004	0.006	−0.100	−0.001	0.002	0.030
$ret_{i,j,[0,5]}$	923	−0.00003	0.007	−0.100	−0.001	0.002	0.030
$ret_{i,j,[0,6]}$	923	0.0003	0.014	−0.100	−0.002	0.002	0.375
SUR	932	−0.0002	0.027	−0.417	−0.002	0.002	0.496
$SUR1_{Weight<Q_{0.25}}$	161	0.0018	0.0416	-0.1006	-0.0023	0.0018	0.4962
$SUR1_{Weight\in[Q_{0.25},Q_{0.5}]}$	205	-0.0021	0.0337	-0.4168	-0.0011	0.0021	0.1743
$SUR1_{Weight\in[Q_{0.5},Q_{0.75}]}$	263	0.0012	0.0225	-0.1504	-0.0025	0.0020	0.1743
$SUR1_{Weight>Q_{0.75}}$	968	-0.0011	0.0079	-0.0419	-0.0017	0.0012	0.0227
$BUSY$	919	0.408	0.492	0.000	0.000	1.000	1.000

Note: The variables in the table are: ETF returns over specified announcement window (-1,...,+6 hours around the announcement), ret; measure of earnings surprise of a stock within a corresponding ETF, SUR; the indicator function that takes the value of 1 if the weight of stock in ETF is in the q^{th} quartile of ETF weights distribution, $I_{W\in q}$; dummy variable that takes the value 1 when the average news score or the number of relevant events on the U.S. market that day is larger than median, $BUSY$. The description of variables is in Section 2.3.1. The sample period is 2016-2017.

Table 2.6: Summary statistics for stock-level analysis

Statistic	N	Mean	St. Dev.	Min	Q_{25}	Q_{75}	Max
$aret_j$	497,538	0.0004	0.019	−0.872	−0.009	0.008	0.536
$aret_i$	497,538	−0.0002	0.006	−0.050	−0.002	0.002	0.043
AIV_i	497,538	−0.245	0.713	−6.334	−0.686	0.229	2.243
AIV_j	497,515	−0.363	0.646	−5.074	−0.720	0.073	2.088
$\frac{1}{P}$	497,538	0.823	43.393	0.00000	0.024	0.184	9,015.100
log(Mkt Cap)	497,538	22.117	1.936	14.275	20.653	23.322	32.067
$ret_{[-1]}$	497,538	0.001	0.023	−1.361	−0.008	0.010	2.324
$ret_{[-3,-2]}$	497,529	0.0003	0.035	−2.335	−0.008	0.009	2.341
$ret_{[-6,-4]}$	497,486	0.002	0.056	−2.362	−0.011	0.015	2.345
Amihud	471,070	0.00000	0.00004	0.000	0.000	0.00000	0.016
log(Amihud)	466,512	−18.003	2.340	−49.451	−19.588	−16.397	−4.146
Δ	463,472	−0.00001	0.003	−0.850	0.000	0.000	0.850
W	497,538	0.167	1.194	0.000	0.010	0.100	99.400

Note: The variables in the table are: Fama-French adjusted cumulative abnormal returns of non-announcing stock, $aret_j$, and ETF, $aret_i$; the abnormal idiosyncratic volatility of ETF, AIV_i, and non-announcing stock in given ETF, AIV_j; the inverse of price of non-announcing stock, $\frac{1}{P}$; the log of market capitalization of non-announcing stock, $log(Mkt\ Cap)$; the lagged returns of non-announcing stock ($ret_{[-1]}$, $ret_{[-3,-2]}$, $ret_{[-6,-4]}$); Amihud illiquidity measure of non-announcing stock, *Amihud*, and the log of it, $log(Amihud)$; intensity of arbitrage, Δ; and the weight of non-announcing stock in ETF, W. The description of variables is in Section 2.3.2. The sample period is 2016-2017.

Chapter 3

Optimal Menu when Agents Make Mistakes

3.1 Introduction

In real life, we often face choices from a discrete menus; for example, when choosing an insurance plan, a school for our children, or a pension fund. When confronted with these important decisions, we often make mistakes for two potential reasons. First, we may misperceive the true properties of alternatives, i.e., we have imperfect information. For example, individuals can be uninformed and underestimate potential cost savings from changing prescription drug plans (Kling et al. 2012), not be fully informed about crucial aspects of an insurance plan (Handel and Kolstad 2015), and, when choosing a car, we may think of fuel costs as scaling linearly in miles per gallon instead of gallons per mile (Allcott 2013). Second, we can misperceive our own tastes, i.e., we have imperfect self-knowledge. For example, individuals overestimate their attendance and their likelihood of cancelling automatically renewed memberships when choosing a gym contract (DellaVigna and Malmendier 2006). We are generally myopic in decision-making, can lack skill predicting our own tastes and risk preferences, and we can be led to erroneous choices thought by fallible memory and incorrect evaluation of past experiences (Kahneman 1994, Heckman, Jagelka, and Kautz 2021).

In the examples above, a government or other social planner can regulate the size of the menu from which consumers choose and the properties of alternatives within it. The social planner cannot possibly know the individual tastes of a particular agent and, hence, is not able to provide the first best alternative for each agent. However, knowing characteristics of the overall population, including probabilities of mistakes and distribution of tastes, he can construct a menu of alternatives, referred to as an *optimal menu*, that maximizes the sum of the expected utilities of agents.

I analyze an optimal menu under the assumptions that agents misperceive either the true properties of available alternatives or their own tastes. In two limiting cases, when the misperception is insignificant or consumers choose an alternative randomly, the optimum menu is identical under both types of misperceptions. For intermediate degrees of rationality, dependence of the optimum choice set on the precision of choice is complex. We use a simple setting and numerical calculations, and demonstrate that when agents misperceive the available options, it is optimal to limit choices when the probability of making mistakes is moderately high. Further, it could be optimal to construct a menu with more distinct alternatives. In contrast, when agents misperceive their own tastes, it is optimal to limit choice only when agents choose randomly, and to propose alternatives that are more similar when there is a greater probability of a mistake.

The intuition behind the results is that, when agents misperceive the properties of alternatives, every additional alternative in the menu has the benefit of providing more choice (matching the agents' taste more precisely) at the cost of increasing the probability and magnitude of mistakes. Thus, the more similar the alternatives are, the more difficult it is for the agent to differentiate between them. Therefore, it could be optimal to construct a menu with more distinct alternatives, to decrease the probability of a mistake, depending on the distribution of tastes in the population. When the probability of a mistake is large, it becomes optimal to remove options that could induce large utility loss, and to leave one option that matches the mean taste in the population.

In contrast, when agents have imperfect self-knowledge, the probability of a mistake depends only on the midpoints between properties of alternatives. Thus, the probability of a mistake would not be decreased if alternatives were differentiated. Moreover, since the probability of mistakes affected by alternatives linearly, it is weakly beneficial to introduce more alternatives into the menu.

The discussion about individuals misperceiving the true properties of alternatives and accordingly failing to choose the best one goes back at least as far as Luce (1959), who analyzes agent choice subject to random noise. Mirrlees (1987) and Sheshinski (2016) study the welfare maximization problem when agents misperceive the true properties of alternatives. They show that, while, the choice should not be limited when the agents are completely rational, the optimum choice-set is a singleton when the probability of a mistake is relatively high. In contrast, this paper focuses on comparing optimal menu allocations in two situations: when the agent misperceives either the true properties of alternatives or her own taste.

In recent years, a growing literature in industrial organization has analyzed the situation in which a firm interacts with boundedly rational agents. For a classic textbook treatment, see Anderson, de Palma, and Thisse (1992); more recent papers include Kamenica (2008), Hefti (2018), Persson (2018), and Gerasimou and Papi (2018). A review of other studies on complexity and manipulation can be found in Spiegler (2016). The main focus of this literature is on the market environment, and the agents' limitations arise solely from misperception of the true properties of available alternatives. This study considers two sources of mistakes and focuses on the welfare maximization problem.

In addition, this paper proposes a new explanatory insight into the choice paradox (Schwartz 2004), i.e., the effect when a larger choice set sometimes decreases the satisfaction of individuals and ultimately can lead to rejection of an offer. This phenomenon has been observed, for example, when consumers purchased jam and chocolate (Iyengar and Lepper 2001) and when they made more important decisions such as a choice of 401k pension plans (Iyengar, Huberman, and Jiang 2004), or decided on participation in an election (Nagler 2015)[1]. Several studies suggest that the existence of the choice paradox and the efficiency of corresponding interventions, such as categorization of goods, depend on whether consumers are familiar with products or not (Chernev 2003, Mogilner, Rudnick, and Iyengar 2008). There are numerous models that attempt to explain this evidence (Irons and Hepburn 2007, Sarver 2008, Ortoleva 2013, Kuksov and Villas-Boas 2010). While my study does not focus on a particular mechanism, it suggests that the existence of this phenomenon and relevant interventions depend on the source of mistakes in the decision making process. Thus, when agents misperceive the true properties of alternatives, we can observe choice overload, and limiting the menu size could be a welfare maximizing intervention. However, when agents have imperfect self-knowledge, we would not observe the choice overload and, hence, should not limit the choice.

The rest of the paper is organized as follows. The next section presents the model setup. Section 3.3 discusses a simple model with two agents, to illustrate the intuition behind the results, and then provides numerical simulations with populations of agents.

[1] Further discussion on empirical evidence when choice opportunities can harm consumer can be found, for example, in Scheibehenne, Greifeneder, and Todd (2010) or Chernev, Bockenholt, and Goodman (2015).

The last section concludes.

3.2 Model

A population of $M \geq 2$ agents chooses from a set of $N \geq 2$ alternatives. The utility of the agent $i \in \{1, ..., M\}$ from the alternative $j \in \{1, ..., N\}$ is $U_i^j = -(t_i - v^j)^2$, where $t_i \in \mathbb{R}$ is the taste (bliss point) of i and $v^j \in \mathbb{R}$ is the property of j. $T \geq 2$ is the number of unique tastes in the population. The agent misperceives parameters of the model. I describe two versions of the model:

– **with misperceived true properties of alternatives:** the agent observes a signal $\vartheta_i^j = v^j + e_i^j$, where v^j is a true property of the option, and noise e_i^j is a random variable drawn from the distribution with mean zero and variance σ_i^j. She chooses the alternative with the signal that is a closest match to her taste[2], i.e., solves the following problem:

$$\max_{j \in \{1,...,N\}} -(t_i - \vartheta^j)^2.$$

– **with misperceived own true taste:** the agent observes a signal $\tau_i = t_i + e_i$, where t_i is the true taste of the agent, and noise e_i is a random variable drawn from the distribution with mean zero and variance σ_i. She chooses the alternative with the property that is a closest match to the signal of her taste, i.e., solves the following problem:

$$\max_{j \in \{1,...,N\}} -(\tau_i - v^j)^2.$$

In both versions of the model, if there are several alternatives that solve the agent's problem, then the agent chooses randomly between them.

The social planner maximizes overall welfare by choosing a number and properties of available alternatives, i.e., the optimal menu:

$$\max_{N, v^j \forall j \in \{1,...,N\}} \sum_{i=1}^{M} \sum_{j=1}^{N} P_i^j U_i^j,$$

where P_i^j is the probability that the agent i chooses option j. I assume that $N \leq T$: the maximum number of options that the social planner could propose is equal to the

[2]For discussion on when this behavior is optimal for the agent, see Weibull, Mattsson, and Voorneveld (2007).

number of tastes in the population.[3]

The problem has the following time-line:

1. The social planner observes (i) distributions of mistakes, and (ii) what the tastes in the population are, and (iii) the number of agents with each taste.

2. He chooses the optimal menu.

3. Agents observe signals.

4. They choose an alternative from the menu.

3.3 Solution

The solution to the welfare maximization problem depends on the size of the noise. Regardless of the source of mistakes, when there is no noise, the social planner creates a menu with alternatives that match tastes perfectly; when noise is infinite, it is optimal to limit choices and provide only one alternative that matches the mean taste in the population. This result is formalized in Propositions 1 and 2.

Proposition 8. *If* $\sigma_i^j = 0$ *or* $\sigma_i = 0$ $\forall (i,j)$*, then* $N = T$*,* $v^j = t_i$*.*

Proof. Since $U_i \leq 0$ $\forall i \Rightarrow \max(\sum_{i=1}^{M}\sum_{j=1}^{N} P_i^j U_i^j) = 0$ which is obtained when $N = T$, $v^j = t_i$. □

Proposition 9. *If* $\sigma_i^j \to \infty$ *or* $\sigma_i \to \infty$ $\forall (i,j)$*, then* $N = 1$ *and* $v^j = \frac{\sum t_i}{M}$*.*

Proof. If $\sigma_i^j \to \infty$ or $\sigma_i \to \infty$, then all alternatives are a priori the same for agents $P_i^j = \frac{1}{N}$. The solution to the welfare maximization problem is $N = 1$ and $v^j = \frac{\sum t_i}{M}$. □

In the next subsection, I illustrate the solution to the model for the intermediate cases using a model with uniformly distributed noise and two agents. Then I show that the results obtained are valid for the larger population of agents with continuous distribution of noise using numerical simulations.

[3] I make this assumption because the welfare function is not monotone in the number of options: for example, if for a given distribution, the optimal number of alternatives is 4, then the solution to the welfare maximization problem automatically includes any number that is divisible by 4.

3.3.1 Two agents

There are two agents, $i \in \{1, 2\}$, with tastes symmetrically allocated around zero, $t_1 = -t_2 < 0$.[4] The social planner could propose at most two options, $j \in \{1, 2\}$. I assume that $v^1 \leq v^2$. The situation when $v^1 = v^2$ is identical to the situation when the social planner proposes only one alternative and limits the agents' choice.

I assume that the noise is uniformly distributed, e_i^j and $e_i \sim U(-b, +b)$. Therefore, the social planner expects that agent 1 chooses the first option with probability P_1^1 and the second option with probability P_1^2. Agent 2 chooses similarly.

In the case of misperceived true properties of alternatives, the probabilities are as follows:

$$\begin{aligned} P_1^1 &= \min\left(1, \max\left(0,\ 1 - 0.5 * (\tfrac{v^1 - v^2 + 2b)}{2b})^2\right)\right), \quad & P_1^2 = 1 - P_1^1. \\ P_2^1 &= \min\left(1, \max\left(0,\ 0.5 * (\tfrac{v^1 - v^2 + 2b)}{2b})^2\right)\right), \quad & P_2^2 = 1 - P_2^1. \end{aligned}$$

In the case of misperceived true own tastes, the probabilities are as follows:

$$\begin{aligned} P_1^1 &= \min\left(1, \max\left(0,\ \frac{\frac{v^1+v^2}{2} - (t_1 - b)}{2b}\right)\right), \quad & P_1^2 = 1 - P_1^1. \\ P_2^1 &= \min\left(1, \max\left(0,\ \frac{\frac{v^1+v^2}{2} - (t_2 - b)}{2b}\right)\right), \quad & P_2^2 = 1 - P_2^1. \end{aligned}$$

To obtain an analytical solution for both cases, I have to make an additional heroic assumption that $v^1 = -v^2$.[5] This symmetry assumption simplifies the characterization of the solution.[6] The solution to the welfare maximization problem is formalized in Propositions 10 and 11.

Proposition 10. *In the case of misperceived true values of alternatives, the welfare maximization problem has the following solution:*

- ***small noise*** *($b \leq |t_i|$): $v^1 = -v^2 = t_1$;*
- ***medium noise*** *($|t_i| < b < 4|t_i|$): $v^1 = -v^2 = \frac{-b^2 - 4bt_1}{3t_1}$;*
- ***large noise*** *($4|t_i| \leq b$): $v^1 = v^2 = 0$.*

Proof. See Appendix 3.A. □

[4]It is without loss of generality, because, for any two distinct tastes one always can re-scale tastes to be symmetrically allocated around zero.

[5]It also could be interpreted as if the welfare function satisfies the Ralwsian principle of social justice, i.e., overall welfare is based on the welfare of society's worse-off member.

[6]Without a symmetry assumption, the solution for the situation when the agent misperceives the true properties of alternatives would be asymmetrical and (possibly) not unique. The solution for the situation when the agent has imperfect self-knowledge would be the same.

Proposition 11. *In the case of misperceived true own tastes, the welfare maximization problem has the following solution:*

– ***small noise*** *($b \leq |t_i|$): $v^1 = -v^2 = t_1$;*

– ***medium and large noise*** *($|t_i| < b$): $v^1 = -v^2 = -\frac{t_1^2}{b}$.*

Proof. See Appendix 3.B. □

Accordingly, when the noise is small ($b \leq |t_i|$), in both cases the social planner proposes options that match the tastes of the agents perfectly, and they choose the option closest to their true taste with certainty. When the noise is significantly large ($|t_i| < b$), then the solution depends on the source of mistakes. If agents misperceive the true properties of alternatives, it is optimal to limit the choice when the noise is finitely large. However, when agents misperceive their tastes, it is optimal to propose two alternatives with different properties for any finite noise.

In addition, if agents misperceive the true properties of alternatives, there exists noise ($|t_i| < b < 2|t_i|$) when the difference in the properties of proposed alternatives increases in the noise, i.e., $\frac{\partial v^1}{\partial b} < 0$ and $\frac{\partial v^2}{\partial b} > 0$. However, if agents misperceive their tastes, the social planner always proposes alternatives that are more similar as the noise becomes greater.

Intuition

The results are driven by the fact that if a taste is unclear, the distance between true taste and the properties of the options is distorted in the same way for all options, while if the properties of the options are unclear, this distortion is different for any option.

In particular, denote t_1 as t and v^1 as v and rewrite the probability that the agent makes the wrong choice (i.e., she chooses the alternative that is not the closest to her true taste) as follows. In the case of misperceived true properties of alternatives:

$$P_1^2 = P_2^1 = 0.5(\frac{2v + 2b}{2b})^2.$$

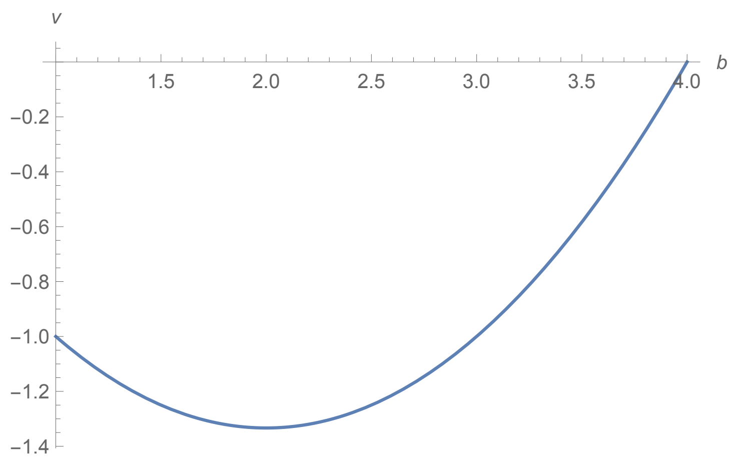
In the case of misperceived true own tastes:

$$P_1^2 = P_2^1 = \frac{\frac{v+(-v)}{2} - t + b}{2b}.$$

Therefore, when the noise originates from the misperception of alternatives, placing options close to each other increases the probability of a mistake, which is a nonlinear function of v. Thus, there is an inverted U-shaped curvilinear relationship between the optimal property of the alternative v and the size of the noise, as depicted in Figure 1. Thus, when the noise is significant, but still small ($|t_i| < b < 2|t_i|$), the social planner wants to distance the properties of alternatives from each other. In this situation, the loss from the decrease in utility, if the correct choice is made, is smaller than the gain from the decrease in the probability of making the wrong choice. However, when the noise is moderately large ($2|t_i| \leq b < 4|t_i|$), it is not profitable to distance the properties of alternatives farther away from each other. The loss from the decrease in utility in the case of the correct choice outweighs the gain from the decrease in the probability of the wrong choice. Therefore, the social planer chooses properties of alternatives closer to each other. When the probability of making the wrong choice is significantly high ($4|t_i| \leq b$), it is optimal to propose alternatives with identical properties.

However, when agents misperceive their tastes, the probability of making a mistake depends linearly on the midpoint between properties of alternatives. Therefore, it is not beneficial to differentiate properties of alternatives, since doing so does not decrease the probability of making the wrong choice. Accordingly, the social planner chooses v by equalizing the marginal gain of locating an option closer to the center for the second agent (reducing the loss in the case of making the wrong choice) and marginal loss for the first agent (reducing the gain in the case of making the correct choice), given the probabilities of making mistakes.

Figure 3.1: Optimal property of the first alternative as a function of b and $t_1 = -1$.



3.3.2 Many agents

Setup

There is a single-peaked population of agents with a variety of tastes $T = 7$. When agents misperceive the true properties of alternatives, e_i^j is assumed to be identically and independently Gumbel distributed. The Gumbel distribution has fatter tails than a Normal distribution; however, the difference between them is often indistinguishable empirically (Train 2002). At the same time, the difference between Gumbel distributed variables, which is used to calculate the probabilities of an agent's choices, follows the Logistic distribution. This significantly simplifies the numerical simulation. Therefore, the probability that agent i chooses option j is:

$$P_i^j = \frac{\exp(U_i^j/\lambda)}{\sum_i^N \exp(U_i^j/\lambda)}.$$

When agents misperceive their own true tastes, e_i is assumed to be identically and independently Logistic distributed.[7] In this case, the probability that agent i chooses option j is:

$$P_i^j = \int_{\frac{v^{j-1}+v^j}{2}}^{\frac{v^j+v^{j+1}}{2}} \frac{\exp(\frac{t_i-v^j}{0.5\lambda})}{0.5\lambda(1+\exp(\frac{t_i-v^j}{0.5\lambda}))^2} dv^j.$$

In both situations, higher values of λ correspond to larger variance and, hence, to a greater probability of making a mistake. I solve for every possible menu size and then select the one that maximizes welfare.[8]

Results

The solution with the optimal number of alternatives and optimal menu allocation is presented in Figures 2-5 for different λ. The grey bars (histogram) correspond to the number of agents with a particular taste. The optimal properties of alternatives are defined by vertical lines. The optimal number of options is stated above the graphs. In some situations, there are fewer vertical lines than the optimal number of alternatives, because there are several identical options that match the same taste. Intuitively, additional options with repeated values increase the probability that agents will choose a particular

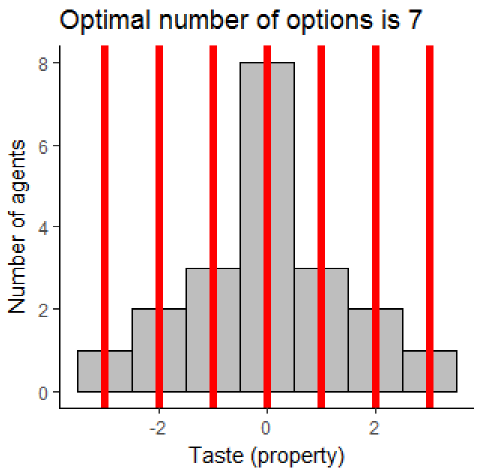
[7]In this case, I do not use the Gumbel distribution, since it is asymmetric. The asymmetry property skews the optimal menu, complicating the visual comparison. However, the qualitative results of the welfare analysis with the Gumbel distribution are identical to the analysis with the Logistic distribution.

[8]Calculations are performed in R using the "optimx" package.

alternative. Thus, when one taste is more salient in the population, it is beneficial to highlight the alternative that matches this taste.[9,10]

Figure 2 shows that, when the noise is small, it is optimal to provide alternatives that match tastes perfectly under both kinds of mistakes.

Figure 3.2: Optimal menu allocation when agents misperceive the true properties of alternatives or their own tastes, and $\lambda = 0.1$. The red lines indicate the optimal properties of alternatives. The histogram shows the distribution of agents.



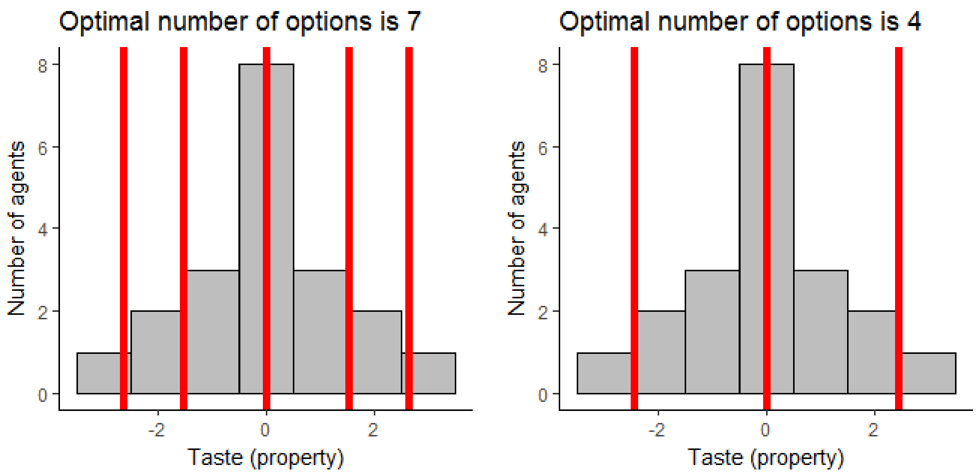
Figures 3 and 4 show the optimal menus for the situation when the noise is significantly large. When agents misperceive the true properties of alternatives, it is optimal to limit the choice (Figures 3). When the probabilities of making mistakes increase, the social planner decreases the menu size. When agents misperceive their own taste, it is not optimal to limit their choice (Figures 4). Thus, the social planner proposes 7 alternatives with unique properties for any noise. When the probabilities of making mistakes increase, he allocates alternatives closer to each other and to the mean taste in the population.

It is worth noticing that the effect of the decrease in the inequality of the tastes is similar to the decrease in noise. Figure 5 shows the optimal menu allocation for different populations of agents with the same variety of tastes $T = 7$, but with lower density of agents with the most frequent (mode) taste $t_{mode} = 0$. In this situation, when agents

[9]Mirrlees (2017) refers to such manipulation as "advertising". One possible type of "advertising" is nudges. For example, it was shown that setting an option as a default increases the probability that this alternative will be chosen. See Thaler and Sunstein (2008) for additional discussion on the topic.

[10]One way to avoid the presence of identical options in the menu is to introduce the following probability function: $P_i^j = \frac{m(j)P_i^j}{\int m(y)P_i^y dy}$, where $m(j)$ is a density of alternatives with identical properties (Mirrlees 2017). This formula relates to the modified multinomial logit model by Matějka and McKay (2015). Accordingly, another possible explanation for the "advertising" effect is prior knowledge of agents about options in a menu.

Figure 3.3: Optimal menu allocation when agents misperceive the true properties of alternatives for different noise ($\lambda = 1$ on the left and $\lambda = 2$ on the right graph). The red lines indicate the optimal properties of alternatives. The histogram shows the distribution of agents.



misperceive the true properties of alternatives (left graph, Figure 5), the social planner proposes more alternatives to agents, compared to the optimal menu for a population with higher density of agents with mode taste (left graph, Figure 3). Similarly, when agents misperceive their own tastes (right graph, Figure 5), the social planner proposes 7 alternatives, but allocates them further away from each other and from the mean taste in the population, compared to the optimal menu for a population with a higher density of agents with mode taste (left graph, Figure 4).

Figure 3.4: Optimal menu allocation when agents misperceive their own tastes for different noise ($\lambda = 1$ on the left and $\lambda = 2$ on the right graph). The red lines indicate the optimal properties of alternatives. The histogram shows the distribution of agents.

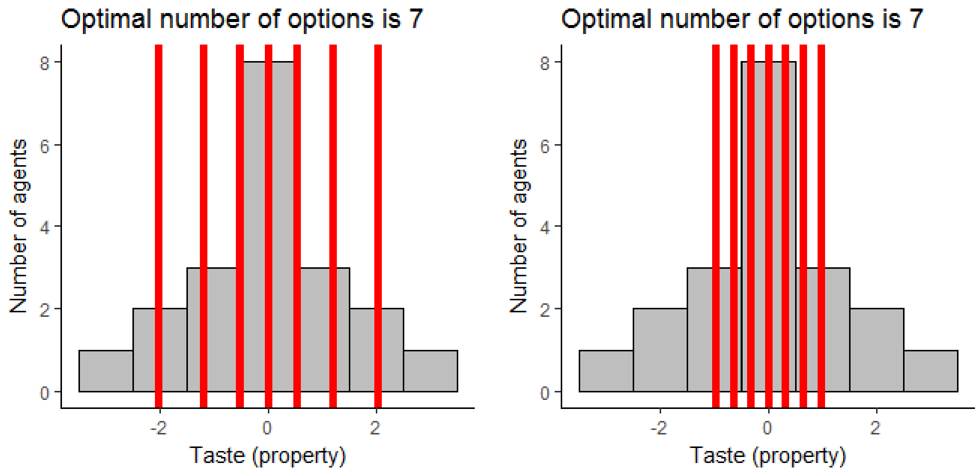
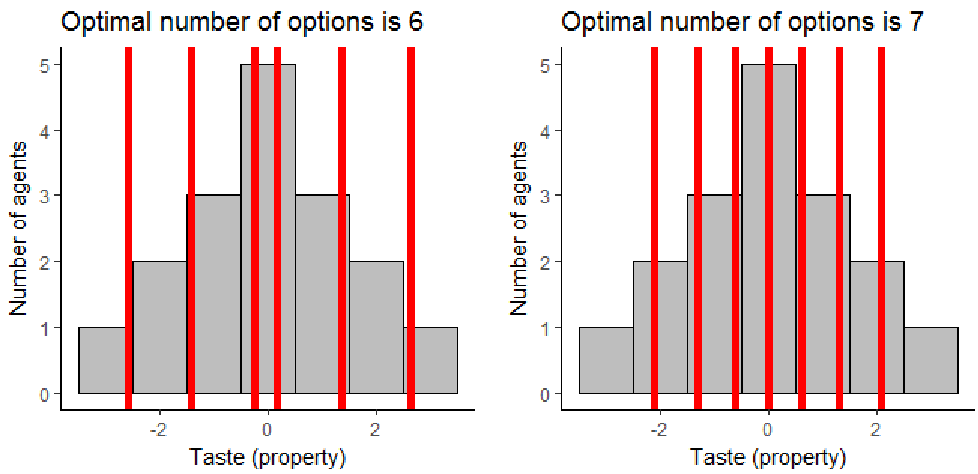


Figure 3.5: Optimal menu allocation when agents misperceive the true properties of alternatives (left graph) or their own tastes (right graph) and $\lambda = 1$. The red lines indicate the optimal properties of alternatives. The histogram shows the distribution of agents.



3.4 Conclusion

Although there is a large body of literature that studies problems with agents who make mistakes, there is still a lack of studies that analyze a discrete choice problem with heterogeneous agents and a social planner. This paper provides the solution to the welfare maximization problem, and shows that if agents misperceive the true properties of alternatives, the optimal menu differs significantly from one when agents misperceive their own tastes. Therefore, this study suggests that, when designing a menu set, one should take into account not only the demand for a particular alternative, but also the probability and source of making a mistake.

3.A Proof of Proposition 3

Because of the symmetry assumption, the welfare maximization problem could be reduced to the choice of one variable $v^1 = v \leq 0$. I denote $t_1 = t < 0$. If $b < |t|$, then the probability of a mistake equals zero and the first best allocation is optimal. Therefore, I consider a situation when $b \geq |t|$ and $0 \leq P_i^j \leq 1\,\forall i, j$. Then the welfare maximization problem is the following:

$$\max_v W(v) = \big\{(1 - 0.5(\frac{2v+2b}{2b})^2) \cdot -2(t-v)^2 + 0.5(\frac{2v+2b}{2b})^2 \cdot -2(t+v)^2\big\}.$$

The derivative with respect to v is the following:

$$0.5(t-v)^2(2b+2v) - 0.5(t+v)^2(2b+2v) - \\ 0.25(t+v)(2b+2v)^2 + 2(t-v)(b^2 - 0.125(2b+2v)^2) = 0.$$

This equation has two solutions:

$$v = 0,$$
$$v = \frac{-b^2 - 4bt}{3t}.$$

Since $v \leq 0$, the second solution exists only for $b \leq 4|t|$. Moreover, when $b = 4|t|$, then $v = 0$ and the two solutions coincide. In this situation the welfare is $W(b = 4|t|) = -2t^2$. At the same time, if one substitutes $v = \frac{-b^2-4bt}{3t}$ into the maximization problem, then $W(b = |t|) = 0$ and $W > -2t^2$ for any $|t| < b < 4|t|$. Therefore, for $b < 4|t|$ the welfare is maximized when $v = \frac{-b^2-4bt}{3t}$; for $b \geq 4t$ it is optimal to provide the menu with two identical alternatives $v^1 = v^2 = 0$. □

3.B Proof of Proposition 4

If the $b < |t|$, then the probability of a mistake equals zero and the first best allocation is optimal. Therefore, I consider a situation when $b \geq |t|$ and $0 \leq P_i^j \leq 1\,\forall i, j$. Then the welfare maximization problem is the following:

$$\max_v \big\{\frac{-t+b}{2b} \cdot -(t-v)^2 + (1 - \frac{-t+b}{2b}) \cdot -(t+v)^2 + \\ \frac{t+b}{2b} \cdot -(t+v)^2 + (1 - \frac{t+b}{2b}) \cdot -(t-v)^2\big\}.$$

The derivative with respect to v is the following:

$$-4(t^2 + bv) = 0.$$

Therefore, the solution to the welfare maximization problem is:

$$v = -\frac{t^2}{b}.$$

□

Bibliography

Acharya, Sushant, and Shu Lin Wee. 2020. "Rational inattention in hiring decisions." *American Economic Journal: Macroeconomics* 12 (1): 1–40.

Ahern, Kenneth R, and Amy K Dittmar. 2012. "The changing of the boards: The impact on firm valuation of mandated female board representation." *The Quarterly Journal of Economics* 127 (1): 137–197.

Allcott, Hunt. 2013. "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market." *American Economic Journal: Economic Policy* 5 (3): 30–66.

Anderson, P. Simon, Andre de Palma, and Jacques-Francois Thisse. 1992. *Discrete Choice Theory of Product Differentiation.* Cambridge, MA: MIT Press.

Antonovics, Kate, and Ben Backes. 2014. "The effect of banning affirmative action on college admissions policies and student quality." *Journal of Human Resources* 49 (2): 295–322.

Appel, Ian R, Todd A Gormley, and Donald B Keim. 2016. "Passive investors, not passive owners." *Journal of Financial Economics* 121 (1): 111–141.

Azrieli, Yaron. 2021. "Constrained versus unconstrained rational inattention." *Games* 12 (1): 3.

Back, Kerry, Kevin Crotty, and Tao Li. 2018. "Identifying information asymmetry in securities markets." *The Review of Financial Studies* 31 (6): 2277–2325.

Bamber, Linda Smith, Orie E Barron, and Douglas E Stevens. 2011. "Trading Volume Around Earnings Announcements and Other Financial Reports: Theory, Research Design, Empirical Evidence, and Directions for Future Research." *Contemporary Accounting Research* 28 (2): 431–471.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilova, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *American Economic Review* 106 (6): 1437–1475.

Basak, Suleyman, and Anna Pavlova. 2013. "Asset prices and institutional investors." *American Economic Review* 103 (5): 1728–58.

Becker, Gary S. 1957. *The economics of discrimination.* University of Chicago press.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi. 2017. "Exchange-traded funds." *Annual Review of Financial Economics* 9:169–189.

———. 2018. "Do ETFs increase volatility?" *The Journal of Finance* 73 (6): 2471–2535.

Ben-David, Itzhak, Francesco A Franzoni, Byungwook Kim, and Rabih Moussawi. 2021. "Competition for Attention in the ETF Space."

Bertrand, Marianne, Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney. 2019. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway." *The Review of Economic Studies* 86 (1): 191–239.

Besley, Timothy, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. "Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden." *American Economic Review* 107 (8): 2204–2242.

Bhattacharya, Ayan, and Maureen O'Hara. 2018. "Can ETFs increase market fragility? Effect of information linkages in ETF markets." *Effect of Information Linkages in ETF Markets (April 17, 2018).*

———. 2020. *Etfs and systemic risks.* CFA Institute Research Foundation.

Bhojraj, Sanjeev, Partha Mohanram, and Suning Zhang. 2020. "ETFs and information transfer across firms." *Journal of Accounting and Economics* 70 (2-3): 101336.

Blankespoor, Elizabeth, Ed deHaan, and Ivan Marinovic. 2020. "Disclosure processing costs, investors' information choice, and equity market outcomes: A review." *Journal of Accounting and Economics* 70 (2-3): 101344.

Bohren, Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope. 2020. "Inaccurate Statistical Discrimination: An Identification Problem."

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and asset prices." *American Economic Review* 103 (3): 623–28.

Caplin, Andrew, and Mark Dean. 2013. "The Behavioral Implications of Rational Inattention with Shannon Entropy."

———. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review* 105 (7): 2183–2203 (July).

Caplin, Andrew, Mark Dean, and John Leahy. 2019. "Rational inattention, optimal consideration sets, and stochastic choice." *The Review of Economic Studies* 86 (3): 1061–1094.

———. 2021. "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy."

Caplin, Andrew, and Daniel Martin. 2017. "Defaults and Attention: The Drop Out Effect." *Revue économique* 68 (5): 747–755.

Card, David, Jochen Kluve, and Andrea Weber. 2017. "What works? A meta analysis of recent active labor market program evaluations." *Journal of the European Economic Association* 16 (3): 894–931.

Chernev, Alexander. 2003. "Product assortment and individual decision processes." *Journal of Personality and Social Psychology* 85 (1): 151–162.

Chernev, Alexander, Ulf Bockenholt, and Joseph Goodman. 2015. "Choice overload: Conceptual review and meta-analysis." *Journal of Consumer Psychology* 25 (2): 333–358.

Chuprinin, Oleg, Arseny Gorbenko, and Chang Mo Kang. 2019. "Rationally Neglected Stocks."

Ciura, Bob. 2016. "How Market Catalysts Affect ETFs." *ETFdb.com.*

Coate, Stephen, and Glenn C. Loury. 1993. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review* 83 (5): 1220–1240.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy* 120 (1): 116–159.

Cover, Thomas M, and Joy A Thomas. 2012. *Elements of information theory.* John Wiley & Sons.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3): 694–719.

DellaVigna, Stefano, and Joshua M Pollet. 2009. "Investor inattention and Friday earnings announcements." *The Journal of Finance* 64 (2): 709–749.

Denti, Tommaso, Massimo Marinacci, and Luigi Montrucchio. 2019. "A note on rational inattention and rate distortion theory." *Decisions in Economics and Finance*, pp. 1–15.

Eckbo, B Espen, Knut Nygaard, and Karin S Thorburn. 2019. "Board Gender-Balancing and Firm Value."

Ernst, Thomas. 2021. "Stock-Specific Price Discovery From ETFs."

Eugene, Fama, and Kenneth French. 1992. "The cross-section of expected stock returns." *Journal of Finance* 47 (2): 427–465.

Fang, Hanming, and Andrea Moro. 2010. "Theories of Statistical Discrimination and Affirmative Action: A Survey." Nber working papers 15860, National Bureau of Economic Research, Inc.

Fedyk, Anastassia. 2021. "Front page news: The effect of news positioning on financial markets."

Figlio, David N, and Bruce A Blonigen. 2000. "The effects of foreign direct investment on local communities." *Journal of Urban Economics* 48 (2): 338–363.

Filippou, Ilias, Arie Gozluklu, and Hari Rozental. 2019. "ETF Arbitrage and International Diversification."

Fosgerau, Mogens, and Nikolaj Nielsen. 2021. "Similarity, discrete choice and rational inattention."

Fosgerau, Mogens, Rajiv Sethi, and Jorgen W Weibull. 2020. "Categorical Screening with Rational Inattention."

Gerasimou, Georgios, and Mauro Papi. 2018. "Duopolistic competition with choice-overloaded consumers." *European Economic Review* 101:330–353.

Glosten, Lawrence R, Suresh Nallareddy, and Yuan Zou. 2016. "ETF activity and informational efficiency of underlying securities."

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating impartiality: The impact of "blind" auditions on female musicians." *American Economic Review* 90 (4): 715–741.

Handel, Benjamin R., and T. Jonathan Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review* 150 (8): 2449–2500.

Heckman, James J, Tomáš Jagelka, and Tim Kautz. 2021. "Some contributions of economics to the study of personality."

Hefti, Andreas. 2018. "Limited attention, competition and welfare." *Journal of Economic Theory* 178:318–359.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh. 2009. "Driven to distraction: Extraneous events and underreaction to earnings news." *The Journal of Finance* 64 (5): 2289–2325.

Hirshleifer, David, and Jinfei Sheng. 2021. "Macro News and Micro News: Complements or Substitutes?"

Hiscox, Michael J, Tara Oliver, Michael Ridgway, Lilia Arcos-Holzinger, Alastair Warren, and Andrew Willis. 2017. "Going blind to see more clearly: Unconscious bias in Australian Public Service shortlisting processes." *Behavioral Economics Team of the Australian Government.*

Huang, Shiyang, Yulin Huang, and Tse-Chun Lin. 2019. "Attention allocation and return co-movement: Evidence from repeated natural experiments." *Journal of Financial Economics* 132 (2): 369–383.

Huang, Shiyang, Maureen O'Hara, and Zhuo Zhong. 2021. "Innovation and informed trading: Evidence from industry ETFs." *The Review of Financial Studies* 34 (3): 1280–1316.

Ibanez, Marcela, and Gerhard Riener. 2018. "Sorting through Affirmative Action: Three Field Experiments in Colombia." *Journal of Labor Economics* 36 (2): 437–478.

Irons, Ben, and Cameron Hepburn. 2007. "Regret theory and the tyranny of choice." *The Economic Records* 86 (261): 191–203.

Israeli, Doron, Charles MC Lee, and Suhas A Sridharan. 2017. "Is there a dark side to exchange traded funds? An information perspective." *Review of Accounting Studies* 22 (3): 1048–1083.

Iyengar, S. Sheena, Gur Huberman, and Wei Jiang. 2004. "How much choice is too much? Contributions to 401(k) retirement plans." In *Pension Design and Structure: New Lessons from Behavioral Finance*, edited by Olivia S. Mitchell and Steve Utkus, 83–95. Oxford University Press.

Iyengar, S. Sheena, and Mark Lepper. 2001. "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of personality and social psychology* 79:995–1006.

John F. Kennedy. 1961. *Executive Order 10925 - Establishing the President's Committee on Equal Employment Opportunity.* The American Presidency Project. https://www.presidency.ucsb.edu/node/237176. Accessed February 2, 2021.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. "A rational theory of mutual funds' attention allocation." *Econometrica* 84 (2): 571–626.

Kahneman, Daniel. 1994. "New challenges to the rationality assumption." *Journal of Institutional and Theoretical Economics* 150:18–36.

Kamenica, Emir. 2008. "Contextual inference in markets: On the informational content of product lines." *American Economic Review* 98 (5): 2127–2149.

Kim, Oliver, and Robert E Verrecchia. 1997. "Pre-announcement and event-period private information." *Journal of accounting and economics* 24 (3): 395–419.

Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel. 2012. "Comparison Friction: Experimental Evidence from Medicare Drug Plans." *The Quarterly Journal of Economics* 127 (1): 199–235.

Kuksov, Dmitri, and J. Miguel Villas-Boas. 2010. "When more alternatives lead to less choice." *Marketing Science* 29 (3): 507–524.

Levy, Ariel, and Offer Lieberman. 2019. "Overreaction of Country ETFs to US Market Returns: Intraday vs. Daily Horizons and the Role of Synchronized Trading."

Li, Jiacui. 2021. "Rational Inattention and Price Underreaction."

Lindbeck, Assar, and Jörgen Weibull. 2020. "Delegation of investment decisions, and optimal remuneration of agents." *European Economic Review* 129:103559.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei. 2020. "Attention management." *American Economic Review: Insights* 2 (1): 17–32.

Luce, R. Duncan. 1959. *Individual Choice Behavior: A Theoretical Analysis.* Wiley.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2021. "Rational Inattention: A Review."

Mackowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal sticky prices under rational inattention." *American Economic Review* 99 (3): 769–803.

Matějka, Filip, and Alisdair McKay. 2015. "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review* 105 (1): 272–298.

Mirrlees, A. James. 1987. "Economic Policy and Nonrational Behavior." UC Berkeley: Department of Economics, UCB.

———. 2017. *Bounded Rationality and Economic Policy.* https://rb.gy/opw1cj. Accessed August 25, 2020.

Mogilner, Cassie, Tamar Rudnick, and S. Sheena Iyengar. 2008. "The mere categorization effect: How the presence of categories increases choosers' perceptions of assortment variety and outcome satisfaction." *Journal of Consumer Research* 35 (2): 202â€"215.

Moro, Andrea, and Peter Norman. 2003. "Affirmative action in a competitive economy." *Journal of Public Economics* 87 (3): 567 – 594.

Nagler, Matthew G. 2015. "Trading off the benefits and costs of choice: Evidence from Australian elections." *Journal of Economic Behavior and Organization* 114:1–12.

Ortoleva, Pietro. 2013. "The price of flexibility: Towards a theory of thinking aversion." *Journal of Economic Theory* 148 (3): 903–934.

Persson, Petra. 2018. "Attention manipulation and information overload." *Behavioural Public Policy* 2 (1): 78–106.

Phelps, Edmund S. 1972. "The statistical theory of racism and sexism." *American Economic Review* 62 (4): 659–661.

Ray, Debraj, and Rajiv Sethi. 2010. "A Remark on Color-Blind Affirmative Action." *Journal of Public Economic Theory* 12 (3): 399–406.

Sarver, Todd. 2008. "Anticipating regret: Why fewer options may be better." *Econometrica* 76 (6): 263–305.

Savor, Pavel, and Mungo Wilson. 2016. "Earnings announcements and systematic risk." *The Journal of Finance* 71 (1): 83–138.

Scheibehenne, Benjamin, Rainer Greifeneder, and M. Peter Todd. 2010. "Can there ever be too many options? A meta-analytic review of choice overload." *Journal of Consumer Research* 37 (3): 409–425.

Schwartz, Barry. 2004. *The paradox of choice: Why more is less.* New York: Harper Collins.

Shannon, Claude Elwood. 1948. "A mathematical theory of communication." *Bell system technical journal* 27 (3): 379–423.

Sheshinski, Eytan. 2016. "A note on income taxation and occupational choice." *Research in Economics* 70 (1): 20–23.

Shim, John J. 2018. "Arbitrage Comovement."

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50 (3): 665–690.

Spiegler, Rani. 2016. "Choice complexity and market competition." *Annual Review of Economics* 8:1–25.

Thaler, H. Richard, and R. Cass Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Train, Kenneth. 2002. *Discrete Choice Methods With Simulation.* Cambridge University Press.

U. S. Commission on Civil Rights. 1977. *Statement on affirmative action.* Clearinghouse Publication no. 54. Washington, DC.

Veldkamp, Laura L. 2006. "Information markets and the comovement of asset prices." *The Review of Economic Studies* 73 (3): 823–845.

Wang, Hua, and Liao Xu. 2019. "Do exchange-traded fund flows increase the volatility of the underlying index? Evidence from the emerging market in China." *Accounting & Finance* 58 (5): 1525–1548.

Weibull, Jörgen W, Lars-Göran Mattsson, and Mark Voorneveld. 2007. "Better may be worse: Some monotonicity results and paradoxes in discrete choice under uncertainty." *Theory and Decision* 63 (2): 121–151.

Weitzman, Martin L. 1974. "Prices vs. Quantities." *The Review of Economic Studies* 41 (4): 477–491.

Xu, Liao, Xiangkang Yin, and Jing Zhao. 2018. "Are Authorized Participants of Exchange-Traded Funds Informed Traders?"

Yang, Yung Chiang, Bohui Zhang, and Chu Zhang. 2020. "Is information risk priced? Evidence from abnormal idiosyncratic volatility." *Journal of Financial Economics* 135 (2): 528–554.