Choosing what to pay attention to

Chad Fulton *

Abstract

This paper studies static rational inattention problems with multiple actions and multiple shocks. We solve for the optimal signals chosen by agents and provide tools to interpret information processing. By relaxing restrictive assumptions previously used to gain tractability, we allow agents more latitude to choose what to pay attention to. Our applications examine the pricing problem of a monopolist who sells in multiple markets and the portfolio problem of an investor who can invest in multiple assets. The more general models that our methods allow us to solve yield new results. We show conditions under which the multimarket monopolist would optimally choose a uniform pricing strategy, and we show how optimal information processing by rationally inattentive investors can be interpreted as learning about the Sharpe ratio of a diversified portfolio.

JEL Classification: D81, D83, G11

Keywords: Rational inattention, information acquisition, price discrimination, portfolio choice

---

*chad.t.fulton@frb.gov. I thank three anonymous referees for their comments. An earlier version of this paper was titled "Mechanics of Linear Quadratic Rational Inattention Tracking Problems". The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System, or anyone else in the Federal Reserve System.

1 Introduction

Economic models with rational inattention incorporate agents who are not perfectly informed about all freely available data, but who can allocate limited information processing capacity to learn about those aspects that are relevant to them. Traditional models with imperfect information also involve agents who are not perfectly informed, but the information available to them – usually formalized through a set of signals about relevant economic variables, each contaminated by some noise – is typically taken as given rather than as chosen by the agent. By contrast, rational inattention gives agents fine control over what they pay attention to – essentially what signals to view – and how much attention they allocate – essentially controlling the variance of the contaminating noise.

To fix ideas, we start by presenting the basic structure of the kind of rational inattention model that we will focus on here. An agent is tasked with choosing an action x, given some economic shocks α and a utility function that depends on both, $u(x, \alpha)$. The agent has some prior uncertainty about the shocks, $P_{-} \equiv Var(\alpha)$, but can choose to learn about them in order to reduce their posterior uncertainty. In our models, optimal learning can be interpreted as the observation of a signal y, so that posterior uncertainty can be represented as $P_{+} \equiv Var(\alpha \mid y)$. The rational inattention tradeoff is that reduced posterior uncertainty about the shocks – also interpreted as the observation of less-noisy signals – allows the agent to better tune their action to achieve higher expected utility, but processing information is costly. The cases we consider have Gaussian shocks and quadratic or exponential utility.

In this paper, we solve the problems in three steps, by (a) determining the optimal level of posterior uncertainty P_{+} (in Propositions 2-3); (b) identifying a corresponding optimal signal vector y (using Proposition 1); and (c) determining the optimal action x conditional on the information received (in Proposition 4). However, the generality of the problems that we consider – allowing for multiple actions, multiple shocks, and multiple signals, and without imposing any independence assumptions – introduces complementarities in information acquisition that must be carefully handled. We will show that these complementarities imply that agents optimally prefer to process information about combinations of the economic shocks, and, moreover, that they may prefer to receive fewer signals than there are shocks. To handle the resulting nonuniqueness and (often) reduced dimension of optimal signals (results introduced in Lemmas 1 and 2), we show how to construct a unique "canonical" signal (in Lemma 2) that always (a) reflects optimal information processing by agents, and (b) is straightforward to interpret in terms of both *what* agents choose pay attention to as well as *how much* attention they pay.

The first application we consider is price setting by an inattentive multimarket monopolist. Their perfect information strategy is to employ price discrimination to extract as much surplus as possible, but a growing literature on behavioral industrial organization, surveyed in Ellison (2006), has begun exploring various deviations from this baseline by incorporating boundedly rational agents. In our model, firms must balance the costs of processing information against increases in profits arising from that information. We show that information-constrained firms may optimally choose to collect information only about aggregate conditions, thus giving up the option of price discrimination. This application is motivated by empirical work showing substantial increases over the past few decades in data collection by firms about their consumers, and by recent experiments of firms in "dynamic pricing", in which consumer characteristics (such as purchasing history) are used to offer a single good to different consumers at different prices (details about these phenomena can be found in Armstrong (2005) and Taylor (2004)).

The second application we consider is a standard portfolio selection problem, but by a rationally inattentive investor. The rational inattention approach to this problem was pioneered by Van Nieuwerburgh and Veldkamp (2010), who impose a fixed set of signals. This assumption eliminates much of the choice of what to pay attention to, since, in their model, each signal must be about the returns of a single asset, and the agent cannot choose to pay attention to signals about portfolios of assets. We extend their model to allow agents to choose which portfolios to learn about. This has a significant impact on results: whereas they find that portfolios chosen by rationally inattentive investors will be underdiversified, in our setting investors will choose to view a signal that allows them to maintain a diversified portfolio. At the same time, we confirm in our setting much of the intuition they develop, so that our models represent complementary approaches.[1]

The rest of the paper proceeds as follows. In section 2, we develop two economic models with rationally inattentive agents, and explain how our approach builds on the existing literature. In section 3, we develop core attention allocation problems that stem from these economic models. In section 4, we solve the rational inattention problems and describe optimal behavior and information processing by agents. Our first application to price-setting is used as a running example throughout the paper, while in section 5 we solve the portfolio choice problem and describe information processing by inattentive investors. Finally, in section 6 we describe how our results fit into dynamic problems, and in section 7, we conclude.

2 Model

We build on two strands of the rational inattention literature that is focused on static problems with Gaussian shocks. The first focuses on problems with quadratic utility (or log-quadratic approximations to more general utility functions) and includes work on price-setting, optimal monetary policy, and consumption dynamics. The second focuses on portfolio choice in the familiar exponential utility case. In each setting, while prior work has solved special cases of the underlying rational inattention problem, the contribution of this paper will be to first extend the analytic solution to a more general formulation, and then second to show how to understand what agents are paying attention to. We now develop the basic problems for each case and position the contribution of this paper in the existing literature.

2.1 Quadratic payoffs

Rational inattention models with quadratic utility and Gaussian shocks have proved to be very popular, both because they are mathematically tractable and also because they produce results that make them comparable to models with signal extraction

[1] In a slightly different setup, Mondria (2010) also notes that when agents are allowed to choose signals they will maintain a diversified portfolio. As described in more detail below, we build on his work by considering a more general model and generating sharper results, particularly in terms of intuition about the chosen signals.

problems.[2] Nonetheless, even this class of problems is sufficiently difficult to solve that analytic solutions have so far remained limited. Our goals are to expand the class of models that can be solved and to provide new tools to interpret information processing by agents.

The model we study begins with some relevant state of the world – a set of fundamental economic shocks, for example – modeled as an exogenous random vector α, drawn from the n-dimensional Gaussian distribution $\mathcal{N}(a_-, P_-)$, where $a_- \in \mathbb{R}^n$ is the prior mean and P_- is the $n \times n$ prior covariance matrix. Our agent takes an action $x \in \mathbb{R}^m$, and receives a payoff $u(x, \alpha)$. Following Sims (2003), we model our agents as rationally inattentive: processing information about the shocks α allows the action x to depend on their realization, at least to some extent, but information processing is costly. Thus, agents have at best imperfect information, and they can only maximize expected payoffs, $E[u(x, \alpha)]$. In the simplest univariate case, $u(x, \alpha) = -\frac{1}{2}(x - \alpha)^2$, and a rationally inattentive agent's goal would be to make their action as close as possible, on average, to the shock.[3] In doing so, they face a trade-off: increased attention implies smaller expected deviations, but is more costly. Here we are interested in studying how agents choose what to pay attention to between multiple shocks, and so we consider the more general quadratic form

$$u(x, \alpha) = -\frac{1}{2}x' M x + c' x + b$$

where M is an $m \times m$ symmetric positive definite matrix, $c = C\alpha + c_0$, C is an $m \times n$ matrix, c_0 is an $n \times 1$ vector, and $b \in \mathbb{R}$. This allows agents to prioritize some variables over others, and accommodates interaction effects. More generally, this could be viewed as a second-order approximation to a more general payoff function.

To specify the costs from information processing, we follow Sims (2003) in quantifying information in terms of "mutual information", which measures the re-

[2]This literature is now quite large, and has applied rational inattention to study, for example, consumption and permanent income (Sims (2003), Luo (2008)), price-setting (Maćkowiak and Wiederholt (2009), Matějka (2016)), business cycle dynamics (Maćkowiak and Wiederholt (2015)), and optimal monetary policy (Paciello and Wiederholt (2014)).

[3]This example is often referred to as a "tracking problem".

duction in uncertainty about the shocks α. When the new information takes the form of a signal vector y, this is denoted $I(\alpha, y)$, or simply κ. This literature has primarily focused on two mechanisms of costly attention. In the first, there is some fixed level of available capacity that cannot be exceeded, so that $\kappa \equiv I(\alpha, y) \leq \kappa^*$. In the second, any amount of information processing capacity can be accessed at a fixed marginal cost, so that the agent pays $f(\kappa) = \lambda\kappa$.

Before formally stating the problem that we will consider in this paper, we introduce several results from Sims (2003) that will considerably simplify its formulation.[4] First, he showed that in this setting, when the agent is optimally processing information, it is as if she is viewing a noisy signal $y \in \mathbb{R}^r$ of the form

$$y = Z\alpha + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \Lambda) \tag{1}$$

where Z is an $r \times n$ matrix and Λ is an $r \times r$ positive definite matrix. The agent's information processing problem thus amounts to the selection of the pair (Z, Λ) that determines a particular signal y, and the problem includes the selection of r, the dimension of the signal. This considerably simplifies matters, as typical Bayesian updating yields an analytic form for the posterior

$$\alpha \mid y \sim \mathcal{N}(a_+, P_+) \tag{2}$$

$$a_+ = a_- + K(y - Za_-) \tag{3}$$

$$P_+ = (P_-^{-1} + Z'\Lambda^{-1}Z)^{-1} \tag{4}$$

where $K = P_-Z'(ZP_-Z' + \Lambda)^{-1}$ is often referred to in dynamic settings as the Kalman gain. As a result of this Gaussian posterior, the quantity of information processed by the agent (measured in *nats*) takes the following particularly simple form $I(\alpha, y) = 0.5(\ln|P_-| - \ln|P_+|)$.[5]

The second important result from Sims (2003) is that the optimal action taken by a rationally inattentive agent in this setting will be of the form $x_+ = E[x^* \mid y]$, where $x^* = M^{-1}(C\alpha + c_0)$ is the perfect information solution. This is the usual cer-

[4] A more general formulation of the basic rational inattention problem can be found in section 3.1 of Sims (2010).

[5] See, for example, Sims (2003).

tainty equivalence result that is obtained with an exogenous Gaussian signal and a quadratic objective function. With this, we can rewrite $E[u(x_+, \alpha)] = -0.5E[(x^* - E[x^* \mid y])' M(x^* - E[x^* \mid y])] + \zeta$, where $\zeta \equiv b + 0.5E[x^{*\prime} M x^*]$ does not depend on the signal y. Combining all of these results with the information constraint and dropping the constant term ζ, we arrive at the agent's problem.

Problem 1

$$\max_{Z \in M_{r,n}, \Lambda \in M_r} -\frac{1}{2} E\Big[(x^* - E[x^* \mid y])' M(x^* - E[x^* \mid y])\Big] - \lambda\kappa \tag{5}$$

subject to

$$x^* = M^{-1}(C\alpha + c_0) \qquad \text{perfect information solution} \tag{6}$$

$$y = Z\alpha + \varepsilon \qquad \text{signal vector} \tag{7}$$

$$\varepsilon \sim \mathcal{N}(0, \Lambda) \qquad \text{noise due to inattention} \tag{8}$$

$$\Lambda \succ 0 \qquad \text{“no forgetting” constraint} \tag{9}$$

$$\kappa \equiv I(\alpha, y) = \frac{1}{2}\Big(\ln|P_-| - \ln|P_+|\Big) \qquad \text{information processed} \tag{10}$$

$$P_+^{-1} = P_-^{-1} + Z'\Lambda^{-1}Z \qquad \text{Bayesian updating} \tag{11}$$

Our notation follows Horn and Johnson (2012), so that $M_{m,n}$ is the set of m-by-n matrices, M_n is the set of n-by-n matrices, and $A \succ B$ denotes $A - B$ positive definite. This formulation is written with the fixed marginal cost of attention in mind, but it can also accommodate the fixed capacity approach by considering λ as a Lagrange multiplier on the constraint $\kappa \leq \kappa^*$. Our application to price-setting by a multimarket monopolist, below, revisits the construction of this kind of problem in a concrete setting.

While the statement of this problem is essentially due to the seminal work of Sims (2003) that introduced rational inattention problems, its solution was not provided there, and the subsequent literature, which we briefly review below, has only focused on certain special cases of this problem. Here, we begin by providing a general solution to this problem, and then we show how to use the solution to examine what agents choose to pay attention to.

The solution that we provide accommodates problems with the following characteristics: (1) shocks α with any dimension n; (2) an action x with any dimension m; (3) signal vectors y with any dimension r; (4) correlation between the shocks α (so that P_{-} is not restricted to be, for example, diagonal) and in information collection (so that Z and Λ are not restricted to be diagonal); and (5) an information constraint modeled in terms of either a fixed capacity or a fixed marginal cost of attention. Thus, the problems that can be solved using our tools go beyond those available in the existing literature, which we now briefly review.

Maćkowiak and Wiederholt (2009), which we will refer to as MW, were among the first to introduce rational inattention in a general equilibrium macroeconomic model. While their overall model is more complex, it includes a core attention problem that is in the form of Problem 1 with $m = 1$, $n = 2$, P_{-} diagonal, and a fixed capacity constraint. They solve this problem in two settings. In their baseline model, they impose independence for signals (so that Z and Λ must be diagonal) and show that the agent will choose two signals, so that $r = 2$. In an extension, they remove this assumption and show that the agent will choose a single signal, so that $r = 1$, and that this signal is of the form "perfect information solution plus noise".[6]

Paciello and Wiederholt (2014), hereafter PW, consider optimal monetary policy with rationally inattentive firms. Firms face one or two shocks that are assumed to be independent, and choose the price of their good. Thus, here, similar to MW, $m = 1$ and $n \leq 2$. Unlike MW, here firms can access information processing capacity at the cost $f(\kappa)$. PW use the fixed marginal costs used in Problem 1, but also introduce a second cost function that introduces additional convexity, which we do not consider. Similar to the results of MW, PW show that when signals are assumed to be independent, agents will choose to observe as many signals as there are shocks, so that $r = n$, while when this assumption is not imposed, agents will choose only a single signal, so that $r = 1$.[7]

[6]Maćkowiak and Wiederholt (2009) solve the attention allocation decision in the context of a general equilibrium model, and they also provide an analytic solution when the two shocks follow AR(1) processes.

[7]Paciello and Wiederholt (2014) also solve this attention allocation decision in a larger equilibrium setup. Additionally, they consider the case that the noise in the signals (ε in our setup) is correlated across firms, but not the case that the shocks (α in our setup) are correlated.

Each of these papers provides analytic solutions to models that featured at most two shocks, assumed to be independent, and a one-dimensional action, and more general results have not been available. For example, Sims (2010) only slightly generalizes PW and MW by increasing the number of shocks to $n > 2$ and is no longer able to present an analytic solution.[8] Analytic solutions to yet more general versions of the problem, for example those including additional choice variables and prior correlation, are similarly not available in the literature. One technical contribution of this paper is to lift these restrictions while still allowing an analytic solution.

In independent and concurrent work, Kőszegi and Matějka (2020) and Dewan (2019) formulate static rational inattention problem that are similar to Problem 1. While we cover some similar ground, there are differences. First, Kőszegi and Matějka (2020) focus on implications for consumption decisions, while Dewan (2019) subsequently considers an extension to Laplacian priors with mean absolute error loss. We focus on using the solution to provide intuition about what agents pay attention to. In addition to the solutions with quadratic payoffs, we also advance a second strand of the literature that considers portfolio choice with exponential utility, by providing an analytic solution also in this case and again describing what rationally inattentive investors pay attention to.[9]

There are also technical differences between our approaches. First, in order to focus on the effect of preferences, Kőszegi and Matějka (2020) explicitly restrict themselves to the case in which the shocks α are uncorrelated and with common variance. In our terminology, this restricts $P_{-} = \sigma^2 I$. As we will show, the key difficulty in solving these problems is handling complementarities in information acquisition, and their restriction precludes an examination of the effect of an important source of complementarities. This is because if shocks are correlated, then information about one shock can provide information about other shocks. The methods that we develop here are required to solve for and interpret the behavior of rationally inattentive agents in the general case with these complementarities. Second,

[8] In particular, Sims (2010) describes the form that the posterior covariance matrix for the shocks (P_{+} in our setup) must take, but can only evaluate it numerically.

[9] Another recent contribution to the rational inattention literature is Kamdar (2018), which uses the method of Kőszegi and Matějka (2020) to study consumer sentiment.

our setup allows us to provide solutions for both the fixed marginal cost and fixed capacity formulations, while Kőszegi and Matějka (2020) focuses only on the fixed marginal cost case and Dewan (2019) focuses only on the fixed capacity case.

There has also been recent work advancing dynamic rational inattention problems with quadratic loss. In the univariate case, Maćkowiak, Matějka, and Wiederholt (2018) provide analytic solutions for a univariate shock that follows an ARMA(p,q) process. In the multivariate case, Miao, Wu, Young, et al. (2019) compute solutions using semidefinite programming methods and Afrouzi and Yang (2019) computes a solution based on first-order conditions. Finally, there have been an increasing number of papers that describe rational inattention problems in terms of the choice of posteriors, as we do here, with Caplin, Dean, and Leahy (2019) and Miao and Xing (2020) as two recent examples.

2.1.1 Multimarket monopolist

Our first application considers price-setting by a rationally inattentive multimarket monopolist. The typical perfect information model assumes that the firm can identify distinct groups with differing demand functions, and can prevent reselling between the groups. The well-known result is that the monopolist optimally charges different prices in each market so as to extract as much surplus as possible. We will relax the assumption of perfect information, instead assuming that the firm must process information about demand in each market in order to discriminate between them, and show that accounting for the firm's optimal attention allocation can imply a wide variety of pricing strategies, including no price discrimination at all.

In this section, we begin by setting up and solving the standard problem faced by a multimarket monopolist in the special cases of perfect information and imperfect information with exogenous information. We then show how the rational inattention problem arises as a natural extension with endogenous information acquisition.

Suppose that a firm faces two markets with linear demand curves $q_i(p_i) = \ell_i - m_i p_i$ for $i = 1, 2$ and constant marginal costs of production γ. The firm's profit function is then, $\pi(p) = \sum_{i=1}^{2}(p_i - \gamma)(\ell_i - m_i p_i)$, which is a quadratic function in the price vector $p = (p_1, p_2)' \in \mathbb{R}^2_{++}$. We assume that the monopolist is inattentive to the levels of demand (ℓ_1, ℓ_2) and the marginal cost γ, which we model as jointly normally distributed, so that $(\gamma, \ell_1, \ell_2)' \equiv \alpha \sim \mathcal{N}(a_-, P_-)$.

Step 1: Perfect information. If the firm knows α with certainty, the problem is standard, $\max_{p\in\mathbb{R}^2_{++}} \pi(p)$, with profit function

$$\pi(p) = -\frac{1}{2}p'Mp + \alpha'C'p + b$$
$$M = \begin{bmatrix} 2m_1 & 0 \\ 0 & 2m_2 \end{bmatrix} \quad C = \begin{bmatrix} m_1 & 1 & 0 \\ m_2 & 0 & 1 \end{bmatrix} \quad b = -\gamma(\ell_1 + \ell_2)$$

The solution can be easily found from the first order condition, $p^* = M^{-1}C\alpha$. Optimal prices in each market are then $p_i^* = 0.5(\gamma + \ell_i/m_i)$, and so the monopolist will price-discriminate unless the markets have identical demand curves.

Step 2: Exogenous information. Suppose now that the firm does not observe α, but does have access to an exogenous signal $y = Z\alpha + \varepsilon$, of the form in equation 1. The firm's problem is then $\max_{p\in\mathbb{R}^2_{++}} E[\pi(p) \mid y]$, and – because the objective is quadratic and the shock is Gaussian – we can again find the solution from the first-order condition. This yields the certainty equivalence result that the optimal price is $p_+ = E[p^* \mid y]$. The firm will try to replicate the perfect information solution as best as possible, given the information available to them. Since the signal is exogenous, there is little more than can be said about this solution in general.

Step 3: Rational inattention. Rationally inattention allows us to endogenize information by allowing firms to choose which signal to receive, subject to information processing costs. A more precise signal can increase expected profits, but requires more attention. As already described, from Sims (2003) we already know that, in this case, the optimal signal chosen by the agent will be of the form of equation 1, and we also know that the optimal policy will be $p_+ = E[p^* \mid y]$. Thus, the rational inattention problem generalizes the exogenous information case by adding an additional decision: the firm must select the nature of the signal by choosing Z and Λ. Following the same steps as in the general case, the firm's rational inattention problem can now be stated.

Problem 2

$$\max_{Z\in M_{r,n},\Lambda\in M_r} -\frac{1}{2}E\Big[(p^* - E[p^* \mid y])' M (p^* - E[p^* \mid y])\Big] - \lambda\kappa$$

subject to $p^* = M^{-1}C\alpha$ and equations 7 - 11

Of course, this is simply a specialization of Problem 1 to the case of the problem faced by the multimarket monopolist. The firm's problem is to select a signal that allows them to set prices to be as close as possible, in a least squares sense, to the perfect information prices p^*. However, through the influence of the weight matrix M, they are more concerned about deviations in markets with more elastic demand. This makes sense – if demand is more sensitive to change in prices, it is more important for the firm to learn about the optimal price.

This example has already gone beyond the problems with analytic solutions considered in the previous literature. First, it includes three shocks ($n = 3$) and two choice variables ($m = 2$), and, second, we have not put any restrictions on the prior covariance matrix P_-, so that the shocks may be correlated. Moreover, while we have focused on an example with two markets to simplify the exposition, generalizing this example to an arbitrary number of markets would be entirely straightforward. Finally, we have not imposed the independence assumption, so that firms optimize both in terms of what to pay attention to and also how much attention to pay.

To illustrate the variety of situations that can arise, we parameterize

$$P_- = 4\sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}$$

and set $m_1 = 0.5$. We will investigate how the solution differs for various values of m_2, σ^2, and ρ. Our baseline case will have identically sloped demand curves, so that $m_2 = m_1 = 0.5$, and uncorrelated demand, so that $\rho = 0$. To avoid complicating the exposition, we will assume throughout that the firm has no prior information differentiating the level of demand between markets, so that $E[\ell_1] = E[\ell_2]$, but we note that this is easy to relax.

2.2 Portfolio choice

The second strand of the literature that we build on considers investment under limited attention. This literature has extended the standard portfolio choice problem with exponential utility and exogenous information to include signal selection through rational inattention. Since payoffs are exponential, the problem does not fit in the framework of the previous section, and so different techniques are required. Similar to the quadratic case, however, the difficulty of these models has so-far limited analytic solutions to special cases. Our goal will again be to expand the class of models that can be solved and provide tools to aid interpretation of results.

Portfolio choice problems with rationally inattentive investors were pioneered by Van Nieuwerburgh and Veldkamp (2010), which we will refer to as VV, and Mondria (2010), with the former considering a partial equilibrium setting and the latter a general equilibrium setting. To focus on the rational inattention problem itself, we will remain in the partial equilibrium setting, and so our model setup follows VV closely. Where we build on their work is in allowing agents to choose more general signals, since VV impose in their model essentially the independence assumption of Maćkowiak and Wiederholt (2009). Notably, Mondria (2010) also relaxed the independence assumption, but was limited to the case with two assets with independent returns. Thus our contribution can alternatively be understood as extending Mondria (2010) to allow for an arbitrary number of assets with correlated returns.[10] After we specify the model, we provide a more detailed comparison with this literature.

We consider the problem of an investor tasked with constructing an optimal portfolio among n risky assets with uncertain returns $\alpha \sim \mathcal{N}(a_{-}, P_{-})$, given one riskless asset with known return $\mathfrak{r}$. We assume that the investor is able to allocate attention to obtain information about the realization of the risky assets. The model consists of two stages. In the first stage, the investor solves a rational inattention problem to optimally process information about asset returns, and in the second

[10] Even though we consider the partial equilibrium case here, the same tools can be used to solve the more general problem considered by Mondria (2010). In a previous version of this paper, we show how to do this, solving the general equilibrium problem in which asset prices are not held fixed.

stage the investor solves a portfolio choice problem conditional on the information collected during the first stage.

The second stage is a standard portfolio optimization problem with exponential utility. Taking as given initial wealth ω_0, the risk-free rate r, a vector of asset prices p, and a signal y about asset returns, the investor must choose a portfolio x of risky assets, in order to $\max_{x \in \mathbb{R}^n} E[-\exp\{-\rho\omega\}|y]$ for a given signal y, subject to $\omega = \omega_0\mathfrak{r} + x'(\alpha - p\mathfrak{r})$, where ρ is a parameter governing risk aversion. The vector $e = \alpha - p\mathfrak{r}$ describes the "excess returns" of the risky assets over an equivalent investment in the riskless asset. In classic versions of this model, the signal was assumed to be exogenous, while in this version, it is chosen by the investor in the first stage.

If the first stage yields a signal that of the form $y = Z\alpha + \varepsilon$ with $\varepsilon \sim N(0, \Lambda)$, as in the quadratic case, equation 1, then the investor's posterior beliefs about asset returns would be Gaussian, $\alpha \mid y \sim N(a_+, P_+)$, as in equation 2. The solution to the second stage would then be entirely standard, $x_+ = \frac{1}{\rho} P_+^{-1}(E[\alpha \mid y] - p\mathfrak{r})$. Broadly speaking, the investor would choose to hold larger quantities of assets whose returns have greater expected value and are less uncertain. Unfortunately, it turns out that the observation of a signal of this form does not represent optimal information processing by the agent, as demonstrated in Jung et al. (2019). Instead, outside the convenient quadratic Gaussian case, fully optimal solutions must be obtained numerically, and attention allocation is no longer easy to work with.[11] As a result, in order to maintain tractability, most the literature on portfolio choice problems with rationally inattentive investors, aside from Jung et al. (2019), has assumed that agents observe a signal of the form of equation 1, and we maintain this assumption here. Thus, while this paper advances this literature by relaxing restrictions on how agents may allocate attention in an important direction, we still impose some restrictions to ensure a tractable model.

The first stage rational inattention problem is then to choose the pair (Z, Λ) that

[11] For example, in many problems with non-quadratic loss or non-Gaussian shocks, including the portfolio choice problem, agents may prefer to receive discretely distributed signals. While this makes models more difficult to work with, Matějka (2016) shows that this feature can explain prices that only move between several fixed values.

determines the signal y, subject to a constraint on information processing.[12]

Problem 3

$$\max_{Z\in M_{r,n},\Lambda\in M_r} E\left[\rho E[\omega \mid y] - \frac{\rho^2}{2}Var(\omega \mid y)\right] - \lambda\kappa \tag{12}$$

where

$$\omega = \omega_0\mathfrak{r} + x_+'(\alpha - p\mathfrak{r}) \qquad \text{end of period wealth} \tag{13}$$

$$x_+ = \frac{1}{\rho}P_+^{-1}(E[\alpha|y] - p\mathfrak{r}) \qquad \text{optimal portfolio} \tag{14}$$

and subject to equations 7 - 11

Each element of the signal vector provides information about the returns of a particular portfolio (i.e. linear combination) of assets, and intuitively there are two aspects to the investor's problem. First, for a signal about any particular portfolio, increased precision reduces the investor's risk but requires additional attention. Second, the agent must select which portfolios they should optimally learn about.

Van Nieuwerburgh and Veldkamp (2010) study this problem, but they require that each signal provides information about only a single asset. This eliminates part of the investor's problem, since they no longer choose which portfolios to learn about. We extend their model by allowing for this choice.[13]

Mondria (2010) studies a general equilibrium version of this problem, in which the vector of asset prices p is determined in equilibrium rather than taken as given, and similarly allows agents to choose signals that contain information about portfolios of assets. To maintain tractability and be able to compute a closed-form solution, he is limited to only two assets, with independent returns, and he shows that investors choose to observe a signal that contains information about both assets.

[12]The "mean-variance" form of the objective function presented in Problem 3 can be derived from the maximization of expected utility in the first stage by assuming that investors have a preference for the early resolution of uncertainty, as in Van Nieuwerburgh and Veldkamp (2010) and Mondria (2010).

[13]Van Nieuwerburgh and Veldkamp (2010) also explore other versions of the problem that we do not consider here, including using an alternative (CRRA) preference specification and an alternative (additive) information cost function.

We extend this result to three-or-more assets with potentially correlated returns.[14]

The literature on rational inattention portfolio choice problems has thus faced essentially the same constraints on analytic solutions as has the macroeconomic rational inattention literature with quadratic payoffs: only two shocks and some restriction on dependence (either the shocks, the signals, or both). We lift these restrictions and present both a general solution and tools for interpreting information processing.

3 Core rational inattention problems

In this section, we start by noting an issue that arises with the problems formulated in the previous section: they do not have unique solutions. This makes them difficult to work with, and so our next step is to reformulate them into what we call the "core rational inattention problems", which do not suffer from this issue. Finally, we show how the solution to each core problem is linked to a set of solutions to the original problems.

3.1 Nonuniqueness of optimal signals

Lemma 1. *There is no unique solution to Problem 1, 2, or 3.*

Proof. All proofs are provided in Appendix A. □

It is straightforward to illustrate this result using the baseline case of the multimarket monopolist problem, and we will do so by presenting several signals that could each represent optimal information processing. To avoid digressing, we simply assert these results for now, though we prove them later.

A solution to this problem turns out to be a noisy signal of the optimal average price across markets, $\bar{p}^* \equiv 0.5(p_1^* + p_2^*)$, and the demand differential between

[14] More recently, Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) consider a similar model to that of Mondria (2010) with more than two assets, but their analysis takes the form of the signal vector as given, and so they do not solve the full rational inattention problem for the optimal signal.

markets, $\tilde{\ell} \equiv \ell_1 - \ell_2$. This signal can be written

$$y_r = \begin{bmatrix} \bar{p}^* \\ \tilde{\ell} \end{bmatrix} + \varepsilon_r = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -1 \end{bmatrix} \alpha + \varepsilon_r, \qquad \varepsilon_r \sim N\left(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right)$$

for some scalars σ_1^2 and σ_2^2. For the moment, we will assume these are both finite. A trivial example of a second optimal signal is any scalar multiple of this signal, for example $y = 2y_r$. This alternative signal still satisfies the problem's constraints, requires the same information processing capacity (see equation 10), and leads to an identical posterior distribution (see equations 2 - 4).

There are more interesting alternative signals that also represent optimal information processing. For example, a signal of the form "perfect information solution plus noise", defined to be

$$y_p = \begin{bmatrix} 1 & \frac{1}{2} \\ 1 & -\frac{1}{2} \end{bmatrix} y_r = p^* + \varepsilon_p, \qquad \varepsilon_p \sim N\left(0, \begin{bmatrix} \sigma_1^2 + \frac{\sigma_2^2}{4} & \sigma_1^2 - \frac{\sigma_2^2}{4} \\ \sigma_1^2 - \frac{\sigma_2^2}{4} & \sigma_1^2 + \frac{\sigma_2^2}{4} \end{bmatrix}\right)$$

can also be a solution to this problem. In this case, since y_p and y_r are related by an invertible linear transformation, if one signal is observed then the other signal can be losslessly recovered. Intuitively, these signals convey the same information, so if one of them is optimal, the other must also be. A similar analysis shows that there is no optimal signal in terms of the fundamental shocks, $y_f = \alpha + \varepsilon$, because there is no nonsingular transformation that would relate y_f to the optimal signal y_p.

There is an important complication that arises as information costs rise. When information costs are low enough, the agent will observe both elements of the signal y_r, and each variance term will correspondingly be finite. However, as costs rise, we will show that at some point the agent will stop observing information about the demand differential, which has the effect of sending $\sigma_2^2 \to \infty$. The effect on y_r or the alternative signal $y = 2y_r$ is straightforward, as this simply implies that agent stops observing the second element of each, but this is not true of the signal y_p. Instead, once information costs are high enough, there ceases to be an optimal signal of the form y_p. Intuitively, this is because at this point the firm wishes to stop collecting information about the demand differential, but a signal about perfect information prices cannot be constructed without this information.

There are three key points to take away from this example. First, we have illustrated how multiple optimal signals can occur, both trivial and non-trivial. Second, we have shown that some signals, such as y_r, are in a sense more useful than others, such as y_p – for example, in how they accommodate changes in information costs. This will be important as we develop the solution to the problem. Finally, this example shows it is not possible to always focus on signals of the form "perfect information solution plus noise" or "fundamental shocks plus noise", even though these have been correct solutions to special cases considered in the previous literature.

In general, this nonuniqueness presents a difficulty for solving problems formulated in terms of choosing a signal. This is true even though the nonuniqueness is in a sense trivial – as each of these equivalent signals is *ex ante* expected to yield the same posterior information set – because there is not even an obvious normalization that would allow one to focus on a specific signal vector in the optimal set. One convenient candidate would normalize $Z = I$, implying a signal of the form "fundamental shocks plus noise", $y_f = \alpha + \varepsilon$, which we refer to as the "fundamentals" signal. If this normalization could be imposed in general, then the problem could be reduced to choosing the noise covariance matrix Λ. However, in most cases agents optimally choose to observe fewer signals than there are shocks. When this happens, the set of optimal signals typically does not include the fundamentals signal, and so this normalization cannot be applied. For example, we noted that the fundamentals signal is not optimal in the example just above.

One way around this problem is restricting the agents' choice of signals, so that the optimal set always includes the fundamentals signal. Indeed this is one benefit of the independence assumption that has been made in the literature. However, this approach has the downside that it usually implies suboptimal information processing by agents. Instead, we will show how a reparameterization of the problem facilitates a general solution without restricting the set of solutions by reparameterizing the problem. Because these reparameterized problems are formulated in a way that is similar to the seminal rational inattention problems described by Sims (2003) and Sims (2010), we call them the "core rational inattention problems".

3.2 Transforming the problem

In this section, we first present the core rational inattention problems, in Problems 4 and 5, and then show how their solutions can be linked to solutions of the original problems, in proposition 1. We also slightly generalize the setting by explicitly incorporating a prior information set, $\mathcal{I}_-$.

Problem 4: Let $\alpha \mid \mathcal{I}_- \sim \mathcal{N}(a_-, P_-)$ where $\mathcal{I}_-$ is some prior information set, and let W be a positive semidefinite matrix.[15] Define the *core static quadratic rational inattention problem* as

$$\begin{aligned}&\min_{P_+ \in M_n} tr(WP_+) + \lambda \left(\ln|P_-| - \ln|P_+|\right)\\ &\text{subject to } P_+ \succeq 0, P_- \succeq P_+\end{aligned}$$

With a slight abuse of notation, λ may represent a fixed marginal cost of information processing or a Lagrange multiplier for the constraint $\frac{1}{2}\left(\ln|P_-| - \ln|P_+|\right) \leq \kappa^*$.

This reparameterizes the quadratic rational inattention problem introduced as Problem 1, with the key benefit that elements of the set of optimal signal vectors that solve Problem 1 leads to the unique posterior covariance matrix that solves Problem 4. Similarly, Problem 5, below, reparameterizes the rational inattention portfolio choice problem that we introduced as Problem 3.

Problem 5: Let $\alpha \mid \mathcal{I}_- \sim \mathcal{N}(a_-, P_-)$ where $\mathcal{I}_-$ is some prior information set, and let W be a positive semidefinite matrix.[16] Define the *core rational inattention portfolio choice problem* as

$$\begin{aligned}&\min_{P_+ \in M_n} tr(-WP_+^{-1}) + \lambda \left(\ln|P_-| - \ln|P_+|\right)\\ &\text{subject to } P_+ \succeq 0, P_- \succeq P_+\end{aligned}$$

While we still allow λ to represent either a fixed marginal cost of attention or a Lagrange multiplier on a fixed capacity constraint, we note that due to proposition 3, below, we will always use the fixed capacity version in practice.

While the original problems selected an optimal signal and then considered the

[15] The formulation of Problem 1 implies $W = C'M^{-1}C$, as in proposition 1.

[16] The formulation of Problem 3 implies $W = P_- + (a_- - p\mathfrak{r})(a_- - p\mathfrak{r})'$, as in proposition 1.

implications for posterior uncertainty, in Problems 4 and 5 we solve for optimal posterior uncertainty and use that to generate a set of valid signals. The links between the solutions to the original and reparameterized problems are formalized below, in proposition 1.

Intuitively, the reparameterization refocuses the problem in such a way as to highlight the core attention allocation decision. To illustrate how this works, we return to the example of the multimarket monopolist. Technically, the reparameterization is straightforward, since

$$-\frac{1}{2}E[(p^*-p_+)'M(p^*-p_+)] \overset{(a)}{=} -\frac{1}{2}E[(\alpha-a_+)'C'M^{-1}C(\alpha-a_+)] \overset{(b)}{=} -\frac{1}{2}tr(WP_+)$$

where we have used $p^* = M^{-1}C\alpha$ and defined $W \equiv C'M^{-1}C$. Combining this with the definition of κ (and multiplying both by -2) yields the objective function given in Problem 4, with the two equalities in the above equation reflecting the two steps taken in refocusing the problem.[17]

First, the basic problem of the multimarket monopolist is to set prices so as to replicate the perfect information solution, p^*, with minimum weighted mean squared error, where the weights reflect the importance to the firm of minimizing the error in each market (in particular, we showed earlier that the firm wishes to track prices more closely in markets that have more elastic demand). This is what is given on the left hand side of (a), above. However, as a rationally inattentive agent, what the firm must actually do is to allocate attention between the fundamental shocks. In moving to the right hand side of (a), we have constructed a new weight matrix, that describes the same tradeoff as before, but now in terms of minimizing errors in each fundamental shock. This new weight matrix

$$W = C'M^{-1}C = \frac{1}{2}\begin{bmatrix} m_1+m_2 & 1 & 1 \\ 1 & 1/m_1 & 0 \\ 1 & 0 & 1/m_2 \end{bmatrix}$$

shows, for example, a steeper demand curve in either market would make it more important for the firm to pay attention to demand in that market, and it would make

[17] A similar derivation for Problems 1 and 3 can be found in the proof of Lemma 1.

it comparatively less important to track costs.

The second step is to reparameterize the problem in terms of the choice of posterior, rather than the choice of signal vector, so as to avoid the issues noted in the previous section. Thus while the left hand side of (b) gives the problem in terms of expectations driven by a signal vector (since $a_+ = E[\alpha \mid y, \mathcal{I}_-]$), on the right hand side of (b) we have rewritten it in terms of the mean squared error driven by the posterior covariance matrix.

3.3 Link to original problems

To link these problems, we need to show (a) how a signal vector that solves Problem 1 or 3 generates a posterior covariance matrix that solves Problem 4 or 5, and (b) how a posterior covariance matrix that solves Problem 4 or 5 generates a set of signals that solve Problem 1 or 3. The first part follows directly from the typical Bayesian updating formula of eq. (4), but we will need some additional results for the second part. In particular, given some prior information set, we need to show how to construct a signal of the form in eq. (1) that yields an arbitrary level of posterior uncertainty. We do this in the next lemma.

Lemma 2. *Let α be an n-dimensional random vector and $\mathcal{I}_-$ be a prior information set, such that $\alpha \mid \mathcal{I}_- \sim \mathcal{N}(a_-, P_-)$, and let P_+ be a positive definite matrix such that $P_- \succeq P_+$.*

- *(a)* Simultaneous diagonalization. *(i) There exists a nonsingular $S \in M_n$ and positive definite diagonal matrix $\Delta^+ \in M_n$ with nonincreasing entries $\delta_i^+, i = 1, \ldots, n$ so that $P_+^{-1} = S'\Delta^+ S$ and $P_-^{-1} = S'IS$. (ii) Moreover, $P_+^{-1} - P_-^{-1} = S'(\Delta^+ - I)S$, with $\Delta^+ - I \succeq 0$. (iii) If the diagonal elements of Δ^+ are unique, then the matrix S is unique. Otherwise, S is unique up to permutations of the rows associated with duplicated elements of Δ^+.*

- *(b)* Canonical shocks. *Define the "canonical shocks" associated with P_- and P_+ as $\beta_c = S\alpha$. This satisfies $Var(\beta_c \mid \mathcal{I}_-) = I$.*

- *(c)* Reduced canonical signal. *Define $r = rk(\Delta^+ - I)$ and let $\Lambda_r^{-1} \in M_r$ be the largest positive definite submatrix of $\Delta^+ - I$. Define the "reduced canonical* *signal" associated with P_- and P_+ as $y_r = \beta_r + \varepsilon_r$ with $\varepsilon_r \sim N(0, \Lambda_r)$, where β_r contains the first r elements of β_c. Then $Var(\alpha \mid y_r, \mathcal{I}_-) = P_+$ and $Var(\beta_c \mid y_r, \mathcal{I}_-) = (\Delta^+)^{-1}$.*

(d) Canonical signal. *Let $\Lambda_c = (\Lambda_r \oplus \infty I_{n-r}) \in M_n$, and define the "canonical signal" to be $y_c = \beta_c + \varepsilon_c$ with $\varepsilon_c \sim N(0, \Lambda_c)$. Then $Var(\alpha \mid y_c, \mathcal{I}_-) = P_+$ and $Var(\beta_c \mid y_c, \mathcal{I}_-) = (\Delta^+)^{-1}$.*

(e) Mutual information. *Define the mutual information between the i-th canonical shock $\beta_{i,c}$ and the i-th element of the canonical signal $y_{i,c}$, conditional on the prior information set $\mathcal{I}_-$, to be $\kappa_i \equiv I(\beta_{i,c}, y_{i,c} \mid \mathcal{I}_-)$. Then $\kappa_i = \ln \delta_i^+$, where δ_i^+ is the i-th diagonal element of Δ^+, and*

$$\kappa \equiv I(\alpha, y_c \mid \mathcal{I}_-) = \sum_{i=1}^{n} I(\beta_{i,c}, y_{i,c} \mid \mathcal{I}_-) = \sum_{i=1}^{n} \kappa_i$$

The insight of this lemma is that it is possible to construct a set of transformed shocks that decompose the information differentiating the posterior from the prior into independent components. This will be key for solving the core rational inattention problems and constructing a normalized signal vector. We call these transformed shocks the "canonical" shocks both because they uniquely decouple information processing and because they share in spirit and methods some similarities to canonical correlations. These shocks then yield a corresponding "canonical signal" that, together with the prior information, generates the desired posterior. Importantly, this result is not limited to the static problems that we consider here, but would also apply to information processing in the dynamic case. Thus, the tools we develop in the rest of this paper for interpreting information processing by rationally inattentive agents apply more broadly.

We now formally link the original and reformulated problems.

Proposition 1.

(a) Suppose P_+ is solution to Problem 4 with $W = C' M^{-1} C$, with C and M defined as in Problem 1 or a solution to Problem 5 with $W = P_- + (a_- - p\mathfrak{r})(a_- - p\mathfrak{r})'$, with p and $\mathfrak{r}$ defined as in Problem 3.

Then any signal of the form of equation 1 *such that* $Var(\alpha \mid y, \mathcal{I}_-) = P_+$ *yields a pair* (Z, Λ) *that solves Problem 1 or Problem 3, respectively. In particular, the canonical signal* y_c *associated with* P_- *and* P_+ *is such a solution.*

(b) Suppose the pair (Z, Λ) *is a solution to Problem 1 or Problem 3, and compute* P_+ *according to eq.* (4).

Then P_+ *is a solution to Problem 4 or Problem 5, respectively, when* W *is as defined in part (a).*

With these results established, we now turn to solving the rational inattention problems.

4 Solution

In this section, we solve the rational inattention problem by (1) determining the posterior covariance matrix P_+ that corresponds to optimal attention allocation; (2) identifying signals that represent optimal information processing; and then (3) computing the optimal action, conditional on the information processed.

4.1 Optimal attention allocation

The chief difficulty in solving the rational inattention problem is that the attention decision for a particular shock is not isolated from the decisions associated with other shocks. This happens for two reasons that are essentially related to the presence of complementarities in information acquisition. First, shocks may be correlated, so that learning about one shock also provides some information about another shock. Second, if any action depends on multiple shocks, this induces a dependency in the value of learning about one shock on the combination of other shocks that are simultaneously being learned about. One or both of these features is present in most problems of interest. For example, the multimarket monopolist problem contains both, since we have allowed levels of demand in the two markets to be correlated through the parameter ρ, and the optimal price in each market depends on both marginal costs and the level of demand in that market.

Our solution to this is to use the insight of lemma 2 to identify a transformation of the fundamental shocks that represent the independent dimensions of uncertainty that the agent wishes to learn about, and use them to decouple the problem. Intuitively, these transformed shocks will have taken into account the complementarities mentioned above, so that the attention allocation decision for each is isolated from the others.

Proposition 2. *Let $\alpha \mid \mathcal{I}_- \sim \mathcal{N}(a_-, P_-)$ and let W be a positive semidefinite matrix, as in Problem 4. Define L to be the lower triangular Cholesky factor of P_- and let QDQ' be the eigendecomposition of the matrix $L'WL$, with eigenvalues d_i arranged in nonincreasing order. Then*

a. Optimal posterior covariance. *The matrix $P_+ = (S'\Delta^+ S)^{-1}$ solves Problem 4, where $S = Q'L^{-1}$ is nonsingular and Δ^+ is a diagonal matrix with entries $\delta_i^+ = \max\{d_i/\lambda, 1\}$, where λ is either the given information cost parameter (in problems with a fixed marginal cost of attention) or is interpreted as a shadow cost (in problems with a fixed capacity of attention), defined below. Moreover, S simultaneously diagonalizes P_+^{-1} and P_-^{-1}, since $P_- = (S'IS)^{-1}$.*

b. Shadow cost for fixed capacity. *Given a fixed capacity of attention κ, the value of λ that solves the problem is*

$$\lambda = \left[e^{-2\kappa}\Pi_{i=1}^{r} d_i\right]^{\frac{1}{r}} \tag{15}$$

if $\kappa > 0$ and is undefined otherwise. The value of r is determined in concert with λ, as follows:

1. *Set $r = n$*
2. *Compute λ as in $eq.$ (15).*
3. *If $d_i > \lambda$ for $i = 1, \ldots, r$, then this pair (r, λ) describes the solution. Otherwise, set $r = r - 1$ and repeat from step (2).*

c. Optimal signal dimension. *The quantity r, defined in lemma 2 and interpreted* *as the number of signals the agent pays attention to, is the integer for which $d_r > \lambda \geq d_{r+1}$.*[18] *Moreover, $r \leq rk(W)$.*

The first part of this proposition can be understood as a generalization of the "reverse water-filling" procedure from the information theory literature, used to solve rate distortion problems with a parallel Gaussian source.[19] This latter procedure is only valid when the weight matrix W is diagonal, which is only the case when, in our terminology, there are no complentarities in information acquisition.[20] Thus another way to view our solution is that by recasting the problem in terms of the canonical shocks, we can recover logic similar to reverse water-filling.[21]

Most importantly, this proposition provides the optimal posterior covariance matrix associated with Problem 4. A second key component, however, is that it constructs the solution in a way that is immediately amenable to lemma 2, and this allows us to directly construct and interpret the canonical shocks and signal.

To illustrate this, we return to the baseline case of the multimarket monopolist problem. After constructing the matrix S using proposition 2, we can compute the canonical shocks, β_c, as shown below. We also include the matrix D, which, we will show, has a useful interpretation as the weight matrix describing information acquisition tradeoffs in terms of the canonical shocks.

$$\beta_c \equiv S\alpha = \begin{bmatrix} \frac{1}{\sqrt{3\sigma^2}}\bar{p}^* \\ \frac{1}{2\sqrt{2\sigma^2}}\tilde{\ell} \\ \frac{1}{\sqrt{6\sigma^2}}\left(\gamma - \frac{1}{2}(\ell_1 + \ell_2)\right) \end{bmatrix} \qquad D = \sigma^2 \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{16}$$

The first two elements of the canonical shock β_c provide information about the optimal average price, $\bar{p}^*$, and the demand differential between markets, $\tilde{\ell}$. These are the two aspects of information that we previously asserted the optimal signal

[18] We can define $d_0 \equiv \infty$ and $d_{n+1} \equiv -\infty$ to encompass degenerate and full rank solutions.

[19] See for example Theorem 10.3.3 and the related discussion from Cover and Thomas (2006).

[20] The method of Kőszegi and Matějka (2020) can also be understood in terms of a generalized reverse water-filling algorithm. As noted earlier, the procedure here is yet more general than theirs, since we do not restrict the prior covariance matrix, which is an important source of complementarities in information acquisition, to be a scalar times the identity matrix.

[21] An alternative geometric intuition for our solution method is in terms of inscribing non-concentric ellipsoids. See Fulton (2017) for details.

(y_r, above) comprises. The third element of β_c is less easy to directly interpret, though we will return to it shortly.

We have already seen two ways of viewing the uncertainty faced by the firm, namely the perfect information prices p^* and the fundamental shocks α, and the transformation from the former to the latter allowed us to formulate the core rational inattention problem. Now given the canonical shocks, we can transform the problem again, and because the canonical shocks represent the independent dimensions of uncertainty that the monopolist wishes to learn about, purged of the complementarities that make information processing difficult to interpret, this transformation will decouple the n-dimensional problem into n 1-dimensional problems. To see this, rewrite

$$E[(\alpha - a_+)'W(\alpha - a_+)] = E[(\beta_c - b_{c,+})'D(\beta_c - b_{c,+})] = \sum_{i=1}^{n} d_i/\delta_i^+ \qquad (17)$$

where $b_{c,+} = E[\beta_c \mid y, \mathcal{I}_-]$. First, notice that while W describes the tradeoffs to information acquisition in terms of the fundamental shocks, the matrix D describes the tradeoffs in terms of the canonical shocks. Second, since D will always be diagonal by definition, the tradeoff for each element of these shocks is independent of the tradeoff for other shocks. A similar result holds for mutual information, since $0.5(\ln|P_-| - \ln|P_+|) = 0.5\sum_{i=1}^{n} \ln \delta_i^+$, and for the "no forgetting" constraint, since $P_- \succeq P_+$ is equivalent to $\delta_i^+ \geq 1, i = 1, \ldots, n$. Intuitively, in terms of the canonical shocks the firm can consider each attention allocation decision independently, since the complementarities in information acquisition have been internalized.

Returning to our example, we can now examine the third canonical shock, which did not have an immediate interpretation before. In particular, we can now see that errors tracking this shock are given zero weight by the firm, since $d_3 = 0$, and so it represents a dimension of the fundamental shocks that the firm will choose to never learn about. In fact, the firm would prefer to *forget* any prior information it had about this shock, but it is constrained from doing so. As a result, defining this shock is important because it captures residual information from the firm's prior that must be accounted for in defining the optimal posterior.

So far we have only considered the baseline case of the multimarket monopolist problem, but it is straightforward to apply the same analysis to other parameterizations of the problem. In Table 1, we showcase the variety of optimal information processing schemes that can arise by deviating from the baseline case for only one of the parameters m_2 or ρ.

Table 1: Information processing by a multimarket monopolist

m_2	ρ	$r \leq$	$\beta_{1,c} \propto$	$\beta_{2,c} \propto$	d_1	d_2
0.5	0	2	$\bar{p}^*$	$\tilde{\ell}$	$6\sigma^2$	$4\sigma^2$
0	0	2	ℓ_2	p_1^*	∞	$5\sigma^2$
∞	0	2	γ	ℓ_1	∞	$4\sigma^2$
0.5	1	1	$\bar{p}^*$	$\cdot$	$10\sigma^2$	0
0.5	-1	2	$\tilde{\ell}$	$\bar{p}^*$	$8\sigma^2$	$2\sigma^2$

This table illustrates how the solution changes in five parameterizations: the baseline, inelastic demand, highly elastic demand, perfect correlation, and perfect inverse correlation.[22] In each case, the canonical shocks have a different definition, describing the independent dimensions of the fundamental shocks that are relevant to the firm given the particular set of parameters.

We now turn to our second application, of portfolio choice, and show how an analogous method can be used to solve the attention allocation problem.

Proposition 3. *Let $\alpha \mid \mathcal{I}_- \sim \mathcal{N}(a_-, P_-)$ and let W be a positive semidefinite matrix, as in Problem 5. Define L to be the lower triangular Cholesky factor of P_- and let QDQ' be the eigendecomposition of the matrix $L^{-1}WL^{-1\prime}$, with eigenvalues d_i arranged in nonincreasing order. Then*

a. Given a fixed marginal cost of attention $\kappa > 0$, the matrix $P_+ = (S'\Delta^+ S)^{-1}$ solves Problem 5, where $S = Q'L^{-1}$ is nonsingular and Δ^+ is a diagonal

[22] In cases where a utility weight d_i becomes arbitrarily large, as in the lines of Table 1 corresponding to perfectly elastic or inelastic demand, it is usually more useful to model the problem using a fixed capacity rather than a fixed marginal cost.

matrix with entries

$$\delta_i^+ = \begin{cases} e^{2\kappa} & i = 1 \\ 1 & i > 1 \end{cases}$$

Thus the solution always has $r = 1$. *Moreover,* S *simultaneously diagonalizes* P_+^{-1} *and* P_-^{-1}, *since* $P_- = (S'IS)^{-1}$.

b. *No solution exists to Problem 5 when given a fixed marginal cost of attention* λ, *because the agent will choose* $\delta_i^+ \to \infty$ *for every* i *that satisfies* $d_i > 0$.

Importantly, in this proposition we are able, as in the quadratic case, to construct a solution that is amenable to lemma 2 and that allows us to construct the canonocal shocks and signals. The main difference between the solutions to the quadratic and portfolio choice problems stems from different returns to scale for information in those problems. The quadratic payoffs case of Problem 4 is characterized by decreasing returns to scale, and so there is an interior solution where the agent balances the marginal benefits of reduced uncertainty about the i-th canonical shock against the marginal cost of increased attention. The agent considers this trade-off separately for each of the n shocks, and so the number of shocks that she pays attention to, r, may be any number from zero to n. Problem 5 is instead characterized by increasing returns to scale, so that solutions will be at the boundaries. This means that, given a fixed information processing capacity, investors will devote all their attention to learn about the returns of one particular portfolio, which we will describe later.[23]

4.2 Optimal signals

As a consequence of proposition 1 and lemma 2, and given the canonical shocks from either proposition 2 or 3, the canonical signal, $y_c = \beta_c + \varepsilon_c$, is always optimal. It is easy to interpret because it is a signal of the canonical shocks, and so the signals are independent and can be analyzed individually. In addition, the analysis

[23] By contrast, if investors were able to observe asset returns to any level of precision for a fixed marginal cost of attention, they would choose to purchase perfect information. In the simple portfolio choice model we consider, this would allow them to make arbitrarily large profits.

of canonical shocks from the previous section applies to the corresponding signals, except that instead of describing attention in terms of the quantity of information processed, $\kappa_i = \ln \delta_i^+$, we now describe it in terms of how noisy each signal is, using $\sigma_i^2 = 1/(1 - \delta_i^+)$.[24]

Returning to the baseline case of the multimarket monopolist problem, it is easy to see that the signal y_r that we originally asserted was optimal is simply a rescaled version of the (reduced) canonical signal. This confirms that it is optimal, and we can now define the noise variance terms, $\sigma_i^2 = 1/(\delta_i^+ - 1)$, where $\delta_i^+ = \max\{d_i/\lambda, 1\}$, $d_1 = 6\sigma^2$, and $d_2 = 4\sigma^2$. When information costs are low enough, in particular $\lambda < 4\sigma^2$, both variance terms will be finite. However, as λ increases to $4\sigma^2$, we have both $\sigma_2^2 \to \infty$ and $\kappa_2 \to 0$. Both of these results have the same interpretation: the agent stops paying attention to the second canonical shock, $\tilde{\ell}$.

4.3 Optimal action

We can now describe the optimal action for rationally inattentive agents, by explicitly computing the posterior expected value of canonical and fundamental shocks.

Proposition 4. *Given an optimal canonical signal y_c, the posterior expected value of the canonical shocks, $b_{c,+} = (b_{1,c,+}, \ldots, b_{n,c,+})'$ is*

$$b_{i,c,+} = \begin{cases} \frac{1}{\delta_i^+} b_{i,c,-} + \left(1 - \frac{1}{\delta_i^+}\right) y_{i,c} & \delta_i^+ > 1 \\ b_{i,c,-} & \delta_i^+ = 1 \end{cases}$$

Then the posterior expected value of the fundamental shocks is $a_+ = S^{-1} b_{c,+}$, and the optimal action by a rationally inattentive agent is

$$x_+ = M^{-1}(C S^{-1} b_{c,+} + c_0)$$

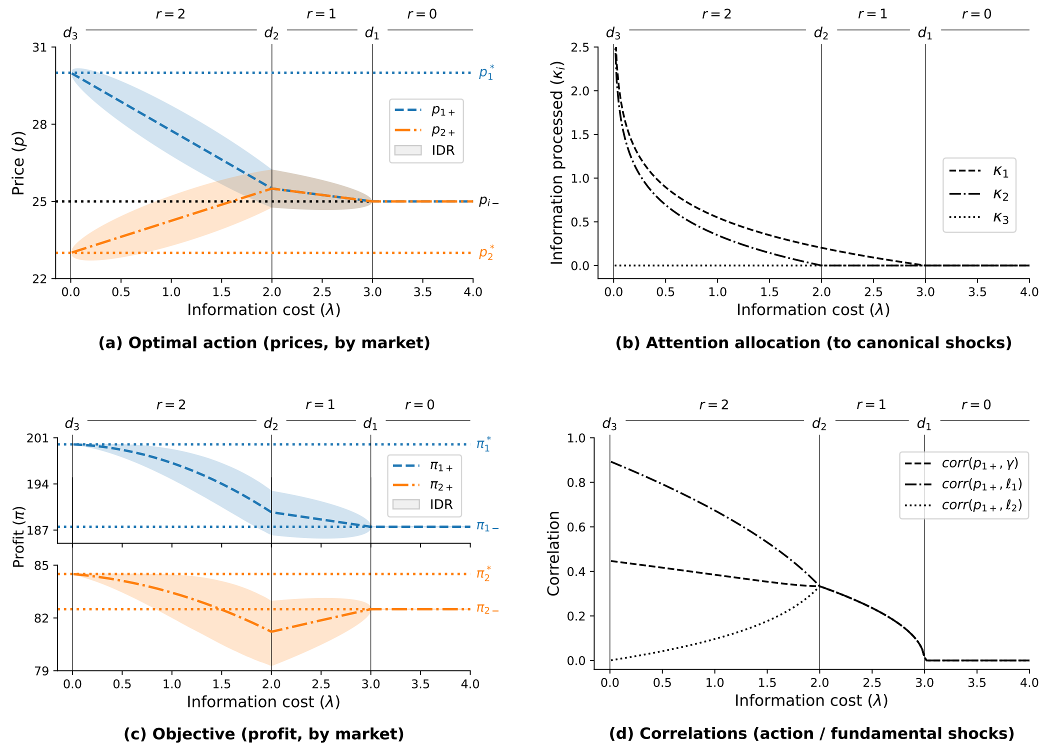
where in the case with quadratic payoffs M and c_0 are given in the problem, and in the case with exponential payoffs, $M \equiv \rho P_+$ and $c_0 \equiv -p\mathfrak{r}$.

[24] This highlights a disadvantage of considering signal vectors, because complete inattention yields $\kappa_i = 0$, but $\sigma_i^2 \to \infty$. As noted earlier, this is a relatively minor inconvenience for the canonical signal, but it can present major difficulties for alternative signal representations.

Here again, the transformation into the space of canonical shocks has decoupled the problem, and this makes Bayesian updating straightforward. The agent processes information about each canonical shock separately, and the posterior for each is simply a weighted average of the prior and the associated canonical signal. The weight depends only on how much attention the agent pays to that shock, as captured by δ_i^+. For those elements about which the agent processes no new information, the posterior is equal to the prior.

With these results, we can now fully examine how the introduction of rational inattention affects firm behavior in our application to the multimarket monopolist problem. To do this, we return to the baseline case and consider how, given a particular realization of the fundamental shocks, changes in the cost of information, λ, affect (a) optimal prices in each market, $p_{i,+}$; (b) the quantities of information processed, κ_i, and (c) profit in each market, $\pi_{i,+}$; and we also show (d) correlations between actions and shocks, $corr(p_{i,+}, \alpha_j)$. Specifically, this exercise sets $m_1 = m_2 = 0.5$, $\rho = 0$, and $\sigma^2 = 0.5$, and sets the prior expectation to be $a_- = (10, 20, 20)$. We present results in fig. 1, where panels (a) - (c) consider the realization $\alpha = (10, 25, 18)$. Note that while the solution to the core rational inattention problems did not depend on the realization of the noise arising from inattention, ε, agents' realized actions and payoffs do depend on it. Therefore, in describing the prices set by firms and the profits obtained, we plot lines showing the expected value and plot shaded areas indicating the interdecile range (IDR) (the range between the 10th and 90th percentiles).

There are clear differences in outcomes across three regions of information costs. First, when information costs are high enough – here, $\lambda > 3$ – the agent chooses to collect no information. Since r is the dimension of the optimal signal vector, this corresponds to the region $r = 0$ in each graph. As we noted earlier, this threshold is determined by the largest loss weight, $d_1 = 6\sigma^2$. In this region, prices must be set using only prior information, and are therefore uncorrelated with the fundamental shocks. Because the agent is already paying no attention, further increases in information costs have no further effect. However, if information costs fall below this threshold, the agent will begin paying attention to the first canonical shock, $\beta_{1,c} \propto \bar{p}^*$, as shown in panel (b). This corresponds to the region $r = 1$. De-



(a) Optimal action (prices, by market)

(b) Attention allocation (to canonical shocks)

(c) Objective (profit, by market)

(d) Correlations (action / fundamental shocks)

Figure 1: Solutions to baseline monopolist problem with varying information costs

spite collecting some information, the firm cannot differentiate between markets, since the chosen signal regards the sum $\gamma + \ell_1 + \ell_2$. As a consequence, the firm cannot pursue price discrimination, and so sets the same price in each market. Since realized overall demand was higher than expected, on average the firm will tend to raise prices relative to their prior. Acquiring this new information increases *overall* expected profits, but, because the realized demand in the second market was below the firm's prior, additional information within the region $r = 1$ actually decreases expected profits *within* the second market.

Finally, if information costs fall below the threshold $d_2 = 4\sigma^2$, the firm will begin to acquire information about the second canonical shock, $\beta_{2,c} \propto \tilde{\ell}$; this corresponds to the region $r = 2$. In this region, the firm collects information that allows them to implement price discrimination, and expected profits begin to rise even in the second market. As information costs fall to zero, the firm moves towards the perfect information solution, and, the correlations between the price set in each market and the level of demand in the other market fall towards zero.

In this way, rational inattention can provide a mechanism for modeling and understanding recently implemented price discrimination strategies by firms that appear to have been made possible by the increased availability of information about their customers. These empirical observations have been documented by, for example, Fudenberg and Villas-Boas (2006), Armstrong (2005) and Taylor (2004), although the theoretical models they employ to explain this behavior are more detailed than ours (for example, considering strategic interactions between firms and consumers). Nonetheless, we have shown that a very simple model of production can generate empirically relevant pricing strategies in a straightforward way when coupled with rational inattention. Moreover, the mechanism here is consistent with observed firm behavior: the increase in information collection arising from advances in electronic monitoring made possible efforts to pursue price discrimination.

5 Application to Portfolio Choice

Our second application is to the portfolio choice problem of Van Nieuwerburgh and Veldkamp (2010). Unlike the case of the monopolist, here we do not need to further parameterize the problem to make substantial progress in interpreting information processing, although to focus on interesting cases, we consider a fixed capacity of attention and we assume that prior expected excess returns are not identically zero.[25] Using proposition 3 and lemma 2, it is straightforward to solve the attention allocation problem.

Proposition 5. Information processing. *Suppose that prior expected excess returns are not all identically zero. Then optimal information processing by a rationally*

[25]If prior expected excess returns are identically zero, then the prior optimal portfolio has no position in any asset and so there is no specific portfolio that the agent wishes to learn about. Instead, they will decompose the assets into a set of independent portfolios with identical variance and learn about the returns of an arbitrary portfolio from this set. Because returns to attention are increasing they will still specialize learning to one portfolio, but no specific portfolio is preferred.

inattentive investor is equivalent to observation of either of the following signals:

1. *A signal about the Sharpe ratio of the prior optimal portfolio:* $y_{\mathcal{S}} = \mathcal{S} + \varepsilon_{\mathcal{S}}$*, with* $\varepsilon_{\mathcal{S}} \sim N(0, 1/(\delta_1^+ - 1))$ *and where* $\mathcal{S} \equiv x'_{-}e/\sqrt{x'_{-}P_{-}x_{-}}$*.*

2. *A signal about returns of the prior optimal portfolio:* $y_x = x'_{-}\alpha + \varepsilon_x$*, with* $\varepsilon_x \sim N(0, x'_{-}P_{-}x_{-}/(\delta_1^+ - 1))$*.*

The first representation is particularly interesting because the Sharpe ratio is a well-known quantity in finance. Intuitively, an asset with a higher Sharpe ratio provides greater compensation for a given level of risk, and this quantity is often used in practice for ranking the performance of assets, portfolios, or investment managers. Amenc, Martellini, and Vaissié (2003) describe it as the most frequently used measure of hedge fund performance.

This result contrasts with that of Van Nieuwerburgh and Veldkamp (2010), who found that the optimal signal would be about the individual asset with the highest prior Sharpe ratio. Our results are different because we allow agents to learn about either assets or portfolios of assets. They choose to learn about the prior optimal portfolio x_{-}, and a second optimal signal representation, given in proposition 5, is a signal about the returns of this portfolio.

Given the optimal information processing strategy described above, we can now describe optimal portfolio choice.

Proposition 6. Optimal portfolio. *Suppose prior expected excess returns are not all identically zero. Then the optimal portfolio for a rationally inattentive investor with fixed information processing capacity* κ *is:*

$$x_{+} = (1 + \chi)x_{-}$$

where $\chi = (\delta_1^+ - 1)(y_{\mathcal{S}}/\mathcal{S}_{-})$*,* $\delta_1^+ = e^{2\kappa}$*, and* x_{-} *is the prior optimal portfolio. The random variable* $y_{\mathcal{S}}$*, defined in proposition 5, is a noisy signal of the Sharpe ratio of the prior optimal portfolio, and* $\mathcal{S}_{-} = E[\mathcal{S} \mid \mathcal{I}_{-}]$ *is the prior expectation of that Sharpe ratio.*

The optimal posterior portfolio is simply a scaled version of the optimal prior portfolio, where the scale factor depends on the information processed about the risk-adjusted returns of the prior optimal portfolio. If the investor's perception of this Sharpe ratio is high, then they will increase their position, particularly if their prior expectation of this quantity was low. This proposition yields several predictions, which we collect in the following corollaries.

Corollary 6.1. Rational inattention and portfolio diversification.

a. Rational inattention does not lead to underdiversification.
b. Additional capacity does not influence portfolio diversification.

Corollary 6.2. Rational inattention and the scale of investment.

a. Information processing decreases purchases of the optimal portfolio if and only if the perceived Sharpe ratio, y_S, is negative.
b. Additional capacity tends to increase the scale of investment.

These results are comforting, since the diversified portfolio chosen by the rationally inattentive investor is consistent with that described by standard portfolio theory. Importantly, they differ from the results of Van Nieuwerburgh and Veldkamp (2010), which were derived under the assumption that agents could not pay attention to portfolios of assets. Nonetheless, we view our models as complementary, since there may be cases in which this assumption is valid. For example, if assets are of very different types, it may make less sense to model an agent as processing information directly about portfolios. On the other hand, when all assets have similar return structures, for example if they are publicly traded equities, then it may make more sense to allow agents to choose portfolios to learn about, as we do here.

6 Dynamic quadratic rational inattention problems

Compared to the static quadratic problem we have considered here, a dynamic version – in which agents make repeated information allocation decisions over time, faced with autocorrelated fundamental shocks – incorporates an additional complementarity to information acquisition: more precise information today can lessen uncertainty tomorrow. This is not a feature of Problem 4, and so proposition 2 cannot be used to solve for the optimal posterior, P_+. Despite this, lemma 2 still applies, and so the analysis and interpretation of information acquisition that we have introduced – that based on canonical signals based on canonical shocks – can still be used. In the future, it will be interesting to combine the tools becoming available to solve for the optimal posteriors in dynamic problems with the tools for analysis and interpretation of agents' behavior that we have developed here.

7 Conclusion

In this paper, we solve two classes of static rational inattention problems and provide new tools to analyze optimal information processing and interpret the behavior of rationally inattentive agents. In particular, we show how to account for complementarities in information acquisition, and so expand analytic results from special cases used in the prior literature to a more general setting. In two examples, taken from behavioral industrial organization and finance, we show why this is important. In each case, the model could not be solved using previously available methods without imposing simplifying assumptions that generate qualitatively different results and imply suboptimal behavior by agents.

Appendices

A Proofs

Lemma A1: Problem 1 is equivalent to

$$\begin{aligned}&\min_{Z\in M_{r,n},\Lambda\in M_r} tr(C'MCP_+) + \lambda\Big(\ln|P_-| - \ln|P_+|\Big)\\ &\text{subject to } \Lambda \succ 0, P_+^{-1} = P_-^{-1} + Z'\Lambda^{-1}Z\end{aligned}$$

Proof: Denote $x_+^* = E[x^* \mid y]$ and $a_+ = E[\alpha \mid y]$. Then we can rewrite the first term of the objective function from Problem 1 as

$$\begin{aligned}-\frac{1}{2}E\Big[(x^* - x_+^*)'M(x^* - x_+^*)\Big] &= -\frac{1}{2}E\Big[(\alpha - a_+)'C'M^{-1}MM^{-1}C(\alpha - a_+)\Big]\\ &= -\frac{1}{2}tr(C'M^{-1}CP_+)\end{aligned}$$

Substituting in for κ in the second term, we have $\lambda\kappa = (\lambda/2)(\ln|P_-| - \ln|P_+|)$. Finally, we multiply this transformed objective by -2 and accordingly change from maximization to minimization.

Lemma A2: Let $e = \alpha - p\mathfrak{r}$ be the vector of "excess returns". Since p and $\mathfrak{r}$ are fixed, the prior distribution for e follows directly from the prior for α, so that $e \sim \mathcal{N}(e_-, P_-)$ where $e_- = a_- - p\mathfrak{r}$. Then Problem 3 is equivalent to

$$\begin{aligned}&\min_{Z\in M_{r,n},\Lambda\in M_r} -tr\Big((P_- + e_-e_-')P_+^{-1}\Big) + \lambda\Big(\ln|P_-| - \ln|P_+|\Big)\\ &\text{subject to } \Lambda \succ 0, P_+^{-1} = P_-^{-1} + Z'\Lambda^{-1}Z\end{aligned}$$

Proof: Notice that the posterior for excess returns is $e \mid y \sim \mathcal{N}(e_+, P_+)$ where $e_+ = a_+ - p\mathfrak{r}$. Then we can rewrite the first term of the objective function from Problem 3 as

$$\begin{aligned}E\Big[\rho E[\omega \mid y] - \frac{\rho^2}{2}Var(\omega \mid y)\Big] &= \rho\omega_0\mathfrak{r} + E\left[e_+'P_+^{-1}e_+ - \frac{1}{2}e_+'P_+^{-1}P_+P_+^{-1}e_+\right]\\ &= \rho\omega_0\mathfrak{r} - \frac{1}{2}n + \frac{1}{2}tr\left(P_+^{-1}(P_- + e_-e_-')\right)\end{aligned}$$

Substituting in for κ in the second term, we have $\lambda\kappa = (\lambda/2)(\ln|P_-| - \ln|P_+|)$. We eliminate the constant term as it does not influence the solution. Finally, we multiply the transformed objective by -2 and accordingly change from maximization to minimization.

A.1 Proof of Lemma 1 Suppose that the pair $(Z \in M_{r,n}, \Lambda \in M_n)$ is a solution to Problem 1 or Problem 3, and let $X \in M_r$ be nonsingular. We claim that the pair $(XZ, X\Lambda X')$ represents another solution to the respective problem. To show this, note that both the former and latter signal lead to the same posterior covariance matrix

$$P_-^{-1} + (XZ)'(X\Lambda X')^{-1}(XZ) = P_-^{-1} + Z'\Lambda^{-1}Z$$

By Lemmas A1 and A2, either of these signals must therefore yield identical objective function values. Moreover, for any X nonsingular, $\Lambda \succ 0 \implies X\Lambda X' \succ 0$. Therefore, if the former signal is feasible under the constraints, then so is the latter. Finally, Problem 2 is a special case of Problem 1 and so this proof applies *a fortiori*.

A.2 Proof of Lemma 2 **Part (a)**: Part (i) follows directly from Theorem 7.6.4 of Horn and Johnson (2012). Part (ii) follows from the assumption that $P_- \succeq P_+$. To prove part (iii), suppose that $\exists\ T \neq S$ such that $P_+^{-1} = T'\Delta^+ T$ and $P_-^{-1} = T'IT$. Then $\exists$ orthogonal matrix V such that $T' = S'V$, and we must have $P_+^{-1} = T'\Delta^+ T = S'V\Delta^+ V'S = S'\Delta^+ S$. This implies $V\Delta^+ = \Delta^+ V$, or $v_{ij}\lambda_i = v_{ij}\lambda_j$ for all $i, j = 1, \ldots, n$. Since $\Delta^+ \succ 0$, $\lambda_i > 0\ \forall\ i$. Then if $\lambda_i \neq \lambda_j$, then it must be that $v_{ij} = 0$ and if $\lambda_i = \lambda_j$ then it must be that $v_{ij} = 1$. Since V is orthogonal, this implies that it must be a permutation matrix, where the only rows that may be permuted are those that correspond to duplicated diagonal elements λ_i.

Part (b): Using the definitions from Part (a) we have $Var(\beta_c \mid \mathcal{I}_-) = SP_-S' = I$.

Part (c): The first part follows directly from equation eq. (4). Then $Var(\beta_c \mid y_r, \mathcal{I}_-) = SP_+S' = (\Delta^+)^{-1}$.

Part (d): Under the convention that $\Lambda_c^{-1} = (\Lambda_r^{-1} \oplus 0_{n-r})$, with 0_{n-r} an $n - r \times n - r$ matrix of zeros, this follows directly from equation eq. (4). Then $Var(\beta_c \mid y_r, \mathcal{I}_-) = SP_+S' = (\Delta^+)^{-1}$.

Part (e): We use the well-known definition of mutual information for jointly Gaussian random vectors, $I(\alpha, y \mid \mathcal{I}_-) = \frac{1}{2}(\ln|P_-| - \ln|P_+|)$ (see for example Sims (2003)). Then $\frac{1}{2}(\ln|P_-| - \ln|P_+|) = \frac{1}{2}(\ln|I| + \ln|\Delta^+|) = \sum_{i=1}^{n} \log_b \delta_i^+$.

A.3 Proof of Proposition 1 The proof for the link between Problems 1 and 4 follows directly from Lemma A1, while the proof for the link between Problems 3 and 5 follows from Lemma A2.

A.4 Proof of Proposition 2

A.4.1 Proof of part (a) We note that we can assume WLOG that P_+ is positive definite, since if it were not the objective function would grow without bound. Define $Z\Delta^+Z'$ to be the eigendecomposition of the matrix $L'P_+^{-1}L$, and apply Lemma 2 (a) to simultaneously diagonalize $P_-^{-1} = S'IS$ and $P_+^{-1} = S'\Delta^+S$. Let $N = (\Delta^+)^{-1}$ and note that $S = Z'L^{-1}$. Applying the result of Lemma 2 part (e), we can rewrite the objective function in Problem 1 as $tr(Z'L'WLZN) - \lambda\sum_{i=1}^{n} \ln n_i$. The decision variables are now the unitary matrix Z and the diagonal elements of N, but Z only appears in the first term. A standard result is that when minimizing the first term over unitary matrices, the optimal Z is exactly the matrix of eigenvectors of $L'WL$, regardless of the diagonal elements of N. Thus, defining QDQ' to be the eigendecomposition of $L'WL$ yields the result that $S = Q'L^{-1}$.

Plugging this in, the problem becomes $\min_{\{n_i\}_{i=1}^{n}} \sum_{i=1}^{n}(d_i n_i - \lambda \ln n_i)$, subject to $n_i \leq 1$ for $i = 1, \ldots, n$. (Remark: we reserve δ_i^+ and n_i^+ for the solution, and so we have used δ_i and n_i here in the problem definition). We can thus consider the solution for each i separately.

If $d_i > 0$, then the objective is convex and the constraint is linear so that the solution, denoted n_i^+, is characterized by the Kuhn-Tucker conditions. The first-order condition yields $n_i = \lambda/d_i$, and the full solution is $n_i^+ = \min\{\lambda/d_i, 1\}$

If $d_i = 0$, then the problem is $\min_{n_i} -\lambda \ln n_i$ and the solution sends $n_i \to \infty$ so that the constraint will be binding and $n_i^+ = 1$.

A.4.2 Proof of part (b) The proof to part (a) is still valid in this case, except that λ is interpreted as a Langrange multiplier so that we must also derive its value at the solution. To do so, note that the associated constraint is $\frac{1}{2}(\log_b |P_-| - \log_b |P_+|) \leq \kappa$ and we can rewrite it as $\frac{1}{2}\sum_{i=1}^{n} \log_b \delta_i^+ \leq \kappa$.

In any solution, all processing capacity will be used, so that this constraint will hold with equality. Define r such that $d_i > \lambda$ for $i = 1, \ldots, r$ and $d_i \leq \lambda$ for $i = r+1, \ldots, n$. From part (a), we will have $\delta_i^+ = 1$ for $i > r$, and so the constraint is $\sum_{i=1}^{r} \log_b \delta_i^+ = 2\kappa$. Then we have

$$\log_b \prod_{i=1}^{r} \frac{d_i}{\lambda} = 2\kappa \implies \lambda^r = b^{-2\kappa} \prod_{i=1}^{r} d_i \implies \lambda = \left[b^{-2\kappa} \prod_{i=1}^{r} d_i \right]^{\frac{1}{r}} \tag{18}$$

Since the choice of r depends on λ, we can compute r in the following way. Initialize $r = n$. First, compute the λ associated with r. If $d_i > \lambda$ for $i = 1, \ldots, r$, then this is the solution. If $\exists i \leq r$ for which $d_i \leq \lambda$, then set $r = r - 1$ and repeat these steps.

A.4.3 Proof of part (c) To show the first part, notice that $\delta_i^+ > 1$ *iff* $d_i > \lambda$. Thus the definition of r as the rank of the matrix $\Delta^+ - I$ is determined by the number of eigenvalues d_i such that $d_i > \lambda$. For the second part, notice that if $rk(W) = k$, then $d_i = 0 \not> \lambda$ for $i > k$, so $rlek$.

A.5 Proof of Proposition 3

A.5.1 Proof of part (a) Here we follow the steps in the proof of Proposition 2 part (a), except that the first term in the objective function is $tr(Z'L^{-1}(-W)L^{-1'}Z\Delta^+)$, so that the optimal Z is the matrix of eigenvectors from $L^{-1}WL^{-1'}$ (multiplication by -1 does not affect the eigenvectors). Thus in this case the definition of S is the same, but we define QDQ to be the eigendecomposition of $L^{-1}WL^{-1'}$. After converting to a maximization problem via multiplication by negative one, we can rewrite Problem 5 as $\max_{\{\delta_i\}_{i=1}^{n}} \sum_{i=1}^{n} d_i \delta_i$, subject to $\delta_i \geq 1$ and $\frac{1}{2}\sum_{i=1}^{n} \log_b \delta_i = \kappa$. Thus the marginal benefit of increasing any particular δ_i is a constant, equal to d_i, but the constraint is concave in δ_i. Thus it will always be optimal for the agent to allocate all attention to a single canonical shock associated with a maximum eigenvalue d_i. Since we have ordered $\{d_i\}$ in nonincreasing order, we can without loss of generalization assume that the agent pays attention to the shock $i = 1$. In other words, in the case that there are multiple eigenvalues that share the maximum value, the agent will choose an arbitrary shock from that group to allocate all of their attention to, and we order that shock first.

A.5.2 Proof of part (b) Rewriting as in part (a), the problem is $\max_{\{\delta_i\}_{i=1}^n} \sum_{i=1}^n (d_i\delta_i - \lambda \ln \delta_i)$, subject to $\delta_i \geq 1$ for $i = 1, \ldots, n$. We can thus consider the solution for each i separately.

If $d_i > 0$, then the objective is convex and there is no interior solution. Since the objective becomes infinite as $\delta_i \to \infty$, there is no solution.

If $d_i = 0$, then the problem is $\max_{\delta_i} -\lambda \ln \delta_i$ and the solution sends $\delta_i \to 0$ so that the constraint will be binding and $\delta_i^+ = 1$.

A.6 Proof of Proposition 4 The first part of this proposition is simply an application of Bayesian updating given the reduced canonical representation, y_r: $a_+ = a_- + P_-S_r'(S_rP_-S_r' + \Lambda_r)^{-1}(y_r - S_ra_-)$. Then, $S_rP_-S_r' = I_r$ and $SP_-S_r' = [I_r; 0]'$. Since $\Lambda_r = (\Delta_r - I)^{-1}$, we have $(S_rP_-S_r' + \Lambda_r)^{-1} = \text{diag}(\{1 - 1/\delta_i^+\}_{i=1}^r)$. Then since $y_{i,c} = y_{i,r}$ for $i = 1, \ldots, r$, we can rewrite $b_{c,+} = Sa_+$ or $b_{i,c,+} = b_{i,c,-} + (1 - 1/\delta_i^+)(y_{i,c} - b_{1,c,-})$ for $i = 1, \ldots, r$ and $b_{i,c,+} = b_{i,c,-}$ for $i = r+1, \ldots, n$.

The second part of this proposition, $a_+ = S^{-1}b_+$, is just an identity resulting from Lemma 2. It is valid since the matrix S is nonsingular.

The third part of the proposition is straightforward, since in both the quadratic and exponential case we showed in the definition of the problem that the optimal action would be $x_+ = M^{-1}c_+$ (for the exponential case, this requires defining $M \equiv \rho P_+$). Then we substitute in as follows: $M^{-1}c_+ = M^{-1}(Ca_+ + c_0) = M^{-1}(CS^{-1}b_{c,+} + c_0)$.

A.7 Proof of Proposition 5 In this case, $L^{-1}WL^{-1\prime} = (I + (L^{-1}e_-)(L^{-1}e_-)')$ and its eigenvectors are the same as those of $(L^{-1}e_-)(L^{-1}e_-)'$. This latter matrix is a rank one matrix, and its only eigenvector associated with a nonzero eigenvalue is

$q_1 = L^{-1}e_-/\|L^{-1}e_-\|$, while the other eigenvectors $q_i, i > 1$ are defined to form an orthonormal basis with q_1. Notice that $\|L^{-1}e_-\| = \sqrt{e_-' P_-^{-1} e_-}$. The eigenvalues of $L^{-1}WL^{-1\prime}$ are $d_1 = (1 + e_-' P_-^{-1} e_-)$ and $d_i = 1,\ i = 2, \ldots, n$. Then the first canonical shock is defined by $s_1' = q_1' L^{-1} = e_-' P_-^{-1}/\sqrt{e_-' P_-^{-1} e_-}$. Recalling that $x_- = \frac{1}{\rho} P_-^{-1} e_-$, we can rewrite $s_1 = x_-/\sqrt{x_-' P_- x_-}$.

The reduced canonical representation is $y_r = s_1'\alpha + \varepsilon_r$ where $\varepsilon_r \sim N(0, 1/(1 - \delta_1^+))$. Two results allow us to to define the signal representations given in the proposition. First, any scalar multiple of this representation is also a feasible representation. Second, any new representation created from y_r by the addition of a constant is also feasible. Then we have $y_x = \sqrt{x_-' P_- x_-} y_r = x_-'\alpha + \varepsilon_x$ where $\varepsilon_x \sim N\left(0, \frac{x_-' P_- x_-}{\delta_1^+ - 1}\right)$, and $y_{\mathcal{S}} = y_r - s_1' pr = \mathcal{S} + \varepsilon_{\mathcal{S}}$ where $\varepsilon_{\mathcal{S}} \sim N(0, 1/(\delta_1^+ - 1))$. Finally, $\mathcal{S}$ is the realized Sharpe ratio associated with the prior optimal portfolio, $\mathcal{S} = x_-' e/\sqrt{x_-' P_- x_-}$.

A.8 Proof of Proposition 6 Recall that, given our assumptions, the solution is $x_+ = \frac{1}{\rho} P_+^{-1} e_+$, where $e_+ = a_+ - p\mathfrak{r}$. Then we have

$$
\begin{aligned}
x_+ &= \frac{1}{\rho} S' \Delta^+ S(a_+ - p\mathfrak{r}) = \frac{1}{\rho} S' S(a_- - p\mathfrak{r}) + \frac{1}{\rho} S' \iota_1 (\delta_i^+ - 1) y_{\mathcal{S}} \\
&= x_- + (\delta_i^+ - 1) \frac{y_{\mathcal{S}}}{\mathcal{S}_-} x_- = (1 + \chi) x_-
\end{aligned}
$$

where $\chi = (\delta_i^+ - 1)\frac{y_{\mathcal{S}}}{\mathcal{S}_-}$ and ι_i is the i-th column of the identity matrix.

A.9 Proof of Corollary 6.1 Both parts of this Corollary follow immediately from Proposition 6, since the effect of rational inattention is entirely captured by the scalar χ.

A.10 Proof of Corollary 6.2 Formally, part (a) simply notes that the sign of χ is entirely determined by the sign of $y_{\mathcal{S}}$, which is the perceived Sharpe ratio of the portfolio x_-. Part (b) notes the effect of additional attention is an increase in the parameter δ_i^+. For any given perceived Sharpe ratio $y_{\mathcal{S}}$, a higher δ_i^+ implies a stronger response (either positive or negative).

References

Afrouzi, Hassan, and Choongryul Yang. 2019. “Dynamic Rational Inattention and the Phillips Curve”. *Available at SSRN 3465793*.

Amenc, Noël, Lionel Martellini, and Mathieu Vaissié. 2003. “Benefits and risks of alternative investment strategies”. *Journal of Asset Management* 4, no. 2 (): 96–118.

Armstrong, Mark. 2005. *Recent Developments in the Economics of Price Discrimination* 0511004. University Library of Munich, Germany.

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

Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory*. 00076. John Wiley & Sons.

Dewan, Ambuj. 2019. *Costly Multidimensional Information*.

Ellison, Glenn. 2006. “Bounded Rationality in Industrial Organization”. In *Advances in Economics and Econometrics*, ed. by Richard Blundell, Whitney K. Newey, and Torsten Persson, 142–174. Cambridge: Cambridge University Press.

Fudenberg, Drew, and J Miguel Villas-Boas. 2006. “Behavior-based price discrimination and customer recognition”. Bibtex*[publisher=Amsterdam: Elsevier], *Handbook on economics and information systems* 1:377–436.

Fulton, Chad. 2017. *Mechanics of Linear Quadratic Gaussian Rational Inattention Tracking Problems* 2017-109. Board of Governors of the Federal Reserve System (U.S.)

Horn, Roger A., and Charles R. Johnson. 2012. *Matrix Analysis*. 32220 Google-Books-ID: O7sgAwAAQBAJ. Cambridge University Press.

Jung, Junehyuk, et al. 2019. “Discrete Actions in Information-Constrained Decision Problems”. *The Review of Economic Studies* ().

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. "A Rational Theory of Mutual Funds' Attention Allocation". *Econometrica* 84, no. 2 (): 571–626.

Kamdar, Rupal. 2018. *The Inattentive Consumer: Sentiment and Expectations*. manuscript.

Kőszegi, Botond, and Filip Matějka. 2020. "Choice Simplification: A Theory of Mental Budgeting and Naive Diversification". Publisher: Oxford Academic, *The Quarterly Journal of Economics* 135, no. 2 (): 1153–1207.

Luo, Yulei. 2008. "Consumption dynamics under information processing constraints". Publisher: Elsevier, *Review of Economic dynamics* 11 (2): 366–385.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2018. "Dynamic rational inattention: Analytical results". *Journal of Economic Theory* 176 (): 650–692.

Maćkowiak, Bartosz, and Mirko Wiederholt. 2015. "Business Cycle Dynamics under Rational Inattention". 00155, *The Review of Economic Studies* 82, no. 4 (): 1502–1532.

— . 2009. "Optimal Sticky Prices under Rational Inattention". 00324, *The American Economic Review* 99, no. 3 (): 769–803.

Matějka, Filip. 2016. "Rationally Inattentive Seller: Sales and Discrete Pricing". Publisher: Oxford Academic, *The Review of Economic Studies* 83, no. 3 (): 1125–1155.

Miao, Jianjun, Jieran Wu, Eric Young, et al. 2019. *Multivariate Rational Inattention*. Boston University-Department of Economics.

Miao, Jianjun, and Hao Xing. 2020. "Dynamic Discrete Choice under Rational Inattention".

Mondria, Jordi. 2010. "Portfolio choice, attention allocation, and price comovement". *Journal of Economic Theory* 145, no. 5 (): 1837–1864.

Paciello, Luigi, and Mirko Wiederholt. 2014. "Exogenous Information, Endogenous Information, and Optimal Monetary Policy". Publisher: Oxford Academic, *The Review of Economic Studies* 81, no. 1 (): 356–388.

Sims, Christopher A. 2003. “Implications of rational inattention”. 01010, *Journal of Monetary Economics* 50, no. 3 (): 665–690.

— . 2010. *Rational Inattention and Monetary Economics*. Handbook of Monetary Economics. 00065. Elsevier.

Taylor, Curtis R. 2004. “Consumer Privacy and the Market for Customer Information”. *The RAND Journal of Economics* 35 (4): 631–650.

Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2010. “Information Acquisition and Under-Diversification”. *The Review of Economic Studies* 77, no. 2 (): 779–805.