Bayesian Learning*

Isaac Baley†
UPF, CREI, BGSE and CEPR

Laura Veldkamp ‡
Columbia University, NBER, and CEPR

July 16, 2021

Abstract

We survey work using Bayesian learning in macroeconomics, highlighting common themes and new directions. First, we present many of the common types of learning problems agents face—signal extraction problems—and trace out their effects on macro aggregates, in different strategic settings. Then we review different perspectives on how agents get their information. Models differ in their motives for information acquisition and the cost of information, or learning technology. Finally, we survey the growing literature on the data economy, where economic activity generates data and the information in data feeds back to affect economic activity.

JEL: D80, D81, D83, D84, E20, E30

Keywords: Bayes' law, passive learning, active learning, signal extraction, information choice, sticky information, rational inattention, experimentation, data economy, coordination games.

*In preparation for the Handbook of Economic Expectations. For useful discussions and feedback, we thank Vladimir Asriyan, Andrés Blanco, Ana Figueiredo, Manolis Galenianos, Benjamin Hébert, Chad Jones, Boyan Jovanovic, Julian Kozlowski, Albert Marcet, Jordi Mondria, Kristoffer Nimark, Luigi Paciello, Lubos Pastor, Luminita Stevens, Robert Ulbricht, Victoria Vanasco, Mirko Wiederholt, and Michael Woodford. Erfan Ghofrani, Ángelo Gutiérrez, Marta Morazzoni, Alejandro Rábano, and Judy Yue provided excellent research assistance.

†Universitat Pompeu Fabra, CREI, Barcelona GSE, and CEPR, isaac.baley@upf.edu.

‡Columbia Graduate School of Business, lv2405@columbia.edu.

1 Introduction

This chapter focuses on Bayesian learning. Learning is the process by which agents form beliefs. While many of the previous chapters consider how to measure beliefs, this chapter uses Bayesian tools to consider how agents form beliefs and the types of consequences these beliefs have on economic outcomes. In one class of models, agents know the true model of the economy and are uncertain only about which realization of the state will be drawn by nature. They use additional pieces of information (e.g., noisy signals) to form expectations about the state. In another class of models, agents are uncertain about the distribution of the state and also use Bayes' law to infer its moments or its shape, using a sample of observations.

Among models with Bayesian learning, there are models of passive learning and models of active learning. In passive learning models, agents are endowed with signals and/or learn as an unintended consequence of observing prices and quantities. One set of examples is models where information is exogenous. Information may be an endowment or it may arrive stochastically. Endogenous information can also be learned passively. For example, information could be conveyed by market prices. This is still passive learning because agents are not exercising any control over the information they observe.

Active learning is intentional. Information is chosen or is the direct result of a choice. This choice might involve purchasing information, choosing how to allocate limited attention or choosing an action, taking into account the information it will generate. Such models go beyond explaining the consequences of having information; they also predict what information agents will choose to have. Models of active learning predict what information agents choose to observe. Because an active-learning model can predict information sets on the basis of observable features of the economic environment, pairing it with a passive-learning model where information predicts observable outcomes results in a model where observables predict observables. Such a model is typically empirically testable.

Economists often use the term learning to refer to a literature in which agents do not use Bayes' Law to form their expectations. One example is adaptive least-squares learning, where agents behave as econometricians, trying to discover the optimal linear forecasting rule for next period's state. Evans and Honkapohja (2001) offer an exhaustive treatment of this literature. That is not our focus.

The chapter is organized as follows. Section 2 introduces a small set of mathematical tools needed to understand the material. Section 3 studies the implications of learning for economic activity given a set of beliefs. We discuss passive learning in signal extraction problems and coordination games with strategic motives in actions and in the use of information. Section 4 discusses several motives for active information acquisition and the most commonly used learning technologies. Finally, Section 5 surveys the growing literature on the data economy, where economic activity generates data and the information in the data feeds back to affect economic activity.

2 Mathematical preliminaries

A few basic concepts and mathematical tools are needed to understand this chapter. Bayes' Law for univariate Normal continuous variables appears many times over. In dynamic settings, this becomes the Kalman filter. For formal derivations and generalizations of Bayes' Law and the Kalman filter, see Liptser and Shiryaev (2001) and Bernardo and Smith (2009).

2.1 Bayesian updating

Bayes' Law. The probability of event A occurring, given that event B occurred is

$$P(A|B) \;=\; \frac{P(B|A)P(A)}{P(B)}, \quad \text{with} \quad P(B) \neq 0. \tag{1}$$

This law comes from the definition of a conditional probability: $P(A|B) = P(A \cap B)/P(B)$.

For continuous random variables with smooth distributions, the probability of any discrete realization is zero. However, Bayes' law can be applied to probability densities as well. Let f be a continuous random variable with a smooth distribution. Then the probability density of event A, given that event B occurred is

$$f(A|B) = \frac{f(B|A)f(A)}{f(B)}, \quad \text{where} \quad f(B) = \int_{-\infty}^{\infty} f(B|A)f(A)\,\mathrm{d}A. \tag{2}$$

Bayes' Law for Normal random variables. Suppose there is an unknown random variable θ and according to agent's prior beliefs $\theta \sim \mathcal{N}(\mu_\theta, \tau_\theta^{-1})$. In other words, before observing any additional information, θ was believed to be μ_θ on average, with a precision of τ_θ. Note that the precision is the inverse of the variance (not the standard deviation). We will work with precisions because doing so generally makes the solutions simpler. Additionally, the agent sees a signal

$$s \;=\; \theta + \eta, \quad \eta \sim \mathcal{N}(0, \tau_s^{-1}). \tag{3}$$

The signal is an unbiased piece of data about θ with precision τ_s and is *conditionally independent* of $\mu_\theta - \theta$. That means that signals and priors are related only because they are both informative about θ, but their errors are independent. Independence implies that $\mathbb{E}[(\mu_\theta - \theta)(s - \theta)] = 0$. Given the prior information and the signal, the agent forms a posterior belief, also called a conditional belief, about the value of θ using Bayes' law:

$$\hat{\theta} \;\equiv\; \mathbb{E}[\theta|s] \;=\; \frac{\tau_\theta \mu_\theta + \tau_s s}{\tau_\theta + \tau_s}. \tag{4}$$

With normal random variables, the posterior belief is simply a weighted average of the prior belief and the signal. Each is weighted by its relative precision. If a signal contains no information about

θ, it would have zero precision. In this case, the posterior belief would be the same as the prior belief. The posterior (or conditional) variance also has a simple form

$$\hat{\Sigma} \equiv \mathbb{V}ar[\theta|s] = \frac{1}{\tau_\theta + \tau_s}. \tag{5}$$

The posterior precision (the inverse of the variance $\hat{\Sigma}^{-1}$) equals the prior precision τ_θ plus the signal precision τ_s. Every additional piece of independent information adds precision to the estimation.[1]

2.2 The Kalman filter

When applied in dynamic models, the formula for Bayesian updating with Normal variables becomes the Kalman filter. This learning applies when agents know the distribution but care about forecasting the actual realization of the variable. We specialize the state θ_t to follow a first order Markov process and let s_t be an unbiased signal about θ_t. The system consists of two equations, one for the hidden state and one for its noisy observation:

$$\theta_{t+1} = \rho\theta_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \tau_\theta^{-1}), \tag{6}$$

$$s_t = \theta_t + \eta_t, \quad \eta_{t+1} \sim \mathcal{N}(0, \tau_s^{-1}). \tag{7}$$

The two shocks ε and η are mutually independent and *i.i.d.* over time. The parameters $\rho \le 1$ and (τ_θ, τ_s) are known and prior beliefs are $\theta_0 \sim \mathcal{N}(\hat{\theta}_0, \Sigma_0)$. For all $t > 0$, let $\hat{\theta}_t$ denote the expectation of θ_t conditional on all the signals s observed up to, but excluding time t: $\hat{\theta}_t \equiv \mathbb{E}[\theta_t|s_0, \ldots, s_{t-1}]$. Also let $\hat{\Sigma}_t \equiv \mathbb{V}ar[\theta_t|s_0, \ldots, s_{t-1}] = \mathbb{E}[(\theta_t - \hat{\theta}_t)^2]$ denote the conditional variance of θ_t. The following three recursive formulas describe how to update the mean and variance of beliefs:

$$\hat{\theta}_{t+1} = \rho\hat{\theta}_t + K_t(s_t - \hat{\theta}_t) \tag{8}$$

$$K_t = \rho\frac{\tau_s}{\hat{\Sigma}_t^{-1} + \tau_s} \tag{9}$$

$$\hat{\Sigma}_{t+1} = \rho^2\frac{1}{\hat{\Sigma}_t^{-1} + \tau_s} + \frac{1}{\tau_\theta}. \tag{10}$$

The term K_t is called the *Kalman gain*. It represents how much weight is put on the new information $(s_t - \hat{\theta}_t)$, relative to the old information in the prior belief $\hat{\theta}_t$ when forming the posterior belief $\hat{\theta}_{t+1}$. The higher the signal precision τ_s, the higher is the weight placed on news. The Kalman gain is the analog of the term $\tau_s/(\tau_\theta + \tau_s)$ in (4). The conditional variance $\hat{\Sigma}_{t+1}$ can be similarly interpreted as the recursive analog of the Bayesian updating formula for posterior variance in (5). It is the inverse of the posterior precision, which equals the sum of the prior and signal precisions.

[1] This observation is not always true. For instance, when agents learn about binomial random variables (e.g., learning a proportion), additional observations may actually reduce precision.

2.3 Learning the distribution of the state

In many setups, agents know the true distribution, but learn about the current state. For some applications, it is important that the distribution of random variables is not known. Pastor and Veronesi (2009) review finance models with parameter learning in more detail. Here, we provide the basics to aid in understanding the papers that follow. We start with a simple case, where an agent knows the distribution is normal, but learns the mean and/or variance by observing a sample of *i.i.d.* realizations of the state.

2.3.1 Learning the mean

Assume that the true distribution of the state is $\theta \sim \mathcal{N}(\mu_\theta, \tau_\theta^{-1})$. Suppose an agent knows the precision τ_θ but ignores the mean μ_θ. She holds prior beliefs about the mean that are Normal: $\mu_\theta \sim \mathcal{N}\left(\mu_0, \tau_0^{-1}\right)$. After observing an *iid* sample of t realizations of θ, which are included into her date-t information set $\mathcal{I}_t = \{\theta_r | r \leq t\}$, the posterior belief is also Normal:[2] $\mu_\theta | \mathcal{I}_t \sim \mathcal{N}\left(\mu_t, \tau_t^{-1}\right)$. The parameters evolve as:[3]

$$(11) \qquad \mu_t = \frac{\tau_0 \mu_0 + t\tau_\theta \overline{\theta}}{\tau_t}, \qquad \tau_t = \tau_0 + t\tau_\theta, \qquad \overline{\theta} = \frac{1}{t}\sum_{r=1}^{t} \theta_r.$$

The posterior belief $\mu_t = \mathbb{E}[\mu_\theta|\mathcal{I}_t]$ is a weighted average of the prior mean μ_0 and the sample mean $\overline{\theta}$. The posterior precision $\tau_t = \mathbb{V}ar[\mu_\theta|\mathcal{I}_t]^{-1}$ is the sum of the prior precision τ_0 and the sample precision $t\tau_\theta$, which grows linearly with the number of observations. As the number of observations increases ($t \to \infty$), the posterior belief converges to the sample mean $\mu_t \to \overline{\theta}$ and precision goes to infinity (uncertainty disappears), and since the sample mean converges to the true mean $\overline{\theta} \to \mu_\theta$, the truth is eventually revealed. In Section 3.1.2, we discuss extensively the literature that applies this type of learning structure to learn about a fixed characteristic (e.g., a worker's ability).

2.3.2 Learning the precision

Next, we consider a setting where the precision is not known. With an *i.i.d.* data sample $\mathcal{I}_t = \{\theta_r | r \leq t\}$, an agent now simultaneously learns about the mean and the precision of θ_t. The standard way to formalize this problem is to use a joint Normal-Gamma distribution. The precision is assumed to follow a Gamma distribution $\tau_\theta \sim \Gamma(\alpha, \beta)$, with density $f(x|\alpha, \beta) \propto (\beta x)^{\alpha-1} e^{-\beta x}$,

[2]When prior and posterior distributions are in the same probability distribution family, like this example, we say that they are conjugate distributions. Working with conjugate distributions is very tractable. Both Normal and Normal-Gamma (used in the next section) are self-conjugate families.

[3]The Kalman filter formulas can be used to learn the mean by setting $\rho = 1$ and $\tau_\theta^{-1} = 0$.

mean α/β and precision β^2/α. Conditional on the precision, the mean is Normal. This formulation is convenient because when a Normal mean and Gamma precision are updated with data drawn from that same type of distribution, the posterior beliefs will also involve a Normally-distributed mean and a Gamma-distributed precision. Given prior beliefs, $\mu_\theta|\tau_\theta \sim \mathcal{N}\left(\mu_0, (\kappa_0\tau_\theta)^{-1}\right)$ and $\tau_\theta \sim \Gamma(\alpha_0, \beta_0)$, the posterior belief about the mean is $\mu_\theta|(\mathcal{I}_t, \tau_\theta) \sim \mathcal{N}\left(\mu_t, (\kappa_t\tau_\theta)^{-1}\right)$, with

$$\mu_t = \frac{\kappa_0\mu_0 + t\bar{\theta}}{\kappa_t}, \qquad \kappa_t = \kappa_0 + t, \qquad \bar{\theta} = \frac{1}{t}\sum_{r=1}^{t}\theta_r. \tag{12}$$

In turn, the parameters governing the precision's posterior distribution $\tau_\theta|\mathcal{I}_t \sim \Gamma(\alpha_t, \beta_t)$ evolve as

$$\alpha_t = \alpha_0 + \frac{t}{2}, \qquad \beta_t = \beta_0 + \frac{1}{2}\left[\sum_{r=1}^{t}\left(\theta_r - \bar{\theta}\right)^2 + \frac{t\kappa_0(\bar{\theta} - \mu_0)^2}{\kappa_t}\right], \tag{13}$$

where $\mathbb{E}[\tau_\theta|\mathcal{I}_t] = \alpha_t/\beta_t$ and $\mathbb{V}ar[\tau_\theta|\mathcal{I}_t]^{-1} = \beta_t^2/\alpha_t$. As the sample size increases ($t \to \infty$), the posterior belief about the mean μ_t in (12) converges to the true value μ_θ and belief uncertainty goes to zero $\kappa_t^{-1} = 0$. Regarding the beliefs about the precision, both α_t and β_t in (13) go to infinity but their ratio α_t/β_t converges to the true precision τ_θ and belief uncertainty goes to zero.

The Normal-Gamma approach is implemented in Cogley and Sargent (2005) to estimate the parameters of a central bank policy rules, and by Weitzman (2007), Bakshi and Skoulakis (2010), and Collin Dufresne, Johannes and Lochstoer (2016) to study asset pricing puzzles. Ghofrani (2021) shows that this framework generates persistent impacts of tail event shocks.

2.3.3 Non-parametric learning

Finally, we consider a case where the functional form of the probability density is not known. Agents use an *i.i.d.* data sample $\mathcal{I}_t = \{\theta_r | r \leq t\}$ to construct a frequentist (not Bayesian) estimate $\hat{f}_t$ of the true density f. A simple approach is to use a Normal kernel density estimator:

$$\hat{f}_t(\theta) = \frac{1}{tb_t}\sum_{s=0}^{t-1}\phi\left(\frac{\theta - \theta_{t-s}}{b_t}\right). \tag{14}$$

Here $\phi(\cdot)$ is the standard normal density function and b_t is the bandwidth parameter. As new data arrives, agents add the new observation to their data set and update their estimates, generating a sequence of beliefs $\{\hat{f}_t\}$. Belief changes tend to be very persistent, even if the θ_t shocks that caused the beliefs to change are transitory. This persistence arises from the martingale property of beliefs: on average, expected future beliefs are the same as current beliefs. As a result, any changes in beliefs induced by new information are approximately permanent. Kozlowski, Veldkamp and Venkateswaran (2020a,b) use this mechanism to generate belief scarring that explains the persistent effects of the Great Recession and the COVID pandemic.

3 Using signals to understand economic activity

This section examines mechanisms though which agents' information affects economic activity. In this section, we take the agents' information sets as given. In other words, learning here is passive. First, we describe signal extraction problems, in which agents try to disentangle permanent from transitory shocks or aggregate from idiosyncratic shocks. Then, we explore coordination games, where strategic motives in actions make the use of information strategic as well.

3.1 Signal-extraction problems

In this environment, an agent's payoff depends on the distance from her action to an unknown stochastic target. This type of quadratic tracking problem is common because it is tractable. Also, one can map many models into this framework, by approximating objectives quadratically.

3.1.1 A tracking problem

The economy is populated by a continuum of agents indexed by $i \in [0,1]$. Every agent chooses a continuous action $a_{it} \in \mathbb{R}$ to minimize the expected distance between her action and an unknown exogenous target a^*_{it} drawn by nature. Each agent solves the following problem:

$$\mathcal{L} = \min_{\{a_{it}\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t (a_{it} - a^*_{it})^2 \Big| \mathcal{I}_{i0}\right], \tag{15}$$

where $\beta < 1$ is the discount factor and $\mathcal{I}_{it}$ denotes agent i's information set at date t. Adding and subtracting the posterior belief $\hat{a}^*_{it} \equiv \mathbb{E}[\hat{a}^*_t | \mathcal{I}_{it}]$ inside the payoff, distributing the expectation, and using the law of iterated expectations, we rewrite the problem in terms of posterior beliefs:

$$\mathcal{L} = \min_{\{a_{it}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}\left[(a_{it} - \hat{a}^*_{it})^2 | \mathcal{I}_{it}\right] + \sum_{t=0}^{\infty} \beta^t \hat{\Sigma}_{it}, \tag{16}$$

where we use the orthogonality of expectional errors $\mathbb{E}[(a_{it} - \hat{a}^*_{it})(a^*_{it} - \hat{a}^*_{it}) | \mathcal{I}_{it}] = 0$ and define the posterior variance $\hat{\Sigma}_{it} \equiv \mathbb{E}[(\hat{a}^*_{it} - a^*_{it})^2 | \mathcal{I}_{it}]$. The problem is equivalent to minimizing the distance between actions and beliefs; the additional term involving the series of conditional variances $\{\hat{\Sigma}_{it}\}$ is a sunk cost that decreases utility but cannot be controlled by the agent because learning is passive. The first order condition implies that the optimal action is the expected value of the target action:

$$a_{it} = \hat{a}^*_{it}. \tag{17}$$

The next two variants of this problem have different stochastic processes for the target action a^*_{it}.

3.1.2 Permanent vs. transitory shocks

In this class of models, the target action is idiosyncratic; it is specific to an individual. But a^*_{it} experiences permanent shocks and transitory shocks. Agents cannot distinguish between these permanent and transitory shocks. Their confusion is what generates interesting learning dynamics.

Suppose the target is an unknown, idiosyncratic, and fixed trait[4]

$$a^*_{it} = \theta_i. \tag{18}$$

The problem is analogous to learning a parameter (as in Section 2.3.1). Agents receive unbiased signals $s_{it} = \theta_i + \eta_{it}$ with noise $\eta \sim_{iid} \mathcal{N}(0, \tau_s^{-1})$. The transitory term prevents agents from backing out the permanent trait. Posterior beliefs are formed using the Kalman formulas in (8), (9), and (10) with $\rho = 1$ and $\tau_\theta^{-1} = 0$. Given initial values $(\hat{a}_{i0}, \hat{\Sigma}_{i0})$, the target forecast and its uncertainty evolve according to

$$\hat{\theta}_{it+1} = \hat{\theta}_{it} + \frac{\tau_s}{\hat{\Sigma}_{it}^{-1} + \tau_s}(s_{it} - \hat{\theta}_{it}); \qquad \hat{\Sigma}_{it+1}^{-1} = \hat{\Sigma}_{it}^{-1} + \tau_s. \tag{19}$$

Early applications of this setup were examined in labor markets. In Jovanovic (1979, 1984), worker-firm match quality—a fixed trait—is an experience good that is gradually revealed by noisy output performance. Learning generates a selection effect: Only the relationships with high match quality survive and job tenure becomes a sufficient statistic for match quality and uncertainty. Besides learning about a match-specific productivity term (Pries and Rogerson, 2005; Nagypál, 2007; Menzio and Shi, 2011), learning can be about innate worker skills in different occupations (Miller, 1984; Neal, 1999; Moscarini, 2001; Groes, Kircher and Manovskii, 2014; Papageorgiou, 2014; Wee, 2016; Baley, Figueiredo and Ulbricht, 2021) or firm characteristics (Borovicková, 2016). In Gonzalez and Shi (2010) and Doppelt (2016), informational dynamics are driven by unemployment, as the posterior probability of being high skilled worsens with the length of unemployment spells.

Learning about fixed characteristics through noisy signals has also been used to examine technology choice (Jovanovic and Nyarko, 1996); entrepreneurship (Minniti and Bygrave, 2001); firm profitability (Pástor, Taylor and Veronesi, 2009); durable consumption (Luo, Nie and Young, 2015); firms' life-cycle (Arkolakis, Papageorgiou and Timoshenko, 2018; Chen, Senga, Sun and Zhang, 2020); and the impact of government policy (Pastor and Veronesi, 2012). Policymakers may also behave as Bayesian agents when learning about climate change parameters (Kelly and Kolstad, 1999) or the trade-off between inflation and unemployment (Cogley and Sargent, 2005; Sargent, Williams and Zha, 2006; Primiceri, 2006).

On the empirical front, Lee and Moretti (2009) use high-frequency data from political prediction markets to show that investors process information contained in polls in a Bayesian way.

[4]More generally, the target may follow a persistent process as in Section 2.2.

Others test learning by exploiting *tenure*—the duration of a relationship—as a proxy for uncertainty. Farber and Gibbons (1996) show that wages of long-tenured workers correlate more with unobserved skills (measured via test scores); Kellogg (2011) show that the productivity of an oil company and its drilling contractor increases with their joint experience; and Botsch and Vanasco (2019) show that loan terms become more correlated with unobserved firm characteristics as the duration of the lender-creditor relationship increases. Lastly, another set of papers exploits dynamic cross-sectional moments in the microdata, such as hazard rates, to discipline the speed of learning and recover the dynamics of information sets. Álvarez, Lippi and Paciello (2011), Baley and Blanco (2019) and Argente and Yeh (2021) use price-adjustment hazards while Borovicková (2016) and Baley, Figueiredo and Ulbricht (2021) use job separation hazards.

Keeping uncertainty alive. According to (10), belief uncertainty (forecast error variance) $\hat{\Sigma}_{it}$ continuously decreases until it reaches a minimum value in the long run. In particular, when agents learn about a fixed characteristic as in (19), uncertainty eventually disappears: $\hat{\Sigma}_\infty \equiv \lim_{t\to\infty} \hat{\Sigma}_{it} = 0$. In the cross-section, differences in uncertainty also disappear. In some setups, this is not a desirable feature. Especially when models aim to explain the cross-sectional differences in uncertainty observed in the data.

The literature proposes various mechanisms to keep uncertainty dynamics active and to generate cross-sectional dispersion in uncertainty. Baley and Blanco (2019) develop a menu cost price-setting model where firm productivity θ_i is subject to occasional shocks (fat-tail risk). Learning about fat-tailed shocks generates uncertainty cycles that translate into cross-sectional dispersion in the frequency of price adjustment and amplify the real effects of monetary policy. Senga (2018) explores the role of uncertainty cycles in a model with heterogenous firms in explaining the level and cyclicality of the cross-sectional dispersion of sales growth. In Baley, Figueiredo and Ulbricht (2021), uncertainty about workers' abilities jumps up when they endogenously switch their occupation and start learning about a new set of occupation-specific abilities. Uncertainty cycles about worker abilities explain features of labor market dynamics at the micro and macro levels.

3.1.3 Aggregate vs. idiosyncratic shocks

Agents can also be confused between aggregate and idiosyncratic factors. Suppose that the target action a^*_{it} is now a linear combination of an aggregate factor common across agents θ_t and an individual factor v_{it} specific to agent i:

$$a^*_{it} = (1-r)\theta_t + r(v_{it} - \theta_t), \quad \text{with} \quad r \in [0,1]. \tag{20}$$

For simplicity, the aggregate shock follows a random walk $\theta_t = \theta_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0, \tau_\theta^{-1})$ and the idiosyncratic shock is *i.i.d* across time and agents $v_{it} \sim \mathcal{N}(0, \tau_v^{-1})$. One noisy signal provides information about the sum of the two components $s_{it} = \theta_t + v_{it} + \eta_{it}$, with a common noise distribution $\eta_{it} \sim \mathcal{N}(0, \tau_s^{-1})$. With one signal and two shocks, agents cannot disentangle the components and mistakenly attribute part of an aggregate shock ϵ_t to an idiosyncratic shock v_{it}.

The confusion between aggregate and idiosyncratic shocks is at the core of the Phelps (1970) and Lucas (1972) island's model. In each island i, a representative agent chooses how much to work a_{it} depending on the demand for her own good v_{it} relative to the aggregate demand θ_t. Setting $r = 1$, agents would like to work more only if they believe their relative price $v_{it} - \theta_t$ is high. Thus the nature of the shock matters. Agents see their own price s_{it}, but cannot tell whether their price is high because nominal demand θ_t is high or because island-specific real demand v_{it} is high. This mechanism gives rise to monetary non-neutrality. When money is abundant, aggregate demand increases, and most agents observe a high price for their good. Since they cannot disentangle the source of higher demand, they work harder, and produce more. Money has real effects.

Hellwig and Venkateswaran (2009) also investigate the implications of signals which combine aggregate and idiosyncratic shocks but in a nominal price-setting context where firms choose their price a_{it} to be close to the target. Setting $r \in [0, 1/2]$ in (20), the optimal price depends positively on aggregate and idiosyncratic factors. In this case, the exact nature of the shock matters little for optimal pricing. When an aggregate shock occurs, firms mistakenly attribute it to a firm-specific shock, but adjust prices nevertheless. This increases the responsiveness to aggregate nominal shocks and reduces monetary non-neutrality. In the same spirit, Venkateswaran (2014) introduces confusion between idiosyncratic and aggregate productivity into a frictional labor market and argues that the increased responsiveness to aggregate shocks explains the large volatility in empirically observed labor market outcomes.

3.2 Using signals in strategic settings

We now consider coordination games. In Section 3.1, agents minimized the distance of actions from an unknown, exogenous state. In contrast, agents now also consider the distance between their action and the average action in the economy—which is an endogenous outcome. We introduce coordination motives through the target action a^*_{it}. It is a linear combination of an exogenous stochastic state θ_t and the average action in the economy a_t:

$$(21) \qquad a^*_{it} = (1-r)\theta_t + r a_t, \quad \text{where} \quad a_t \equiv \int_0^1 a_{it}\,\mathrm{d}i, \quad \text{and} \quad r \in [-1, 1].$$

The parameter r governs the type of strategic interaction. If $r = 0$ the optimal action is independent of the actions of others, as in the models from Section 3.1.2. When $r \neq 0$, actions become strategic. If $r > 0$ there is strategic complementarity, as the optimal action is increasing in the actions of others. If $r < 0$ there is strategic substitutability, as the optimal action is decreasing in the actions of others.

Next, we examine how coordination motives in actions generate coordination motives in the use of information. We make this point in a passive learning model where information is exogenous. Section 4.6 revisits these models in the context of active learning models.

3.2.1 A beauty-contest with exogenous signals

We simplify the tracking problem in (15) to a static model. Each agent chooses their action a_i to minimize the expected distance to the common target $a^* = (1-r)\theta + ra$, where $a \equiv \int_0^1 a_i \,\mathrm{d}i$ is the average action and θ is an unknown, exogenous state. Each agent solves the following problem:

$$\mathcal{L} \;=\; \min_{a_i} \;\; \mathbb{E}\left[(a_i - (1-r)\theta - ra)^2 \middle| \mathcal{I}_i\right] \tag{22}$$

The order of events is as follows. Nature draws the state θ from a normal distribution $\mathcal{N}(\mu_\theta, \tau_\theta^{-1})$ with mean μ_θ and precision τ_θ. These parameters are common knowledge and summarize all prior public information. Second, each agent receives a public signal z and private signal s that reveal additional information about the state: $z = \theta + \eta_z$ with $\eta_z \sim \mathcal{N}(0, \tau_z^{-1})$ and $s_i = \theta + \eta_{s,i}$ with $\eta_{s,i} \sim_{iid} \mathcal{N}(0, \tau_s^{-1})$, independent of s_j and z. Signals' precisions τ_z and τ_s are equal across agents. Finally, given their information set $\mathcal{I}_i = \{z, s_i\}$, each agent forms beliefs about θ and a, chooses an action a_i, and payoffs are realized. We look for a symmetric Nash equilibrium to solve the game.

Beliefs and equilibrium. According to the first order condition, the optimal action is a convex combination of the belief about the state and the belief about the average action

$$a_i \;=\; (1-r)\mathbb{E}[\theta|\mathcal{I}_i] \;+\; r\mathbb{E}[a|\mathcal{I}_i]. \tag{23}$$

Averaging across agents, we get the average action as a function of the average beliefs

$$a \;=\; (1-r)\overline{\mathbb{E}}[\theta] \;+\; r\overline{\mathbb{E}}[a], \quad \text{where} \quad \overline{\mathbb{E}}[\cdot] = \int_i \mathbb{E}[\cdot|\mathcal{I}_i]\,\mathrm{d}i. \tag{24}$$

The aggregate action a can be described as an infinite sum of higher-order expectations. To see this, recursively substitute for a on the right side of (24) to get $a = \sum_{k=1}^{\infty}(1-r)r^{(k-1)}\overline{\mathbb{E}}^k[\theta]$, where the superscript k represents the k^{th}-order average expectation. For example, $\overline{\mathbb{E}}^1[\theta] = \overline{\mathbb{E}}[\theta]$ is the average belief about θ, while $\overline{\mathbb{E}}^2[\theta] = \overline{\mathbb{E}}[\overline{\mathbb{E}}^1[\theta]]$ is the average belief about the average belief of θ, and so forth. Working with this infinite sum is complex. To avoid this, we follow Morris and Shin (2002) by conjecturing and then verifying a symmetric strategy.

Before we continue, we solve the full information problem where the realization of θ is known. The optimal action is $a_i = (1-r)\theta + ra$. Integrating across agents yields $a = (1-r)\theta + ra$, or simply $a = \theta$, which implies that $a_i = \theta$ for all i. This is the unique Nash equilibrium.

Heterogenous incomplete information. To compute the optimal action in (23) requires forming beliefs about the state θ and the average action a. To form beliefs about θ we use Bayes' law:

$$\mathbb{E}[\theta|\mathcal{I}_i] = \frac{\tau_\theta \mu_\theta + \tau_z z + \tau_s s_i}{\tau_\theta + \tau_z + \tau_s}, \qquad \mathbb{V}ar[\theta|\mathcal{I}_i]^{-1} = \tau_\theta + \tau_z + \tau_s. \tag{25}$$

The best estimate is a convex combination of the prior mean μ_θ, the public signal z, and the individual signal s_i, with weights equal to their relative precisions. The posterior precision equals the sum of the precisions. Since precision is equal across agents, we denote it by $\Sigma^{-1} \equiv \mathbb{V}ar[\theta|\mathcal{I}_i]^{-1}$. Defining relative precisions as $\alpha_\theta = \Sigma\tau_\theta$, $\alpha_s = \Sigma\tau_s$, $\alpha_z = \Sigma\tau_z$, where $\alpha_\theta + \alpha_z + \alpha_s = 1$, we write the expected state as: $\mathbb{E}[\theta|\mathcal{I}_i] = \alpha_\theta\mu_\theta + \alpha_z z + \alpha_s s_i$. To form beliefs about the average action $\mathbb{E}[a|\mathcal{I}_i]$, we guess and verify that the individual action is linear in the signals $a_i = \mu_\theta + \gamma_z(z - \mu_\theta) + \gamma_s(s_i - \mu_\theta)$, where the coefficients γ_z and γ_s are to be determined. Integrating the guess across agents, using the fact that the mean of the private signals s_i equals the true state θ, and taking expectations:

$$\mathbb{E}[a|\mathcal{I}_i] = \mu_\theta + \gamma_z(z - \mu_\theta) + \gamma_s(\mathbb{E}[\theta|\mathcal{I}_i] - \mu_\theta) \tag{26}$$

Substituting the beliefs about the state (25) and the average action (26) into the first order condition (23), rearranging terms, and matching coefficients, we obtain the optimal weights on the public and the private signals

$$\gamma_z = \frac{\alpha_z}{1 - \alpha_s r}, \qquad \gamma_s = \frac{\alpha_s(1 - r)}{1 - \alpha_s r}. \tag{27}$$

We verify the conjecture that the action is linear in signals by checking that the weights on the prior μ_θ is $\gamma_\theta = 1 - \gamma_z - \gamma_s = \frac{\alpha_\theta}{1-\alpha_s r}$. Finally, substituting the optimal action and the equilibrium average action into the loss function (22), we obtain

$$\mathcal{L} = (1 - r)^2 \left(\frac{\gamma_\theta^2}{\tau_\theta} + \frac{\gamma_z^2}{\tau_z} \right) + r^2 \frac{\gamma_s^2}{\tau_s}. \tag{28}$$

The expected loss decreases in the precision of both signals. Without additional externalities in payoffs, more information is always welfare improving.[5]

Optimal use of information. There are two key features of the solution. First, due to Bayesian updating, optimal actions a_i weight signals $\{z, s\}$ according to their precision. If signals are too noisy relative to the prior, the weight on the prior γ_θ dominates. If the public signal is very noisy relatively to the private signal $\alpha_z < \alpha_s$, then the weight on the public signal is smaller $\gamma_z < \gamma_s$ and actions will not move much with z. As private heterogeneous signals become more important,

[5] With information externalities, more information can be welfare reducing (Morris and Shin, 2002; Angeletos and Pavan, 2007).

dispersion in actions increases.[6] The opposite happens if the public signal is relatively more precise than the private signal. Second, the weight agents put on the public signal when forming their action γ_z is increasing in the value of coordination r. Agents who want to do what others do make their actions more sensitive to information that others know. Whenever there is strategic complementarity in actions ($r > 0$), we have $\gamma_z > \alpha_z$, meaning that agents' actions react more to changes in public information than their beliefs do. Conversely, when there is substitutability in actions ($r < 0$), agents weight private signals more in their actions than in their beliefs.

Responsiveness to shocks. To describe the effects of information and coordination motives on aggregate outcomes, we define the *responsiveness to shocks* as the covariance of the average action with the state, normalized by fundamental volatility:

$$\frac{\mathbb{C}ov[a,\theta]}{\mathbb{V}ar[\theta]} \;=\; \gamma_z + \gamma_s \;=\; \frac{\alpha_z + \alpha_s(1-r)}{(1-\alpha_s r)}. \tag{29}$$

It is equal to the sum of the weights on signals and thus depends on relative precisions. Additionally, this covariance is a measure of informativeness of actions and could in principle be used to measure how much information the average agent has.

In the extreme case with perfect information, all agents set their action equal to the known state $a = \theta$ (there is no cross-sectional dispersion); responsiveness is highest at a value of 1 and expected welfare losses are zero $\mathcal{L} = 0$. In the other extreme with complete ignorance (both private and public signals have zero precision), actions equal the prior $a = \mu_\theta$ and do not correlate with the state (but are equal to each other). Responsiveness is lowest at a value of 0. Even if everyone takes the same action, utility losses arise because actions are far away from the state.

Between these two informational extremes, the strength of strategic motives r matters for the optimal use of information and the implied responsiveness to shocks. In particular, the responsiveness measure in (29) decreases with r. We use this fact in the discussion that follows.

3.2.2 Strategic complementarity and aggregate inertia

With strategic complementarity ($r > 0$), agents want to do what others do, so they make their actions more sensitive to the information that others know. To achieve this goal, agents actions strongly comove with public information. For $r > 0$, the optimal weighting in (27) sets $\gamma_z > \alpha_z$ so that actions react more to public signals z than beliefs. Moreover, γ_z increases with the value of coordination r. With extreme complementarity ($r = 1$), all agents take the same action as welfare only depends on the closeness to others; agents completely ignore their private information ($\gamma_s = 0$) as it would only bring their choices apart. The dependence on public information generates

[6] Drenik and Perez (2020) exploit a historical episode—the manipulation of inflation statistics in Argentina—to show that a reduction in public signal precision τ_z increased the weight γ_s on private signals when forecasting inflation and generated larger cross-sectional price dispersion.

aggregate inertia or delays in the adjustment of aggregate variables to shocks. Even if agents' private information tells them to adjust to changing economic conditions, they wait for others to do so. Thus responsiveness to shocks is low.

Strategic complementarity arises naturally in Bertrand (price) competition, as firms have incentives to coordinate price-setting. They set a higher price if the competitor's price is higher and vice versa. Woodford (2003) introduces price complementary in the Lucas (1972) islands' model discussed in the previous section. Optimal prices not only depend upon the state of nominal demand θ_t but also on the average level of prices charged by others a_t. Other settings that feature strategic complementarity include increasing returns to aggregate investment, technology spillovers, or speculative attacks.

3.2.3 Strategic substitutability and aggregate volatility

With strategic substitutability ($r < 0$), agents want to do the opposite of what others do. They weight private signals more in their actions than in their beliefs $\gamma_s > \alpha_s$ and their actions strongly move with private information. Information substitutability generates overreaction in the adjustment of aggregate variables to shocks, or *aggregate volatility*. That is, responsiveness is high.

In which environments is it natural to observe strategic substitutability? Market clearing is one mechanism that generates strategic substitutability through the equilibrium movement of prices. For instance, when firms compete by choosing quantities through Cournot competition, if a firm increases its production, its good becomes more abundant and its price goes down. Therefore, others wants to produce less when one firm produces more. Similarly, consumers want to buy goods that others do not want to buy because the goods others demand will be more expensive. The same logic applies to financial investment, as investors want to buy assets with low demand, low price and high return. Hiring decisions in frictional labor markets also feature strategic substitutability. The greater the aggregate number of vacancies posted in the economy, the lower is the incentive of an individual firm to post vacancies. This mechanism is examined in Venkateswaran (2014) to explain labor market volatility. Lastly, models with returns to specialization are also situations where agents want to behave differently from other agents.

The role of preferences. The simple games in this section considered either complementarity or substitutability motives. When both motives are present, the relationship between information and actions is more nuanced. Baley, Veldkamp and Waugh (2020) make this point in a general equilibrium international trade model where domestic firms choose how much to export based on their beliefs about foreign exports. Market clearing through the terms of trade introduces substitutability. Preference for a balanced consumption bundle introduces complementarity. The effect of information in export decisions depends on the relative strength of these two forces, which are encoded in agents' preferences.

4 Information choice and learning technologies

Active learning means that agents make choices to influence their future information sets. While in the passive learning models in Section 3, signals were taken as given, now signals depend on choices. We present the two most commonly used learning technologies: sticky information, where there is infrequent acquisition of perfect information, and rational inattention, where there is frequent acquisition of noisy information. Finally, we discuss returns to scale in information acquisition and learning specialization.

4.1 Sticky information

Sticky information, also known as *inattentiveness*, is a learning technology in which most of the time, agents get no information flow; but occasionally, they observe the entire history of events. It is a lumpy informational flow, with periods of inaction followed by bursts of information processing. In settings where an agent has to exert some effort to observe information, but that information is not difficult to process, this technology makes sense. Examples include checking one's bank balance, looking up a sports score, or checking the current temperature. Dynamic models with information choice are notoriously hard to solve. Inattentiveness simplifies these problems by making the history of learning choices irrelevant, each time an agent decides to learn.

4.1.1 A beauty contest with infrequent information updating

The following model introduces infrequent information updating to the tracking problem in (15). There is a continuum of agents $i \in [0, 1]$. Each agent chooses her action a_{it} to minimize its distance from an unknown stochastic target a^*_t that an agent with full information would set:

$$\mathcal{L} = \min_{\{a_{it},\, U_{it}\}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left((a_{it} - a^*_t)^2 + U_{it}\kappa_{it} \right) \Big| \mathcal{I}_{it} \right], \tag{30}$$

where $U_{it} \in \{0, 1\}$ is its decision to update the information set $\mathcal{I}_{it}$ at a cost $\kappa_{it} > 0$. The target is a convex combination of an exogenous unobserved state θ_t and the average action a_t:

$$a^*_t = (1 - r)\theta_t + r a_t, \quad \text{with} \quad a_t = \int_i a_{it}\, \mathrm{d}i. \tag{31}$$

The state follows a random walk $\theta_t = \theta_{t-1} + \varepsilon_t$ with *iid* innovations $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. The following information cost, observed at the beginning of period t, nests two specifications in the literature:

$$\kappa_{it} = \begin{cases} \kappa & \text{with prob.} \quad 1 - \lambda \\ 0 & \text{with prob.} \quad \lambda \end{cases} \qquad \text{with} \quad \lambda \in [0, 1]. \tag{32}$$

Setting $\kappa = \infty$ and $\lambda > 0$, this cost structure generates the (passive learning) sticky information model in Mankiw and Reis (2002), where information arrives freely at an exogenous constant rate. Setting $\kappa > 0$ and $\lambda = 0$, this cost structure embodies active learning. It generates the costly information updating in Reis (2006a,b), where agents face a fixed observation cost.

Information dynamics. An agent who last updated in period $\hat{\tau}$ enters period t with an information set that contains state realizations from every period up to and including $\hat{\tau}$: $\mathcal{I}_{i(t-1)} = \mathcal{I}_{\hat{\tau}} = \{\theta_\tau\}_{\tau=0}^{\hat{\tau}}$. If an agent chooses to update in the current period ($U_{it} = 1$), her new information set will contain all state realizations up to and including the current state: $\mathcal{I}_{it} = \mathcal{I}_t = \{\theta_\tau\}_{\tau=0}^{t}$. If the agent does not update in the current period ($U_{it} = 0$), she will not observe any new information, not even endogenous information like the average action $\left(\mathcal{I}_{it} = \mathcal{I}_{i(t-1)} \right)$. Individual information sets $\mathcal{I}_{it}$ evolve according to:

$$\mathcal{I}_{it} = \begin{cases} \mathcal{I}_{\hat{\tau}} = \{\theta_\tau\}_{\tau=0}^{\hat{\tau}} & \text{if} \quad U_{it} = 0 \\ \mathcal{I}_t = \{\theta_\tau\}_{\tau=0}^{t} & \text{if} \quad U_{it} = 1. \end{cases} \tag{33}$$

Equilibrium and optimal choices. An equilibrium is a sequence of information choices by every agent $\{U_{it}\}$, and actions $\{a_{it}\}$, that are $\mathcal{I}_{it}$–measurable and maximize (30), taking as given the choices of all other agents. The first-order condition dictates that agent i who last updated at date $\hat{\tau}$, sets her action equal to its expected target at time t:

$$a_{it} \;=\; \mathbb{E}\left[a_t^* | \mathcal{I}_{\hat{\tau}}\right] \;=\; (1-r)\mathbb{E}\left[\theta_t | \mathcal{I}_{\hat{\tau}}\right] + r\mathbb{E}\left[a_t | \mathcal{I}_{\hat{\tau}}\right]. \tag{34}$$

Since the state is a random walk, $\mathbb{E}\left[\theta_t | \mathcal{I}_{\hat{\tau}}\right] = \theta_{\hat{\tau}}$. We guess and verify that the average action is also a random walk, $\mathbb{E}\left[a_t | \mathcal{I}_{\hat{\tau}}\right] = a_{\hat{\tau}}$. Let $\lambda_{t,\hat{\tau}}$ denote the measure of agents who last updated in period $\hat{\tau} \leq t$. Then the average action a_t is a weighted sum of the expected target of all agents: $a_t \;=\; \sum_{\tau=0}^{t} \lambda_{t,\tau} \mathbb{E}\left[a_t^* | \mathcal{I}_\tau\right] = \sum_{\tau=0}^{t} \lambda_{t,\tau} \left((1-r)\theta_\tau + r a_\tau\right)$. Recursively substituting in for a_τ reveals that the average action is a weighted sum of all past innovations:

$$a_t \;=\; \sum_{\tau=0}^{t} \frac{\Lambda_{t,\tau}(1-r)}{1 - r\Lambda_{t,\tau}} \varepsilon_\tau, \tag{35}$$

where $\Lambda_{t,\hat{\tau}} \;\equiv\; \sum_{\tau=\hat{\tau}}^{t} \lambda_{t,\tau}$ denotes the measure of agents who have last updated between dates $\hat{\tau}$ and t. Substituting (35) into (31) tells us that the target action is $a_t^* = \sum_{\tau=0}^{t} \frac{1-r}{1-r\Lambda_{t,\tau}} \varepsilon_\tau$. Agents who last updated at date $\hat{\tau}$ set their action to $a_{it} = \mathbb{E}\left[a_t^* | \mathcal{I}_{\hat{\tau}}\right] = \sum_{\tau=0}^{\hat{\tau}} \frac{1-r}{1-r\Lambda_{t,\tau}} \varepsilon_\tau$. Their expected one-period loss (which agents compare to the information processing cost) depends on all the innovations since the last update:

$$\mathcal{L}_{t,\hat{\tau}} \;\equiv\; \mathbb{E}\left[\left(\mathbb{E}\left[a_t^* | \mathcal{I}_{\hat{\tau}}\right] - a_t^* | \mathcal{I}_{\hat{\tau}}\right)^2\right] \;=\; \sum_{\tau \;=\; \hat{\tau}+1}^{t} \left(\frac{1-r}{1-r\Lambda_{t,\tau}}\right)^2 \sigma^2. \tag{36}$$

The longer it has been since an agent updated her information, the higher are the incentives to update in the current period. If information arrives exogenously, firms have no choice but to update at rate λ. In contrast, when information is actively chosen by paying the fixed cost κ, the updating policy consists of threshold dates such that agents who last updated at date $\hat{\tau} < \tau_t^*$ update at date t, while those who last updated at $\hat{\tau} > \tau_t^*$ find not updating strictly optimal. One equilibrium consists of staggered updating, meaning that all firms update after a fixed number of periods T and each period a fraction $1/T$ of the firms updates (Reis, 2006b).

Expression (36) highlights the updating complementarity. For any $\tau = \hat{\tau} + 1, ..., t$, the one-period loss increases with the strategic motive parameter, that is, $\partial \mathcal{L}_{t,\hat{\tau}} / \partial \Lambda_{t,t-\tau} > 0$, if and only if $r > 0$. When actions are complements ($r > 0$), there is complementarity in information acquisition: the more agents are aware of a shock that has occurred since the agent last updated, the higher the per-period loss of not being aware of this shock. The complementarity in updating information delays adjusting to changing economic conditions, or *inertia*. When actions are strategic substitutes ($r < 0$), the converse is true. This general principle discussed in the static games of Section 3.2 re-appears in dynamic settings.

4.1.2 Applications of sticky information

Sticky information has been extensively used to explain the observed inflation inertia, as price-setting firms slowly update their information about money supply and demand. In Mankiw and Reis (2002) and Ball, Mankiw and Reis (2005) firms passively update information at random dates. In Reis (2006b) the adjustment dates are actively chosen by paying an observation cost. Álvarez, Lippi and Paciello (2016) generalize this setup allowing for heterogeneity in observation costs.

From the household side, sticky information has been put forward as an explanation for the equity premium puzzle (Gabaix and Laibson, 2001) and the excess sensitivity and excess smoothness puzzles of aggregate consumption (Reis, 2006a; Carroll *et al.*, 2020). Auclert, Rognlie and Straub (2020) show that embedding sticky expectations in a heterogeneous agent new Keynesian model reconciles the micro and macro responses to monetary policy shocks.

Inattentiveness is often mixed with adjustment costs in actions, combining information updating with tools from the s-S literature. Bonomo and Carvalho (2004) and Álvarez, Lippi and Paciello (2011, 2017) study price-setting problems where firms pay an observation cost to discover their target price and pay a menu cost to change their price. Similarly, Álvarez, Guiso and Lippi (2012) and Abel, Eberly and Panageas (2013) study portfolio choice where investors pay an observation cost to reveal the value of a risky asset and pay a transaction cost to adjust their portfolio. In these papers, small information and adjustment costs generate infrequent adjustment, yielding long periods of inertia.

On the empirical side, Klenow and Willis (2007) test inattentiveness models of price-setting by asking whether information revealed in past periods acts as a shock to prices in the current period. In a similar exercise with asset prices, Hong, Torous and Valkanov (2007) and Cohen and Frazzini (2008) find that industry information affects the market index value with a lag. Andrade and Le Bihan (2013) document that professional forecasters fail to systematically update their forecasts and disagree when updating, all of which suggests inattention.

4.2 Rational inattention

The idea that economic agents have limited ability to process information, or to pay attention, is often referred to as *rational inattention.* Following Sims (2003), rational inattention has taken on a more specific meaning. Models that use rational inattention either bound the amount of information or charge agents a utility cost for information, where the amount of information is measured according to how much it reduces entropy.

A large subset of the literature simplifies the problem by allowing agents to directly choose the precision with which they observe an exogenously specified set of Normal signals. It turns out that with a quadratic payoffs and Normal priors, Normal signals are optimal. We lay out such a quadratic-normal model, in order to convey the main ideas from this literature.

4.2.1 Measuring information: entropy and mutual information

The standard measure of the quantity of information in information theory is Shannon *entropy* (Cover and Thomas, 1991). Entropy measures the amount of uncertainty in a random variable. For a random variable θ with density function f, entropy is defined as:[7]

$$\mathcal{E}(\theta) \equiv -\mathbb{E}[\ln(f(\theta))]. \tag{37}$$

Sims (2003) proposed modelling the informational content of a signal s about θ as the reduction in entropy achieved by conditioning on the additional information provided by the signal. This measure of uncertainty reduction is known as *mutual information.* It is defined as:

$$I(\theta, s) \equiv \mathcal{E}(\theta) - \mathcal{E}(\theta|s), \tag{38}$$

where the second term is conditional entropy: $\mathcal{E}(\theta|s) = \mathcal{E}(\theta, s) - \mathcal{E}(s)$. The expectation in $\mathcal{E}(\theta, s)$ is taken over the realizations of (θ, s). With a Normal state $\theta \sim \mathcal{N}(\mu_\theta, \tau_\theta^{-1})$ and a Normal signal $s \sim \mathcal{N}(0, \tau_s^{-1})$, mutual information takes a simple form:

$$I(\theta, s) = \frac{1}{2} \ln\left(1 + \frac{\tau_s}{\tau_\theta}\right). \tag{39}$$

[7] By using the natural logarithm, we express information units in *nats*, as opposed to *bits*, in which case, the logarithm has base 2.

Mutual information reflects the ratio of the posterior precision to the prior precision $(\tau_\theta + \tau_s)/\tau_\theta$. Mutual information increases with signal precision, as it generates a larger reduction in uncertainty.

4.2.2 A tracking problem with noisy information acquisition

Consider a repeated tracking problem. The agent chooses the action a that minimizes the expected distance to an *i.i.d.* state $\theta \sim \mathcal{N}(0, \tau_\theta^{-1})$. She receives a noisy signal $s = \theta + \eta$, with precision τ_s. There are two constraints governing how the agent can choose signal precision. First is the capacity constraint. It takes the form of an upper bound $\kappa > 0$ on the mutual information of priors and prior plus signals. Second is a "no forgetting" constraint that requires mutual information to be non-negative. The agent can increase capacity κ by paying a proportional utility cost $c\kappa$. The agent solves the following problem:

$$\begin{aligned} \mathcal{L} \; = \; \min_{\{a,\kappa\}} \quad & \frac{1}{2}\,\mathbb{E}\left[(a-\theta)^2\,|s\right] + c\kappa \\ s.t. \quad & 0 \;\leq\; I(\theta, s) \;\leq\; \kappa, \end{aligned} \tag{40}$$

where $I(\theta, s)$ is the mutual information for Normal state and signals in (39). The solution to the problem takes place in two stages. In the first stage, the agent chooses how much attention to allocate to θ by choosing the total processing capacity κ. This choice determines the optimal signal precision τ_s. In the second stage, the agent receives a noisy signal s with the precision proportional to the attention allocated in the first stage and chooses the action.

To solve the model, we work backwards. Suppose the agent receives signal s with precision τ_s. Conditional on this signal, the agent chooses the optimal action $a^* = \mathbb{E}[\theta|s] = \frac{\tau_s}{\tau_\theta+\tau_s}s$. The expected loss implied by the optimal action is $\mathbb{E}\left[(a^*-\theta)^2\right] = (\tau_\theta + \tau_s)^{-1}$. Since the expected loss decreases with signal precision, the capacity constraint will always bind and we set $\tau_\theta + \tau_s^* = \tau_\theta e^{2\kappa}$. Plugging the optimal action and binding capacity constraint into (40), and taking the first order condition with respect to κ, we find the optimal attention capacity and the implied optimal signal precision. And conditional on these choices, we find the optimal action:

$$\kappa^* \;=\; \frac{1}{2}\ln\left(\frac{1}{\tau_\theta c}\right); \qquad (\tau_s^* \;+\; \tau_\theta)^{-1} \;=\; c; \qquad a^* \;=\; \max\left\{(1-\tau_\theta c)\,s, 0\right\}. \tag{41}$$

This example illustrates key tradeoffs of rational inattention models. The agent chooses signal precision by trading the costs of acquiring information with the benefits from better information. The agent increases attention κ^* if the state's volatility τ_θ^{-1} is high and if the marginal cost of acquiring information c is low. If the marginal cost is too large, the agent acquires no information and sets the action equal to the prior mean (zero in this example).

Tracking multiple states. Assume the agent tracks two $i.i.d.$ states $(\theta_1, \theta_2) \sim \mathcal{N}(0, \Sigma)$, where Σ denotes the prior variance-covariance matrix. The agent chooses an action a to minimize the distance to both states subject to a bound on mutual information:

$$\mathcal{L} \;=\; \min_{\{a,\kappa\}} \; \frac{1}{2}\mathbb{E}\left[(a-\theta_1-\theta_2)^2\right] \;+\; c\kappa \tag{42}$$

$$s.t. \quad 0 \;\leq\; \frac{1}{2}\ln\left(\frac{\det \Sigma}{\det \hat{\Sigma}}\right) \;\leq\; \kappa, \tag{43}$$

where $\hat{\Sigma}$ denotes the posterior covariance matrix. Mutual information with multivariate normal variables reflects the ratio of the determinants of the prior and the posterior variances.

The problem of tracking two states has a simple solution when the agent is allowed to choose the variance-covariance structure of the signals. To see this, define the target $\theta^* = \theta_1 + \theta_2$, and assume the agent receives a single noisy signal of this target. The problem can be restated as a single-state problem, with the optimal allocation of attention taking the same form as (41). This one signal allows the agent to achieve the same expected loss as two independent signals, while requiring lower mutual information.

Rationally inattentive agents generally prefer signals about a linear combination of the payoff-relevant states. However, restricting the set of signals to be independent and associated to a specific state is a plausible economic constraint, in many settings. Following Mackowiak and Wiederholt (2009), suppose the states are independent from each other and the agent receives two independent signals, $s_1 = \theta_1 + \eta_1$, and $s_2 = \theta_2 + \eta_2$, with respective precisions τ_{s1}, τ_{s2}. The prior Σ and posterior $\hat{\Sigma}$ variances are diagonal matrices and the entropy constraint simplifies to:

$$\frac{\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{22}^{-1}}{\tau_{\theta 1}\tau_{\theta 2}} \leq e^{2\kappa}, \qquad \text{where} \quad \hat{\Sigma}_{ii}^{-1} = \tau_{\theta_i} + \tau_{si} \quad \text{for} \quad i = 1, 2. \tag{44}$$

Expected loss associated with the action are $\mathbb{E}\left[(a^* - \theta)^2\right] = \hat{\Sigma}_{11} + \hat{\Sigma}_{22}$. If $c \neq 0$, total capacity κ can be chosen and the problem reduces to two independent single-state problems like (40). The optimal allocation of attention for each state is given by (41), replacing by the corresponding precision of each state. In the dual problem where total capacity is fixed, the capacity constraint always binds. In this case, more attention to one state reduces the attention allocated to the other state. The ratios of posterior to prior precisions are increasing in the other state's precision:

$$\frac{\tau_{s1}^* + \tau_{\theta 1}}{\tau_{\theta_1}} \;=\; e^{\kappa}\sqrt{\frac{\tau_{\theta 2}}{\tau_{\theta 1}}}; \qquad \frac{\tau_{s2}^* + \tau_{\theta 2}}{\tau_{\theta_2}} \;=\; e^{\kappa}\sqrt{\frac{\tau_{\theta 1}}{\tau_{\theta 2}}}. \tag{45}$$

These expressions highlight a key lesson from multivariate rational inattention models: Agents optimally pay more attention to the more volatile state, as it generates larger welfare losses.

4.2.3 Applications of rational inattention

Following Sims (2003), the literature on rational inattention has rapidly expanded. One of the earliest applications is the price-setting model by Mackowiak and Wiederholt (2009). Its core resembles the two-shock model in (42), where θ_1 represents monetary shocks and θ_2 idiosyncratic productivity shocks. Rational inattentive firms optimally pay more attention to idiosyncratic shocks as these are relatively more volatile than aggregate shocks, and thus under-react to monetary policy shocks. In this setting, Paciello and Wiederholt (2014) study optimal monetary policy and Mackowiak and Wiederholt (2015) introduce rationally inattentive households. As with inattentiveness (Section 4.1), rational inattention gives rise to nominal rigidities and price inertia.

Unlike the simple static problems we presented, many applications feature dynamic settings where states evolve persistently over time. This is technically challenging. Mackowiak, Matějka and Wiederholt (2018), Miao, Wu and Young (2019), and Afrouzi and Yang (2021a) study dynamic inattention problems and propose algorithms to solve them. Another strand of the literature provides axiomatic foundations (Caplin and Dean, 2015; de Oliveira, Denti, Mihm and Ozbek, 2017; Ellis, 2018; Hébert and Woodford, 2019) and generalizations of entropy costs (Caplin, Dean and Leahy, 2017; Hébert and Woodford, Forthcoming).

Applications of rational inattention include portfolio allocation (Mondria, 2010); mutual fund management (Kacperczyk, Van Nieuwerburgh and Veldkamp, 2016); discrimination of minorities (Barto, Bauer, Chytilov and Matějka, 2016); electoral competition (Matějka and Tabellini, 2016); international trade (Dasgupta and Mondria, 2018); insurance choice (Brown and Jeon, 2020); marriage markets (Cheremukhin, Restrepo-Echavarria and Tutino, 2020); hiring decisions (Acharya and Wee, 2020); migration (Porcher, 2020; Bertoli, Moraga and Guichard, 2020); consumption (Luo, 2008; Kszegi and Matějka, 2020); expectation formation (Fuster, Perez-Truglia, Wiederholt and Zafar, 2020; Gutiérrez-Daza, 2021); and price-setting (Woodford, 2009; Stevens, 2020; Turén, 2020; Yang, 2020; Afrouzi and Yang, 2021b). Mackowiak, Matějka and Wiederholt (2020) provide a comprehensive review of this literature.

4.2.4 Linear cost of signal precision

Mutual information in (38) is one of several informational costs $I(\theta, s)$ employed in the active learning literature. A popular alternative is to specify costs that are linear in signal precision. In the case of two states and two signals, this constraint takes the form

$$\hat{\Sigma}_1^{-1} + \hat{\Sigma}_2^{-1} \leq \kappa. \tag{46}$$

Comparing (46) with (44), we see that entropy constrains the product of precisions, while the linear constraint bounds the sum of precisions.

While the entropy technology represents a process of increasingly refined search, the linear technology models search as a sequence of independent explorations. Van Nieuwerburgh and Veldkamp (2010) and Myatt and Wallace (2012) use linear constraints to jointly study information acquisition and investment decisions in financial markets. Hébert and Woodford (Forthcoming) show that the linear constraint (46) can be obtained by assuming neighborhood-based information costs that capture notions of perceptual distance. Departures of mutual information on this direction have important welfare implications in general equilibrium, as examined by Angeletos and Sastry (2019) and Hébert and La'O (2020).

4.3 Other learning technologies

Agents' learning technology could involve a combination of both sticky information (inattentiveness) and noisy information acquisition (rational inattention). Agents may pay a fixed cost at discrete dates to observe perfect information about the current and past states of the economy. But in between these updates, they may observe a flow of noisy information about the state. For example, Bonomo, Carvalho, Garcia, Malta and Rigato (2021) develop a hybrid price-setting model with features from both learning technologies, and Coibion and Gorodnichenko (2012) consider both learning technologies and show that survey data favor models of noisy information.

A different learning technology developed in Woodford (2009), Stevens (2020) and Khaw, Stevens and Woodford (2017) is a two-stage form of rational inattention: a decision whether to adjust and then how much to adjust. Although their two-stage adjustment is infrequent like sticky information, the decision to adjust is nonetheless based on a continuous flow of information. The experimental data in Khaw, Stevens and Woodford (2017) supports this two-stage technology.

4.4 Information choice as a source of inequality

Information choice may exacerbate initial differences across agents when there are increasing returns in information. Increasing returns to information refers to the idea that an entity with more data values additional data more. Often, the reason is that when a firm or an economy gets more information, it grows, invests more or takes on more risky actions. But agents with larger investment or risky positions value information more. Thus more information raises the value of acquiring more information. In data economics, this same force is called the "data feedback loop."

Increasing returns in information appear naturally in production settings (Wilson, 1975). The more a firm learns to improve its technology, the more it wants to produce and the more it produces, the more valuable information becomes. Firms that initially operate on a large scale are more likely to acquire information, produce with better technology, and grow faster than small firms. In the previous case, there was increasing returns to data acquisition. Increasing returns can also arise in data production. In Begenau, Farboodi and Veldkamp (2018), big data disproportionately benefits large firms. Because they have more economic activity, large firms produce more data. Abundant data improves investors' forecasts, which reduces investors' uncertainty and lowers the large firm's cost of capital. Lower investment costs enable large firms to grow even larger.

Information has increasing returns in portfolio problems because it can be used to evaluate one share of an asset or many shares of an asset. When a decision maker has lots of an asset, information about the asset's payoff is more valuable. Investors that are initially wealthier acquire more information because they have more asset value to apply that information to and they will also earn higher returns on their investments. Poor individuals may stay poor while the rich get richer. This mechanism has been shown to account for inequality in portfolio holdings of risky assets (Peress, 2004) and the increase in capital income inequality due to investor sophistication (Kacperczyk, Nosal and Stevens, 2019) and financial innovation (Mihet, 2021).

4.5 Learning what others know

Agents can acquire information by observing the behavior of others. Agents with information can share that information with others through their actions. Social learning may generate behavior that resembles irrationality, such as herds, bubbles, booms and crashes, but that is completely rational. For a comprehensive discussion of social learning and its consequences we refer to Chamley (2004) and Goyal (2011). Here we describe two recent evolutions of social learning.

Local learning. In certain settings, it is natural to observe the behavior of other agents that are geographically, culturally, socially, or economically "close". Moreover, agents may actively choose who forms part of the network of connections from which they obtain their information from (Herskovic and Ramos, 2020). The local learning that emerges generates similar beliefs and actions within members of the same group but different beliefs and actions across different groups. In the aggregate, local learning may slow down the transmission of information.

The following papers examine local learning mechanisms. In Conley and Udry (2010) farmers learn a new agricultural technology from other farmers in their village; in Fogli and Veldkamp (2011) women learn the effects of maternal employment on children by observing nearby employed women; in Buera, Monge-Naranjo and Primiceri (2011) countries learn the impact of market-oriented policies from the experience of similar countries; in Galenianos (2013) firms learn an applicants suitability for a job (match quality) when the applicant is referred to the firm; in Fernandes and Tang (2014) exporters learn the returns to exporting in foreign markets from neighbouring firms' export performance; in Figueiredo (2018) high-school students learn the college premium from the wages of college-educated workers in their neighborhood; in Boerma and Karabarbounis (2021) households learn the returns to entrepreneurship from their dynasty's entrepreneurial experience.

News media. Agents also learn what others know through news media. News media is a technology for aggregating and sharing information. One service that the news media provides is to select what events to report, playing a big role as a source of *selected* information. In particular, media may strategically choose to report unusual or extreme events to increase the number of its users (Nimark and Pitschner, 2019). By focusing on certain observations, media generates bias even among rational readers. In turn, this bias may increase volatility and can lead to aggregate fluctuations (Nimark, 2014; Chahrour, Nimark and Pitschner, 2019).

4.6 Information choice in strategic settings

In settings with strategic behavior in actions, as in the coordination games of Section 3.2, an agent's choice to acquire information depends on others' information acquisition. Hellwig and Veldkamp (2009) and Hellwig, Kohls and Veldkamp (2012) introduce information choice in the beauty contest game in (22). Agents choose their actions a_i to match an unknown target that depends on the exogenous state θ and the average action in the economy $a = \int_i a_i \, \mathrm{d}i$. But before playing the action game, agents choose how much to pay to acquire a common signal $z \sim \mathcal{N}(\theta, \tau_z^{-1})$ and/or a private signal $s \sim \mathcal{N}(\theta, \tau_s^{-1})$. The cost of information acquisition $\kappa(\tau_z, \tau_s)$ is increasing, convex, and twice differentiable in signal precisions. Each agent solves:

$$\mathcal{L} \;=\; \min_{a_i, \tau_z, \tau_s} \; \mathbb{E}\left[(1-r)(a_i-\theta)^2 + r(a_i-a)^2 \Big| z, s_i\right] \;+\; \kappa(\tau_z, \tau_s), \tag{47}$$

The model is solved by backward induction. Taking agents' information as given, one solves the action game and computes expected utility, as a function of information precision. That expected utility is the objective in the first stage information choice game. With this objective, one can solve for optimal information choices.

The key result is that strategic motives in actions generate strategic motives in information choice. Information changes the economy's *responsiveness to shocks*, defined in (29) as the covariance of the average action with the state, normalized by fundamental volatility: $\mathbb{C}ov[a,\theta]/\mathbb{V}ar[\theta]$. That is what makes information more or less valuable.

When actions exhibit complementarity ($r > 0$) and other agents have precise information (high $\tau_z + \tau_s$), responsiveness is high. When the average action and the state covary, the agent faces more payoff uncertainty because if he chooses an action that turns out to be far away from θ, it will also be far away from a and he will be penalized twice. This added utility risk raises the value of accurate information. Information acquisition is complementary. Correlation in information choice induces further correlation in actions, such as financial investment (Veldkamp, 2006), production (Veldkamp and Wolfers, 2007), and price-setting (Gorodnichenko, 2008).

Conversely, when actions are substitutes ($r < 0$) and other agents have precise information (high $\tau_z + \tau_s$), responsiveness is again high, meaning that if the agent chooses an action that turns out to be far away from θ, it will also be far away from a. But in this case, that covariance reduces payoff uncertainty: Taking an action that is far away from a confers a utility benefit, while being far away from the state θ incurs a utility cost. The cost and benefit partially cancel each other. The risk of being far from the state θ hedges the risk of taking an action that is close to a. This hedging makes the variability of overall utility less. When others know more, the state and average action are more aligned and offset each other more effectively. The offset dampens utility fluctuations more. Less utility risk lowers the value of information. Thus information is a strategic substitute because its value is less when others acquire more of it. Exploring strategic substitutability in information has a long tradition in the portfolio choice literature, starting with Grossman and Stiglitz (1980).

5 Theories of the data economy

Models of the data economy are learning models, like the ones we have examined so far. The key difference with this class of models is the two-way feedback between information and economic activity. Economic activity generates data and the information in the data feeds back to affect economic activity. In one class of models, data is modeled as ideas or knowledge. In another class, data is information that reduces uncertainty and guides decision-making. Both can speak to long-run growth and business cycle fluctuations.

5.1 Experimentation

Data economy models are examples of a broader class of models with active experimentation. Active experimentation means that an agent chooses an action that may generate information. The value of the information is explicitly incorporated into the agent's choice problem. Such problems often produce feedback between economic activity and information. Agents control the information flow (e.g., signal quality) through their actions. These actions, in turn, depend on the information agents learn.

In a class of models called bandit problems, all actions generate equally precise signals, but the decision is whether to act or not. In another class of models with experimentation, the signal precision depends on the agent's action. As a simple example, consider the quadratic tracking problem (16) in which the optimal action is set equal to a belief and thus depends on signal precision. In some cases, precision τ_s increases in the distance from the myopic target action a^*_{it}:

$$\tau_s = \phi(a_{it} - a^*_{it})^2, \qquad \phi > 0. \tag{48}$$

Experimentation is costly as agents deviate from their target to obtain information. At the same time, experimentation brings benefits in the shape of more precise future information. A feedback loop between actions and information often arises. The optimal experimentation strategy solves a fixed point problem that balances the current costs from deviating from the target against the benefits of better decision-making in the future.

Active learning through experimentation in optimal control problems in macroeconomics arises in Prescott (1972). Since Rothschild (1974), experimentation has been widely applied to price-setting models in which a monopolist learns about uncertain demand. Firms use their prices to learn about demand's slope or intercept (Balvers and Cosimano, 1990; Mirman, Samuelson and Urbano, 1993; Keller and Rady, 1999; Willems, 2017). In these models, firms are willing to produce at negative revenue, in order to obtain better information. Bachmann and Moscarini (2011) and Argente and Yeh (2021) build general equilibrium versions of this framework. Other applications of experimentation include investment and growth (Bertocchi and Spagat, 1998), optimal monetary policy (Wieland, 2000; Svensson and Williams, 2007), job mobility (Pastorino, 2009), and occupational choice (Antonovics and Golan, 2012).

5.2 Data and growth

We begin by exploring the connection between data and long-run growth. Consider an economy with a continuum of firms $i \in [0, 1]$. Each firm i produces output y_{it} using labor l_{it} and idiosyncratic productivity A_{it}:

$$y_{it} = A_{it} l_{it}^{\alpha}, \quad \alpha \leq 1. \tag{49}$$

Data D_{it} is generated as a by-product of economic activity. Data is generated through own output, with a "data-savviness" parameter z_i, or is produced by other firms B_{it}:

$$D_{it} = z_i y_{it} + B_{it}. \tag{50}$$

A data-savvy firm harvests lots of data per unit of output. $B_{i,t}$ captures the fact that data is a non-rival good: the information produced by the activity of one firm can be used by others. Two main approaches have been explored to study the impact of data on productivity A_{it}: data as knowledge and data as information. These approaches differ starkly in their implications for long-run growth.

5.2.1 Data as knowledge

The first approach considers data as knowledge. We present a simplified version of the model by Jones and Tonetti (2020). Data D_{it} improves the quality of ideas, directly increasing firm productivity A_{it}. Data relevance for productivity is mediated by the parameter η.

$$A_{it} = D_{it}^{\eta}. \tag{51}$$

Data from other firms B_{it}, produced with their output, may increase firm i's data at the *non-rivalry* rate $\tilde{z}_i$.

$$B_{it} = \int \tilde{z}_i y_{it} \, \mathrm{d}i. \tag{52}$$

If $\tilde{z}_i = 0$, then data is rival and data from other firms cannot be used by firm i. Higher values of $\tilde{z}_i$ indicate that firm i obtains more data generated by the production of the rest of the firms. Substituting B_{it} into (50), assuming symmetry ($z_i = \tilde{z}_i = z$ and $y_{it} = y_t$) and using the fact that firm i has measure zero yields $D_{it} = zy_t$. Then substituting A_{it} and D_{it} into (49) and rearranging we obtain:

$$y_t = z^{\frac{\eta}{1-\eta}} \, l_t^{\frac{\alpha}{1-\eta}}. \tag{53}$$

For $\eta > 1-\alpha$, data production leads to increasing returns and long-run growth. This approach of equating data with knowledge also appears in Cong, Xie and Zhang (2020), where data is used for R&D in an endogenous growth model, and by Abis and Veldkamp (2021), who study the impact of artificial intelligence (AI) in the investment management industry.

5.2.2 Data as information

The second approach considers data as information that reduces uncertainty and guides decision-making. Farboodi and Veldkamp (2021) consider data as information that is used in forecasting an optimal production technique. Firms face a signal extraction problem as in section (3.1). They choose a production technique a_{it} to match the optimal technique $a_t^* = \theta_t + \varepsilon_t$ that consists of persistent θ_t and transitory ε_t components. Better forecasts of the optimal technique increase productivity A_{it}:

$$A_{it} = \bar{A} - (a_{it} - \theta_t - \varepsilon_t)^2. \tag{54}$$

Firms receive a noisy signal s_{it} about θ_t. Its precision increases with data D_{it}, capturing data-driven improvements in forecasting:

$$s_{it} = \theta_t + \eta_{it}; \qquad \eta_{it} \sim \mathcal{N}(0, (\tau_s D_{it})^{-1}). \tag{55}$$

As in (50), data D_{it} comes from own production and data shared by others. This structure generates a *data-feedback loop*. Large firms produce more data, which allows them to improve the estimates of the optimal technique, increasing output and future data. Nevertheless, since reducing a forecast error has a bounded value, data producing cannot lead to long-run growth (the highest productivity, obtained with zero uncertainty, is $\bar{A}$).

Farboodi, Mihet, Philippon and Veldkamp (2019) apply a similar information structure to show that data production induces larger firms in steady state. Besides data arising from own production, data can be purchased in the market B_{it}. Differences in data savviness z_i can induce some firms to specialize in data production and grow faster than the rest.

5.3 Data and economic fluctuations

When analyzing business cycles, data informs about the current state of the economy, which is usually aggregate productivity. The feedback loop between data and economic activity amplifies or helps propagate the business cycle. Booms are times of high activity and information production.

To examine how data propagates business cycles, we make two assumptions. First, to produce, firms must pay a random idiosyncratic cost v_{it} which is *i.i.d.* across firms and time. Second, θ_t is an aggregate productivity shock identical across firms. It follows a two-state Markov process between a good state θ_g and a bad state θ_b, with $\theta_g > \theta_b$.

$$y_{it} = \theta_t l_{it}^{\alpha} - v_{it}; \qquad v_{it} \sim_{iid} \mathcal{N}(0, \sigma_v^2); \qquad \theta_t \in \{\theta_g, \theta_b\}. \tag{56}$$

Firms observe v_{it} but ignore θ_t. They receive a public signal s_t with a precision that increases with the number of active firms n_t in the economy:

$$s_t = \theta_t + \varepsilon_t; \qquad \varepsilon_t \sim \mathcal{N}(0, (n_t \tau_s)^{-1}). \tag{57}$$

A firm chooses to produce if its belief about productivity is high relative to the fixed production cost. When lots of firms believe productivity is at the good state θ_g, there is high economic activity (large n_t) and information production (high signal precision). The opposite happens if firms believe productivity is low.

This structure generates an asymmetric flow of information over the cycle. When the economy is at the good state, signal precision is very high. According to Bayesian updating, firms put a high weight to unexpected news when updating their beliefs. Therefore, at the peak of the business cycle, an exogenous change from good times θ_g to bad times θ_b triggers a rapid adjustment in firms' beliefs and leads to an abrupt downward adjustment in production. In contrast, when times are bad, scarce information and high uncertainty slow down belief updating. If the economy goes back to the good state, output will transition slowly.

This mechanism was proposed by Veldkamp (2005) to explain why many asset markets exhibit slow booms and sudden crashes and by Van Nieuwerburgh and Veldkamp (2006) to understand business cycles asymmetries, with slow expansions and sudden recessions. This mechanism is tested empirically by Ordonez (2013). Quantitative versions are developed by Saijo (2017) and Fajgelbaum, Schaal and Taschereau-Dumouchel (2017) where the level of aggregate investment determines the amount of information available to firms. In Straub and Ulbricht (2018), the ability of investors to learn about firm-level fundamentals declines during financial crises, generating negative spillovers from financial distress onto the real economy. In all of these papers, there is a two-way interaction between the level of economic activity and aggregate uncertainty.

6 Conclusion

As the economy transforms itself from a physical production economy to a knowledge economy, understanding learning becomes more central to economics. Learning is the process whereby information is transformed into knowledge. While we have described models in terms that suggest human beings are doing the learning, it may also be that, in the future, machines do some of this learning for us. That does not make these problems less relevant. Machines will also work to solve signal extraction problems and algorithms will need to choose what data to process. The magnitude of the constraints may be quite different. But as data grows in abundance and value, understanding what signal extraction can reveal, the costs and benefits of this knowledge and how it affects aggregate economic activity has never been a more urgent endeavor.

References

Abel, A. B., Eberly, J. C. and Panageas, S. (2013). Optimal inattention to the stock market with information costs and transactions costs. *Econometrica*, **81** (4), 1455–1481.

Abis, S. and Veldkamp, L. (2021). The changing economics of knowledge production. *Available at SSRN 3570130*.

Acharya, S. and Wee, S. L. (2020). Rational inattention in hiring decisions. *American Economic Journal: Macroeconomics*, **12** (1), 1–40.

Afrouzi, H. and Yang, C. (2021a). *Dynamic Rational Inattention and the Phillips Curve*. Tech. rep., CESifo.

— and — (2021b). *Selection in Information Acquisition and Monetary Non-Neutrality*. Tech. rep., Columbia University.

Álvarez, F., Guiso, L. and Lippi, F. (2012). Durable consumption and asset management with transaction and observation costs. *American Economic Review*, **102** (5), 2272–2300.

—, Lippi, F. and Paciello, L. (2011). Optimal price setting with observation and menu costs. *The Quarterly Journal of Economics*, **126** (4), 1909–1960.

—, — and — (2016). Monetary shocks in models with inattentive producers. *The Review of economic studies*, **83** (2), 421–459.

—, — and — (2017). Monetary shocks in models with observation and menu costs. *Journal of the European Economic Association*, **16** (2), 353–382.

Andrade, P. and Le Bihan, H. (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, **60** (8), 967–982.

Angeletos, G.-M. and Pavan, A. (2007). Efficient use of information and social value of information. *Econometrica*, **75(4)**, 1103–1142.

— and Sastry, K. (2019). *Inattentive Economies*. Working Paper 26413, National Bureau of Economic Research.

Antonovics, K. and Golan, L. (2012). Experimentation and job choice. *Journal of Labor Economics*, **30** (2), 333–366.

Argente, D. and Yeh, C. (2021). *Product Life Cycle, Learning, and Nominal Shocks*. Tech. rep., Penn State.

Arkolakis, C., Papageorgiou, T. and Timoshenko, O. A. (2018). Firm learning and growth. *Review of Economic Dynamics*, **27**, 146–168.

Auclert, A., Rognlie, M. and Straub, L. (2020). *Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model*. Working Paper 26647, National Bureau of Economic Research.

Bachmann, R. and Moscarini, G. (2011). Business cycles and endogenous uncertainty. In *2011 Meeting Papers*, Society for Economic Dynamics, vol. 36, pp. 82–99.

Bakshi, G. and Skoulakis, G. (2010). Do subjective expectations explain asset pricing puzzles? *Journal of Financial Economics*, **98** (3), 462–477.

Baley, I. and Blanco, A. (2019). Firm uncertainty cycles and the propagation of nominal shocks. *American Economic Journal: Macroeconomics*, **11** (1), 276–337.

—, Figueiredo, A. and Ulbricht, R. (2021). *Mismatch cycles*. Tech. rep., Barcelona GSE Working Paper 1151.

—, Veldkamp, L. and Waugh, M. (2020). Can global uncertainty promote international trade? *Journal of International Economics*, **126**, 103347.

Ball, L., Mankiw, N. G. and Reis, R. (2005). Monetary policy for inattentive economies. *Journal of monetary economics*, **52** (4), 703–725.

Balvers, R. J. and Cosimano, T. F. (1990). Actively learning about demand and the dynamics of price adjustment. *The Economic Journal*, **100** (402), 882–898.

Barto, V., Bauer, M., Chytilov, J. and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, **106** (6), 1437–75.

Begenau, J., Farboodi, M. and Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, **97**, 71–87.

Bernardo, J. M. and Smith, A. F. (2009). *Bayesian theory*, vol. 405. John Wiley & Sons.

Bertocchi, G. and Spagat, M. (1998). Growth under uncertainty with experimentation. *Journal of Economic Dynamics and Control*, **23** (2), 209–231.

Bertoli, S., Moraga, J. F.-H. and Guichard, L. (2020). Rational inattention and migration decisions. *Journal of International Economics*, **126**, 103364.

Boerma, J. and Karabarbounis, L. (2021). *Reparations and Persistent Racial Wealth Gaps*. Tech. rep., National Bureau of Economic Research.

Bonomo, M. and Carvalho, C. (2004). Endogenous time-dependent rules and inflation inertia. *Journal of Money, Credit and Banking*, pp. 1015–1041.

—, —, Garcia, R., Malta, V. and Rigato, R. (2021). Persistent monetary non-neutrality in an estimated model with menu costs and partially costly information. *Available at SSRN 2755119*.

Borovicková, K. (2016). *Job flows, worker flows and labor market policies*. Tech. rep., New York University.

Botsch, M. and Vanasco, V. (2019). Learning by lending. *Journal of Financial Intermediation*, **37**, 1–14.

Brown, Z. Y. and Jeon, J. (2020). Endogenous information acquisition and insurance choice. *University of Michigan and Boston University Working Paper*.

Buera, F. J., Monge-Naranjo, A. and Primiceri, G. E. (2011). Learning the wealth of nations. *Econometrica*, **79** (1), 1–45.

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, **105** (7), 2183–2203.

—, — and Leahy, J. (2017). *Rationally inattentive behavior: Characterizing and generalizing Shannon entropy*. Tech. rep., National Bureau of Economic Research.

Carroll, C. D., Crawley, E., Slacalek, J., Tokuoka, K. and White, M. N. (2020). Sticky expectations and consumption dynamics. *American Economic Journal: Macroeconomics*, **12** (3), 40–76.

Chahrour, R., Nimark, K. P. and Pitschner, S. (2019). *Sectoral media focus and aggregate fluctuations*. Boston College.

Chamley, C. (2004). *Rational Herds: Economics Models of Social Learning*. Cambridge University Press, 1st edn.

Chen, C., Senga, T., Sun, C. and Zhang, H. (2020). *Uncertainty, Imperfect Information, and Expectation Formation over the Firms's Life Cycle*. Tech. rep., CESifo.

Cheremukhin, A., Restrepo-Echavarria, P. and Tutino, A. (2020). Targeted search in matching markets. *Journal of Economic Theory*, **185**, 104956.

Cogley, T. and Sargent, T. J. (2005). The conquest of us inflation: Learning and robustness to model uncertainty. *Review of Economic Dynamics*, **8** (2), 528–563.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, **63** (4), 1977–2011.

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

Collin Dufresne, P., Johannes, M. and Lochstoer, L. A. (2016). Parameter learning in general equilibrium: The asset pricing implications. *American Economic Review*, **106** (3), 664–98.

Cong, L. W., Xie, D. and Zhang, L. (2020). Knowledge accumulation, privacy, and growth in a data economy. *Privacy, and Growth in a Data Economy (October 8, 2020)*.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American economic review*, **100** (1), 35–69.

Cover, T. M. and Thomas, J. A. (1991). Entropy, relative entropy and mutual information. *Elements of information theory*, **2** (1), 12–13.

Dasgupta, K. and Mondria, J. (2018). Inattentive importers. *Journal of International Economics*, **112**, 150–165.

de Oliveira, H., Denti, T., Mihm, M. and Ozbek, K. (2017). Rationally inattentive preferences and hidden information costs. *Theoretical Economics*, **12** (2), 621–654.

Doppelt, R. (2016). The hazards of unemployment, working Paper.

Drenik, A. and Perez, D. J. (2020). Price setting under uncertainty about inflation. *Journal of Monetary Economics*, **116**, 23–38.

Ellis, A. (2018). Foundations for optimal inattention. *Journal of Economic Theory*, **173**, 56–94.

Evans, G. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press, 1st edn.

Fajgelbaum, P. D., Schaal, E. and Taschereau-Dumouchel, M. (2017). Uncertainty traps. *The Quarterly Journal of Economics*, **132** (4), 1641–1692.

Farber, H. S. and Gibbons, R. (1996). Learning and wage dynamics. *The Quarterly Journal of Economics*, **111** (4), 1007–1047.

Farboodi, M., Mihet, R., Philippon, T. and Veldkamp, L. (2019). Big data and firm dynamics. In *AEA papers and proceedings*, vol. 109, pp. 38–42.

— and Veldkamp, L. (2021). *A Growth Model of the Data Economy*. Tech. rep., National Bureau of Economic Research.

Fernandes, A. P. and Tang, H. (2014). Learning to export from neighbors. *Jornal of International Economics*, **94** (1), 67–84.

Figueiredo, A. (2018). Information frictions in education and inequality. In *2018 Meeting Papers*, 804, Society for Economic Dynamics.

Fogli, A. and Veldkamp, L. (2011). Nature or nurture? learning and the geography of female labor force participation. *Econometrica*, **79** (4), 1103–1138.

Fuster, A., Perez-Truglia, R., Wiederholt, M. and Zafar, B. (2020). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *The Review of Economics and Statistics*, pp. 1–54.

Gabaix, X. and Laibson, D. (2001). The 6d bias and the equity-premium puzzle. *NBER macroeconomics annual*, **16**, 257–312.

Galenianos, M. (2013). Learning about match quality and the use of referrals. *Review of Economic Dynamics*, **16** (4), 668–690.

Ghofrani, E. (2021). *Learning with uncertainty's uncertainty*. Tech. rep., UPF.

Gonzalez, F. M. and Shi, S. (2010). An equilibrium theory of learning, search, and wages. *Econometrica*, **78** (2), 509–537.

Gorodnichenko, Y. (2008). *Endogenous information, menu costs and inflation persistence*. Tech. rep., National Bureau of Economic Research.

Goyal, S. (2011). Learning in networks. In *Handbook of Social Economics*, vol. 1, Elsevier, pp. 679–727.

Groes, F., Kircher, P. and Manovskii, I. (2014). The u-shapes of occupational mobility. *The Review of Economic Studies*, **82** (2), 659–692.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, **70** (3), 393–408.

Gutiérrez-Daza, A. (2021). *Inattentive Inflation Expectations*. Tech. rep., UPF.

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

— and Woodford, M. (2019). *Rational inattention when decisions take time*. Tech. rep., National Bureau of Economic Research.

— and Woodford, M. (Forthcoming). Neighborhood-based information costs. *American Economic Review*.

Hellwig, C., Kohls, S. and Veldkamp, L. (2012). Information choice technologies. *American Economic Review*, **102** (3), 35–40.

— and Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies*, **76**, 223–251.

— and Venkateswaran, V. (2009). Setting the right prices for the wrong reasons. *Journal of Monetary Economics*, **56**, S57–S77.

Herskovic, B. and Ramos, J. (2020). Acquiring information through peers. *American Economic Review*, **110** (7), 2128–52.

Hong, H., Torous, W. and Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, **83** (2), 367–396.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, **110** (9), 2819–58.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy*, **87** (5), 972–990.

— (1984). Matching, turnover, and unemployment. *Journal of political Economy*, **92** (1), 108–122.

— and Nyarko, Y. (1996). Learning by doing and the choice of technology. *Econometrica*, **64**, 1299–1310.

Kacperczyk, M., Nosal, J. and Stevens, L. (2019). Investor sophistication and capital income inequality. *Journal of Monetary Economics*, **107**, 18–31.

—, Van Nieuwerburgh, S. and Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, **84** (2), 571–626.

Keller, G. and Rady, S. (1999). Optimal experimentation in a changing environment. *The Review of Economic Studies*, **66** (3), 475–507.

Kellogg, R. (2011). Learning by drilling: Interfirm learning and relationship persistence in the texas oilpatch. *The Quarterly Journal of Economics*, **126** (4), 1961–2004.

Kelly, D. L. and Kolstad, C. D. (1999). Bayesian learning, growth, and pollution. *Journal of economic dynamics and control*, **23** (4), 491–518.

Khaw, M. W., Stevens, L. and Woodford, M. (2017). Discrete adjustment to a changing environment: Experimental evidence. *Journal of Monetary Economics*, **91**, 88–103.

Klenow, P. J. and Willis, J. L. (2007). Sticky information and sticky prices. *Journal of Monetary Economics*, **54**, 79–99.

Kozlowski, J., Veldkamp, L. and Venkateswaran, V. (2020a). *Scarring body and mind: the long-term belief-scarring effects of Covid-19*. Tech. rep., National Bureau of Economic Research.

—, — and — (2020b). The tail that wags the economy: Beliefs and persistent stagnation. *Journal of Political Economy*, **128** (8), 2839–2879.

Kszegi, B. and Matějka, F. (2020). Choice Simplification: A Theory of Mental Budgeting and Naive Diversification*. *The Quarterly Journal of Economics*, **135** (2), 1153–1207.

Lee, D. S. and Moretti, E. (2009). Bayesian learning and the pricing of new information: Evidence from prediction markets. *American Economic Review*, **99** (2), 330–36.

Liptser, R. S. and Shiryaev, A. N. (2001). *Statistics of random processes: I. General theory*, vol. 1. Springer Science & Business Media.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of economic theory*, **4** (2), 103–124.

Luo, Y. (2008). Consumption dynamics under information processing constraints. *Review of Economic dynamics*, **11** (2), 366–385.

—, Nie, J. and Young, E. R. (2015). Slow information diffusion and the inertial behavior of durable consumption. *Journal of the European Economic Association*, **13** (5), 805–840.

Mackowiak, B., Matějka, F. and Wiederholt, M. (2020). Rational inattention: A review. *CEPR Discussion Papers*, (15408).

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

— and — (2015). Business Cycle Dynamics under Rational Inattention. *Review of Economic Studies*, **82** (4), 1502–1532.

Mackowiak, B., Matějka, F. and Wiederholt, M. (2018). Dynamic rational inattention: Analytical results. *Journal of Economic Theory*, **176**, 650–692.

Mankiw, G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *Quarterly Journal of Economics*, **117**, 1295–1328.

Matějka, F. and Tabellini, G. (2016). Electoral competition with rationally inattentive voters. *Journal of the European Economic Association*.

Menzio, G. and Shi, S. (2011). Efficient search on the job and the business cycle. *Journal of Political Economy*, **119** (3), 468–510.

Miao, J., Wu, J. and Young, E. (2019). *Multivariate Rational Inattention*. Boston University - Department of Economics - Working Papers Series WP2019-07, Boston University - Department of Economics.

Mihet, R. (2021). *Financial Technology and the Inequality Gap*. Swiss Finance Institute Research Paper Series 21-04, Swiss Finance Institute.

Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political economy*, **92** (6), 1086–1120.

Minniti, M. and Bygrave, W. (2001). A dynamic model of entrepreneurial learning. *Entrepreneurship theory and practice*, **25** (3), 5–16.

Mirman, L. J., Samuelson, L. and Urbano, A. (1993). Monopoly experimentation. *International Economic Review*, pp. 549–563.

Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, **145** (5), 1837–1864.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, **92** (5), 1521–1534.

Moscarini, G. (2001). Excess Worker Reallocation. *The Review of Economic Studies*, **68** (3), 593–612.

Myatt, D. P. and Wallace, C. (2012). Endogenous information acquisition in coordination games. *The Review of Economic Studies*, **79** (1), 340–374.

Nagypál, É. (2007). Learning by doing vs. learning about match quality: Can we tell them apart? *The Review of Economic Studies*, **74** (2), 537–566.

Neal, D. (1999). The complexity of job mobility among young men. *Journal of labor Economics*, **17** (2), 237–261.

Nimark, K. P. (2014). Man-bites-dog business cycles. *American Economic Review*, **104** (8), 2320–67.

— and Pitschner, S. (2019). News media and delegated information choice. *Journal of Economic Theory*, **181**, 160–196.

Ordonez, G. (2013). The asymmetric effects of financial frictions. *Journal of Political Economy*, **121** (5), 844–895.

Paciello, L. and Wiederholt, M. (2014). Exogenous Information, Endogenous Information, and Optimal Monetary Policy. *Review of Economic Studies*, **81** (1), 356–388.

Papageorgiou, T. (2014). Learning your comparative advantages. *The Review of Economic Studies*, **81** (3), 1263–1295.

Pástor, L., Taylor, L. A. and Veronesi, P. (2009). Entrepreneurial learning, the ipo decision, and the post-ipo drop in firm profitability. *The Review of Financial Studies*, **22** (8), 3005–3046.

Pastor, L. and Veronesi, P. (2009). Learning in financial markets. *Annu. Rev. Financ. Econ.*, **1** (1), 361–381.

— and — (2012). Uncertainty about government policy and stock prices. *The journal of Finance*, **67** (4), 1219–1264.

Pastorino, E. (2009). Learning in labor markets and job mobility. *Unpublished manuscript, Department of Economics, University of Iowa.*

Peress, J. (2004). Wealth, information acquisition, and portfolio choice. *The Review of Financial Studies*, **17** (3), 879–914.

Phelps, E. S. (1970). Introduction: The new microeconomics in employment and inflation theory. *Microeconomic foundations of employment and inflation theory*, **1**, 23.

Porcher, C. (2020). *Migration with Costly Information*. Tech. rep., Working Paper, Princeton.

Prescott, E. C. (1972). The multi-period control problem under uncertainty. *Econometrica: Journal of the Econometric Society*, pp. 1043–1058.

Pries, M. and Rogerson, R. (2005). Hiring policies, labor market institutions, and labor market flows. *Journal of Political Economy*, **113** (4), 811–839.

Primiceri, G. E. (2006). Why inflation rose and fell: policy-makers' beliefs and us postwar stabilization policy. *The Quarterly Journal of Economics*, **121** (3), 867–901.

Reis, R. (2006a). Inattentive consumers. *Journal of monetary Economics*, **53** (8), 1761–1800.

— (2006b). Inattentive producers. *The Review of Economic Studies*, **73** (3), 793–821.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory*, **9** (2), 185–202.

Saijo, H. (2017). The uncertainty multiplier and business cycles. *Journal of Economic Dynamics and Control*, **78**, 1–25.

Sargent, T., Williams, N. and Zha, T. (2006). Shocks and government beliefs: The rise and fall of american inflation. *American Economic Review*, **96** (4), 1193–1224.

Senga, T. (2018). *A New Look at Uncertainty Shocks: Imperfect Information and Misallocation*. Working Papers 763, Queen Mary University of London, School of Economics and Finance.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50(3)**, 665–90.

Stevens, L. (2020). Coarse pricing policies. *The Review of Economic Studies*, **87** (1), 420–453.

Straub, L. and Ulbricht, R. (2018). Endogenous uncertainty and credit crunches. *Available at SSRN 2668078*.

Svensson, L. E. and Williams, N. M. (2007). *Bayesian and adaptive optimal policy under model uncertainty*. Tech. rep., National Bureau of Economic Research.

Turén, J. (2020). *State-dependent attention and pricing decisions*. Working paper, Pontificia Universidad Católica de Chile.

Van Nieuwerburgh, S. and Veldkamp, L. (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics*, **53(4)**, 753–772.

— and Veldkamp, L. (2010). Information acquisition and portfolio under-diversification. *The Review of Economic Studies*, **77(2)**, 779–805.

Veldkamp, L. (2005). Slow boom, sudden crash. *Journal of Economic Theory*, **124(2)**, 230–257.

— (2006). Information markets and the comovement of asset prices. *The Review of Economic Studies*, **73** (3), 823–845.

— and Wolfers, J. (2007). Aggregate shocks or aggregate information? costly information and business cycle comovement. *Journal of Monetary Economics*, **54**, 37–55.

Venkateswaran, V. (2014). Heterogeneous information and labor market fluctuations, working Paper.

Wee, S. L. (2016). Delayed learning and human capital accumulation: The cost of entering the job market during a recession. *Unpublished manuscript*, **18**.

Weitzman, L., Martin (2007). Subjective expectations and asset-return puzzles. *American Economic Review*, **97** (4), 1102–1130.

Wieland, V. (2000). Monetary policy, parameter uncertainty and optimal learning. *Journal of Monetary Economics*, **46** (1), 199–228.

Willems, T. (2017). Actively learning by pricing: a model of an experimenting seller. *The Economic Journal*, **127** (604), 2216–2239.

Wilson, R. (1975). Informational economies of scale. *The Bell Journal of Economics*, **6**, 184–95.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, p. 25.

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

Yang, C. (2020). *Rational inattention, menu costs, and multi-product firms: Micro evidence and aggregate implications*. Tech. rep., Federal Reserve Board.