Rational Inattention: A Review*

Bartosz Maćkowiak† Filip Matějka‡ Mirko Wiederholt¶

May 3, 2021

Abstract

We review the recent literature on rational inattention, identify the main theoretical mechanisms, and explain how it helps us understand a variety of phenomena across fields of economics. The theory of rational inattention assumes that agents cannot process all available information, but they can choose which exact pieces of information to attend to. Several important results in economics have been built around imperfect information. Nowadays, many more forms of information than ever before are available due to new technologies, and yet we are able to digest little of it. Which form of imperfect information we possess and act upon is thus largely determined by which information we choose to pay attention to. These choices are driven by current economic conditions and imply behavior that features numerous empirically supported departures from standard models. Combining these insights about human limitations with the optimizing approach of neoclassical economics yields a new, generally applicable model.

---

*This paper appeared in a much earlier form as CEPR Discussion Paper No. 13243, "Rational Inattention: A Disciplined Behavioral Model"

†European Central Bank and CEPR (bartosz.mackowiak@ecb.int)

‡CERGE-EI and CEPR (filip.matejka@cerge-ei.cz). This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreements No 678081 and No 101002898).

§Sciences Po and CEPR (mirko.wiederholt@gmail.com)

¶We thank Steven Durlauf, four anonymous referees, Michal Bauer, Andrew Caplin, Mark Dean, Stefano DellaVigna, Dmitry Taubinsky, and Peter Zorn for very useful comments and discussions, and Vladimír Novák, Andrei Matveenko, Klára Svitáková, Sergei Mikhalischev, Michal Šoltés, Rastislav Rehák, and Andrew Hannon for exemplary research assistance.

1 Motivation

> In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.
>
> *Herbert A. Simon (1971), pp. 40-41*

Every decision situation comes with a choice of attention. Agents always face the fundamental trade-off between processing more information to improve decisions and saving on the mental effort of doing so. Humans cannot process all available information; yet they can choose how to deal with this cognitive limitation. The theory of rational inattention is likely to become an important part of economics because it formalizes this idea. Rational inattention advances the earlier literature on information acquisition by relaxing assumptions of what information can be acquired. It also brings classical economics and behavioral economics closer together.

We present the theory of rational inattention (Section 2), survey applications of the theory by field (Section 3), review the existing empirical evidence (Section 4), and discuss policy implications (Section 5). Throughout the paper, we mention what we believe are the most fruitful steps for future research in this area.

Economics is about adjustments to scarcity. Rational inattention studies adjustments to scarcity of attention. Understanding how people select, summarize, and digest the abundant available information is key to understanding many phenomena in economics. Several important results in economics, even whole subfields, have been built around the assumption of imperfect or asymmetric information. Nowadays, many more forms of information than ever before are available due to new technologies, and yet we are able to digest little of it. Which form of imperfect information we possess and act upon is thus largely not determined by which information is given to us, but by which information we choose to attend to.

The way people deal with the abundant information has far reaching implications for:

- macroeconomics, because it forms our expectations, and thus affects the dynamics of prices, consumption, and investment;

- finance, because it determines investors' beliefs about asset returns, which in turn affect portfolio allocations and asset prices;
- labor economics, because it affects and directs the searches of both firms and job applicants;
- behavioral economics, because it determines what simplifications we use and thus may explain systematic biases in decision making;
- political economy, because we pay little attention to facts when our personal stakes are low.

Let us give a few examples of the implications of rational inattention that we will describe in greater detail below. In macroeconomics, rational inattention can explain why slow adjustment of expectations and actions, such as prices and consumption, to some shocks can co-exist with flexible reactions to other shocks. Rational inattention also predicts how the responses to shocks depend on the environment, in particular on monetary policy, yielding lessons for optimal policy. In finance, rational inattention tells us why portfolio under-diversification can be optimal, rather than an anomaly, in a world with abundant information, and why contagion between markets with unrelated fundamentals can arise. In labor economics, rational inattention leads to a theory of attention discrimination, a form of bias that exacerbates statistical discrimination and taste-based discrimination. This new theory has implications for how the society can counteract discrimination in the labor market and other settings. In political economy, rational inattention explains how systematic distortions can arise in a democracy when selectively ignorant voters interact with politicians. Many insights from rational inattention have implications across fields. For instance, rational inattention makes it precise why and in what circumstances private agents and policy makers under-prepare for a rare event, be it a global financial crisis or a deadly pandemic, which leads to lessons for how the society can prepare for and respond to a rare event.

The theory of rational inattention (following the seminal work of Christopher A. Sims, 2003) provides a model of how cognitively limited people simplify and summarize available information. It describes behavior that seems error-prone, yet the form of mistakes in final actions is subject to agents' choice; it is driven by agents' preferences and the stochastic properties of the environment. Rational inattention is motivated by the observation that people often cannot avoid mistakes due to lack of information, but they can choose what to think about, what to pay attention to and to what level of detail, i.e., what type of mistakes to minimize. People are inattentive; psychology and behavioral economics have been very successful in showing that humans' cognitive limitations are important for economic outcomes. But how do agents deal with their own cognitive limitations when they are aware of them? How do firms act or how should policy makers act when facing such agents? The next step is thus to study how the economic actors adjust to such frictions. We will argue that rational inattention is a suitable model for doing so. See also Sims (2010), Wiederholt (2010), and Veldkamp (2011) for reviews of the earlier development of the literature.[1]

2 Theoretical framework

Rational inattention, henceforth RI, builds on the observation that humans cannot pay full attention to all available information, but can choose to pay more attention to more important things.

How does RI fit within existing theories? There exist several well-established models of imperfect information. Many of them are based on imperfect information of an exogenously given form; this is not RI. A large class of models describe information acquisition – agents' choices of costly signals. RI belongs to this family of models but its distinctive feature is that agents can choose to acquire signals of any form. Any information is available, yet costly to process. We discuss this in more depth in Section 2.4.

Consider a manager who sets a price to maximize profit. The optimal price depends on the state of the world x, which describes the current market conditions (e.g., elasticity of demand and marginal input cost). If x is observed, then an optimizing manager chooses deterministically the price y that maximizes profit,

$$\text{perfect information: } x \to y(x).$$

If the manager instead gets noisy information about x, then she chooses y that maximizes expected profit. The form of noisy information that she gets exogenously determines what posterior beliefs she may hold and what prices she sets, i.e., it determines the distribution $f(y|x)$,

$$\text{noisy information: } x \to f(y|x).$$

RI, however, allows for a more endogenous approach. The RI manager acts as if she were *choosing*

[1] Gabaix (2019) surveys a related concept of inattention called "sparsity". We discuss its connection to rational inattention in Section 2.4.

$f(y|x)$.[2] The choice of f reflects the manager's decision of what information to receive, and describes the form of pricing mistakes that she makes.

If she chooses not to get any information, then she selects an $f(y|x)$ that does not depend on x. In fact, $f(y|x)$ is degenerate so that she selects one constant y for all x. If she pays positive attention, i.e., gets some information, then $f(y|x)$ varies across x. And if, for instance, she pays more attention to x of low values, e.g., states of low demand, then pricing is more precise at these levels of demand and $f(y|x)$ is tighter for low x. Making more accurate choices, however, takes more effort, and thus more concentrated $f(y|x)$ are associated with a higher cost.

$$\text{RI: } x \to f(y|x), \text{ where } f(y|x) \text{ is chosen optimally.}$$

In a series of papers, Christopher A. Sims put forth two main cornerstones of RI as a model of *processing of available information:*

1. *The idea of selective and costly attention*: Available information is not internalized information. In principle we can have the whole Internet at our disposal, yet we choose to process only a very limited amount of this information; we choose what questions we ask our friends, or what to read about in the news.

2. *A convenient modeling framework:* A combination of the flexible choice of information with a specific form of an entropy-based cost function. Sims (2003) formulates a dynamic model where a single agent chooses how much information to process about different Gaussian shocks. Sims (2006) emphasizes that in practice it is not only the amount of information that agents choose but also the nature of information, both of which can be modeled by the choice of $f(y|x)$ subject to the cost of information. We show below that the flexibility of f, perhaps surprisingly, often leads to higher tractability. While initially the models used an entropy-based cost of information, the approach is also applicable with other types of the cost.

In the purest form of rational inattention, an optimizing agent can choose from an unrestricted set of Blackwell experiments $f(y|x)$, subject to a cost of information.

In the rest of this section, we first formulate a general static model and discuss the main properties of its solution. Next, we discuss the assumptions of rational inattention in more detail.

[2] She chooses what Blackwell experiment to run.

Finally, we present results for a dynamic model and a model with a multi-dimensional state or action.

2.1 Static model

Here we describe the general static model of choice under RI (see Matějka & McKay, 2015).[3] The unknown random state is x, and the agent's prior belief is given by a pdf $g(x)$. The agent faces a two-stage decision problem:

(i) *What to pay attention to*: The agent selects an information strategy to refine her belief about the state. This is described by what signals s the agent gets for a given realized state x, i.e., by a distribution $f_{sx}(s|x)$.

(ii) *What action y to take*: This is a standard choice under uncertainty with the beliefs generated in the first stage via Bayesian updating.

The objective is to maximize the expectation of $U(y, x)$ less the cost of information $C(f_{sx})$, which is a function of the information strategy. The timing is as follows:

1. The agent chooses the information strategy to maximize the expectation of utility less the cost of information while considering the action strategy she applies later.

2. The agent receives a signal s, the cost of information is incurred, and her posterior is formed.

3. The agent chooses an action y. Because the action does not affect the cost of information, she chooses y to maximize the expectation of $U(y, x)$ given the posterior.

While the agent chooses two strategies, information (i) and action (ii) strategies above, it turns out that a joint distribution $f(y, x)$ describes both of the strategies as in Sims (2003) and Kamenica & Gentzkow (2011). If optimal then the two strategies must be such that no two signals in step 2 lead to the same action in step 3, otherwise the agent would be wasting costly information by distinguishing between two signals that do not change her actions.[4] We can thus make a one-to-one association between s and y, and use $f(y, x)$ only.

[3] See also Caplin & Dean (2015) where such a model is derived from revealed preference principles.

[4] This is subject to Blackwell monotonicity of the cost of information.

The agent's problem then is:

$$\max_{f} \int U(y,x) f(y,x)\, dx\, dy - C(f), \tag{1}$$

$$\text{subject to} \int f(y,x)\, dy = g(x)\,, \quad \forall x, \tag{2}$$

where the first term in (1) is the expectation of U, and $C(f)$ is the cost of information.[5] The constraint (2) captures Bayesian rationality, requiring the consistency of prior and posterior beliefs.

While the cost of information $C(f)$ could in principle take many different forms, following Sims (2003) we now use $C(f) = \lambda I(y;x)$,[6] where $I(y;x)$ is the Shannon mutual information between the random variables y and x (Cover & Thomas, 2006). Letting $p(y)$ denote the marginal of y, mutual information is defined as

$$I(y;x) \equiv H(x) - E[H(x|y)] = \int f(y,x) \log\left(\frac{f(y,x)}{g(x)p(y)}\right) dxdy, \tag{3}$$

where $H(x)$ is entropy of the random variable x.[7] In Section 2.4 we discuss the choice of a cost function in more detail. For the moment, it suffices to say that $I(y;x)$ measures the expected uncertainty reduction about x due to knowledge of y, which is a common way to measure the amount of information processed about x.

2.2 Solution, implications, and optimal biases

The first order condition to (1)-(2) implies that the behavior is probabilistic and follows a logit model (Matějka & McKay, 2015). For an unknown state x the distribution of actions is given by:

$$f(y|x) = \frac{p(y)e^{U(y,x)/\lambda}}{\int_z p(z)e^{U(z,x)/\lambda} dz}. \tag{4}$$

To connect this formula to the applied literature on discrete choice, let us state the resulting choice probabilities in the case when the action set is discrete, $y \equiv i \in \{1, ..., N\}$:

$$P(i|x) = \frac{e^{\frac{U(i,x)+\alpha(i)}{\lambda}}}{\sum_{j=1}^{N} e^{\frac{U(j,x)+\alpha(j)}{\lambda}}}, \tag{5}$$

[5]For simplicity here we use the imprecise notation using probability distribution functions only. See Jung et al. (2019) for a formulation in terms of probability measures.

[6]Here we use a linear function of I for simplicity, other functions are possible, too. In fact, Sims (2003) used a hard constraint on mutual information, $I(y;x) \leq \kappa$, and a more general convex cost is also reasonable in many settings.

[7]The entropy of x is $H(x) = -\int f(x) log(f(x))dx$.

where $\alpha(i) = \lambda \log P(i)$ and $P(i) > 0$ is the marginal probability of the choice i.

Solutions to RI problems thus take a convenient analytical form and yet can exhibit rich behavioral properties. Utility enters in the logistic way, and all effects of beliefs and the choice of information are summarized by the additive shifters $\alpha(i)$. The quantities $\alpha(i)$ reflect *biases* towards action i. The biases are independent of state x, but endogenous to the prior knowledge and preferences that determine the agent's choice of attention. See Matějka & McKay (2015) and Caplin et al. (2019) for how to solve for $P(i)$ or $\alpha(i)$.[8] The full description of behavior is particularly simple for a low number of possible actions or for quadratic utility when the prior g is Gaussian, see the examples in Section 2.3.

As we will demonstrate in the examples, the implied features of the behavior of RI agents are the following. Some of the features are direct implications of (4) and (5), while others are driven by the choices of the unconditional probabilities $p(y)$ and the resulting biases. At the end of this subsection we discuss which of these implications are inherent to RI only and which to a more general class of models of information acquisition.

(F1) *Stochastic choice:* RI agents make random mistakes. The mapping from the state x to the action y is not deterministic.

(F2) *Stakes and lower cost of information increase responsiveness:* Scaling up the utility function or down the cost of information λ implies that the action y becomes more responsive to the state x.

(F3) *Logistic choice* with biases is the exact optimal form of behavior for any utility function and prior beliefs, and it is implied by the entropy-based cost of information. This makes the RI model tractable, and potentially amenable to empirical applications. Consider a discrete choice problem. If all alternatives $i = 1, ..., N$ happen to be equally attractive a priori, then the probability of choosing a specific alternative i conditional on state x takes exactly the standard logit form. If one

[8] The solution in the discrete action case satisfies

$$\int_x \frac{e^{U(i,x)/\lambda}}{\sum_{j=1}^N P(j) e^{U(j,x)/\lambda}} g(x) dx = 1, \tag{6}$$

for all i such that $P(i) > 0$. Notice that this condition says that the posterior distribution $P(x|i)$ needs to integrate to 1. Caplin et al. (2019) show that for $P(i) = 0$ the LHS of the equation above is less than one and that the two conditions together are sufficient and necessary.

alternative i seems more attractive a priori, then the agent chooses to pay attention in such a way that the bias towards this alternative $\alpha(i)$ has the same effect as a positive utility shock $\alpha(i)$ in each state.[9] This bias towards a specific alternative is endogenous and changes with the choice set, the prior, the utility function, and the cost of information.

The sophistication in the choice of α distinguishes RI-logit from the standard logit or from any information acquisition of a fixed form. Comparative statics are very different anytime the prior (and preferences) or the choice set change, because α changes too.

(F4) *Gaussian signals:* In a continuous choice problem with an unbounded action set, $y \in R$, quadratic preferences $U(y, x)$, and Gaussian prior uncertainty $g\left(x\right)$, Gaussian signals are optimal. A large part of the literature following Sims (2003) works with Gaussian signals, because they often yield tractable solutions - the action y depends on the state x linearly. In general, the form of signals depends on the form of $U(y, x)$ and reflects what types of imperfections in beliefs it is most important to refine.

(F5) *Magnified relative elasticities:* RI agents pay attention to important variables, which in relative terms makes them even more important than under perfect information.

(F6) *More attention to more volatile variables:* RI agents pay more attention when prior uncertainty is larger.[10]

(F7) *Categorization, discreteness, and consideration sets:* RI agents most often find it optimal to contemplate a low number of actions only.[11] This is the case even for continuous action sets, where the resulting set of possible actions is discrete - for instance when a price setter can choose any price, but keeps alternating between two fixed levels, e.g., a regular and a sale price. If the agent chooses to focus on some actions only, she does not waste information capacity on small movements, and is thus less likely to make larger errors. For a sufficiently high cost of information or a sufficiently strong prior, the agent may even consider only a single action.

(F8) *Violations of revealed preference:* RI can imply choices that are seemingly irrational. This can be driven by the fact that changing the choice set can induce RI agents to pay attention

[9]See Fosgerau et al. (2020a) for generalizations of the choice formula for costs beyond the mutual information.

[10]See for instance Maćkowiak & Wiederholt (2009) where price-setters pay more attention and respond more strongly to firm-specific shocks because they are more volatile than aggregate shocks.

[11]Matějka (2016), Caplin et al. (2019), Jung et al. (2019), Stevens (2020).

differently, albeit optimally. If signals are endogenous to what options are presented to the agent, then transitivity in choice can be violated.[12]

(F9) *Posterior invariance:* As long as the number of possible states is no larger than the number of alternatives, then the set of possible posteriors the agents can acquire is often independent of small changes to the prior. This feature is useful for solving models with RI.[13]

(F10) *Multi-dimensional simplification - indexation*: If agents need to pay attention to several shocks and choose multiple actions, then RI models what kind of simplified representation of this high-dimensional environment the agents use. A consumer may choose not to attend to all prices of all products, but might compare only close substitutes, because that is the most useful piece of information determining what to buy, keeping the total consumption fixed. Or an investor agent pays attention to a particular linear combination of the asset prices only, i.e., to an endogenously constructed index, and then purchases or sells the whole portfolio given by the index.[14]

Feature (F1) would hold in any model of imperfect information with noisy signals, and (F2) emerges in most models of costly information acquisition. The other features need more flexibility of what type of information can be acquired. Features (F5), (F6) and (F8) require a finer model of costly information acquisition where the agent can choose to acquire signals of different precision on different shocks. Finally, (F3), (F4), (F7), (F9), and (F10) require the high flexibility of information choice under RI and (F3) and (F4) require the specific cost function. For other costs of information, (F3) would instead state some other form of under-reaction to shocks. RI thus shares some of the main features with other models of costly information acquisition, but gives less discretion to the modeler over what information is available.

2.3 Examples

We now present two canonical examples that highlight the interaction between preferences and the form of attention that agents choose to pay.

[12]Woodford (2012), Matějka & McKay (2015).

[13]Caplin et al. (2017).

[14]Fulton (2017), Kőszegi & Matějka (2020), Miao et al. (2020).

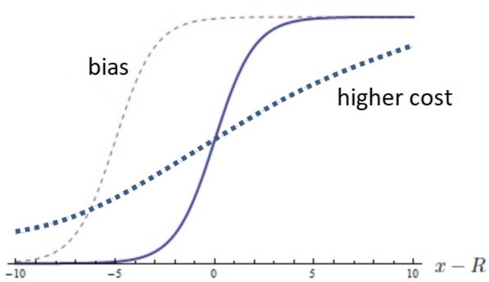

Figure 1: Probability of accepting the candidate, $P(1|x)$. Solid: symmetric attention, dotted: symmetric but less attention, dashed: biased attention.

2.3.1 Discrete choice: logit and biased logit

We start with a simple binary-choice problem. Consider an employer who is facing an applicant for a job opening, and is deciding whether to hire the applicant ($i = 1$) or not ($i = 0$). The unknown state x summarizes the applicant's quality and has a prior distribution $g(x)$. Utility from accepting the applicant is equal to x ($U(1, x) = x$), while utility from rejecting the applicant equals a known reservation utility R ($U(0, x) = R$).

The employer chooses to process information about the applicant that results in a behavior according to (5):

$$P(1|x) = \frac{e^{\frac{x+\alpha(1)}{\lambda}}}{e^{\frac{x+\alpha(1)}{\lambda}} + e^{\frac{R+\alpha(0)}{\lambda}}}. \tag{7}$$

To fully characterize the choices, we still need to find $\alpha(1) = \lambda \log P(1)$, where $P(1)$ is the unconditional probability of accepting the applicant. This is easy in the case the prior $g(x)$ is symmetric about R, i.e. $g(R + x) = g(R - x)$. In this case it must a priori be as likely that the applicant is accepted as rejected, $\alpha(1) = \alpha(0)$, and the choice behavior takes the standard logit form:

$$P(1|x) = \frac{e^{x/\lambda}}{e^{x/\lambda} + e^{R/\lambda}}. \tag{8}$$

(F1) *Stochastic choice*: The solid line in Figure 1 shows the probability (8) of accepting the candidate conditional on the state. Under unlimited information-processing capacity, the probability jumps from 0 to 1 at $x = R$. Under rational inattention, the probability of hiring is smoothly increasing in the state. Choice is stochastic conditional on the state and firms make mistakes. However, the employer pays attention in such a way that it rarely misses great opportunities ($x >> R$)

and rarely makes horrible hires ($x << R$), but frequently makes mistakes when confronted with a marginally profitable opportunity or a marginally unprofitable opportunity (x close to R). The signal that generates these choice probabilities is a signal that announces whether x exceeds R and is more likely to be correct when the absolute value of $x - R$ is larger.

(F2) *Stakes and cost of information*: Scaling down stakes (multiplying $U(1, x)$ and $U(0, x)$ by the same constant smaller than one) and scaling up the cost of information λ moves the response closer to a constant, see the less steep dotted line in Figure 1. This is because employers choose to pay less attention to the state and thus are more likely to make mistakes conditional on the state.

(F3) *Biased logit*: In this example with a symmetric prior, choice behavior takes the standard logit form, but if $g(x)$ is asymmetric, which means that the employer a priori believes that the applicant's value is more likely to be either above or below the reservation utility, then the endogenous biases α enter. The reason for the asymmetric prior could be that the employer has already conditioned on known characteristics of the applicant such as gender or race. The exponents in the logit formula now include $x + \alpha(1) - \alpha(0)$. The employer acts as if the true utility provided by the applicant were different by a shifter $\alpha(1) - \alpha(0)$. The dashed line in Figure 1 shows the resulting choice probability if the bias towards accepting the candidate were positive. The choice behavior becomes logit with bias.

For some $g(x)$ it is easy to solve for α analytically.[15] Furthermore, computing α numerically is usually straightforward. In Figure 1, we have shifted the prior to the right. The employer now a priori believes that the applicant's quality is more likely to be above the reservation utility. The bias becomes strictly positive, $\alpha(1) - \alpha(0) > 0$, and the probability of hiring is increased for each x.

To illustrate the *magnification* effects of RI, feature (F5) in Section 2.2, it is useful to plot the bias directly. Figure 2 shows the bias for the example given in footnote 15 as a function of $E[x] - R$. If the expected value of quality of a group relative to R is increased, then not only x itself increases their chances of being accepted, but the additional bias $\alpha(1) - \alpha(0)$ towards that group

[15] For example, for $x \in \left\{R - \frac{1}{2}, R + \frac{1}{2}\right\}$, we get from (6):

$$P(1) = \max\left(0, \min\left(1, -\frac{e^{\frac{R+\frac{1}{2}}{\lambda}}\left(-e^{\frac{1}{\lambda}} + e^{\frac{R+\frac{1}{2}}{\lambda}} - g(-\tfrac{1}{2}) + g(-\tfrac{1}{2})e^{\frac{1}{\lambda}}\right)}{\left(e^{\frac{1}{\lambda}} - e^{\frac{R+\frac{1}{2}}{\lambda}}\right)\left(-1 + e^{\frac{R+\frac{1}{2}}{\lambda}}\right)}\right)\right), \qquad P(0) = 1 - P(1). \tag{9}$$

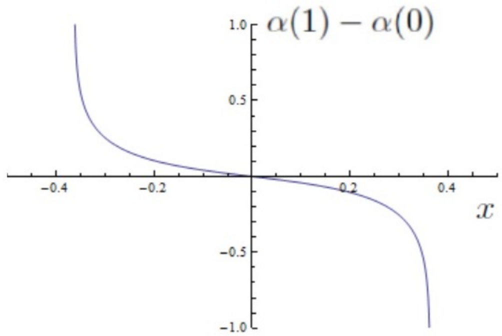

Figure 2: Attention-driven bias towards accepting the candidate.

is also higher. In Figure 2, if the values in such a group are increased relative to R by 0.2, then the attention-driven bias equals about 0.1, and for the change relative to R of 0.35, the additional bias is 0.5.

If the expected value relative to R is sufficiently high or low (about 0.4 in the figure), then the employer chooses to pay no attention, and simply accepts or rejects the applicant based on the prior. The employer does not consider the other alternative at all (feature F7 in Section 2.2).

The RI behavior takes a logit-form, but with adaptive biases that summarize *heuristics* that the agent chooses to use, i.e., strategies based on less-than-full information that are optimally tailored to the environment. The sophistication in the choice of α distinguishes RI-logit from the standard logit or from any information acquisition of a fixed form. Comparative statics are thus very different anytime the prior (and preferences) or the choice set change.

Debreu (1960) criticized the standard logit-behavior using a thought experiment with two duplicate alternatives (the "red-bus, blue-bus problem"). He pointed out that in this case the property of independence from irrelevant alternatives (IIA) is unappealing - if people choose the train and the bus both with the probability of 1/2, then after addition of the second bus the probabilities should not be 1/3 for all three alternatives, but 1/2 for the train and 1/4 for each of the buses. The IIA does not hold in the RI model. The sophisticated RI agent would have no reason to distinguish between the two buses, and the resulting probabilities would be 1/2, 1/4 and 1/4, exactly as Debreu found preferable.[16]

[16]See Matějka & McKay (2015) for a more detailed discussion of this example.

2.3.2 Quadratic-Gaussian case

We now explore a situation where actions are continuous, and the losses from imperfect information depend on the conditional variance only.

Consider a manager who sets a price y to maximize profit subject to unknown current market conditions, x (Maćkowiak & Wiederholt, 2009, Wiederholt, 2010). Let the utility be

$$U(y,x) = -r\Big(ax - y\Big)^2, \tag{10}$$

which can be derived from a log-quadratic approximation of the profit function, where ax is the log of the target price. The agent needs to pay attention to x, which can be a deviation in marginal cost and translates into an optimal deviation in price. The parameter a denotes the elasticity of the target price to the shock x - under perfect information it would equal the elasticity of y to x. The parameter r scales stakes.

Let the agent's prior $g(x)$ be Gaussian:

$$x \sim N(0, \sigma_x^2).$$

This is a popular specification because with quadratic preferences, Gaussian prior uncertainty, and an unbounded action set, Gaussian signals are optimal. For quadratic $U(y,x)$, the distribution proportional to $e^{U(y,x)/\lambda}$ is Gaussian. Therefore, if $p(y)$ is Gaussian, then equation (4) implies that $f(y,x)$ is jointly Gaussian as well. This is a fixed-point problem: take $p(y)$ and generate $f(y,x)$ via equation (4). The marginal of the resulting $f(y,x)$ must be the pdf $p(y)$ above. It is straightforward to show that such $p(y)$ exists, and uniqueness implies that this must be the only solution.[17] Hence, Gaussian signals are the unique solution (feature F4 in Section 2.2).

Therefore, the objective (1) with the utility (10) takes the form:

$$\max_{\sigma_{x|s}^2 \leq \sigma_x^2} E_x\Big[E_s\big[-ra^2(x - E[x|s])^2\big]\Big] - \lambda I\left(y;x\right) = \max_{\sigma_{x|s}^2 \leq \sigma_x^2}\Big(-ra^2\sigma_{x|s}^2 - \frac{\lambda}{2}\log\frac{\sigma_x^2}{\sigma_{x|s}^2}\Big), \tag{11}$$

where $\sigma_{x|s}^2$ denotes the posterior variance. The first term on both sides of the equation is the expected utility, and the second is the cost of information. The entropy $H(\cdot)$ of a random variable drawn from a normal distribution with variance σ^2 is $\frac{1}{2}\log(2\pi e\sigma^2)$.

[17] See Matějka & McKay (2015) for uniqueness statements.

Upon reception of a signal (which here has the form $s = x + \epsilon$ with a Gaussian ϵ), the action $y = aE[x|s]$ maximizes the expectation of (10) for any given posterior belief. Bayesian updating with Gaussian prior uncertainty and signals delivers linear dependence of $E[x|s]$ on x,

$$E[x|s] = (1-\xi)\bar{x} + \xi s = \xi(x+\epsilon),$$

where $\bar{x}$ is the prior mean of x and the weight on the signal, $\xi \equiv \left(1 - \sigma^2_{x|s}/\sigma^2_x\right) \in [0,1]$ reflects the chosen level of attention.[18] Therefore,

$$y = (a\xi)x + (a\xi)\epsilon, \tag{12}$$

where $(a\xi)\epsilon$ is the resulting noise in actions. Notice that (12) describes a jointly-normal distribution of y and x, the object $f(y|x)$ discussed above. If $\xi = 1$, then the agent pays full attention, and thus $y = ax$; $\xi = 0$ means no attention and no response to x.

We can now rewrite the problem (11) in terms of the choice variable ξ.

$$\max_{\xi\in[0,1]}\left(-ra^2(1-\xi)\sigma^2_x - \frac{\lambda}{2}\log\frac{1}{1-\xi}\right). \tag{13}$$

The solution is

$$\xi = max\left(0, 1 - \frac{\lambda}{2ra^2\sigma^2_x}\right). \tag{14}$$

The formula (14) together with (12) illustrates some of the general features of the solutions to RI problems that were summarized in Section 2.2:

(F1 and F3) *Stochastic under-reaction*: Realized prices move on average less than optimal prices because the prior matters. If $\lambda > 0$, then the agent under-responds to the realization of x, because $a\xi < a$. She does not get perfect information about x and thus puts a positive weight on the prior knowledge. This effect drives Sims' initial motivation for RI as a micro-foundation for sluggish behavior.

(F2) *Stakes and cost of information:* Higher stakes r and a lower cost λ increase responsiveness.

(F5) *Magnified relative elasticities:* RI magnifies differences in responsiveness to different shocks. Consider two different products with elasticities $a_1 > a_2$; under RI the relative elasticities are $\frac{a_1\xi(a_1)}{a_2\xi(a_2)} > \frac{a_1}{a_2}$. The elasticity under RI is $a\xi$, and since ξ is increasing in a, the realized elasticity is convex in the elasticity under perfect information.

[18] $\xi = \frac{\sigma^2_x}{\sigma^2_x+\sigma^2_\epsilon}$; and the posterior uncertainty is $\sigma^2_{x|s} = \frac{\sigma^2_x\sigma^2_\epsilon}{\sigma^2_x+\sigma^2_\epsilon}$.

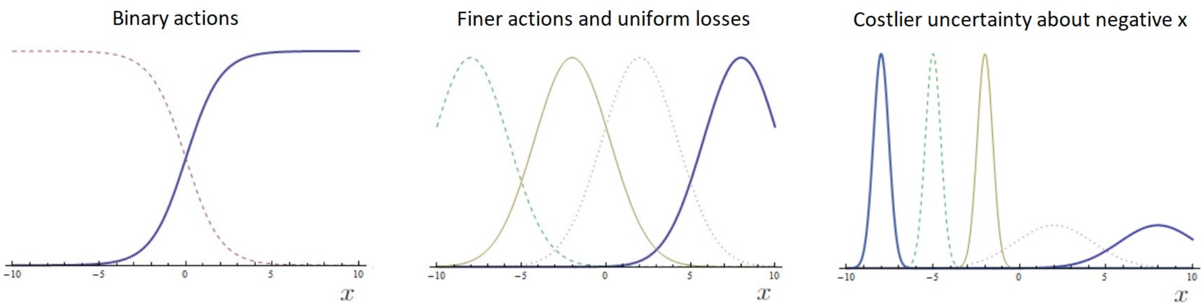

Figure 3: Collections of posteriors for various types of choice problems.

(F6) *Uncertainty increases responsiveness:* The higher the uncertainty about the target price, the more attention the price-setter pays to shocks and the more elastic the response is, i.e., ξ is increasing in the prior uncertainty σ_x^2.

2.3.3 General preferences

In general, the form of signals depends on the form of $U(y, x)$. It reflects potential losses from misinformation about x and thus what types of imperfections in beliefs it is most important to refine. When actions are binary, as in Section 2.3.1, then the agent chooses to collect information that yields threshold-like posterior beliefs, see the left exhibit in Figure 3. Let x measure the utility gain from choosing one alternative over the other. The agent then either finds out that x is most likely positive, and chooses that alternative, or negative, and chooses the other. Any additional information including information on the size of the difference in utilities is wasteful.

In other cases (such as when utility is quadratic and the action can be selected from a continuous interval, as in Section 2.3.2) RI agents do choose to pay attention to the level of x. The optimal signals then typically lead to posteriors such as those in the middle exhibit in Figure 3, where the agent can distinguish between different levels of x albeit imperfectly.

Notice however, that in the case of quadratic utility, the losses are given only by posterior variance of x, the level of x itself does not affect how significant a given level of uncertainty is, and thus the agent pays equal attention to all levels of x. The situation is different if the losses are higher for negative x. For instance, if x is an unknown bank account balance, then the RI agent might choose to pay more attention to negative balances, when she could be subject to high fees. She would, for instance, read in more detail if the balance started with a minus sign, or were written in red ink. The resulting posterior beliefs would be more precise in such cases, see the right exhibit in Figure 3. In general, there are as many posteriors as different actions, and they have a form $\propto e^{U(y,x)/\lambda}$.

2.4 Discussion

In this subsection we answer several questions that we have encountered over the years. Some of the questions address what rational inattention is and what it is useful for, while others are criticisms of RI.

2.4.1 Assumptions of RI

What is the definition of RI? Is the use of entropy crucial? We think of the model consisting of equations (1)-(3) as the benchmark static RI model. This model formalizes the following three main assumptions of RI (where we view the first two as the main assumptions).

(A1) *Information is available in a wide variety of forms.* This is reflected in model (1)-(3) by the agent's ability to shape signals in any way she chooses, i.e., she can choose any Blackwell experiment $f(y|x)$.

(A2) *Agents choose information optimally*, reflected by the optimization of the distribution f.

(A3) *The cost of information is measured by mutual information $I(y;x)$.* This assumption is not crucial, and can be relaxed.

While the model consisting of equations (1)-(3) is the "pure form" of RI, models with cost functions other than mutual information and models with some restrictions on available signals can also be viewed as RI.

We conjecture that RI models best describe repetitive decisions with a great deal of available information. In such decision situations not only can agents choose any pieces of information that they wish, but they are also more likely to use the optimal information strategy. On the other hand, decision situations that the model fits less clearly are new and quick one-time decisions, because they probably feature violations of (A2). If information is available in fewer forms (violation of (A1)), this can be incorporated by adding a constraint to model (1)-(3) specifying the available signals (see for instance Maćkowiak & Wiederholt, 2009, or Van Nieuwerburgh & Veldkamp, 2010).

When is information available in a wide variety of forms, and when is it not? Humans often have flexibility in shaping the information that they get, but sometimes they cannot shape it in exactly the way they would like. The assumption (A1) is the main distinction between RI and other models of information acquisition, which typically allow for one form of information only, see Section 2.4.3.

For many types of information we have the whole Internet at our disposal, where the flexibility is large. Such information includes movements in financial markets, most macroeconomic or even political data. On a more individual level, employers can get information about job applicants from structured CVs, and we can all judge the state of our finances using apps on our cellphones. Very often we can also ask questions of exactly our choice, which determine the form of answers we can get be it from an expert advisor or a sales representative. RI could also potentially be a proxy for directed thinking; the formation of signals is then internal, within the agent's mind.

In some situations, however, information is not available or it is available only in a particular form. Think about future financial shocks that have not been realized yet, or about a product that you contemplate buying on a flea market where you need to rely only on a visual perception of its exterior. When products are not presented on the internet, we often need to journey to the store at a considerable cost to acquire information. Sims (2010) gives the example of a prospector who wants to find out if there is oil underground or not. The prospector needs to drill a test well and see. No other signal is available, and the main cost is the physical cost of drilling a test well. A model of information acquisition with an observation cost is more appropriate in this case. Agents can not always obtain information in exactly the optimal form given by equation (4).

How can RI agents choose the information strategies optimally when they are cognitively limited? We consider RI to be an "as-if model" or a benchmark that applies well in repeated choice situations, or in choices over the long term. In these cases, the agent thinks about the optimal strategy once, and then applies it many times with little additional effort. Alternatively, it can be a strategy that the agent gradually learned through experience or stumbled upon it due to some evolutionary reasons.

It is likely that when it comes to our everyday consumption decisions, we know what information is useful for us to decide well. For instance, in a country with low and stable inflation, it makes sense to pay closer attention to own nominal income than to general inflation, which would matter as much in perfect-information models. When an HR manager inspects hundreds of CVs a day, she knows which pieces of information on the CV are the most important, and perhaps looks at those only. Similarly, when driving a car we know where to look for traffic signs, and which of them to follow more closely. But what if a car in front of us punctures a tire and spins out of control? It is unlikely that we will quickly choose precisely those pieces of information that would be most useful to assess what to do in that situation.

How can RI agents have accurate priors? How can they know the correct model and the distribution of shocks? For many settings, this seems a good benchmark. But importantly, RI does not require correct priors. The prior $g(x)$ does not have to equal the true distribution of x. If the agents have incorrect priors, then the model works just as well. But with RI the incorrect priors affect the choices of attention strategy, too, and can thus be even more detrimental.

Why use entropy? Is the use of entropy not arbitrary? We view entropy as a good benchmark, which is supported by theoretical arguments. At the same time, entropy is certainly not universally valid and recent theoretical and empirical work has been exploring alternatives.

While we do not intend to argue for a universal application of entropy, we think that entropy is an appealing benchmark, similar to the Cobb-Douglas production function. The main reasons for that are: (i) entropy allows for tractability, (ii) most of its qualitative properties are reasonable (more precision at a higher cost), and thus many qualitative implications of the model are independent of this particular choice, (iii) the foundation on optimal coding and also the axiomatic foundations of entropy suggest that it is a suitable function for processing of available information (Shannon, 1948, Cover & Thomas, 2006).

The description of the cost function in assumption (A3) is related to what information is available. In fact, the entropy-based cost is exactly in line with the assumptions of full availability and flexibility, (A1) and (A2). It was shown in the literature on information theory[19] that mutual information (3) is exactly proportional to the expected number of signals or symbols that need to be received to acquire the desired information. If the incurred cost is then proportional to this number

[19] See MacKay (2003) for a textbook presentation of optimal coding and information theory.

(due to effort devoted or time spent), then it is proportional to expected reduction of entropy, too. The important assumption is the possibility of optimal coding of information, for which (A1) and (A2) are the necessary prerequisites.

Entropy is also linked to the optimal sequencing of such independent pieces of information, which is related to (A2). In information theory, if a message is coded optimally, then entropy describes the best and also an achievable bound of what can be communicated in an interval of time (see for instance MacKay, 2003).

The axioms that entropy is built on are monotonicity, continuity, and some form of the possibility of independence of different pieces of information (Shannon, 1948). The remaining axiom, which distinguishes entropy from other reasonable cost functions, says that the cumulative amount of information is the same in the following two cases: (i) the agent finds out directly which state has occurred, and (ii) the agent learns which subgroup of states has occurred, and then finds out which state within the subgroup has occurred. This axiom provides certain additive properties that help with tractability, and it is linked to the availability of information (assumption (A1)). The axiom implies that the agent can obtain some information about a subgroup without getting more information about states within the subgroup. In reality, this kind of information may not always be available.

Entropy as a cost function is thus likely to be reasonable when information is presented in an optimal way, when we can ask questions of our choice, when a product's structured description gives prominence to important features or when the buyer knows the structure and can choose what to look at, etc.

On the other hand, some early pieces of empirical evidence suggest that other cost functions might be more appropriate especially in the case of perceptual situations. Woodford (2012) discusses how in an experiment by Shaw & Shaw (1977) subjects make larger errors when a signal appeared in an unusual location, while with entropy, errors would be independent of the likelihood of the location. This experiment highlights violation of optimality of information (A1, A2), or of the optimal coding, since here subjects need to scout the whole screen with their eyes, while an optimal signal might be the signal in the middle of the screen. Similarly, Caplin & Dean (2014) and Dean & Neligh (2019) run a perceptual experiment, where subjects need to assess a number of balls. Distinguishing between 49 and 51 balls is much more difficult than between 10 and 90. Entropy does not apply if the agent needs to count the balls one by one while knowing the total number is one of the two possibilities only. Entropy would apply better if the agent could ask someone knowledgeable "there are more balls of which color?" or if the number of balls with different color were written on a screen for the agent to see. Both cases would then be equally difficult to assess, since they allow for the optimal piece of information given the task at hand.

In practice, information is not always fully available, and often we get it from perceptual senses, too, e.g., by judging sizes of houses when we infer how wealthy a neighborhood is. It is not clear yet what cost function is the most appropriate one. Moreover, most likely there are different cost functions appropriate in different choice situations. Caplin & Dean (2015), Matějka & McKay (2015), De Oliveira et al. (2017) and Ellis (2018) provide axiomatic foundations of RI using various revealed preference approaches.[20] Caplin et al. (2017) provide a characterization of a generalized model of RI with state-dependent stochastic choice data. In computer science, it was shown that entropy-cost emerges as an achievable bound in repetitive information processing (Shannon, 1948, Cover & Thomas, 2006). Hébert & Woodford (2019) and Morris & Strack (2019) generalize these findings and relate a micro-founded cost function to the cost of sequential sampling.[21]

Does it matter if one uses entropy or another cost function? It obviously matters for fine details of the decision making. But for many qualitative implications it does not. Often a model relies on a mechanism such as "agents get more information about more important shocks, and thus they respond to them more than to less important shocks," and predictions of this kind will be largely invariant to substituting other reasonable cost functions for entropy.

The cost function can make a difference in a strategic setup. Morris & Yang (2019) find that how costly it is for agents to distinguish between nearby states is crucial for multiplicity of equilibria. Van Nieuwerburgh & Veldkamp (2010) show that whether investors under-diversify or not depends on the form of the cost function they face, i.e., if the cost is sufficiently convex then focusing on more assets can be detrimental. Angeletos & Sastry (2019) show that certain amendments of the Welfare Theorems hold for competitive markets with RI agents if attention costs satisfy an invariance condition, which is satisfied by the entropy-based cost function.

[20]See also a related work of Manzini & Mariotti (2014).

[21]See also Pomatto et al. (2018), Hébert & Woodford (2020), Zhong (2017), and Cerreia-Vioglio et al. (2020).

2.4.2 Specific modeling choices and common issues of RI

An endowment of information-processing capacity vs. a cost function Sims (2003) uses a capacity constraint, i.e., a strict limit on mutual information $I(y;x) \leq \kappa$, while most of the recent literature uses a linear cost in $I(y;x)$, as we do here in (1). In a static setup, these problems are dual to each, and thus the solution for a fixed budget also takes the form of (4), only λ is then the endogenous shadow cost of information. Allowing agents to choose the total amount of information, not just its form, seems more realistic or in line with the idea of endogenous information, and also allows for more tractability, because then λ in the logit formula (4) is exogenous. At the same time, in practice due to fatigue or decreasing marginal returns on time devoted to other activities, the implied appropriate cost can be convex in entropy, or in another measure of information.

For instance, Mondria & Quintana-Domeque (2013) show that investors' limited information capacity can generate financial contagion, which is driven by attention relocated from a healthy market to a distressed one. They use a fixed limit on the amount of investors' information, while with the linear cost the two markets would be independent.

In equilibrium analysis à la Grossman & Stiglitz (1980), do RI agents observe prices? The purest form of RI treats prices just as any other piece of available information. Agents can choose to look at them in more or less detail, or not at all. If market prices move continuously, we do not follow them perfectly every millisecond, even if the numbers are right in front of us on a computer screen.

While some papers study the implication of consumers' inattention to prices (Matějka, 2015, Maćkowiak & Wiederholt, 2015), other papers with financial applications do assume that information contained in equilibrium prices can be acquired and processed for free to highlight that acquiring and processing this kind of information is easier.

Is noise in signals independent across agents? RI models with multiple agents typically assume that signal noise is independent across agents. The assumption of independent noise seems appealing when the main information constraints are cognitive, i.e., driven by mental limitations of agents' minds, and different agents can thus process the same headline differently. A different question is whether agents have an incentive to learn about endogenous outcomes that are orthogonal to the exogenous state. For example, do agents want to learn about other agents' signal noise?

Hellwig & Veldkamp (2009), Afrouzi (2016), Denti (2019), and Hébert & La'O (2020) study this question in detail.

How do inattentive agents satisfy a budget constraint? In existing models with RI, a budget constraint typically holds because a residual variable adjusts to satisfy it.[22] For example, a household imperfectly aware of its income chooses consumption, and the household's saving (the balance in its checking account) adjusts so that the period budget constraint holds. Some substantive predictions may depend on whether one assumes that the residual variable is saving or consumption. Another possibility is to explicitly model insolvency and bankruptcy. In the simplest way, the utility of exceeding the budget constraint can be set to minus infinity, or it can be reflected in the agent's objective by describing more some sophisticated institutional features of insolvency or bankruptcy. The optimal information acquisition strategy $f(y, x)$ then reflects the need to satisfy the constraint in line with the degree of the disutility from exceeding it. Finally, the modeler can impose a hard budget constraint as a restriction on $f(y|x)$. The model (1)-(3) is solved with this constraint in Sims (2006).

Several papers assume that agents obtain independent signals about independent sources of randomness. Is this consistent with the idea of a flexible choice of information? Maćkowiak & Wiederholt (2009) or Van Nieuwerburgh & Veldkamp (2010) use this assumption. It is true that this is a departure from the assumption (A1). For instance, the agent needs to get separate signals on x_1 and x_2, from which she can infer $x_1 + x_2$, but cannot get a signal directly on $x_1 + x_2$ only. This assumption can be plausible in some contexts, it can make equilibrium analysis more tractable, and substantive conclusions may change little if it is dropped (see, for example, Maćkowiak & Wiederholt, 2015, Section 6).

Would one not get the same outcomes if one simply assumed the information structure that RI agents choose in equilibrium? In a non-strategic setup, the answer is yes, but RI agents typically choose signals that are different from the ones assumed in models with an exogenous information structure. Moreover, comparative statics are different, because as the decision problem changes, the optimal allocation of attention changes.

Is RI not a vacuous model that is so flexible it can explain anything? Can it be rejected? Yes, it can be rejected as it provides numerous testable implications that were

[22] The same point applies to other standard constraints.

presented in this section. See also Matějka & McKay (2015), who characterize the model with mutual information, and Caplin & Dean (2015), who show how several features of the model can be tested even for a broader class of cost functions. In principle, the cost function can be identified from sufficiently detailed choice data.

2.4.3 Connections to other approaches

Is RI not the same as information acquisition, which has been around for a long time? We think that RI advances the literature on information acquisition in a similar way to how rational expectations advanced the early literature on dynamic models with a constant marginal propensity to consume.

Most existing models of information choice make very restrictive assumptions. For instance, a common assumption is that agents learn nothing about the realization of the payoff-relevant state and, if they choose to pay a fixed cost, they learn everything. As another example, the island model of Lucas (1973) assumes that the only information that price setters see about current conditions is the price in their specific market (all other information has an infinite cost). Such restrictions can sometimes be convenient because they lead to transparent and tractable models. On the other hand, humans can often in practical situations get information of more forms than of just one (and choose which), and model outcomes and policy implications in the existing literature may depend critically on the very restrictive assumptions. By modeling information choice as flexible, RI aims to reach more robust conclusions.

The friction that RI formalizes is agents' limited ability to process freely available, easily accessible information (see the opening quote by Herbert A. Simon on page 3). It is not the lack of publicly available information and not the monetary cost of acquiring information. Moreover, the earlier literature on information acquisition places many restrictions on the information that agents can acquire and thereby on the posteriors that agents can achieve, while RI gives agents flexibility through assumption (A1).

For instance, in Diamond (1971) buyers have to search for prices observing them one by one, which then leads to the famous Diamond paradox and monopoly equilibrium prices. RI buyers could instead choose to compare prices of several stores at once, even if imperfectly, which would lead to positive markups that would be increasing in the cost of information (Matějka & McKay, 2012). Both types of the assumptions are somewhat extreme in this case. While rigid assumptions of what form information can take more often lead to various paradoxes, in practice unless shopping on the Internet, prices about all products are certainly not equally and simultaneously available as RI assumes.

Similarly, an investor in the model of Grossman & Stiglitz (1980) can choose to observe an asset return at a cost or not at all, while RI investor could choose a precision of signals but also whether the signal should be on one asset or on performance of some index of interest (Van Nieuwerburgh & Veldkamp, 2009). For some questions of interest the simple cost of observation is a sufficient modeling device, but for some others such as for studying finer heterogeneity across investors, or the co-movement of allocations, it is not. Yang (2020) finds that flexibility of information can lead to very different optimal design of financial securities. The reason is that flexible choice of security with inflexible information (e.g., with observation cost only) generates optimal securities that heavily exploit the limits of information acquisition.

In the sticky information model of Mankiw & Reis (2002), agents get perfect information infrequently and no information in between. This assumption yields a tractable model that can be used to study the dynamics of the macroeconomy (see also Reis, 2006a, and Reis, 2006b). The decision making at the level of an individual, however, is very different from RI. RI agents typically choose to collect information gradually in a dynamic setting, and they differentiate between attention to different variables.

Why connect RI to behavioral economics when RI agents are so ultra rational? We do not necessarily think that realistic human beings solve complicated optimization problems. The assumption of optimality of information subject to an explicit cognitive constraint yields an "as-if" model that lets us study the endogeneity of the behavioral aspects of decision making. In practice, many decisions seem sub-optimal, but the errors can still be subject to choices - one can often work hard to decrease errors, e.g., think harder about the problem at hand, but there is a trade-off, such process is costly. This is a very broad point that makes RI an inevitable step in this field to explore.

In fact, some of the recent models in behavioral economics do not directly work with the notion of imperfect information, but their motivation as well as implications are very similar to those of RI.

In the model of sparsity (Gabaix, 2014), agents choose costly loads of responses to shocks. In the case of quadratic preferences and Gaussian prior uncertainty, sparsity is very similar to RI, because RI also predicts costly linear loads and also the possibility of complete inattention, see equations (12)-(14). The main differences arise when the optimal action depends on multiple shocks - in that case while sparsity assumes a particular base of shocks, RI derives it. See for instance the solutions' feature (F10) in Section 2.2: let $p(1)$ and $p(2)$ be prices of different goods, while in sparsity agents respond to p_1 and p_2, perhaps with different strength, RI agent can choose to respond directly to $(p_1 - p_2)$. See Section 2.6 below. RI is invariant to transformations of variables of the state-space, while sparsity is not. In general, predictions of sparsity and RI can be quite different, because predictions of RI go beyond linear loads, i.e., the logit model and discreteness of actions.

A model of focusing (Kőszegi & Szeidl, 2013) is also closely related to RI. It assumes that agents that assess different alternatives with multiple attributes put more weight on attributes that differ across various alternatives more. The model can explain several behavioral puzzles. This type of behavior can also be microfounded by RI. See the implication (F6) in Section 2.2, i.e., that higher volatility draws attention. RI would imply the model of focusing in ex-ante manner, since expected differences matter for allocation of attention. If RI agents expect that a particular attribute has higher dispersion, then they choose to pay more attention to it, which results in the higher load and thus also a higher weight on it.

Of course, there are many findings in behavioral economics that RI cannot replicate, and even attention is often driven by forces that are better understood outside of a rational model.

2.5 Dynamic model

Sims (2003) studied dynamic RI problems. RI generates inertia in actions, but the information choice of RI agents with memory is also forward-looking.

We present results on dynamic RI problems by starting from a simple example and then discussing generalizations. We build on the quadratic-Gaussian (QG) example of Section 2.3.2. The agent's flow utility in every period $t = 1, 2, \ldots$ depends on the action y_t and the state x_t, $U(y_t, x_t) = -r(x_t - y_t)^2$. The payoff-relevant state, x_t, follows a Gaussian stochastic process, e.g., a Gaussian AR(p) process. The agent's information set in period t consists of the new signal received in period t and the previous period's information set due to memory, $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{s_t\}$.

Let $x^t = (x_{1-p}, \ldots, x_0, x_1, \ldots, x_t)$ denote the history of payoff-relevant states including initial conditions, and let $s^t = (s_1, \ldots, s_t)$ denote the history of signals. The agent chooses an information strategy – a distribution of s_t given x^t – and an action strategy – an action y_t for given $\mathcal{I}_t$. The optimal information strategy maximizes the expected discounted sum of payoffs less the cost of information, $C(f)$:

$$E\left[-\sum_{t=1}^{\infty} \beta^t r \left(x_t - y_t\right)^2\right] - C(f)$$

subject to the law of motion for the state

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \nu_t,$$

with $\nu_t \sim i.i.d.N\left(0, \sigma_\nu^2\right)$, the definition of the information set

$$\mathcal{I}_t = \mathcal{I}_0 \cup \{s_1, \ldots, s_t\},$$

and the optimal action strategy

$$y_t = E\left[x_t | \mathcal{I}_t\right].$$

Popular specifications of the information cost in the dynamic setting are $C(f) = \lambda lim_{T\to\infty} \frac{1}{T} I\left(x^T; s^T\right)$ (e.g., Section 4 of Sims, 2003) and $C(f) = \lambda \sum_{t=1}^{\infty} \beta^t I\left(x^t; s_t | \mathcal{I}_{t-1}\right)$ (e.g., Sims, 2010) where $\lambda > 0$ is the information cost parameter.

Maćkowiak & Wiederholt (2009) assume that (x_t, s_t) follows a stationary Gaussian process. They also assume that, after the agent has chosen the information strategy in period $t = 0$, the agent receives a long sequence of signals in period zero such that the conditional second moments of the state vector $(x_t, \ldots, x_{t-p+1})$ given $\mathcal{I}_t$ are independent of time. Under these two assumptions, several different formulations of the dynamic RI problem are equivalent. The two information cost functions mentioned in the previous paragraph coincide (Maćkowiak et al., 2018) and the decision problem stated here is identical to the decision problem in Section 4 of Sims (2003).

If the state follows an AR(1) process ($p = 1$) the solution is very simple. The optimal signal is $s_t = x_t + \epsilon_t$ with $\epsilon_t \sim i.i.d.N\left(0, \sigma_\epsilon^2\right)$ (Maćkowiak & Wiederholt, 2009). That is, the RI agent behaves as if he or she observes the *current* payoff-relevant state with i.i.d. noise and takes the action $E\left[x_t | \mathcal{I}_t\right] = K s_t + (1 - K) E\left[x_t | \mathcal{I}_{t-1}\right]$ where K is the Kalman gain. Delay in actions arises due to the weight on the prior, which is due to the noise in the signal. The variance of the noise in

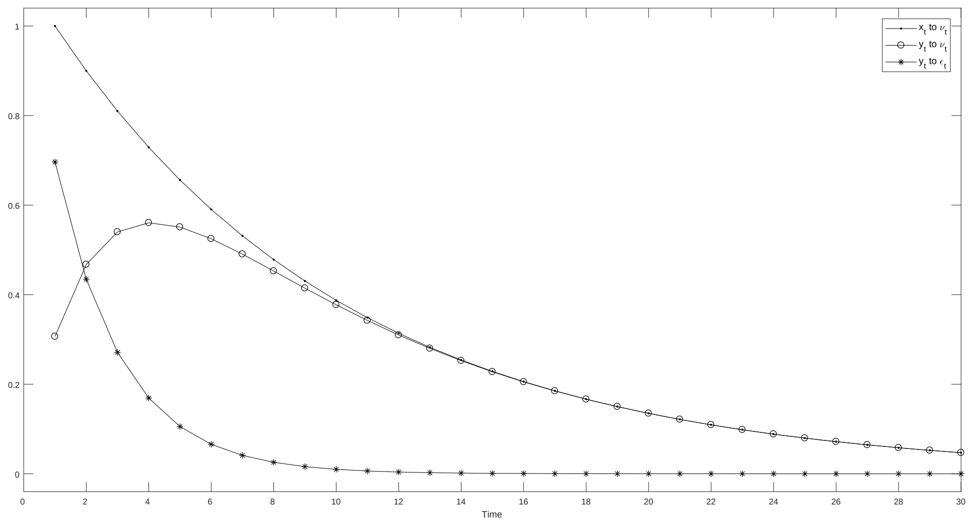

Figure 4: Dynamic model, AR(1) example

the signal, σ_ϵ^2, and the resulting Kalman gain, K, depend on the optimal level of attention, which in turn depends on the parameters $r, \phi_1, \sigma_\nu^2, \lambda$.

Figure 4 shows a typical example in the AR(1) case. The impulse response of the action to an innovation in the state is hump-shaped. The impulse response to noise follows an AR(1) process. As in the static model, the action is dampened – its response to a change in the state is weaker than under perfect information. In the dynamic model the action is also *delayed* – its strongest response occurs later than with perfect information. Another feature of the dynamic model is that mistakes are *persistent* (the impulse response of the error, $x_t - y_t$, to ν_t follows an AR(1) process). Stakes and prior uncertainty determine how much dampening there is, as in the static model, but also the extent of the delay and the autocorrelation of the error. The insight that stakes and prior uncertainty affect attention and thereby the persistence of mistakes has played a prominent role in the literature on how people form expectations (see Section 4.2).

If the state follows an AR(p) process, the optimal signal is a *one-dimensional* signal on the *state vector* $(x_t, \ldots, x_{t-p+1})$ and typically the signal weights on all elements of the state vector are non-zero (Maćkowiak et al., 2018). This result illustrates some general features of rational inattention. Agents focus on important variables–here state variables rather than higher lags of x_t. Agents engage in dimensionality reduction–the optimal signal is one-dimensional even though the state vector in period t is p-dimensional. Surprisingly, information choice is also forward-looking. The only payoff-relevant state in period t is x_t, but the optimal signal typically also has non-zero weights on $x_{t-1}, \ldots, x_{t-p+1}$ because those elements of the state vector help to predict future payoff-relevant states.[23]

Afrouzi & Yang (2020) relax the assumption in Maćkowiak & Wiederholt (2009) and Maćkowiak et al. (2018) that agents receive a long sequence of signals after having chosen the information strategy in period zero. The authors provide very fast code to compute the transitional dynamics in conditional second moments and the limiting steady state.

In the example covered above, there are no state variables that are affected by the agent's *own* past choices ("endogenous state variables"). Section 5 of Sims (2003), Sims (2010), and Miao et al. (2020) study dynamic RI problems with endogenous state variables, e.g., consumption-saving problems. One of the recurring themes is that with high attention cost or low stakes the optimal signal is a low-dimensional signal on the state vector with noise.

For a general non-quadratic setup, Steiner et al. (2017) show that the logit behavior also emerges in the dynamic case.

$$f(y_t|x^t, y^{t-1}) = \frac{e^{U(y_t,x^t)/\lambda + \alpha(y_t|y^{t-1})}}{\int_z e^{U(z,x^t)/\lambda + \alpha(z|y^{t-1})} dz}, \tag{15}$$

where t in the superscripts denotes the whole history until period t. The difference from the static version is that the biases $\alpha(y_t|y^{t-1})$ depend on how likely y_t is conditional on the history of actions taken until the current period. But again, each action history is associated with only one posterior belief, which reduces dimensionality of the problem. The linear entropy-based cost of information is also a benchmark case that abstracts from additional incentives to either smooth information acquisition over time or to bunch it.

In sum,

1. RI generates delay in expectations and actions, but information choice is also forward-looking,

2. the optimal signal is typically a low-dimensional signal on the state vector, and

3. the logit behavior also emerges in the dynamic setting.

[23] Maćkowiak et al. (2018) also study the case where the state follows an MA(q) process or an ARMA(p,q) process. Jurado (2020) solves for the optimal signal weights in closed form in the AR(2), MA(1), and ARMA(1,1) case.

2.6 Multi-dimensional model

To illustrate static RI with multi-dimensional state or action, we extend the example of Section 2.3.2.

Consider a manager who has to set a single price y. The target price is now a function of two normally distributed random variables x_1 and x_2:

$$U(y,x) = -r\Big(a_1x_1 + a_2x_2 - y\Big)^2,$$

where $a_1x_1 + a_2x_2$ is the target price and the two-dimensional state $x = (x_1, x_2)'$ has a multivariate normal distribution with a diagonal variance-covariance matrix.

The model again takes the form of (1)-(3). If there is no restriction on the available signals, the optimal signal has the form $s = a_1x_1 + a_2x_2 + \epsilon$ with a Gaussian ϵ. The target price $a_1x_1 + a_2x_2$ is the only payoff-relevant aspect of the state. If we use a transformation of the state-space such that $\tilde{x} = a_1x_1 + a_2x_2$, then the problem takes the form of the one-dimensional example in Section 2.3.2

However, some authors have argued that in the multi-dimensional case, there may exist additional restrictions on the set of available signals. For example, in Maćkowiak & Wiederholt (2009), x_1 are idiosyncratic conditions and x_2 are aggregate conditions, and it is assumed that the signal s can be partitioned into one subvector s_1 that only contains information about idiosyncratic conditions and another subvector s_2 that only contains information about aggregate conditions. That is, paying attention to idiosyncratic conditions and paying attention to aggregate conditions are independent activities. In this case, the optimal signals have the form $s_1 = x_1 + \epsilon_1$ and $s_2 = x_2 + \epsilon_2$ where ϵ_1 and ϵ_2 are mutually independent Gaussian noise terms. The action under RI is

$$y = E[a_1x_1 + a_2x_2|s] = a_1\xi_1(x_1 + \epsilon_1) + a_2\xi_2(x_2 + \epsilon_2),$$

with

$$\xi_1 = max\Big(0, 1 - \frac{\lambda}{2ra_1^2\sigma_{x_1}^2}\Big), \xi_2 = max\Big(0, 1 - \frac{\lambda}{2ra_2^2\sigma_{x_2}^2}\Big).$$

The agent is paying more attention to idiosyncratic conditions than to aggregate conditions, if idiosyncratic conditions are more important or more volatile, $a_1^2\sigma_{x_1}^2 > a_2^2\sigma_{x_2}^2$. Prices then respond strongly to idiosyncratic shocks and weakly to aggregate shocks. Hence, the model generates micro flexibility and aggregate stickiness of prices through the optimal allocation of attention. It turns out that results are similar when firms can pay attention to variables such as quantity sold, wage bill, and total factor productivity, because none of these variables is the optimal linear combination of idiosyncratic and aggregate conditions, and price setters choose to pay more attention to those variables that are more driven by idiosyncratic conditions.

Finally, let us explore how the RI agent behaves if she has to set prices y_1, y_2 of two products facing two random shocks x_1, x_2 (see Kőszegi & Matějka, 2020). Let the profit take a form:

$$U(y,x) = -\Big(x_1 - y_1\Big)^2 - \Big(x_2 - y_2\Big)^2 - by_1y_2, \tag{16}$$

where the term $-by_1y_2$ summarizes interactions between the two products. If $b > 0$ then the products are strategic substitutes, for $b < 0$ they are complements.

What is the optimal flexible information strategy here? Clearly, to set two prices the agent might optimally want to get (at least) two signals. In this case, the transformation $x^+ = x_1 + x_2, x^- = x_1 - x_2$ converts the objective (16) to another with two quadratic terms as losses from misperceptions of x^+ and of x^-, but without an interaction term. The optimal price y_1 under perfect information is

$$y_1 = \beta_1 x^- + \beta_2 x^+.$$

The responsiveness to x^-, $|\beta_1|$, is greater than to x^+, $|\beta_2|$, if and only if the products are substitutes.[24]

The transformation was chosen in such a way that the interaction term in the objective disappears, and thus information acquisitions about x^- and x^+ are independent. The two-dimensional problem thus simplifies to two independent uni-dimensional choices (information on x^- to determine the difference $y_1 - y_2$, and information on x^+ to determine $y_1 + y_2$). From now on, we can use the same techniques as in the one-dimensional Example 2.3.2. Under RI, we thus get:

$$E[y_1|x_1, x_2] = \beta_1\xi(\beta_1)x^- + \beta_2\xi(\beta_2)x^+,$$

where $\xi(\beta_i)$ are attention weights that are given by the formula (14) as in the one-dimensional case. Figure 5 shows a contour plot of utility when x_1, x_2 are misperceived. It shows that for $b > 0$ the steepest descent of losses is in the direction of $(-1, 1)$, i.e., along a varying x^-, which is thus the most important component to pay attention to.

[24] $\beta_1 = 1/(2 + b)$ and $\beta_2 = 1/(b - 2)$.

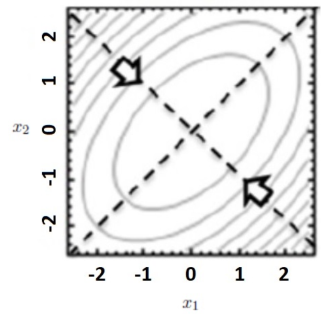

Figure 5: Utility losses from misperceptions of (x_1, x_2) for $b > 0$.

Relative elasticities to different types of shocks, x^- vs x^+, are again magnified, feature (F5) of the static problem. If the cost of information is sufficiently high, then for $b > 0$ agent chooses to get information on x^- only - both prices y_1 and y_2 are then perfectly negatively correlated as both respond to x^- only. The RI agent does not always pay attention to all dimensions of the state-space, but chooses only the most important ones. The magnification thus takes a form of (F10): simplification and indexation. The index is determined by the optimal strategy of attention.

We can always decompose a multi-dimensional problem into simpler independent choices, even in more complicated choice settings than this illustrative one. To do this we can use a method similar to principal component analysis that is called "reversed water-filling".[25] The components of uncertainty about x would correspond to the axes of ellipses given by quadratic losses as in Figure 5, and the RI agent would choose how much attention to pay to each component. However, the chosen components would be endogenous transformations of the underlying shocks, e.g., $x^- = x_1 - x_2$.

3 Applications of the theory of RI by field

This section reviews the rapidly growing set of applications of RI in several fields, from macroeconomics to political economy. Section 4 focuses on the empirical evidence and Section 5 on policy implications.

[25] See Kőszegi & Matějka (2020) for the application to RI.

3.1 Macroeconomics

When Sims (1998) proposed the idea of rational inattention, his motivation was macroeconomics. Sims considered a conventional dynamic stochastic general equilibrium (DSGE) model with various forms of slow adjustment of nominal and real variables. He concluded that *multiple* sources of slow adjustment were necessary for the model to match the inertia in macroeconomic data.[26] Sims conjectured that the inertia in the data could instead be understood as the result of a *single* new source of slow adjustment, rational inattention. Sims's hypothesis has defined a research agenda.

Much of the subsequent work focuses on how firms set prices, a key question in macroeconomics. This research is motivated by both microeconomic data on prices and the behavior of the aggregate price level that Sims (1998) emphasized, e.g., the fact that at the micro level prices change frequently and by large amounts while the aggregate price level responds slowly to shocks (Maćkowiak & Wiederholt, 2009), the fact that price distributions are discrete and prices tend to fluctuate between a regular price and a sales price (Matějka, 2016), and the fact that pricing policies appear to be coarse and updated infrequently (Stevens, 2020).

This literature begins with Woodford (2003). He assumes that price setters observe nominal aggregate demand with idiosyncratic noise, interprets the noise as resulting from limited attention, and shows that nominal shocks have strong and persistent real effects. The idea is that macroeconomic data is publicly available with little delay, but most agents presumably have little incentive to track it carefully; as a result, prices respond slowly to nominal shocks (interest rate shocks or money supply shocks) and nominal shocks have real effects.

Maćkowiak & Wiederholt (2009) study price setting under rational inattention, subject to the constraint that paying attention to aggregate conditions and paying attention to idiosyncratic conditions are separate activities.[27] To match the large average absolute size of price changes in the micro data, idiosyncratic volatility in the model has to be an order of magnitude larger than aggregate volatility. Firms then allocate almost all attention to idiosyncratic conditions.[28] Hence, prices react strongly and quickly to idiosyncratic shocks, but only weakly and slowly to aggregate

[26]Later, Christiano et al. (2005), Smets & Wouters (2007), and many others have confirmed Sims's finding in more formal analysis.

[27]This constraint is relaxed later in that paper.

[28]In addition, feedback effects arise: An individual firm finds it optimal to allocate little attention to the aggregate economy in part because other firms do the same.

shocks. Thus, the aggregate price level responds slowly to shocks. The model can match the empirical finding that at the micro level prices change frequently and by large amounts but the aggregate price level responds slowly to shocks.[29] This paper exploits (F4)-(F6).

Matějka (2016) studies price setting under rational inattention without approximating the profit function.[30] As a result, firms find it desirable to choose prices from a finite set – even though shocks are distributed continuously. Discrete price adjustment is optimal despite the absence of any physical cost of price adjustment. In addition, prices endogenously fluctuate between two levels – a regular price and a sales price.[31] This paper exploits (F7).

Stevens (2020) presents evidence that firms use coarse pricing policies that are updated infrequently and consist of a small menu of prices. She builds a model that can endogenously generate such pricing policies. In her model there is a fixed cost of reviewing a pricing policy and the choice of a price within a pricing policy is made under RI.[32] This paper also exploits (F7).

Other recent papers on price setting under rational inattention include Pasten & Schoenle (2016), Afrouzi (2016), Afrouzi & Yang (2020), and Turén (2020). For example, Afrouzi & Yang (2020) argue that rational inattention can explain the flattening of the Phillips curve in recent decades.

Several papers argue that RI can help understand consumption data. Luo (2008) studies a permanent income model with quadratic utility. He assumes that agents observe permanent income with i.i.d. noise. One can show that this signal is optimal under RI, because permanent income is the only state variable. His closed-form solution helps understand how RI affects the impulse responses of consumption to income shocks. He shows that RI can be a potential explanation for two empirical puzzles: the excess smoothness puzzle and the excess sensitivity puzzle.[33] Sims (2006)

[29] In a representative study, Klenow & Kryvtsov (2008) report that half of all non-housing consumer prices collected by the Bureau of Labor Statistics in order to calculate the consumer price index last less than 3.7 months and, conditional on the occurrence of a price change, the average absolute size of the price change is about 10 percent. See also Bils & Klenow (2004) and Nakamura & Steinsson (2008).

[30] For tractability, Maćkowiak & Wiederholt (2009) work with a quadratic approximation to the firms' profit function. With Gaussian shocks, the distribution of prices under rational inattention is then also Gaussian.

[31] In complementary work, Matějka (2015) demonstrates that a perfectly informed firm moves prices discretely if it faces a consumer who is subject to rational inattention.

[32] Her model builds on Woodford (2009).

[33] Luo (2008) also compares RI to habit formation and to signal extraction with exogenous variance of noise. Subsequent work explores interactions of RI with robustness and recursive preferences, see Luo et al. (2012), Luo &

and Tutino (2013) analyze consumption-saving choices under RI with non-quadratic utility. The main prediction of the latter model is that consumption responses to wealth shocks are asymmetric, with negative shocks producing faster and stronger reaction than positive shocks.

Another set of questions in macroeconomics is: What is the source of the business cycle? How do business cycle shocks propagate? How can policy affect the propagation of shocks?

Maćkowiak & Wiederholt (2015) return to the original conjecture of Sims (1998) that the inertia in aggregate data could be understood as the result of a single source of slow adjustment – rational inattention. The model is close to a New Keynesian model, except that it discards *all* sources of slow adjustment that usually are in New Keynesian models (Calvo pricing, habit formation in consumption, Calvo wage setting), instead featuring rational inattention on the side of firms and households as the only source of slow adjustment. Firms set prices subject to rational inattention. Households make consumption decisions subject to rational inattention. In equilibrium, households pay little attention to the real interest rate, because fluctuations in the real interest rate are modest and small deviations from the consumption Euler equation are inexpensive in utility terms. The model matches the impulse responses to a monetary policy shock and to a technology shock from a standard vector autoregression, confirming Sims's (1998) conjecture. This paper exploits (F4)-(F6).

Maćkowiak & Wiederholt (2020) study a Real Business Cycle (RBC) model with rational inattention. The introduction of rational inattention on the firm side addresses two weaknesses of the baseline RBC model: It raises the persistence of employment, investment and output growth and it creates co-movement after news shocks.

Insights about policy emerge (Paciello & Wiederholt, 2014). Attention to some shocks is good. Attention to other shocks is bad. Policy can affect the incentives to pay attention to shocks. In a standard business cycle model, quick price responses to productivity shocks are good, while quick price responses to markup shocks are bad. The central bank can affect price setters' incentives to pay attention to shocks through its interest rate policy. This is an implication of (F5)-(F6). At the optimal policy, the central bank discourages firms from paying attention to markup shocks.

Ilut & Valchev (2020) study a general equilibrium incomplete markets model in which agents use costly reasoning effort to update their perception of the optimal policy function. The model has empirically desirable properties: the marginal propensity to consume is higher, hand-to-mouth

Young (2016) and Luo et al. (2017).

status is more frequent and persistent, and there is more wealth inequality than in the standard model. Other work on business cycles under RI includes Zorn (2020), who studies investment under RI, Ellison & Macaulay (2019), who explain how RI can lead to unemployment traps, and Kamdar (2019), who argues that RI can rationalize the co-movement of expectations of real activity and expectations of inflation in survey data.

RI can be tested using survey data on expectations. RI passes this test, while the full information rational expectations (FIRE) model fails. Coibion and Gorodnichenko (2012, 2015) document that in the data expectations deviate systematically from full information rational expectations. The average forecast across agents of various macroeconomic variables underreacts to shocks to the economy. If a shock raises inflation for some time, the average inflation forecast of agents increases by less than actual future inflation. Relatedly, the ex-post average forecast error is predictable with the ex-ante average forecast revision. If inflation is rising and forecasts are being revised up, the subsequent average forecast error tends to be positive. Rational inattention implies exactly the systematic deviations from full information rational expectations found in the data.

Several conceptual issues arise when rational inattention is applied to macroeconomics. Each paper mentioned in this subsection confronts at least some of these issues. Macroeconomic models are dynamic, agents interact, and agents tend to take multiple actions in a world with various shocks. In equilibrium, the optimal attention allocation of one agent depends on the attention allocation of other agents.

3.2 Finance

The theory of RI has been used to address core questions in finance: How come most investors hold so much of their wealth in domestic assets? What explains the lack of diversification more generally? If the answer is information, what explains the lack of information flows? What are possible channels of contagion? Do mutual fund managers provide valuable services for their clients?

In today's globalized financial markets, investors can choose among a wide array of assets and a large amount of information relevant for a portfolio decision arrives continuously. RI makes the plausible assumption that investors cannot keep track of all information necessary for an optimal portfolio choice. Rationally inattentive investors develop strategies for processing information that leave them systematically oblivious to some data. New predictions for portfolios and asset prices arise that offer answers to the aforementioned questions.

Van Nieuwerburgh & Veldkamp (2010) show that RI can rationalize investing in a diversified fund and a concentrated set of assets, an allocation often observed. In the simplest version of their model, there are multiple risky assets with independent returns. Investors can process information on the future value of these assets before forming a portfolio. On the one hand, there is a force towards learning about a single asset, because the more an investor learns about an asset, the more the investor holds of the asset, which makes it even more valuable to learn about the asset. On the other hand, there is a force towards learning about many assets, because the investor would like to hold a diversified portfolio. For certain objective functions, the first effect dominates and the investor concentrates all learning on a single asset – the asset with the highest Sharpe ratio – trades heavily in that asset and holds a diversified portfolio on the side. In Van Nieuwerburgh & Veldkamp (2009) and Mondria & Wu (2010), investors endowed with a small home information advantage specialize in home information and tilt their portfolios to home assets ("home bias"). While information advantages have been used to explain a range of phenomena in finance (see Van Nieuwerburgh & Veldkamp, 2009), RI suggests why information advantages can persist even though information is publicly available.[34]

When investors are rationally inattentive, asset prices can comove excessively, relative to the covariance of their fundamentals. In Mondria (2010), investors choose to track a linear combination of asset payoffs rather than individual asset payoffs, which magnifies the comovement of asset prices, because good news about one asset is partly attributed to the other asset. In Peng & Xiong (2006), investors decide to engage in category learning, i.e., they tend to process more market- and sector-wide information than firm-specific information.

Mondria & Quintana-Domeque (2013) propose a model of financial contagion and test two key predictions of the model. Investors hold shares in two markets with uncorrelated fundamentals. If returns in market A become more volatile, investors divert attention to it from market B. Uncertainty rises and prices drop in market B. Financial contagion arises due to attention reallocation.

Since processing information is costly, mutual fund managers can provide valuable services for their clients. Kacperczyk et al. (2016) model mutual fund managers who can process information about future stock returns. The optimal attention allocation varies with the business cycle. In

[34] See also Luo (2010).

the data, the aggregate risk in stock returns rises in recessions relative to the idiosyncratic risk. Managers thus attend to macroeconomic information more carefully in recessions than in booms. The elevated price of risk in recessions magnifies this effect. Managers choose their portfolios accordingly, focusing on "market timing" in recessions and on "stock picking" in booms. By allocating attention rationally, managers outperform "unskilled investors" more in recessions than in booms. The authors find support for all key predictions of the model in the data on actively managed equity mutual funds.

RI has also been applied in corporate finance to study optimal security design. Results in the previous literature may depend on what form of information outside investors were assumed to be able to acquire. In the RI approach to corporate finance, outside investors can acquire information flexibly. This leads to more general conclusions about which characteristics of an investment project make debt finance or equity finance optimal (Yang, 2020, and Yang & Zeng, 2018).

3.3 Strategic interactions

Economists have been interested in how endogenizing information sets affects outcomes in games such as "beauty contests" and "global games" (see Angeletos & Lian, 2016, for a review). The price setting model in Maćkowiak & Wiederholt (2009) is a beauty contest. The optimal price of a firm is a linear function of an exogenous fundamental and the average price. Prices in the model are strategic complements, and this leads to strategic complementarity in attention choices ("a firm pays more attention if other firms pay more attention"). Hellwig & Veldkamp (2009) consider a more general beauty contest with information choice. While strategic complementarity in actions induces strategic complementarity in information acquisition, strategic substitutability in actions leads to strategic substitutability in information acquisition.[35]

Yang (2015) studies a global game under RI. Two players take a binary action ("invest" or "not invest"), actions are strategic complements, and the players design optimal private signals about the fundamental. When attention is inexpensive, the players can coordinate to achieve "approximate common knowledge," which leads to multiple equilibria.[36] Denti (2019) formulates a general game-theoretic model with rationally inattentive players who choose the type of information as well as the

[35] See also Myatt & Wallace (2011) and Colombo et al. (2014).

[36] See also Morris & Yang (2019).

correlation of their information with the information of others. This is an active research area.[37]

Some authors analyze the interactions between perfectly informed sellers and rationally inattentive buyers. Matějka & McKay (2012) study equilibrium prices in markets with many sellers. In Matějka (2015), a monopolistic seller chooses a discrete pricing strategy when facing customers who are subject to RI. Martin (2017) investigates how firms set prices when customers are rationally inattentive to information about product quality.

Matyskova (2018) and Bloedel & Segal (2018) study persuasion with RI receivers. Senders commit to signals, but receivers acquire only noisy representations of their choice of these signals. Lipnowski et al. (2020) explore what information a well-intentioned principal chooses to send to an RI agent. It turns out that when the state is multi-dimensional, then the principal chooses not to disclose some information.

A few papers study optimal contracts when at least one party's attention is scarce. Lindbeck & Weibull (2020) analyze the problem of a principal who considers delegating an investment decision to an agent with greater expertise who is subject to RI. The design of a contract affects not only the agent's investment decision but also her choice of what information to acquire. Li & Yang (2020) characterize, in a principal-agent model with moral hazard, the optimal design of a wage schedule and a technology to monitor an employee's effort by a rationally inattentive firm. See also Yang (2020) and Yang & Zeng (2018) who study optimal security design under RI.

Ravid (2019) applies RI to bargaining. A central feature of high-stakes bargaining situations, for instance in the context of mergers and acquisitions, is that they involve costly attention. One needs to analyze intricate product details, and read and understand long contracts. In Ravid's model, a seller makes repeated offers to a rationally inattentive buyer. The seller knows the product's quality while the buyer is uncertain about it. The buyer needs to pay attention to the product's quality and the seller's offers. RI causes a delay in trade and generates a significant surplus to the buyer, in contrast to the standard results in the bargaining literature. With RI the buyer's valuation of the good across time becomes stochastic, and the seller loses monopoly power. Ravid et al. (2020) study a static bilateral trade when the buyer can choose to receive any form of a signal on the

[37]Afrouzi (2016) solves a model of price setting under RI where oligopolistic competition incentivizes firms to monitor other firms' beliefs. Kondor & Zawadowski (2019) study investors who decide whether to enter a market, are uncertain about their competitive advantage relative to other investors, and can reduce this uncertainty subject to RI.

object's uncertain value. The cost of information is more general here.[38]

RI has also been applied to organizational economics. A production process in a firm typically consists of a number of complementary tasks, such as engineering, purchasing, manufacturing, and marketing, that must be coordinated. Information within a firm is dispersed, and employees' attention is scarce. Coordination involves instructing an agent responsible for a particular task either to stick to a pre-specified plan, or to respond to new private information and communicate it to co-workers with some precision. Dessein et al. (2016) show that optimal coordination involves concentrating on a limited number of tasks. Scarce attention requires setting priorities within a firm and engenders specialization.

3.4 Behavioral economics

RI is a behavioral model with a special discipline. It describes error-prone behavior, yet the form of mistakes is subject to agents' choice; it is driven by agents' preferences and the stochastic properties of the environment. Informally speaking, this is motivated by the fact that people often cannot avoid mistakes, but they can choose what to think about and to what level of detail, i.e., what type of mistakes to minimize. People are inattentive; psychology and behavioral economics have been very successful in showing that humans' cognitive limitations are important for economic outcomes (Kahneman, 1973, DellaVigna, 2009, Handel & Schwartzstein, 2018). But how do agents deal with their own cognitive limitations when they are aware of them?

RI leads to choice behavior that seems imperfect to an outside observer, which is of great interest to behavioral economics. Moreover, it is important for many policy considerations, since RI describes how the behavioral imperfections change if the choice situation changes. RI does not model the procedural details of decision making, but it does describe what choices would be achievable given certain limits to cognition. It combines the insights of psychology with the successful optimizing framework of economics. RI can therefore be viewed as a model of a dual system, similarly as in Kahneman's (2011) "Thinking Fast and Slow". Fast thinking is based on an application of heuristic thinking that is optimal for a range of situations where agents do not explore the details of the current situation, while slow thinking also involves information acquisition about the current situation. RI models the formation of optimal heuristics.[39]

[38] See also Ravid (2020) and Roesler & Szentes (2017).

[39] Gigerenzer et al. (1999) define heuristics as efficient cognitive processes, conscious or unconscious, that ignore

Following on (F1)-(F10) in Section 2, RI provides a range of implications that may be of interest to behavioral economists. These features include random choice (Matějka & McKay, 2015, Caplin & Dean, 2015, Fosgerau et al., 2020a) that follows heuristics such as categorization, summarization, and mental accounting (Kőszegi & Matějka, 2020, Dewan, 2020, Novak et al., 2020), small sets of considered alternatives (Jung et al., 2019, Caplin et al., 2019), and the endogenous form of inertia in dynamic problems (Sims, 2003, Maćkowiak & Wiederholt, 2009, Steiner et al., 2017). Importantly, the RI behavior can violate revealed preference (Woodford, 2012, Matějka & McKay, 2015) and yet many forms of errors can be studied in terms of different forms of the cost of information, e.g. Caplin et al. (2018). Many of the behavioral features above are described in detail in Section 2.

Behavior that resembles mental accounting emerges because RI agents ignore some aspects of the multi-dimensional choice problems. For instance, consumers do not pay attention to all details of price changes (Kőszegi & Matějka, 2020). If the agent's problem is to choose the consumption levels of many goods with different degrees of substitutability, then she may create mental budgets for more substitutable products (e.g., entertainment). This is because the agent chooses to pay attention to price differences of similar goods only, and not to their absolute prices. On the other hand, when the goods are complements, then the same model generates naive diversification. The consumer then pays attention to price-indices of the group of goods and never to differences between single prices. She then fixes the structure of the basket of products of this group, i.e., proportions of the goods in the basket, and chooses how much of the basket to purchase.

Slight modifications of the basic RI model described can also explain phenomena such as reference-dependence and decoy effects (Woodford, 2012, Woodford, 2014), or overweighing of small probability events as in prospect theory (Steiner & Stewart, 2016). In other literatures, closely related approaches based on optimal allocation of cognitive or energetic resources also explain several puzzles in decision making such as over-representation of extreme events, e.g., Christie & Schrater (2015), Griffiths et al. (2015) and Lieder et al. (2018).

Because RI is optimization based, some phenomena studied in behavioral economics are clearly unrelated to RI. RI ignores complex procedural details of human decision making. On the other hand, humans solve trade-offs of what to focus on that do affect many properties of our decisions. An avenue for future research is not only to explain some behavioral phenomena through the lenses

part of the information to save effort. They argue that heuristics are ecologically rational to the degree that they are adapted to the structure of the environment.

of RI, but also study interaction of RI with them, e.g., RI with self-control problems or social preferences. An example is Pagel (2018) who estimates a portfolio-allocation model and finds that both costly attention and loss-aversion are needed to simultaneously explain consumption and wealth accumulation over the life cycle.

3.5 Labor

Labor economics has traditionally emphasized information frictions, e.g., there is a vast literature on search for suitable matches going all the way back to McCall (1970) and there is a large literature on statistical discrimination going back to Phelps (1972) and Arrow (1973).

Attention choice can exacerbate discrimination based on an observable characteristic such as ethnicity, gender, or history of unemployment. Bartoš et al. (2016) find in a field experiment that scarce attention affects the selection decisions of human resources managers in the labor market and of landlords in the rental housing market. Suppose that a manager or a landlord, who must decide how carefully to read an applicant's resume, know the applicant's ethnicity. In a highly selective cherry-picking market (think of markets for many types of labor), a negatively stereotyped group gets less attention – their resumes are read less carefully – than others. The expected benefit of extra information about a member of that group is low. In a thin lemon-dropping market (think of many rental housing markets), a negatively stereotyped group gets *more* attention. The expected benefit of extra information about them is high because all others are likely to be accepted. In *both* markets, attention choice increases the probability that a member of the negatively stereotyped group will be rejected. This "attention discrimination" reinforces discrimination in the subsequent selection decisions. The authors discuss implications for the persistence of discrimination, the duration of unemployment, returns to human capital accumulation, and policy initiatives such as name-blind resumes or blind auditioning.

Matveenko & Mikhalishchev (2019) use RI to explore the implications of quotas, i.e., regulated shares of applicants from various groups, for the attention discrimination. It turns out that the RI-logit model extends naturally to incorporate quotas in a similar way as it incorporates biases due to asymmetric priors. Matveenko & Mikhalishchev (2019) then show that a suitably selected quota can achieve perfect meritocracy, i.e., that the choice of accepted applicants depends on their individual quality only and not on their group membership, or that it can improve quality of selected applicants or even fully correct for employers' incorrect priors. In a closely related work, Fosgerau et al. (2020b) also study discrimination of job applicants by inattentive employers. This paper also uses the RI-logit model and within it the authors manage to study the effects of statistical discrimination, differential screening costs, prejudice as well as the equilibrium outcomes with applicants' choice of investment in human capital.

The discrete choice model with RI resulting in (5) can be interpreted as a version of a noisy directed-search model. This is because agents choose what exact pieces of information to acquire based on their prior knowledge and also conditional on the true state. Such procedure is equivalent to sequential acquisition of very small pieces of information, which is studied in information theory. RI is thus naturally applicable to situations where an agent uncertain about the characteristics of others seeks to form a match such as a job, a rental housing contract, or marriage. The difference from traditional search models is that the RI agent can choose information in a much finer way, does not eliminate all uncertainty about an employer or an applicant at once and does so more selectively. Traditional search models apply better when there is a fixed cost of visiting the employer or arranging a meeting with applicants, for instance, when the agent cannot choose to pay a portion of this cost only to receive partial information of her choice. On the other hand, RI seems to describe well situations when the main costs are those of conducting the interview or exploration on the Internet, when selective and partial information is at hand.

Cheremukhin et al. (2020) consider a model in which agents are heterogeneous in their type and simultaneously search for a match. Individuals know the distribution of types in the market, but there is noise – an agent cannot locate a potential partner with certainty. Each individual chooses a probability distribution for meeting different types, where paying more attention increases the probability of a match whose expected benefit is higher. The authors find that the model can rationalize the observed aggregate matching rates in the U.S. marriage market based on education, race, income, and age.

RI can help understand the dynamics of labor market variables such as unemployment in the data. Acharya & Wee (2020) introduce RI to a Diamond-Mortensen-Pissarides search and matching model. Following a random match with a job seeker, a firm makes a hiring decision under RI. Firms subject to RI become more selective in hiring in a recession, even if the average quality of job seekers rises. The model can explain the persistent fall in matching efficiency and the similarly long-lasting rise in unemployment in the aftermath of the Great Recession.[40]

3.6 Trade, migration, and development

Dasgupta & Mondria (2018) study an RI version of a multi-country, Ricardian model of trade. Importers choose from which country to purchase a product, and they are only imperfectly aware of prices of all potential trade partners. With perfect information, importers from country A purchase more from country B if prices in country B, inclusive of trade costs, decline. An additional effect arises under RI. Importers from country A respond by paying more attention to country B after a price decrease there, thereby boosting the volume of trade by more than in the perfect information work-horse trade model of Eaton & Kortum (2002). Attention choice magnifies the effect of trade costs or prices on trade flows. Even though Dasgupta & Mondria (2018) do not assume idiosyncratic taste shocks, the implied demand of RI firms has a logit-based form similar to that in Eaton & Kortum (2002) because of equation (5).

Bertoli et al. (2020) develop a model of migration with rational inattention. They obtain a closed-form expression for the migration gravity equation. Porcher (2020) estimates a quantitative, dynamic model of migration with RI and local information sharing, using data on internal migration flows in Brazil. Both papers exploit the fact that under the assumptions on the prior g introduced in Dasgupta & Mondria (2018) and Brown & Jeon (2020) there is a closed-form solution for the $\alpha\,(i)$ in equation (5). In this literature on migration $P(i|x)$ is the probability of moving to location i in state of the world x.[41]

Naeher (2020) argues that there can exist attentional barriers to technology adoption. The main idea is that modern technologies often require users to make complex choices about the parameters of usage. When attention is costly, agents will make mistakes in usage. The anticipation of those mistakes can lead to rational non-adoption of technologies that are in principle profitable.

3.7 Political economy

Most voters know little about what policies politicians propose and implement (Carpini & Keeter, 1996), even though this information is publicly available. At the same time, *some* voters are highly

[40]See also Maccuish (2019) who develops a life-cycle model with RI, with a focus on the retirement decision and an application to UK data.

[41]See also Jiang et al. (2020) who use RI to study to a route choice problem that travelers face.

attentive to *some* political issues. How does the selective ignorance of voters affect policy choices and platforms by politicians? How does it affect voters' beliefs?

Matějka & Tabellini (2016) show that RI implies that systematic distortions arise in the democratic process. Due to the selective attention of voters, politicians are motivated to run on platforms that are Pareto inefficient. RI empowers voters with strong preferences as well as small groups, even if they fail to coordinate or engage in lobbying. The reason is that an issue gets more attention from a voter if he or she has higher stakes in it. Voters with higher stakes pay more attention and thus react more strongly to a pleasing proposal from a politician. In response, politicians focus on controversial issues, target policies to specific groups (for instance through tax credits or transfers), and underfund public goods that benefit everyone a little.[42] Information acquisition by heterogeneous voters can also drive polarization of politicians' platforms, see Yuksel (2018) and Hu & Li (2018). The Internet can strengthen these phenomena by providing a finer granularity of information, allowing voters to focus even more on narrow topics of their particular interest.[43]

Nimark & Sundaresan (2019) and Novak et al. (2020) show how the endogeneity of information acquisition can generate polarization of beliefs in the society. Finally, Novak et al. (2020) show that if voters choose between two politicians, then the voters tend to categorize a continuous state space of policies in such a coarse way that guides them only to a binary decision which of the politicians is more favorable to them. Posterior beliefs about the underlying policies of different voters can diverge even in expectation.

4 Empirical relevance

When Sims (2003) wrote, even the basic notion that scarcity of attention matters for economic outcomes was controversial. This seems well-established by now (see Handel & Schwartzstein, 2018, for a review). Consequently, this section focuses on the empirical work on *optimal* inattention and its implications for beliefs and actions. There is a large empirical literature on behavioral economics, and on behavioral inattention in particular, that we do not cover (see DellaVigna, 2009, and Gabaix, 2019, for surveys).

[42] Similar mechanism and outcomes arise when voters choose media sources that selectively design information that they provide (Perego & Yuksel, 2018).

[43] See also Martinelli (2006) who studies how much information a rational voter acquires given a general cost function.

Rational inattention seems to be ripe for an empirical boom. The empirical potential of the theory is due to at least three reasons:

1. *RI implies a range of testable predictions.* Attention choices, and hence beliefs and actions, vary systematically with payoffs and stochastic properties of the environment.

2. *New technologies make measurement of attention possible* (for instance, Google Analytics and web presence patterns).

3. *The predicted behavior follows a generalized logit*, the model already popular in the vast empirical literature on discrete choice.

In Section 2.2 we describe ten basic features of behavior under RI. The next three subsections, summarizing research that tests the theory of RI with data on actions, beliefs and attention, draw mostly on features (F1), (F2), (F5), and (F6). These features can also emerge in other models of costly information acquisition that allow for a somewhat flexible choice of attention, e.g., they can emerge in models where the agent can choose to acquire signals of different precision on different shocks. The fourth subsection is devoted to laboratory experiments, which allow for finer testing of the models. Papers discussed there explore also (F3), (F4), (F7), (F9), and (F10) that require the high flexibility of RI. The subsequent subsection focuses on the identification of preferences and beliefs in discrete choice RI models.

4.1 Explanations of puzzling actions

One strategy for testing RI is to use data on economic actions. In RI models economic conditions affect the incentives to pay attention, and thereby the speed of response of actions to shocks as well as the covariance of actions with shocks. One can test these predictions. Many papers discussed in Section 3 compare predictions of RI models to data on actions. Here we focus on papers that formally test predictions.

In Maćkowiak & Wiederholt (2009) firms pay more attention to idiosyncratic conditions than to aggregate conditions when the former are more volatile than the latter. Prices therefore respond faster to idiosyncratic shocks than to aggregate shocks. In the data, sectoral conditions are more volatile than aggregate conditions. Maćkowiak et al. (2009) use U.S. sectoral price data to test the model. It turns out that prices respond faster to sectoral shocks than to aggregate shocks.

Moreover, the speed of response of prices to sectoral shocks rises with the volatility of sectoral conditions. Finally, the speed of response of prices to aggregate shocks falls with the volatility of sectoral conditions.[44]

Kacperczyk et al. (2016) test their model of attention and portfolio choice by exploiting time-series rather than cross-sectional variation in the incentives to pay attention. The aggregate risk in stocks rises in recessions, and therefore the model predicts more attention to the macroeconomy in recessions than in booms. The authors study the universe of actively managed U.S. equity mutual funds. The covariance of the funds' portfolios with an aggregate payoff shock rises in recessions. The covariance of the funds' portfolios with asset-specific payoff shocks increases in booms. Other predictions of the model also find support in the data.[45]

Taubinsky & Rees-Jones (2018) document in an online shopping experiment that the average under-reaction to not-fully-salient sales taxes falls when the subjects are informed that the sales tax is equal to triple the standard value. They also find significant individual differences in the extent of under-reaction (which can be thought of as arising due to the noise in signals), and accounting for this heterogeneity increases the efficiency cost of taxation, because a product is no longer allocated to the consumer who benefits from it the most.[46]

The authors above use econometric tests, and link them to the theoretical features (F1), (F2), (F5), and (F6). There is also a much larger literature that uses RI to explain puzzling actions less formally. Most of these findings are discussed in Section 3, and those papers also address empirical relevance of the features that depend on the finer choice of attention, i.e., (F3), (F7), and also of (F10).

4.2 Predictions regarding beliefs

Another strategy for evaluating RI models is to look at data on beliefs. Economists have more and more survey data on expectations and beliefs, and they can infer expectations from market data. This line of work is especially promising for macroeconomics. The success or failure of key macro theories may depend on how expectations are modeled (see Coibion et al., 2018a, on the Phillips curve).

[44] In related work, Zorn (2020) studies investment choices and Gondhi (2020) labor input and investment choices.

[45] See also Kacperczyk et al. (2014) and Abis (2017).

[46] See also Morrison & Taubinsky (2020).

Expectations in the data deviate systematically from the benchmark of perfect information rational expectations. Expectations change slowly. The average forecast across agents of a macroeconomic variable under-reacts to news. If a shock raises inflation, for instance, the average forecast of inflation increases by less than actual inflation (Coibion & Gorodnichenko, 2012). The average forecast error in a cross-section of agents is of the same sign as the ex-ante revision in the average forecast. If inflation is rising and forecasts are being revised up, the average forecast error is likely to be positive – on average, agents tend to under-predict inflation (Coibion & Gorodnichenko, 2015). RI implies exactly this pattern of systematic deviations from perfect information rational expectations. Furthermore, the systematic deviations in the data are sizable, consistent with moderate attention to the macroeconomy by most agents.[47]

The degree of sluggishness in expectations seems to depend on economic conditions in ways predicted by RI. Expectations adjusted faster in the Great Inflation period of the 1970s and early 1980s than during the subsequent Great Moderation. This finding fits the prediction that the optimal degree of attention to the macroeconomy rises with aggregate volatility. Similarly, expectations are less sluggish in recessions, when there is more uncertainty about the economy, than on average over the business cycle (Coibion & Gorodnichenko, 2015).

While prior evidence focuses on professional forecasters and households, Coibion et al. (2018b) survey agents who actually set prices in the economy, managers in New Zealand firms. Respondents report not only their expectations about the future, but also their beliefs concerning recent economic conditions. The beliefs about publicly available information are widely dispersed, especially about inflation data. The accuracy with which data are perceived appears to be systematically related to incentives. Managers have more accurate views about the macroeconomy when they expect to change prices in the near future, face more competitors, or operate with a steeper profit function. Subjects make smaller errors tracking macro variables that they have reported as being important for their business decisions.

Several papers go a step further by exposing a random subset of respondents in a survey to publicly available information and by measuring subjects' beliefs before and after (e.g., Armantier et al., 2016, Cavallo et al., 2017, Armona et al., 2019, Coibion et al., 2019, Fuster et al., 2019, and Roth & Wohlfart, 2020, use this method to study the formation of household expectations,

[47]See also Andrade & Le Bihan (2013) and Andrade et al. (2016).

while Coibion et al., 2018b, and Coibion et al., 2020, use this method to study the formation of firm expectations). This is a very active area of research. One of the main findings in that literature is that agents incorporate the information into their beliefs, even though the information had been publicly available before the survey, suggesting that agents had been inattentive to this information before the information treatment. Some papers also study the effect of the information treatment on actions. For example, Coibion et al. (2018b) provide managers in their survey with information about recent macroeconomic variables, the value of the central bank's inflation target, and so on. They find that managers who lower their inflation expectations reduce employment and investment relative to what they were planning to do before the information was provided and relative to the control group. Finally, a few papers study whether the extent to which provided information is incorporated into beliefs depends on the cost and benefit of processing information. For example, Fuster et al. (2019) find that low-numeracy individuals incorporate the provided information less into their beliefs. Roth et al. (2020) find that respondents who learn of a higher exposure to unemployment risk during recessions increase their demand for an expert forecast about the likelihood of a recession.

Gaglianone et al. (2019) also provide evidence that the choice of attention results from a cost-benefit optimization. They estimate a model where professional forecasters face a fixed cost of updating their forecasts and an entropy-based cost of processing information. The authors exploit a unique dataset of Brazilian forecasters who update their inflation forecasts at days of their choice and participate in a monthly forecasting contest. They find that on the monthly contest day the fraction of updaters and the updaters' forecast accuracy both rise sharply.

These findings constitute direct evidence on beliefs. They are consistent with the idea that macroeconomic variables tend to move slowly *because* people *choose* to adjust beliefs slowly.

4.3 Direct measurements of attention choices in the field

Recent advances in information technology have made it possible for researchers to observe directly attention choices and how these affect beliefs and actions.

Bartoš et al. (2016) monitor attention choices in a correspondence field experiment (Bertrand & Mullainathan, 2004). The authors respond by email to job ads and apartment rental ads, randomly varying names of fictitious applicants. A job application contains a hyperlink to a resume, while a response to a housing ad includes a hyperlink to the applicant's personal website. The authors observe whether employers and landlords access the resume or the website, the intensity of information acquisition, and subsequent actions (selection decisions) of the subjects. The employers and landlords engage in "attention discrimination" towards a negatively stereotyped group of applicants, in line with predictions of an RI model. See Section 3.5. The subjects fail to follow a fixed rule of "less attention to a negatively stereotyped group." They appear capable of learning what use of information is optimal in their different circumstances. Moreover, their actions depend on the preceding information choices.[48]

Mondria et al. (2010) use data on internet search queries to study the link between attention and the home bias in portfolios of U.S. investors. The idea is to measure the attention allocated to a country by the number of times this country comes up as part of the answer to a search query. Causality runs both ways. Investment in a given country increases attention to it. Strikingly, attention to a given country raises investment in it. Using instrumental variables, the authors find that if all countries received as much attention as the United States, the home bias in the portfolio of a typical U.S. investor would fall significantly. Similarly, De los Santos et al. (2012) employ internet search data to test models of consumer search, Sicherman et al. (2016) investigate financial attention using panel data on investor online account logins, and Gu & Stangebye (2018) identify information costs using Bloomberg news-heat data (using the identification to discipline a model of sovereign default).

Song & Stern (2021) construct a text-based measure of firm attention. They document that firm attention to macroeconomic news is countercyclical. Furthermore, differences in attention across firms lead to asymmetric responses to monetary policy: expansionary monetary shocks raise stock returns of attentive firms more than those of inattentive firms, while contractionary shocks lower returns of attentive firms by less. These findings are consistent with the idea that uncertainty triggers attention and that attention matters for choices.

[48] See also Huang et al. (2020) who study attention discrimination in retail lending. They observe how much time lending officers devote to different types of applicants for loans, and also how the effort as well as the approval rates depend on how scarce time is for the officers on a particular day.

4.4 Experimental data from the lab

Laboratory experiments can be an attractive tool for testing RI. They allow for controlling for preferences, prior beliefs, or even some features of the cost of information, which may not be possible in the field. It is often important to study fine properties of decision making to distinguish RI from other earlier approaches to information acquisition.[49]

Caplin & Dean (2014) and Dean & Neligh (2019) study details of the behavior implied by RI. They characterize the choice behavior of RI using two conditions of revealed preference and introduce an experiment that allows for testing the conditions and stochastic properties of choice. Subjects are presented with a hundred balls of two colors (red and blue) on a screen. A random state is described by the number of blue balls. The balls are displayed and subjects take an incentivized action. An important contribution is that the experiment yields "state-dependent stochastic choice data". As is discussed in Section 2.2, the joint distribution of states and actions is a very useful object in the theory of RI as it allows for inference about attention. They find that the subjects' attention changes with the stakes, in line with the qualitative predictions of RI. However, certain choices of subjects would be better explained by a different cost function than the entropy-based one. One possible reason for this is that in this case the information is presented in a way that makes distinguishing between distant states cheaper than between nearby states (49 vs 51, or 10 vs 90 balls), which is not in line with the assumptions of entropy.

Ambuehl et al. (2020) test their theoretical findings regarding how agents that are heterogeneous in costs of information respond to changes in incentives. They confirm that more inattentive agents respond to a priori incentives more. This is important because while inattention implies lower ex post responsiveness, changing prior information about incentives can have a negative selection effect by motivating agents who make uninformed decisions.

Novak et al. (2020) design an experiment that shows that subjects perceive random states as if they were endogenously pooled together in line with RI. The novelty of their design is that they allow subjects to choose information structures explicitly. There is a safe action and a risky action. There are three possible states, where the risky action delivers a low, medium, and high payoff,

[49] Camerer & Johnson (2004) survey earlier information acquisition measures in the lab, and how they allow us to understand human decision making better. More recently, Gabaix et al. (2006) use a mouse-tracking device to study the adaptation of attention. They found that a cost-benefit model of the endogenous allocation of attention explains data well.

respectively. There are two information strategies to choose from. One strategy identifies the low state quite precisely and does not distinguish much or at all between medium and high. The other strategy focuses more on the high state and does not distinguish much between low and medium. They find, in line with the theory, that the subjects' choice of the information strategy depends on the payoff of the safe action. If it is lower than the medium payoff of the risky action, then subjects tend to choose the first information strategy, i.e., a signal that pools the two higher states together, and vice versa. The authors also discuss implications for polarization of beliefs and status-quo bias.[50]

Finally, a few papers study dynamic decision-making in the lab. Khaw et al. (2017) asked their subjects to estimate the proportion of red vs. green balls in a hidden box, which changes once in a while. The only information the subjects get is a single draw of one ball from the box in each period. The paper confirms that the behavior features state-dependent adjustment implied by RI, which is much more discrete than the range of actions (Matějka, 2016, Jung et al., 2019, Stevens, 2020). On the other hand, some features of the behavior can be better explained by a slightly modified version of RI, which does not allow for continuous adjustment of information strategies (Woodford, 2008, Stevens, 2020). Note that here too the set of possible signals was not unrestricted. Matyskova et al. (2020) show in a lab that the choice of information strategies in RI can explain some features of endogenous formation of habits.

4.5 Inference about unobservable characteristics

An econometrician can use the RI-logit formula of Matějka & McKay (2015) or its generalized versions of Fosgerau et al. (2020a) to infer preferences, or other unobservable characteristics, from choices. The vast applied micro literature following McFadden (1974) accomplishes this task under the assumption of perfect information. In that literature the logit formula reflects unobserved taste heterogeneity. The RI model of discrete choice makes the same inference possible assuming imperfect information of a general form. The challenge is that an econometrician must disentangle preferences $U(i, x)$ from prior beliefs $\alpha(i)$ in (5).[51]

[50] Cheremukhin et al. (2011) analyze existing lab data of repeated binary choices, and conclude that the RI model accounts for the distribution of errors that subjects make better than other random choice models (with fixed noise).

[51] See for instance Carvalho & Silverman (2019) who use this method to study complexity of financial products. Note that $\alpha(i)/\lambda$ equals the log of the ex ante probability of choosing option i.

Caplin et al. (2016) describe one of the possible tractable ways for an econometrician to proceed. Heterogeneous consumers enter a market with a discrete set of available alternatives and imperfect information concerning which option is best for them. They freely observe past market shares (social learning) and can also acquire additional private information about the options, subject to RI. After a finite number of periods, the model converges to a steady state in which the market shares and the ex ante probabilities of choosing the different options coincide: $\alpha(i)/\lambda$ becomes observable. The paper also gives applied examples and analyzes the welfare implications of the model.

Joo (2020) and Brown & Jeon (2020) use a more flexible approach. Joo (2020) assumes that some manipulations change the beliefs, and thus the biases $\alpha(i)$, but not preferences. He uses A.C.Nielsen supermarket data to estimate preferences for laundry detergents. The data includes volumes sold of packages at various prices and of different sizes, but also volumes when the packages were displayed differently, for instance. This variation in the end allows for inference of what portion of quantity surcharge, i.e., puzzling increase of per unit price of large packages, is due to inattention rather than preference for larger packages.[52] Brown & Jeon (2020) also estimate (5), and study how individuals choose among notoriously complicated health insurance plans. They assume that while premiums are observed, and thus their changes affect $\alpha(i)$, the expected out-of-pocket costs are not, and need to be paid attention to. The paper explores endogeneity of attention to various elements of plans in detail and finds that lowering the number of plans that are offered to customers or shifting costs to premiums, which are easier to compare, can increase welfare.

4.6 Summary and challenges

The empirical literature on RI is beginning to grow rapidly. Its first findings typically provide support for the hypothesis of *rational* inattention: Attention choices respond to incentives, and beliefs and actions change in line with the theory. Economic consequences are potentially important. For example, think of discrimination in the labor market, the real effects of nominal disturbances, and portfolio choices.

On the other hand, some papers present findings that are not in line with rational adjustment of

[52] Regarding biases arising from the assumption of perfect information in the dynamic case, see also the discussion of Rust's (1987) dynamic logit model in Steiner et al. (2017).

attention. This may or may not be because the choice situations studied in those works are not in line with the assumptions of RI, as we discuss in Section 2.4.1. Information might not be available in a sufficient variety of forms, the choice environment might be more complex than what it seems, or deviations from rational choice might dominate. After all, our everyday attention is often drawn to news or commercials without us choosing that. Sicherman et al. (2016) investigate financial attention using panel data on daily investor online account logins. They find support for selective attention to portfolio information, but account logins fall after market declines and investors also pay less attention when the VIX volatility index is high. Similarly, Olafsson & Pagel (2017) find that individuals are considerably more likely to log into their accounts when they expect a higher balance. This is likely more consistent with Ostrich effects rather than with rational allocation of attention. In the lab, Khaw et al. (2017) and Dean & Neligh (2019), for instance, find that some other cost functions than entropy-based explain choices better, especially in perception tasks.

We view RI to be a very useful starting point for the modeling of information acquisition when there are no clear limits to what information is available and what not. Yet, it clearly does not apply in all situations and we still do not know enough about its explanatory power in a variety of choice situations.

5 Policy implications

To understand the implications of rational inattention for economic policy, it is useful to start from some basic principles.

1. RI agents make mistakes. The actions of RI agents differ from the actions the same agents would have taken under beliefs that incorporate all publicly available information. One implication is that the equilibrium allocation of an otherwise frictionless economy is not necessarily Pareto efficient.

2. Biases are adaptive. The behavior of RI agents is erroneous, but is not fixed. Changes in prior beliefs about properties of the economy (e.g., policy) change attention and thus can have large effects.

3. Equilibrium versus efficient allocation of attention. The private value of attention may differ from the social value of attention.

4. Cognitive costs enter welfare. It is not only the resulting actions that matter, but cognitive (attention) costs, too, unless attention is an endowment and has no other use.

5. Inattention to policy changes the effects of policy.

6. Incentives to pay attention are heterogeneous in the population.

A few papers formally study optimal policy under RI or policy implications of RI. Most of this work is in macroeconomics. We therefore first illustrate the aforementioned principles with papers from macroeconomics and then turn to papers in other fields.

Standard business cycle models feature multiple sources of slow adjustment to match empirical findings on slow responses of prices, consumption, and investment to shocks. Maćkowiak & Wiederholt (2015) solve a business cycle model where RI by firms and households is the only source of inertia. They show that the conduct of monetary policy can have a large effect on the equilibrium allocation of attention and thus on the response of the economy to shocks (Principle 2).[53]

Paciello & Wiederholt (2014) study optimal monetary policy in a business cycle model with RI on the side of price setters in firms. There are two types of aggregate shocks: shocks that cause *efficient* fluctuations under unlimited attention (e.g., technology shocks), and shocks that cause *inefficient* fluctuations under unlimited attention (e.g., markup shocks). In response to the first type of shock, the central bank would like to replicate the allocation under unlimited attention, but limited attention introduces mistakes (Principle 1) and attention is costly (Principle 4). The optimal monetary policy is to fully stabilize the *profit-maximizing* price in response to the shock, implying that firms do not have to change prices in response to the shock. In response to the second type of shock, the central bank would like to avoid the allocation under unlimited attention. The optimal monetary policy is to reduce the variance of the *profit-maximizing* price in response to the shock such that firms pay no attention to the shock (Principle 2) and thereby the equilibrium allocation of attention equals the socially optimal allocation of attention (Principle 3).

Rational inattention also has several implications for central bank communication. Inattention to central bank communication is probably the simplest explanation for limited power of forward guidance (Principle 5, see Wiederholt, 2015, and Angeletos & Lian, 2018).[54] Furthermore, rational inattention predicts heterogeneous attention across the population; financial market participants

[53]See Paciello (2012) for analytical results in a special case.

[54]See also Gaballo (2016).

are likely to attend to every nuance of central bank communication, while ordinary people will most likely pay little attention to even simple policy announcements (Principle 6, see Sims, 2010).

Bartoš et al. (2016) develop the theory of attention discrimination (see Section 3.5), test it in several field experiments (see Section 4.3) and discuss its policy implications. Attention discrimination exacerbates statistical discrimination and taste-based discrimination. Policies that affect prior beliefs (and thus statistical discrimination) or preferences (and thus taste-based discrimination) can have large effects because they also change attention discrimination (Principle 2). Moreover, to reduce statistical discrimination it does not suffice to simply provide more individual-specific information, because decision makers may choose not to read it. Instead, according to RI, discrimination can in some cases be reduced by using name/race/sex-blind forms or by quotas for minorities in the early stages of interviews. Name-blind forms change beliefs about the applicant, and put minorities and majorities on the same footing regarding the attention paid to them. Quotas for the early stages can also ensure that the same attention is paid to minorities and majorities, while leaving the final choice of who to hire, for instance, completely unconstrained.

Maćkowiak & Wiederholt (2018) use RI theory to study whether private-sector agents and policy makers have enough incentives to prepare for rare events. Under limited liability, the equilibrium preparation for rare events is smaller than the efficient preparation for rare events (Principle 3).[55] This result can be viewed as a justification for the living will mandate in the Dodd-Frank Act, which requires that each systemically important financial institution prepares a "living will," a plan for an orderly resolution of that institution in the event of its failure.

Angeletos & Sastry (2019) study the efficiency of markets when agents are rationally inattentive and there are no other distortions (no market incompleteness, market power, externality in the utility and production function, or nominal rigidity). Prices are assumed to adjust to clear markets. If the attention choice of one agent does not affect the attention costs of other agents (a condition that is satisfied by entropy-based attention costs), the equilibrium is constrained efficient. Relatedly, Hébert & La'O (2020) study how the properties of agents' information cost function relate to the properties of equilibria for a class of large games featuring strategic interaction and endogenous information acquisition. They assume efficiency in the use of information and study efficiency in the acquisition of information. The entropy-based cost function implies that equilibria exhibit zero

[55] See also Colombo et al. (2014).

non-fundamental volatility and are constrained efficient, whereas other cost functions may lead to non-fundamental volatility or inefficiency. An important property of the entropy-based cost function is again that the information choice of one agent does not affect the information cost of another agent.

We believe that one of the important next steps in the RI literature is to study the implications of RI for other policy questions.

6 Conclusion and future steps

Decisions depend on the subjective perception of the state of the world, not the objective state of the world. The theory of rational inattention assumes that agents choose what kind of information to process and that they are fairly unconstrained in doing so. The resulting beliefs feature complete ignorance to some information, inertia, discreteness, and multi-dimensional simplification. In this review, we summarized the theory, existing applications, existing empirical work, and known policy implications.

This is obviously an ongoing research project. The development, the application and the testing of the theory is happening right now and this survey is only a progress report. We would like to emphasize that we find the applications and the empirical work very important. Ultimately a theory gains relevance not because of its elegance but because it helps to match data in a variety of different fields, its empirical support, and its usefulness in addressing policy questions.

There are some big, conceptual challenges that lie ahead. Joint measurement of the state of the world, attention, beliefs, and actions would be ideal. Thinking carefully how one can measure a subset of those variables and how this measurement can be used to test the theory is already an important step. As researchers become more confident that certain phenomena are indeed driven by RI, it will also become more important to think carefully about policy implications.

References

Abis, S. (2017). Man vs. machine: Quantitative and discretionary equity management. *Unpublished working paper. Columbia Business School.*

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

Afrouzi, H. (2016). Strategic inattention, inflation dynamics, and the non-neutrality of money. *Unpublished working paper. Columbia University.*

Afrouzi, H. & Yang, C. (2020). Dynamic rational inattention and the phillips curve. *Unpublished working paper. Columbia University and UT Austin.*

Ambuehl, S., Ockenfels, A., & Stewart, C. (2020). Who opts in? *Unpublished working paper. University of Zurich, University of Cologne, and University of Toronto.*

Andrade, P., Crump, R. K., Eusepi, S., & Moench, E. (2016). Fundamental disagreement. *Journal of Monetary Economics*, 83, 106–128.

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

Angeletos, G.-M. & Lian, C. (2016). Incomplete information in macroeconomics: Accommodating frictions in coordination. In *Handbook of macroeconomics*, volume 2 (pp. 1065–1240). Elsevier.

Angeletos, G.-M. & Lian, C. (2018). Forward guidance without common knowledge. *The American Economic Review*, 108(9), 2477–2512.

Angeletos, G.-M. & Sastry, K. (2019). Inattentive economies. *NBER Working Paper No. 26413.*

Armantier, O., Nelson, S., Topa, G., Van der Klaauw, W., & Zafar, B. (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics*, 98(3), 503–523.

Armona, L., Fuster, A., & Zafar, B. (2019). Home price expectations and behaviour: Evidence from a randomized information experiment. *The Review of Economic Studies*, 86(4), 1371–1410.

Arrow, K. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in Labor Markets*. NJ: Princeton University Press.

Bartoš, V., Bauer, M., Chytilová, J., & Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *The American Economic Review*, 106(6), 1437–1475.

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

Bertrand, M. & Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review*, 94(4), 991–1013.

Bils, M. & Klenow, P. J. (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112(5), 947–985.

Bloedel, A. W. & Segal, I. R. (2018). Persuasion with rational inattention. *Available at SSRN 3164033*.

Brown, Z. Y. & Jeon, J. (2020). Endogenous information and simplifying insurance choice. *Unpublished working paper. University of Michigan and Boston University.*

Camerer, C. F. & Johnson, E. J. (2004). Thinking about attention in games: backward and forward induction. *The Psychology of Economic Decisions*, 2, 111–129.

Caplin, A., Csaba, D., Leahy, J., & Nov, O. (2018). Rational inattention, competitive supply, and psychometrics. *NBER Working Paper No. 25224*.

Caplin, A. & Dean, M. (2014). Revealed preference, rational inattention, and costly information acquisition. *NBER Working Paper No. 19876*.

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

Caplin, A., Dean, M., & Leahy, J. (2017). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *NBER Working Paper No. 23652*.

Caplin, A., Dean, M., & Leahy, J. (2019). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3), 1061–1094.

Caplin, A., Leahy, J., & Matějka, F. (2016). Rational inattention and inference from market share data. *Unpublished working paper. NYU and CERGE-EI.*

Carpini, M. X. D. & Keeter, S. (1996). *What Americans Know about Politics and Why It Matters.* Yale University Press.

Carvalho, L. & Silverman, D. (2019). Complexity and sophistication. *NBER Working Paper No. 26036.*

Cavallo, A., Cruces, G., & Perez-Truglia, R. (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, 9(3), 1–35.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., & Rustichini, A. (2020). Multinomial logit processes and preference discovery: inside and outside the black box. *arXiv e-prints*, (pp. arXiv–2004).

Cheremukhin, A., Popova, A., & Tutino, A. (2011). Experimental evidence on rational inattention. *Federal Reserve Bank of Dallas Working Paper No. 1112.*

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

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1), 1–45.

Christie, S. T. & Schrater, P. (2015). Cognitive cost as dynamic allocation of energetic resources. *Frontiers in Neuroscience*, 9, 289.

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

Coibion, O. & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *The American Economic Review*, 105(8), 2644–78.

Coibion, O., Gorodnichenko, Y., & Kamdar, R. (2018a). The formation of expectations, inflation, and the phillips curve. *Journal of Economic Literature*, 56(4), 1447–91.

Coibion, O., Gorodnichenko, Y., & Kumar, S. (2018b). How do firms form their expectations? new survey evidence. *The American Economic Review*, 108(9), 2671–2713.

Coibion, O., Gorodnichenko, Y., & Ropele, T. (2020). Inflation expectations and firm decisions: New causal evidence. *The Quarterly Journal of Economics*, 135(1), 165–219.

Coibion, O., Gorodnichenko, Y., & Weber, M. (2019). Monetary policy communications and their effects on household inflation expectations. *NBER Working Paper No. 25482*.

Colombo, L., Femminis, G., & Pavan, A. (2014). Information acquisition and welfare. *The Review of Economic Studies*, 81(4), 1438–1483.

Cover, T. M. & Thomas, J. A. (2006). *Elements of Information Theory*. Hoboken, NJ: Wiley.

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

De los Santos, B., Hortaçsu, A., & Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review*, 102(6), 2955–80.

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

Dean, M. & Neligh, N. (2019). Experimental tests of rational inattention. *Unpublished working paper. Columbia University and Chapman University.*

Debreu, G. (1960). Review of individual choice behavior by R. D. Luce. *The American Economic Review*, 50(1), 186–188.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2), 315–72.

Denti, T. (2019). Unrestricted information acquisition. *Unpublished working paper. Cornell University.*

Dessein, W., Galeotti, A., & Santos, T. (2016). Rational inattention and organizational focus. *The American Economic Review*, 106(6), 1522–36.

Dewan, A. (2020). Costly multidimensional information. *Unpublished working paper. NYU.*

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory*, 3(2), 156–168.

Eaton, J. & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5), 1741–1779.

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

Ellison, M. & Macaulay, A. (2019). A rational inattention unemployment trap. *CEPR Discussion Paper No. 13761.*

Fosgerau, M., Melo, E., de Palma, A., & Shum, M. (2020a). Discrete choice and rational inattention: A general equivalence result. *International Economic Review, doi.org/10.1111/iere.12469.*

Fosgerau, M., Sethi, R., & Weibull, J. (2020b). Costly screening and categorical inequality.

Fulton, C. (2017). Mechanics of linear quadratic gaussian rational inattention tracking problems. *Finance and Economics Discussion Series 2017-109, Washington: Board of Governors of the Federal Reserve System.*

Fuster, A., Perez-Truglia, R., Wiederholt, M., & Zafar, B. (2019). Expectations with endogenous information acquisition: An experimental investigation. *NBER Working Paper No. 24767.*

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4), 1661–1710.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Foundations and Applications 1*, volume 2 (pp. 261–343). Elsevier.

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. E. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *The American Economic Review*, 96(4), 1043–1068.

Gaballo, G. (2016). Rational inattention to news: the perils of forward guidance. *American Economic Journal: Macroeconomics*, 8(1), 42–97.

Gaglianone, W., Giacomini, R., Issler, J., & Skreta, V. (2019). Incentive-driven inattention. *CEPR Discussion Paper No. 13619.*

Gigerenzer, G., Todd, P. M., & The ABC Research Group (1999). *Simple Heuristics That Make Us Smart.* Oxford University Press.

Gondhi, N. (2020). Rational inattention, misallocation, and asset prices. *Unpublished working paper. INSEAD.*

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7(2), 217–229.

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

Gu, G. W. & Stangebye, Z. R. (2018). Costly information and heteroskedasticity in sovereign risk spreads. *Unpublished working paper. University of California, Santa Cruz and Notre Dame University.*

Handel, B. & Schwartzstein, J. (2018). Frictions or mental gaps: what's behind the information we (don't) use and when do we care? *Journal of Economic Perspectives*, 32(1), 155–78.

Hébert, B. & Woodford, M. (2020). Neighborhood-based information costs. *Unpublished working paper. Stanford University and Columbia University.*

Hébert, B. M. & La'O, J. (2020). Information acquisition, efficiency, and non-fundamental volatility. *NBER Working Paper No. 26771.*

Hébert, B. M. & Woodford, M. (2019). Rational inattention when decisions take time. *NBER Working Paper No. 26415.*

Hellwig, C. & Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies*, 76(1), 223–251.

Hu, L. & Li, A. (2018). The politics of attention. *arXiv preprint arXiv:1810.11449.*

Huang, B., Li, J., Lin, T.-C., Tai, M., & Zhou, Y. (2020). Attention discrimination in retail lending.

Ilut, C. & Valchev, R. (2020). Economic agents as imperfect problem solvers. *NBER Working Paper No. 27820.*

Jiang, G., Fosgerau, M., & Lo, H. K. (2020). Route choice, travel time variability, and rational inattention. *Transportation Research Part B: Methodological*, 132, 188–207.

Joo, J. (2020). Rational inattention as an empirical framework: Application to the welfare effects of new product introduction. *Unpublished working paper. University of Texas at Dallas.*

Jung, J., Kim, J. H., Matějka, F., & Sims, C. A. (2019). Discrete actions in information-constrained decision problems. *The Review of Economic Studies*, 86(6), 2643–2667.

Jurado, K. (2020). Rational inattention in the frequency domain. *Unpublished working paper. Duke University.*

Kacperczyk, M., Van Nieuwerburgh, S., & Veldkamp, L. (2014). Time-varying fund manager skill. *The Journal of Finance*, 69(4), 1455–1484.

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

Kahneman, D. (1973). *Attention and Effort.* New Jersey: Prentice Hall.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Macmillan.

Kamdar, R. (2019). The inattentive consumer: Sentiment and expectations. *Unpublished working paper. Indiana University Bloomington.*

Kamenica, E. & Gentzkow, M. (2011). Bayesian persuasion. *The American Economic Review*, 101(6), 2590–2615.

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

Klenow, P. J. & Kryvtsov, O. (2008). State-dependent or time-dependent pricing: Does it matter for recent us inflation? *The Quarterly Journal of Economics*, 123(3), 863–904.

Kondor, P. & Zawadowski, A. (2019). Learning in crowded markets. *Journal of Economic Theory*, 184, Article 104936.

Kőszegi, B. & Matějka, F. (2020). Choice simplification: A theory of mental budgeting and naive diversification. *The Quarterly Journal of Economics*, 135(2), 1153–1207.

Kőszegi, B. & Szeidl, A. (2013). A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1), 53–104.

Li, A. & Yang, M. (2020). Optimal incentive contract with endogenous monitoring technology. *Theoretical Economics*, 15(3), 1135–1173.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review*, 125(1), 1–32.

Lindbeck, A. & Weibull, J. (2020). Delegation of investment decisions, and optimal remuneration of agents. *European Economic Review*, Article 103559.

Lipnowski, E., Mathevet, L., & Wei, D. (2020). Attention management. *American Economic Review: Insights*, 2(1), 17–32.

Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review*, 63(3), 326–334.

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

Luo, Y. (2010). Rational inattention, long-run consumption risk, and portfolio choice. *Review of Economic Dynamics*, 13(4), 843–860.

Luo, Y., Nie, J., Wang, G., & Young, E. R. (2017). Rational inattention and the dynamics of consumption and wealth in general equilibrium. *Journal of Economic Theory*, 172, 55–87.

Luo, Y., Nie, J., & Young, E. R. (2012). Robustness, information–processing constraints, and the current account in small open economies. *Journal of International Economics*, 88(1), 104–120.

Luo, Y. & Young, E. R. (2016). Induced uncertainty, market price of risk, and the dynamics of consumption and wealth. *Journal of Economic Theory*, 163, 1–41.

Maccuish, J. H. (2019). Costly attention and retirement puzzles. *Unpublished working paper. University College London.*

MacKay, D. J. (2003). *Information Theory, Inference and Learning Algorithms.* Cambridge University Press.

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

Maćkowiak, B., Moench, E., & Wiederholt, M. (2009). Sectoral price data and models of price setting. *Journal of Monetary Economics*, 56, S78–S99.

Maćkowiak, B. & Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *The American Economic Review*, 99(3), 769–803.

Maćkowiak, B. & Wiederholt, M. (2015). Business cycle dynamics under rational inattention. *The Review of Economic Studies*, 82(4), 1502–1532.

Maćkowiak, B. & Wiederholt, M. (2018). Lack of preparation for rare events. *Journal of Monetary Economics*, 100, 35–47.

Maćkowiak, B. & Wiederholt, M. (2020). Rational inattention and the business cycle effects of productivity and news shocks. *Unpublished working paper. European Central Bank and Sciences Po.*

Mankiw, N. G. & Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117(4), 1295–1328.

Manzini, P. & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3), 1153–1176.

Martin, D. (2017). Strategic pricing with rational inattention to quality. *Games and Economic Behavior*, 104, 131–145.

Martinelli, C. (2006). Would rational voters acquire costly information? *Journal of Economic Theory*, 129(1), 225–251.

Matějka, F. (2015). Rigid pricing and rationally inattentive consumer. *Journal of Economic Theory*, 158, 656–678.

Matějka, F. (2016). Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, 83(3), 1125–1155.

Matějka, F. & McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *The American Economic Review*, 102(3), 24–29.

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

Matějka, F. & Tabellini, G. (2016). Electoral competition with rationally inattentive voters. *Unpublished working paper. CERGE-EI and Bocconi University.*

Matveenko, A. & Mikhalishchev, S. (2019). Attentional role of quota implementation. *CERGE-EI Working Paper No. 645.*

Matyskova, L. (2018). Bayesian persuasion with costly information acquisition. *Unpublished working paper. University of Bonn.*

Matyskova, L., Rogers, B., Steiner, J., & Sun, K.-K. (2020). Habits as adaptations: An experimental study. *Games and Economic Behavior*, 122, 391–406.

McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, 84(1), 113–126.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). New York: Academic Press.

Miao, J., Wu, J., & Young, E. (2020). Multivariate rational inattention. *Unpublished working paper. Boston University, Zhejiang University and University of Virginia.*

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

Mondria, J. & Quintana-Domeque, C. (2013). Financial contagion and attention allocation. *The Economic Journal*, 123(568), 429–454.

Mondria, J. & Wu, T. (2010). The puzzling evolution of the home bias, information processing and financial openness. *Journal of Economic Dynamics and Control*, 34(5), 875–896.

Mondria, J., Wu, T., & Zhang, Y. (2010). The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics*, 82(1), 85–95.

Morris, S. & Strack, P. (2019). The wald problem and the relation of sequential sampling and ex-ante information costs. *Available at SSRN 2991567*.

Morris, S. & Yang, M. (2019). Coordination and continuous stochastic choice. *Available at SSRN 2889861*.

Morrison, W. & Taubinsky, D. (2020). Rules of thumb and attention elasticities: Evidence from under- and overreaction to taxes. *Unpublished working paper. UC Berkeley*.

Myatt, D. P. & Wallace, C. (2011). Endogenous information acquisition in coordination games. *The Review of Economic Studies*, 79(1), 340–374.

Naeher, D. (2020). Technology adoption under costly information processing. *Unpublished working paper. University College Dublin*.

Nakamura, E. & Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics*, 123(4), 1415–1464.

Nimark, K. P. & Sundaresan, S. (2019). Inattention and belief polarization. *Journal of Economic Theory*, 180, 203–228.

Novak, V., Matveenko, A., Ravaioli, S., et al. (2020). The status quo and belief polarization of inattentive agents: Theory and experiment. *Unpublished working paper. University of Copenhagen, CERGE-EI and Columbia University*.

Olafsson, A. & Pagel, M. (2017). The ostrich in us: Selective attention to financial accounts, income, spending, and liquidity. *NBER Working Paper No. 23945*.

Paciello, L. (2012). Monetary policy and price responsiveness to aggregate shocks under rational inattention. *Journal of Money, Credit and Banking*, 44(7), 1375–1399.

Paciello, L. & Wiederholt, M. (2014). Exogenous information, endogenous information, and optimal monetary policy. *The Review of Economic Studies*, 81(1), 356–388.

Pagel, M. (2018). A news-utility theory for inattention and delegation in portfolio choice. *Econometrica*, 86(2), 491–522.

Pasten, E. & Schoenle, R. (2016). Rational inattention, multi-product firms and the neutrality of money. *Journal of Monetary Economics*, 80, 1–16.

Peng, L. & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563–602.

Perego, J. & Yuksel, S. (2018). Media competition and social disagreement. *Unpublished working paper. Columbia University and UC Santa Barbara.*

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

Pomatto, L., Strack, P., & Tamuz, O. (2018). The cost of information. *arXiv preprint arXiv:1812.04211.*

Porcher, C. (2020). Migration with costly information. *Unpublished working paper. Princeton University.*

Ravid, D. (2019). Bargaining with rational inattention. *Available at SSRN 2957890.*

Ravid, D. (2020). Ultimatum bargaining with rational inattention. *The American Economic Review*, 110(9), 2948–63.

Ravid, D., Roesler, A.-K., & Szentes, B. (2020). Learning before trading: on the inefficiency of ignoring free information. *Available at SSRN 3317917.*

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

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

Roesler, A.-K. & Szentes, B. (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review*, 107(7), 2072–80.

Roth, C., Settele, S., & Wohlfart, J. (2020). Risk exposure and acquisition of macroeconomic information. *Unpublished working paper. University of Warwick and University of Copenhagen.*

Roth, C. & Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *Unpublished working paper. University of Warwick and University of Copenhagen.*

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.

Shaw, M. L. & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance*, 3(2), 201–211.

Sicherman, N., Loewenstein, G., Seppi, D. J., & Utkus, S. P. (2016). Financial attention. *The Review of Financial Studies*, 29(4), 863–897.

Simon, H. A. (1971). Designing organizations for an information-rich world.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356.

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

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *The American Economic Review*, 96(2), 158–163.

Sims, C. A. (2010). Rational inattention and monetary economics. In B. M. Friedman & M. Woodford (Eds.), *Handbook of Monetary Economics*, volume 3 (pp. 155–181). Elsevier.

Smets, F. & Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *The American Economic Review*, 97(3), 586–606.

Song, W. & Stern, S. (2021). Firm inattention and the efficacy of monetary policy: A text-based approach. *Unpublished working paper. University of Michigan.*

Steiner, J. & Stewart, C. (2016). Perceiving prospects properly. *The American Economic Review*, 106(7), 1601–31.

Steiner, J., Stewart, C., & Matějka, F. (2017). Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica*, 85(2), 521–553.

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

Taubinsky, D. & Rees-Jones, A. (2018). Attention variation and welfare: Theory and evidence from a tax salience experiment. *The Review of Economic Studies*, 85(4), 2462–2496.

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

Tutino, A. (2013). Rationally inattentive consumption choices. *Review of Economic Dynamics*, 16(3), 421–439.

Van Nieuwerburgh, S. & Veldkamp, L. (2009). Information immobility and the home bias puzzle. *The Journal of Finance*, 64(3), 1187–1215.

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

Veldkamp, L. (2011). *Information Choice in Macroeconomics and Finance.* Princeton University Press.

Wiederholt, M. (2010). Rational inattention. *The New Palgrave Dictionary of Economics (Online Edition).*

Wiederholt, M. (2015). Empirical properties of inflation expectations and the zero lower bound. *Unpublished working paper. Goethe University Frankfurt.*

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In P. Aghion, R. Frydman, E. Phelps, J. Stiglitz, & M. Woodford (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps.* Princeton and Oxford: Princeton University Press.

Woodford, M. (2008). Inattention as a source of randomized discrete adjustment. *Unpublished working paper. Columbia University.*

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

Woodford, M. (2012). Inattentive valuation and reference-dependent choice. *Unpublished working paper. Columbia University.*

Woodford, M. (2014). Stochastic choice: An optimizing neuroeconomic model. *The American Economic Review*, 104(5), 495–500.

Yang, M. (2015). Coordination with flexible information acquisition. *Journal of Economic Theory*, 158, 721–738.

Yang, M. (2020). Optimality of debt under flexible information acquisition. *The Review of Economic Studies*, 87(1), 487–536.

Yang, M. & Zeng, Y. (2018). Financing entrepreneurial production: Security design with flexible information acquisition. *The Review of Financial Studies*, 32(3), 819–863.

Yuksel, S. (2018). Specialized learning and political polarization. *Unpublished working paper. UC Santa Barbara.*

Zhong, W. (2017). Optimal dynamic information acquisition. *Available at SSRN 3024275.*

Zorn, P. (2020). Investment under rational inattention: Evidence from us sectoral data. *Unpublished working paper. LMU Munich.*