

Rational Inattention: A Disciplined Behavioral Model

Bartosz Maćkowiak*

Filip Matějka[†]

Mirko Wiederholt^{‡§}

May 18, 2018

Abstract

A recent growing body of studies shows that many important phenomena in economics are, or can be, driven by the fact that humans cannot digest all available information, but they can choose which exact pieces of information to attend to. Such phenomena span macroeconomics, finance, labor economics, political economy, and beyond. People's choices of what information to attend to, i.e., what *optimal heuristic* to use, are driven by current economic conditions and determine the form of mistakes that they make. Combining these behavioral insights together with optimizing approaches of classical economics yields a new generally applicable model. The implied behavior features numerous types of empirically supported departures from existing classical models, is potentially highly practical for answering policy questions, and motivates further empirical work. One distinction from most models in behavioral economics is that this model allows for studying the *adaptation of agents' behavioral biases* due to changes in policy or economic conditions.

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

[†]CERGE-EI and CEPR (filip.matejka@cerge-ei.cz)

[‡]Goethe University Frankfurt and CEPR (wiederholt@wiwi.uni-frankfurt.de)

[§]We thank Vladimir Novak and Andrei Matveenko 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

This survey paper argues that rational inattention matters. It is likely to become an important part of Economics, because it bridges a gap between classical economics and behavioral economics. Actions look behavioral, since agents cannot process all available information; yet agents optimize in the sense that they try to deal optimally with their cognitive limitations - hence the term "rational inattention." We show how rational inattention describes the adaptation of agents' behavioral biases due to policy and other changes of the economic environment. Then, we survey the existing literature, and discuss what the unifying mechanisms behind the results in these papers are. Finally, we lay out implications for policy, and propose 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 summarize, filter, and digest the abundant available information is key to understanding many phenomena in economics. Several crucial findings in economics, even some whole subfields, have been built around the assumptions of imperfect or asymmetric information. However, nowadays, many more forms of information than ever before are available due to new technologies, 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 drives the dynamics of prices, consumption or investment;
- finance, because it determines the form of investors' uncertainty about asset returns, and drives the patterns of portfolio allocation;

- labor economics, because it affects and directs the searches of both firms and job applicants;
- behavioral economics, because it determines what heuristics in choice we use and thus it drives our systematic biases in decision making;
- political economy, because we pay little attention to facts when our personal stakes are low.

For instance, people might choose to be more attentive to the macroeconomy in times of crisis, and thus change their behavior more swiftly if taxes or interest rates change. This in turn affects the asymmetry of the business cycle as well as the efficiency of both fiscal and monetary policies in such times. Meanwhile on the labor market, in times of high unemployment and high supply of workers, firms might choose not to read the CVs of unemployed or minority applicants at all, which would then exacerbate long-term unemployment and discrimination. And finally, if investors cannot follow the details of current market conditions in each country, they might instead pay attention to regional indices only. Crises in one country would then not be distinguished from a crisis in another country within the same region, and the investors might sell the whole regional portfolio.

The theory of rational inattention, henceforth RI, (following the seminal work of Christopher A. Sims, 2003), provides a model of how cognitively limited people simplify and summarize available information. RI is a disciplined behavioral model. 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, RI 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. 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. RI is a suitable model for doing so. It combines the insights of psychology together 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", with the difference being that in RI both types of

thinking are optimal. The fast one 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.¹

In Section 2 we describe the theoretical framework of RI and its main implications. Section 3 surveys the most important findings in the theoretical literature. Section 4 discusses existing empirical results and related challenges. Section 5 discusses policy implication, and in Section 6 we highlight what we think are the most fruitful next steps and the broad unresolved questions in the literature on RI.

2 Theoretical framework

Rational Inattention builds on a natural assumption: agents cannot pay full attention to all available information, but can choose to pay more attention to more important things. The literature on RI was started by Sims in a series of papers (Sims 1998, 2003, 2006). The initial motivation was to lay the foundations for a new type of dynamic macroeconomic models, in which all types of inertia would be driven by a unifying source - optimal inattention on the side of households and firms. Over time rational inattention has been embraced by many other fields. A key feature of RI models is that behavioral biases change with stakes and the environments.

Consider a manager who sets a price to maximize profit. The optimal price y 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 the price that maximizes profit. If instead the manager gets noisy information about x , then she chooses y that maximizes expected profit. The form of noisy information that she gets exogenously determines the prices that she might set, i.e., the distribution $f(y|x)$.

RI, however, allows for a more flexible approach. The RI manager acts as if she were *choosing* $f(y|x)$. The choice reflects the manager's choice of what information to receive, and results in the form of pricing mistakes that she makes. For instance, if she chooses not to get any information,

¹Gigerenzer et al. (1999) define heuristics as efficient cognitive processes, conscious or unconscious, that ignore 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.

then she chooses one fixed y for any x . If she pays more attention to lower x , e.g., states of lower demand, then pricing is more precise at these levels of demand, i.e., $f(y|x)$ is tighter. Making more precise decisions, however, takes more effort, and more concentrated $f(y|x)$ are associated with a higher cost.

Sims put forth two main cornerstones of RI.

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.

Models of information acquisition have been around for a long time, but RI is its own particular version - the novelty lays in the assumption that agents can choose to get information of any form from an unrestricted menu of signals.

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 single agents choose how much information to get about different sources of Gaussian shocks. Sims (2006) then emphasizes that in practice it is not only the amount of information that agents choose, but also the nature of information. For instance, households may, due to credit constraints, want to be more attentive to negative than to positive income shocks, which would then drive asymmetry of their consumption patterns. Sims (2006) presents a model, where agents can choose any form of signals, not just Gaussian.

This basic framework of RI may also be useful for studying of the implications of other types of cognitive limitations beyond limited attention; perhaps with a different form of the cost function of precision of choices.

In the rest of this section, we first present an illustrative example, then we formulate a general static model, and discuss its properties as well as an extension to a dynamic model.

2.1 A simple example

This example illustrates the basic elements of the modeling framework of RI. Consider again the manager who sets a price to maximize profit, but to do so she needs to pay costly attention to the

current market conditions, which are unknown (Wiederholt, 2010). How large the profit stakes and uncertainty are affects her choice of how much to attend to this problem, which in turn determines how responsive the price is to changes of the market conditions.

Objective, actions and uncertainty. Let the agent maximize the expectation of the objective function

$$U(y, x) = -r(ax - y)^2 \tag{1}$$

less the cost of attention. The objective (1) can be thought of as an approximation of the profit function around its maximum. y is the agent's action, price to set. ax is the target, where x is random and the agent needs to pay costly attention to it (it can be a deviation in marginal cost, which translates into an optimal deviation in price). The parameter a denotes 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 us assume that the agent faces Gaussian uncertainty about x :

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

which coincides with the agent's prior knowledge.

Costly attention. In this example, we assume that to refine her knowledge about x , the price-setter needs to receive Gaussian signals,²

$$s = x + \epsilon, \text{ where } \epsilon \sim N(0, \sigma_\epsilon^2). \tag{2}$$

Precision of the signal is subject to the agent's choice. The more attention he pays, the more precise signal he receives, which helps him target the optimal y better. However, more precise signals are also more costly. The cost function most typically used in RI is the following:

$$cost = \lambda\kappa,$$

where $\lambda > 0$ is a parameter, a unit cost of information, and κ is the chosen amount of information. This amount is measured as the expected reduction of uncertainty, which is expressed by entropy,

²Later on we show that while Gaussian signals are optimal in this setup, i.e., for this objective and prior uncertainty, they are not optimal in general.

due to the acquisition of signal s . The entropy $H(\cdot)$ of a random variable drawn from a normal distribution of variance σ^2 is $\frac{1}{2} \log(2\pi e\sigma^2)$.

$$\kappa = H(x) - E[H(x|s)] = \frac{\log 2\pi e\sigma_x^2}{2} - \frac{\log 2\pi e\sigma_{x|s}^2}{2}.$$

We will discuss this particular modeling choice in Section 2.2.

Decision problem. The agent thus faces two choices in succession.

- (i) How much attention to pay - the choice of posterior variance $\sigma_{x|s}^2$.
- (ii) How to act upon posterior belief - the choice of optimal price y conditional on signal s .

Step (ii) is easy. $y = aE[x|s]$ maximizes the expectation of (1) for a given posterior belief.

Therefore, the choice of attention in (i) is given by:

$$\max_{\sigma_{x|s}^2 \geq \sigma_x^2} E_x \left[E_s \left[-ra^2(x - E[x|s])^2 \right] \right] - \lambda\kappa = \max_{\sigma_{x|s}^2 \geq \sigma_x^2} \left(-ra^2\sigma_{x|s}^2 - \frac{\lambda}{2} \log \frac{\sigma_x^2}{\sigma_{x|s}^2} \right). \quad (3)$$

Solution. 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 the weight on the signal, $\xi \equiv \left(1 - \sigma_{x|s}^2/\sigma_x^2\right) \in [0, 1]$ reflects the chosen level of attention.³

Therefore,

$$y = (a\xi)x + \eta. \quad (4)$$

Notice that (4) 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 . η is the resulting noise in actions, which has zero mean.

We can now rewrite the problem (3) in terms of the choice variable ξ .

$$\max_{\xi \in [0,1]} \left(-ra^2(1 - \xi)\sigma_x^2 - \frac{\lambda}{2} \log \frac{1}{1 - \xi} \right). \quad (5)$$

The solution is

$$\xi = \max\left(0, 1 - \frac{\lambda}{2ra^2\sigma_x^2}\right). \quad (6)$$

The main qualitative implications are:

³ $\xi = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2}$; and the posterior uncertainty is $\sigma_{x|s}^2 = \frac{\sigma_x^2\sigma_\epsilon^2}{\sigma_x^2 + \sigma_\epsilon^2}$.

(i) *Under-reaction*: realized prices move less than optimal prices. If $\lambda > 0$, then the agent under-responds to shocks in 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.

(ii) *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 ξ is increasing in a , too, and thus the realized elasticity is convex in the elasticity under perfect information.

(iii) *Uncertainty increases responses*: 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 .

(iv) *Stakes and cost of information*: similarly, higher stakes r and lower cost λ increase responses.

These features were first described in Maćkowiak & Wiederholt (2009), where price-setting firms choose to allocate attention across firm-specific and aggregate shocks. There, firms respond relatively much more strongly to the firm-specific shocks, which are more important and more volatile.

2.2 Assumptions of RI

The example explored *how much* or *which* shocks to pay attention to, but RI allows for a more subtle modeling of attention: *how*, i.e., to *what exact features* of shocks the agents pay attention to.

In a realistic price-setting problem, Gaussian and thus symmetric uncertainty may not be the most useful one. What if the price-setter was choosing which one of two products to put on sale and run a TV add for? Given this binary problem, finding out which product faces higher elasticity of demand might be sufficient as opposed to gauging what the exact elasticities are. RI allows for the acquisition of the most useful pieces of limited information.

RI is built on a few assumptions, which distinguish it from other models of information acquisition. Here we highlight the decision situations that RI is a good fit for, and which not. RI fits best situations in which information is fully available, the agents are able to choose which of its pieces

to acquire, and they do so optimally.

1. *Information is available in all forms.* This is reflected by the agent's ability to shape signals in any way she chooses, e.g., binary or Gaussian.

Situations that fit this assumption are those with information in many forms, from many sources, such as in many macroeconomic applications. It is a similar case when much more information is provided than the agent can process, e.g., in descriptions of products or even the records of political candidates. Some researchers consider RI to be a proxy of directed thinking (in the agent's mind) as well. The formation of signals is then internal.

2. *Agents choose the information to be processed optimally.* This is reflected by optimization over what signals to acquire, i.e., what to pay attention to. Obviously, this assumption faces the criticism that the agent is cognitively limited, yet she picks the strategy optimally.

We consider RI to be a benchmark that applies well in repeated situations, or in choices over the long term. In this case, 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 stumbled upon due to evolutionary reasons.

A good example is a household's repeated consumption decisions following simple rules of thumb, which are near-optimal. Similarly, HR managers reading the CVs of many applicants every day are likely to adopt an optimal strategy of what parts of the CV to inspect in most detail.

On the other hand, one-time quick decisions do not fit the RI framework well. Consider an agent being presented with a near-crash situation in a car for the first time - it is unlikely that she will first think about what information to process, and then pick the limited amount of it optimally.

3. *The cost of information is measured by mutual information.* This cost is microfounded when the limits are the difficulties of processing and understanding, i.e., reading information.

The main reasons for using the entropy-based function are the following. First, it is tractable. The second is that 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. And third, the axiomatic foundations of entropy are aligned well with the processing of available information (Shannon, 1948; Cover & Thomas, 2006).⁴

To summarize, RI models describe repetitive decisions with a great deal of available information well. Decision situations fitting RI entail information acquisition about macroeconomic conditions, about political candidates, about job candidates, or about product characteristics, for instance. Decision situations that the model fits less clearly are: quick and one-time decisions (e.g., emergency or lab), or when information needs to be acquired through a rigid technology (e.g., thermometer or oil-drilling).

One of the important implications of these points is empirical validity: there is very high value to testing RI in the field, since decision situations in the lab do not always fit the setup for which RI is meant to be applied to. Often, in the lab the agent does not have enough time or proper motivation to design the optimal strategy (as in repeated decisions in real life) and moreover, information is also often available in many fewer forms inside of the lab.

⁴Informally put, Shannon (1948) asked a question: how much information does an agent receive if she finds out which of N states, each with probability p_i , occurred? Shannon was looking for a function $H(p_1, \dots, p_N)$ with the following three properties:

(i) H is continuous; (ii) $H(\frac{1}{N}, \dots, \frac{1}{N})$, if states are uniformly distributed, then H is increasing in N , i.e., the agent learns more if she discovers the true state in the case that there are more states to begin with;

The third assumption distinguishes entropy from other reasonable functions: (iii) Independence from intermediate steps: the cumulative amount of information is the same if the agent finds out directly which of the states it is as if she first finds out which subgroup it belongs to, and then which of the states it is within the subgroup.

If all three conditions are satisfied, then the unique function is entropy $-\sum_i p_i \log p_i$. Entropy is a good measure of cost when all pieces of information are equally costly. This is for instance when information is already acquired by someone, and presented on a table in front of the agent, on the Internet, or in case the agent can ask someone knowledgeable. The only cost is in asking questions, reading text, viewing pictures, or reading and understanding digits.

These properties imply that entropy does not depend on a metric, i.e., the distance between states does not matter. With entropy, it is as difficult to distinguish the temperature of $10^\circ C$ from $20^\circ C$, as $1^\circ C$ from $2^\circ C$. In each case the agent needs to ask one binary question, resolve the uncertainty of one bit. If, however, the agent needed to use a thermometer with inherent additive noise of a given size, then it is clear that distinguishing the more distant states $10^\circ C$ and $20^\circ C$ would be easier - reduction of entropy is not a good measure of information in that case.

Finally, one can also use different functions of entropy. In this survey, we focus on the cost linear in entropy. The literature on RI has, however, also used other cost functions; a convex function of the amount of attention, i.e., of entropy, or even with a limited total attention, and the agent has a choice of how to relocate it across shocks.

2.3 Static model

Here we describe the general static choice model under RI. The unknown random state is x , and the agent's prior belief is given by a pdf $g(x)$. The timing is:

1. The agent chooses what information about x to process. This is described by what signals the agent gets for a given realized state x , i.e., by a distribution $f_{sx}(s|x)$. She maximizes the expectation of utility (see step 3 below) less the cost of information;
2. The agent receives a signal s , and forms a posterior belief $f_{xs}(x|s) = f_{sx}(s|x)g(x)/p(s)$, where $p(s)$ is the pdf of signals;
3. The agent chooses an action y to maximize the expectation of utility $U(y, x)$.

The agent chooses two strategies: an information strategy in step 1 above, and an action strategy in step 3. It turns out that a joint distribution $f(y, x)$ describes both of the strategies as in Sims (2003) and Kamenica & Gentzkow (2011). 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. We can thus make a one-to-one association between s and y , and use $f(y, x)$ only. The two-stage optimization process is formally defined in Matějka & McKay (2015).

The agent's problem then is:

$$\max_f \int U(y, x) f(y, x) dx dy - C(f), \tag{7}$$

$$\text{subject to } \int f(y, x) dy = g(x), \quad \forall x. \tag{8}$$

where the first term in (7) is the expectation of U , and $C(f)$ is the cost of information.⁵ The constraint (8) captures Bayesian rationality, requiring the consistency of prior and posterior belief.

While the cost of information $C(f)$ could in principle take any form, following Sims (2003) we use $C(f) = \lambda I(y; x)$, where $I(y; x)$ is the Shannon mutual information between the random variables y and x . It is the expected reduction of entropy of beliefs about x upon processing information, and choosing y .⁶ Letting $p(y)$ denote the marginal of y , mutual information is defined as

$$I(y; x) \equiv H(x) - E[H(x|y)] = \int f(y, x) \log \left(\frac{f(y, x)}{g(x)p(y)} \right) dx dy. \tag{9}$$

⁵For simplicity here we use the imprecise notation using probability distribution functions only.

⁶The entropy of x is $H(x) = - \int f(x) \log(f(x)) dx$.

2.4 Solution, implications, and optimal heuristics

The first order condition to (7)-(8) implies that the behavior is probabilistic and follows a logit model (Matějka & McKay, 2015). For a true, but perhaps unknown, state of nature x , RI agents choose an action y with the conditional probability:

$$f(y|x) = \frac{e^{U(y,x)/\lambda + \alpha(y)}}{\int_z e^{U(z,x)/\lambda + \alpha(z)} dz}, \quad (10)$$

where $\alpha(y) = \log(p(y))$ if $p(y) > 0$, and zero otherwise. For the connection of this formula to the applied literature on discrete choice, let us express and emphasize the resulting choice probabilities in case the action set is discrete, $i \in \{1, \dots, N\}$, where $\alpha_i = \log(\int_x P(i|x)g(x)dx)$.

$$P(i|x) = \frac{e^{U(i,x)/\lambda + \alpha_i}}{\sum_j e^{U(j,x)/\lambda + \alpha_j}}. \quad (11)$$

The α_i in (11) reflects *biases* towards action i , i.e., it shifts the choice probabilities towards this action if it appears to be a good candidate a priori. The biases α describe the heuristics that the agent chooses to use, and reflect the agent's choice of attention. The biases are independent of state x ; it is only prior knowledge and preferences that determine how the agent chooses to approach the problem at hand, i.e., what pieces of uncertainty to resolve. See Matějka & McKay (2015) and Caplin et al. (2016) for how to solve for α given prior $g(x)$. 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.

The implied features of the behavior of RI agents are:

(i) *Stochastic choice*: RI agents make mistakes. If an action is ever selected ($p(y) > 0$) then it can be selected at any state. The shape of the cost function implies that ruling out some states with certainty is very costly on the margin; many other micro-founded cost functions would share this property;

(ii) *Logistic choice* with heuristics is optimal for any preferences and prior beliefs, which makes the RI model tractable, and potentially amenable to empirical applications. If an option i seems attractive a priori, then the agent chooses to pay attention in such a way that the bias towards this option α_i has the same effect as a positive utility shock α_i in each state. The exponents additively separate the effects of preferences, U , and beliefs, α , on the choice. This property rests on the entropy-based cost of information.

The RI-logit is, however, different from the standard logit, because the RI agent’s heuristics are optimally tailored to the environment. For instance, if an alternative i is good on average, perhaps often better than most other alternatives, but it is dominated by $j \neq i$ in all states of the world, then $\alpha_i = -\infty$. The agent never chooses a dominated i .

The sophistication extends beyond this simple example. For a given heuristic, and a chosen bias α , the RI behavior satisfies independence from irrelevant alternatives (IIA). However, it does not need to satisfy IIA across choice problems. For instance, Debreu (1960) criticized logit-behavior using a thought experiment with two duplicate options (“red-bus, blue-bus problem”). He argued that adding a duplicate option should decrease the probability of the duplicated one only, which is not what IIA implies. On the other hand, if a duplicate option is added, the agent chooses to use a different heuristic (ignores the duplicate completely), and IIA does not hold (Matějka & McKay, 2015);

(iii) *Linear-quadratic preferences and Gaussian prior uncertainty*, as in Section 2.1, imply that Gaussian signals as in (2) are in fact the optimal ones that solve the general model, see (10) with a quadratic loss in the exponent.

When the losses from misperception depend on the size of misperception only, e.g., on $(x - y)^2$, and $y \in \mathbb{R}$ then it is intuitive that symmetric signals with equal precision across states are optimal. If, on the other hand, the agent chooses between binary alternatives, then she needs to find out which of the two alternatives provides a higher utility, i.e., whether the difference between their utilities is positive or negative only. A Gaussian signal is wasteful in this case, since it provides more refined information on what the difference approximately is;

(iv) *Categorization, discreteness, and consideration sets*. RI agents most often find it optimal to contemplate a low number of actions only (Jung et al., 2015). 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 (Matějka, 2016; Stevens, 2017). The resulting discreteness is an alternative explanation for inaction that we see in the data, i.e., inaction of prices or portfolio allocations. Typically, such an inaction would be explained by an adjustment cost, while here it is an optimal response to the choice of what information to acquire. If the agent chooses to focus on two distant levels of prices, she does not waste information capacity (time and mental effort) on small movements, and is thus less likely

to make larger errors.

Matějka & McKay (2015) show how to solve for α using an alternative formulation of the original problem, and Caplin et al. (2016) provide sufficient and necessary conditions for what actions are considered. Again, RI functions as a magnifying force - if an alternative seems bad a priori, then the agent might choose not to process more information about this alternative, and thus never select it;

(v) *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 approach the problem differently, albeit optimally. If signals are endogenous to what options are presented to the agent, then transitivity in choice can be violated (Woodford, 2015; Matějka & McKay, 2015);

(vi) *Posterior invariance.* Caplin & Dean (2015) show that 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 independent of small changes to the prior. This feature is useful for solving models with RI;

(vii) *Multi-dimensional simplification - indexation.* If agents need to pay attention to several shocks and choose multiple actions, then RI models what simplified representation of the high-dimensional environment they use. Consider an investor who needs to get information about prices of numerous assets, and then she decides how much to buy of each. For an LQ-Gaussian problem, if the cost of information is high, then the agent pays attention to a particular linear combination of the prices only - i.e., to an endogenously constructed index, which is determined by correlations of prices. In turn, she then purchases or sells the whole portfolio determined by the index (Peng, 2005).

Similarly a consumer does not attend to all prices of all products, but compares only close substitutes, which then results in behavior resembling mental accounting. A solution method is based on the decomposition of losses into losses from deviations along principal components. The agent then chooses how much attention to pay to each of the components, see Sims (2003) and (Kőszegi & Matějka, 2017). For instance, when choosing how much of two imperfectly substitutable goods to consume, the components are the relative prices $(p_1 - p_2)$ of the two goods and price level $p_1 + p_2$, with the optimal consumption of good 1 c_1 equal to $-a_1(p_1 - p_2) - a_2(p_1 + p_2)$, where elasticities $a_1 > a_2$ are given. Under RI, however, the expected consumption will be (following the

example in Section 2.1)

$$E[c_1] = -a_1\xi(a_1)(p_1 - p_2) - a_2\xi(a_2)(p_1 + p_2)$$

where $\xi(a_i)$ are attention weights, which are increasing in elasticities. Relative elasticities are magnified just like in Section 2.1). For sufficiently high costs, the consumer's optimal heuristic is to pay attention and respond to $(p_1 - p_2)$ only.

2.5 Dynamic model

Sims (2003) studied a dynamic RI problem. RI describes inertia in choice, but the inertia is not driven by explicit adjustment costs, but by the fact that agents do not think it is worthy to get enough new information that would justify a change in action.

If the problem is dynamic, then the agent faces a more complicated problem. She is faced with a multi-dimensional problem (there are shocks in multiple periods), but also while thinking about what information to pay attention to today she also needs to consider its use tomorrow. To forecast future states, it may not be optimal to get information about today's state, e.g., price level, but a trend, instead, such as about inflation.

If x_t follows a random process, then in each period the agent processes information, forms a posterior belief, chooses an action y_t , and updates her beliefs going to the next period. The cost of information is again proportional to (9), but now the uncertainty is about the whole history of states

$$I\left((y^t \equiv (y_t, y_{t-1}, y_{t-2}, \dots)); x^t \equiv (x_t, x_{t-1}, x_{t-2}, \dots)\right).$$

We will build on the LQ-Gaussian example. Let the utility be quadratic, $E[-(x_t - y_t)^2]$, and the random process be moving-average Gaussian:

$$x_t = \sum_{\tau=0}^{\infty} a_{\tau} \nu_{t-\tau}, \tag{12}$$

where $\nu_{t-\tau}$ are iid drawn from $N(0, 1)$. The optimal signals are again Gaussian, and y_t follows a moving-average Gaussian process, too, which is the exact analog of the linear relation (4) in the static case using the insight that Gaussian signals are optimal.

$$y_t = \sum_{\tau=0}^{\infty} b_{\tau} \nu_{t-\tau} + \sum_{\tau=0}^{\infty} c_{\tau} \epsilon_{t-\tau}. \tag{13}$$

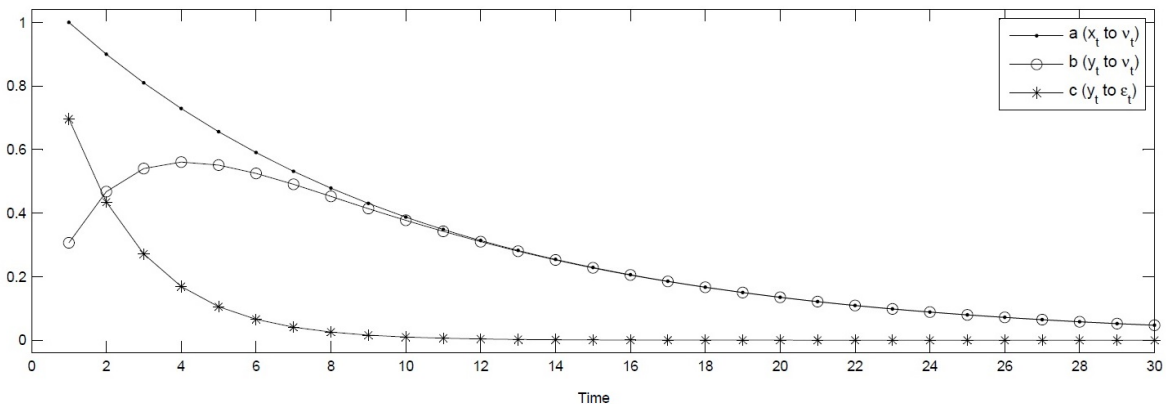


Figure 1: Dynamic model, AR(1) example

Here, b_τ plays the role of $a\xi$ in the static case. It measures how strongly y_t responds to past shock $\nu_{t-\tau}$, while $c_\tau \epsilon_{t-\tau}$ is the noise in signals, $\epsilon_{t-\tau} \sim N(0, 1)$.

Sims (2003) solves this problem numerically; he does so by expressing the information constraint in frequency domain (\tilde{b} and \tilde{c} are Fourier transforms of b and c):

$$I = -\frac{1}{2} \int_{-\pi}^{\pi} \log \left(1 - \frac{1}{1 + |\tilde{c}|^2 / |\tilde{b}|^2} d\omega \right), \quad (14)$$

while the objective is (this is the analog to choosing ξ in the static problem (5)):

$$\max_{\{b_\tau, c_\tau\}_{\tau=0}^{\infty}} \left(\sum_{s=0}^{\infty} -(a_s - b_s)^2 - c_s^2 \right) - \lambda I. \quad (15)$$

This formulation highlights that the agent would like to choose little noise c , while the response b is as close to the target process a as possible. However, processes that are responsive at high frequencies must also feature high-frequency randomness c , otherwise the cost of information (14) would be too high.

Figure 1 presents a solution in the case when x_t follows an AR(1) process, which means that shocks ϵ_t vanish exponentially. Response of y_t is gradual, represented by a gradual convergence of series b to the series a in the figure. Noise c is the largest early on, as implied by (14), and then it vanishes.

Maćkowiak & Wiederholt (2009) then find analytical solutions for x_t following an AR(1) process. For AR(1) the agent does not have any motivation to get information on anything other than the current state since the past states do not enter the current objective and neither do they help

forecast future states. Maćkowiak et al. (2016) show that a general AR(p) case can be studied as an optimal Kalman filter with only one one-dimension signal $s_t = \sum_{\tau=0}^{p-1} w_{\tau} x_{t-\tau}$ on a linear combination of the current and p past states x . The agent chooses the weights w_{τ} , subject to the information constraint.

This problem is closely related to a multi-dimensional static problem, where instead of considering what information about many past states to get, the agent considers what information about many dimensions of the current state to pay attention to. In the case that the agent were choosing two-dimensional actions (e.g., how much to consume and how much to work), then she would get two signals, unless the cost of information were high, in which case she would acquire one signal only, and then there would be perfect co-movement in the two actions.

For a general non-quadratic dynamic setup, Steiner et al. (2017) show that logit behavior also emerges in the dynamic case. The difference from the static version is that in the dynamic case 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. The resulting behavior resembles a habit formation.

The main qualitative features of the RI behavior in dynamic setups are the following:

(i) *Smooth impulse responses and noise in high-frequencies.* Gradual information acquisition implies that actions are biased towards expected optimal actions, and do not respond to shocks immediately. If the target action is persistent, then the resulting actions will feature inertia (Sims, 2003; Maćkowiak & Wiederholt, 2009);

(ii) *The effect of stochastic properties - comparative statics.* The agent responds faster to more important variables, and to variables which are driven by larger shocks. Furthermore, an increase in the persistence of the variable tracked by the agent (holding its variance constant) leads to smaller errors (Maćkowiak & Wiederholt, 2009);

(iii) *Logistic behavior.* The choice behavior is closely related to the dynamic logit estimation of adjustment cost in Rust (1987). While the term of the bias in RI $-\alpha(y_t|y^{t-1})$ would be interpreted as an adjustment cost, if the modeler did not allow for RI, the term is not fixed - it is endogenous to the whole history of actions (Steiner et al., 2017). The adjustment cost would seem particularly large if some trajectory of actions were highly unlikely ex ante, but perhaps due to probabilities of the necessary shocks leading to such actions, not due to the cost of adjustment.

2.6 Differences from related approaches

Signal-extraction (Lucas, 1973): this shows that imperfect information can drive the real effects of monetary policy. While signal-extraction assumes a particular set of available signals, RI derives it from first-principles, and thus comparative statics are different, e.g., when the volatility of certain variables increases.

Sticky-information (Mankiw & Reis, 2002): agents get perfect information infrequently, and none in-between. This results in a very useful and tractable model. RI can, in contrast, explain differential responses to different shocks. Decision-making on the individual level is completely different.

Saliency (Bordalo et al., 2012): behavioral assumptions of what draws attention. RI probably applies better to repeated choices; saliency is better motivated by psychology for quick choices; simple, exogenously given.

Focusing (Kőszegi & Szeidl, 2013): this model can explain several behavioral puzzles. It assumes that agents put more weight on attributes that differ across various alternatives more; this behavior can be microfounded by RI.

Sparsity (Gabaix, 2014): in sparsity, agents choose costly loads of responses. In the case of quadratic preferences and Gaussian prior uncertainty, sparsity is very similar to RI, because RI already predicts linear loads (equation (4)) and corner solutions (equation (6)). In this case, the main differences arise when the optimal action depends on multiple shocks - more on this below. 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.

3 Survey part

Ever since Sims' theoretical work (Sims, 2003, 2006), a growing number of fields have embraced it - macroeconomics was the first, and other fields have also become fertile targets.

3.1 Macroeconomics

When Sims (1998) proposed the idea of rational inattention, his motivation was macroeconomics. Sims considered a conventional 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.⁷ 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. His hypothesis has defined a research agenda.

A key question in macroeconomics is how firms set prices. Answering this question is important for understanding the dynamics of inflation and the effects of monetary policy. Studies of micro price data find that individual prices change fairly frequently and by large amounts.⁸ Rational inattention explains why the price level can respond slowly to nominal shocks even though individual prices change frequently. In Woodford (2002), firms observe aggregate nominal demand with noise. Nominal shocks have strong and persistent real effects. In effect, Woodford assumes that firms pay little attention to the aggregate economy. Maćkowiak and Wiederholt (2009) show under what circumstances firms find it optimal to pay little attention to the aggregate economy. Firms face a trade-off between paying attention to aggregate nominal demand and firm-specific cost or demand conditions. 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. This implies that firms allocate almost all attention to idiosyncratic conditions.⁹ As a result, prices respond strongly and quickly to idiosyncratic shocks, but only weakly and slowly to nominal shocks. Nominal shocks have strong and persistent real effects.

Rational inattention can also explain why many individual prices in the data remain constant for some time. Matějka (2016) studies price setting under rational inattention without approximating the profit function.¹⁰ In his model, prices move discretely – the firm finds it desirable to choose prices from a finite set – even though shocks are distributed continuously. Discrete price adjustment

⁷Since then, Christiano et al. (2005), Smets and Wouters (2007), and many others have confirmed Sims’s finding in more formal analysis.

⁸In a representative study, Klenow and 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 and Klenow (2004) and Nakamura and Steinsson (2008).

⁹In 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.

¹⁰For tractability, Maćkowiak and 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.

is optimal despite the absence of any physical cost of price adjustment.¹¹

Rational inattention is consistent with a number of facts about micro price data (Matějka, 2016, Stevens, 2017). Most strikingly, prices in a rational inattention model tend to fluctuate back and forth between a small number of values, and are more likely to be below the modal price than above it, replicating a regularity in the data stressed by Eichenbaum et al. (2011). At the same time, nominal shocks continue to have strong and persistent real effects. Stevens (2017) builds on the model of Woodford (2009) in which a firm decides when to review its price facing a fixed cost of a price review, and the firm makes the decision whether to review its price subject to rational inattention.

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? A DSGE model with rational inattention can replicate basic features of the business cycle. Maćkowiak and Wiederholt (2015) construct a DSGE model with rational inattention on the side of both firms and households. 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 as the only source of slow adjustment. In particular, households decide how much to consume and save subject to rational inattention.¹² Small deviations from the consumption Euler equation are inexpensive in utility terms and therefore households choose to pay little attention to the real interest rate, implying that consumption responds slowly to an innovation in monetary policy. The model matches the impulse responses to a monetary policy shock and to a technology shock from a standard vector autoregression.¹³

Insights about policy emerge (Paciello and 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

¹¹In complementary work, Matějka (2015) demonstrates that a perfectly informed firm moves prices discretely if it faces a consumer who is subject to rational inattention.

¹²See also Luo (2008) and Tutino (2013) for the analysis of consumption dynamics under rational inattention.

¹³One feature of the empirical impulse responses is that the price level responds faster to aggregate technology shocks than to monetary policy shocks. The model replicates this finding. In postwar U.S. data, aggregate technology shocks appear to be much larger than monetary policy shocks. Firms in the model therefore decide to pay more attention to aggregate technology shocks than to monetary policy shocks, and as a result, prices respond faster to aggregate technology shocks than to monetary policy shocks.

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. At the optimal monetary policy, the central bank fully discourages firms from paying attention to markup shocks.

Rational inattention can be tested using data on expectations of economic agents. Rational inattention passes this test, while the full-information rational expectations 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 a macroeconomic variable under-reacts to news. For instance, if a shock raises inflation for some time, the average forecast of inflation increases by less than actual inflation. Moreover, 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 will tend to under-predict inflation. Rational inattention implies exactly the systematic deviations from full-information rational expectations found in the data.¹⁴

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, and therefore agents solve dynamic rational inattention problems (see Section 2.5). Typically, there are multiple shocks and each agent takes multiple actions. Often, strategic interactions between agents arise.

3.2 Finance

RI has proved to be useful in explaining some of the existing puzzles regarding the portfolio allocation of investors. In a changing world, investors can not keep track of all market movements necessary to make an optimal choice for their portfolio. They therefore use strategies for acquiring information, which leave them systematically oblivious to some market movements. RI models do exactly this and can thus provide a bridge between classical and behavioral finance.

RI predicts observable patterns of portfolio investments and returns. Kacperczyk et al. (2016) study how mutual fund managers allocate their attention to future asset values, and how these

¹⁴The same patterns are present in the data for survey-based and market-based measures of inflation expectations, for expectations of macroeconomic variables other than inflation, among consumers and professional economists, and in different countries. See Coibion and Gorodnichenko (2012, 2015).

change with the state of the business cycle. In the data, recessions are associated with higher aggregate volatility. RI investors thus allocate more attention to aggregate shocks in recessions and relatively more to idiosyncratic shocks in booms (as in Section 2.1). The increasing price of risk in recessions further magnifies this effect.

Kacperczyk et al. (2016) find in data, consistent with their theory, that investors time the market more in recessions, which is that they decrease or increase their whole portfolio positions. On the other hand, they focus on picking single stocks more in booms.

RI can also speak to financial contagion. Consider investors operating on two markets. If one market is hit by a shock and returns become more volatile, the investors need to pay more attention to it. But if investors' total attention is limited, then this implies less attention and thus increased uncertainty together with lower prices on the second market. Financial shock transmits to the second market purely due to attention reallocation (Mondria & Quintana-Domeque, 2013).

A simplified strategy can also entail paying attention to an index of assets, as in point (vii) in Section 2.5, rather than to each asset separately, resulting in an amplified co-movement of allocations and prices. It can also generate an optimal strategy of ignoring some assets altogether, if the investor knows too little about them to begin with, and thus too much attention would be required to decrease the risk sufficiently. RI explains when under-diversification of assets can emerge and how this depends on the form of the cost of information (Van Nieuwerburgh & Veldkamp, 2010). Similarly, it can also explain home-bias (Van Nieuwerburgh & Veldkamp, 2009; Mondria & Wu, 2010), and co-movement of asset prices (Peng, 2005; Mondria, 2010; Mondria & Quintana-Domeque, 2013).

3.3 Theory and behavioral economics

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 does. RI does not model the procedural details of decision making, but it does describe what choices would be achievable given certain limits to cognition.

For instance, RI can help understand how people choose from menus of pension plans. What information do agents collect? What is the effect of the quality of distinct plans, and what is the

effect of agents' beliefs about the whole menu? Should a government authority provide targeted recommendations, should it ensure unrestricted competition, or should it regulate quality? How do firms tailor their menus of products to attract cognitively limited agents?

Behavioral properties. The behavioral properties of the RI model are described in Section 2 in more detail. It features random choice that follows heuristics such as categorization, summarization, and mental accounting (Matějka & McKay, 2015; Caplin & Dean, 2015; Matějka & Koszegi, 2017), small sets of considered alternatives (Jung et al., 2015; Caplin et al., 2016), and the endogenous form of inertia in dynamic problems (Sims, 2003; Maćkowiak & Wiederholt, 2009; Steiner et al., 2017). Slight modifications of the RI model described here can also explain phenomena such as reference-dependence or decoy effects (Woodford, 2014, 2015).

Strategic considerations. The fact that the choice of heuristics is subject to the environment and the task at hand is important for models with strategic considerations. RI has proved to be tractable even then, either with the LQ-Gaussian specification, or with discrete choice. With Gaussian noise, Bayesian updating amounts to linear weighting, and if there are N alternatives to choose from, then all the implied behavioral biases in the logit form (11), are determined by the N quantities $\{\alpha\}_{i=1}^N$ only. This is unlike in Bayesian models with exogenous signals, where the dimensionality of possible behavior is as large as the dimensionality of prior beliefs, which is infinite for a continuously distributed state.

The first paper with RI agents and strategic interactions was an application to macroeconomics in Maćkowiak & Wiederholt (2009). There, agents choose to pay little attention to aggregate movements, since they are not volatile. This in turn depresses the responsiveness of aggregate variables to shocks, and further decreases incentives to pay attention to them. In equilibrium, aggregate variables only mildly respond to shocks.

Hellwig & Veldkamp (2009) studied investors who face a coordination problem and can choose how correlated their knowledge should be with the knowledge of others (i.e., they choose whether to get information from public or private sources). This will determine whether their portfolio choices are similar to those of others or not - RI can exacerbate herd-like behavior.

Alternatively, Yang (2015a) studies the case where investors can choose and coordinate not what sources of information to pay attention to (i.e., what correlation with others), but what type of

information to attend to (i.e., what characteristics of returns). As explained in point (iii) in Section 2.4, if investors choose between discrete alternatives such as invest vs. not invest, then they tend to allocate their attention in a more partition-like manner. Such an attitude saves on information, but the type of partition that the investors use depends due to the strategic considerations on partitions of others. This can generate a multiplicity of equilibria and result in large market swings. Denti (2017) formulates a general game-theoretic model with RI players who choose both the type of information as well as the correlation structure of their information with the information of others.

Optimal contracts and delegation are another avenue studied using RI (Yang, 2015b; Lindbeck et al., 2017). The design of a contract by a principle then does not only affect the agent’s effort or investment choice, but also her choice of what information to collect before taking the action. Lindbeck et al. (2017) find that, for instance, limited liability on the agent’s side induces the principal to choose a contract where the agent’s marginal incentives are globally misaligned. They conclude that RI is easier to work with than explicit signal-extraction models; this is due to the low number of endogenous biases that fully summarize the choice of information.

A related branch of papers studies interactions between sellers and RI buyers. Matějka & McKay (2012) study equilibrium prices in markets with many sellers. Matějka (2015) shows that a monopolistic seller tends to choose coarse pricing strategies when facing RI buyers. The reason is that such buyers can better observe prices alternating between a few values only, and given concerns for risk thus choose to purchase more. Martin (2017) then studies strategic pricing with RI to product’s quality both theoretically as well as experimentally.

Ravid (2014) is the first application of RI to bargaining; it solves the case of repeated strategic interactions. He finds that buyers’ RI introduces a delay in trade and generates a significant surplus to the buyer (as opposed to Rubinstein bargaining). This is because buyers’ valuation of the good across periods of bargaining is due to inattention partly random. This introduces Coasian dynamics, and the seller loses monopoly power and needs to make more pleasing proposals.

Characterization. Finally, a currently very active line of research on RI looks into its axiomatic foundations as well as generalizations. While some of the features of the behavior of RI agents are robust and driven by the qualitative features of the model only (i.e., more information about more important parts of the uncertainty), some depend on detailed specifics of the model. For instance, the logit form clearly depends on the specific entropy-based form of the cost function.

Sims (2003) introduced the big idea that the stochastic choice $f(y|x)$ could be selected very flexibly depending on the choice at hand. The restriction placed on the selection of $f(y|x)$ are then determined by the particular cognitive limitations at hand. To go forward, Sims put forth a particular entropy-based cost function, which is motivated by information theory, but the model can be applicable more broadly, when a different cost function would apply.

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 (see for instance Section 2.2). De Oliveira et al. (2014); Caplin & Dean (2015); Matějka & McKay (2015); Ellis (2013) provide axiomatic foundations of RI using various revealed preference approaches.¹⁵ Caplin et al. (2017) provide 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 et al. (2016) and Morris & Strack (2017) generalize these findings and relate a micro-founded cost function to the cost of sequential sampling.

3.4 Labor

An important part of labor economics studies the effects of information frictions on the functioning of the labor markets. RI has the potential to microfound the matching function between firms and applicants. It drives what applications firms pay attention to, and vice versa, and how this depends on the state of the economy and the current market conditions.

Bartoš et al. (2016) show in a field experiment that endogenous information allocation drives the choices of HR managers in labor markets (and of landlords in rental markets). For instance, a different name or recent unemployment induces the HR manager to read a job application and a CV in more or less detail, affecting the resulting probability of invitation to a job interview.

On a highly selective market, where one star out of many applicants is selected, a negatively stereotyped group is given less attention than others (e.g., their CV is read less). They are then even more likely not to be accepted, and thus additional information is less likely to be valuable. However, on a thin market (e.g., on some rental markets), the negatively stereotyped group is given relatively more attention since other groups are a priori highly likely to be accepted and thus the

¹⁵See also a related work of Manzini & Mariotti (2013).

benefits from paying attention to them are smaller.

Cheremukhin et al. (2016) use the logit-based model of RI, and microfound a matching function, which has the feature of a directed search. The resulting matches are stochastic, but the probability of them occurring is affected by the benefits of the match, as in (11). In this model, actors do not distinguish agents on the other side of the market ex ante, but it could be extended to encompass the effects of type heterogeneity on attention allocation, as in Bartoš et al. (2016). Similarly, it could imply various comparative statics such as that the efficiency of matches could decrease in recessions. If many unqualified candidates are applying for a position, then it is costlier for the HR managers to find a quality match and they might shy away from ex ante risky candidates even more.

3.5 Trade

RI can be useful in understanding trade flows, because it provides a tractable model of directed search for trade partners in distant countries. In the first such model (Dasgupta et al., 2014), inattentive importers look for the cheapest products to import. Prices are random in foreign countries, but the importers know the potential costs of logistics, which induce them to explore nearby countries more than others. The resulting model using RI has a form of logit-based demands similar to the work-horse trade model of Eaton & Kortum (2002), yet the costs of logistics have an amplified effect due to the endogenous search, which yields demands skewed towards nearby countries more than in Eaton & Kortum (2002).

The reason for this effect is that the cost of logistics affects not only the total cost, but also attention allocation and the resulting biases, α_i in equation (10), because the costs of logistics are known ex ante and guide importers in what countries to explore in more detail. Importers look for products in distant countries less, since they expect that they are less likely to find a cheap product there, which further magnifies the expectation of unlikeliness of import from distant locations.

3.6 Political economy

Finally, political economy has always seemed to be potentially a very fruitful target to apply RI to. A great deal of information about politics is available, yet we do not gather much hard data - perhaps because the stakes from our voting decisions are very low. Voters seem to be highly

inattentive: they do not follow in detail what politicians propose or how they act (Carpini & Keeter, 1996). Yet they can choose what to be attentive to. In turn, politicians seem to be aware of these facts and at least partially tailor their behavior to them in order to gain advantage in elections.

Martinelli (2006) studies how much information a rational voter choosing between two alternatives would acquire. If voters have a given level of information, albeit noisy, then the more voters there are the more informed the aggregate decision is. With RI the result is the opposite - the larger the election is, the noisier is the outcome; see how attention decreases with lower stakes r in equation (6).¹⁶ An increasing number of voters decreases the probability that a voter is pivotal, which lowers incentives for information acquisition sufficiently fast.

Taking a positive level of attention as given, Matějka & Tabellini (2016) study the implications of voters' attention allocation across different policy issues. They show that RI empowers voters with extreme preferences and small groups. This is because each issue is paid more attention by voters who care more about it, and thus under RI these voters respond to pleasing proposals by politicians relatively more than other voters (see the magnification of relative elasticities under RI explained in Section 2.1). If the Internet provides finer granularity of information, then this only increases these inefficiencies since different voters can then focus even more on a narrow issue of their particular interest. Similarly, divisive issues attract the most attention and public goods are underfunded.

4 Empirical relevance

The model of RI is well suited for a boom in empirical work, which has not yet occurred. The empirical potential is due to at least three reasons:

1. The *emergence of new technologies*, which make measurement of attention possible (e.g., google analytics, web presence patterns);
2. The *model of RI provides a broader range of testable hypotheses* than most behavioral models. The behavior under RI is adaptive - the heuristics that agents use change if incentives,

¹⁶Martinelli (2006) studies the problem for a general cost of information, and finds that this effect holds if the derivative of the cost at zero precision is infinite.

stochastic properties of the environment, or beliefs about payoffs change;

3. The *implied behavior takes the logit-based form* (11), which is almost the same structure as what the vast empirical literature on discrete choice already uses.

Let us note that when Sims (2003) was published, even the notion of attention to be sufficiently limited that it would matter for economic outcomes was controversial. This is a well established fact now; see DellaVigna (2007) and Handel & Schwartzstein (2017) for good reviews. Below, we thus focus on the implications of the selectivity of attention. These empirical studies found evidence that people design strategies that deal with scarcity of attention in line with RI, yet much more work remains to be done.

The findings below highlight that the approaches of economics can, in certain cases, be as insightful as, or complimentary to, those of psychology. In psychology, attention has long been acknowledged as one of the main drivers of features of choice (Kahneman, 1973). But unlike in RI, it is typically not considered to be selective in some optimal sense, but rather automatic (often sub-optimally) and stimuli-driven. The findings in psychology are most often based on one-off choice situations in the lab, exactly those situations that RI is not best suited for. But once people make repeated choices over the long term and in information-rich environments, the adaptability results become more pronounced.

The basic testable implications of endogenous attention are described in Section 2.1. More refined testable implications are described in Section 2.4.

The qualitative feature number (i) in Section 2.1, i.e., under-reaction, is shared by both endogenous attention as well as pure inattention. The other three properties are inherent to RI, and not to exogenously given imperfect information. They are driven by the adjustment of attention due to its different benefits; (ii) magnified relative elasticities, (iii) uncertainty increases responses, (iv) the effect of stakes and costs. Under RI, beliefs have a bigger effect than under exogenous information, they enter not only as an additional signal to form posterior knowledge, but they also affect the form of all additional information that agents choose to get.

4.1 Explanations of puzzling choices

One strategy for testing rational inattention models is to look at incentives to pay attention and actions. Stochastic properties of the environment, for instance, influence the incentives to pay

attention and should therefore affect the covariance of actions with shocks and the speed of response of actions to shocks.

Recall the model of price setting in Maćkowiak & Wiederholt (2009). When idiosyncratic conditions are more volatile than aggregate conditions, firms pay more attention to idiosyncratic conditions than to aggregate conditions. Thus, prices respond faster to idiosyncratic shocks than to aggregate shocks. Maćkowiak et al. (2009) study U.S. sectoral price data to test predictions of this model. Prices indeed respond faster to idiosyncratic shocks than to aggregate shocks. Furthermore, prices indeed respond more slowly to aggregate shocks in sectors with a higher volatility of idiosyncratic shocks.

Kacperczyk et al. (2016) build a model of mutual funds' attention allocation and test its predictions by exploiting time-series rather than cross-sectional variation in the incentives to pay attention. In the data, recessions are times when aggregate volatility rises. Hence, mutual funds should reallocate attention to aggregate shocks in recessions. Studying the universe of actively managed U.S. equity mutual funds, the authors find that the covariance of each fund's portfolio holdings with the aggregate payoff shock indeed rises in recessions.

4.2 Predictions regarding beliefs

Other empirical papers study the implications of RI for the formation of expectations. The form of the evolution of beliefs about the future is of crucial interest to macroeconomics, since the dynamics of inflation and many other aggregate variables are driven by it. It is also closely linked to the initial motivation for RI in Sims (2003), who argued that expectations are not driven by all shocks that we can observe, but only by those that we pay attention to.

Coibion & Gorodnichenko (2012) used The Survey of Professional Forecasters, and show that expectations change slowly, which is in line with imperfect information models. Moreover, in line with the predictions of RI, they find that beliefs about some variables and in specific times are revised more slowly than otherwise.

They find that the highest information frictions are associated with less volatile macroeconomic variables, point (iii) above, also that beliefs about most variables adjust more slowly during the period of the Great Moderation (between about 1985 and 2007), and that the rigidity of expectations drops in recessions (when volatility is higher). If a macroeconomic variable, say inflation, has little

volatility, then the forecasters do not need to pay much attention to it, and thus they respond slowly when a shock does occur.

This is an important finding also methodologically - they confirmed that RI can explain not only the resulting actions, as in the previous section, but they also provided initial suggestive evidence that the mechanism operates through an endogenous formation of beliefs that people hold (in line and perhaps due to endogenous information acquisition). See also Andrade & Le Bihan (2013) who use data from ECB Survey of Professional Forecasters.

4.3 Direct measurements of attention in the field

The papers discussed in this section addressed the step preceding the belief formation. They measured the attention itself, not the resulting beliefs. This class of papers emphasizes the experimental potential of the recent development of information technologies.

Mondria et al. (2010) use data on search queries on the Internet and study the effect of rational inattention on home bias in finance.¹⁷ Home bias means that investors invest too much in their home country, where they have other sources of exposure, and thus they do not diversify as much as they could. They combine a dataset with over 20 million web search queries together with data on U.S. holdings of foreign securities. They present evidence in the favor of joint causality; more attention drives investment, but more investment increases attention, too.

First, agents allocate more attention to countries whose assets make up a greater share of their portfolios. Second, international investors favor assets from more familiar countries. Investors increase their holdings of a particular country assets in response to an exogenous increase in the information they have about that country. Finally, they estimate that if all countries received the same level of attention by U.S. investors as the U.S., then the average home bias would fall from 85.2% to 57.3%.

To measure attention on an individual level in the field, Bartoš et al. (2016) integrate tools to monitor information acquisition into correspondence field experiments (Bertrand & Mullainathan, 2004). They send emails responding to apartment rental advertisements and to job openings, and randomly vary the names of fictitious applicants and the quality of applicants. Email applications for a job opening contain a hyperlink to a resume. Similarly, in the housing market landlords can

¹⁷See also Van Nieuwerburgh & Veldkamp (2009) and Mondria & Wu (2010).

click on a hyperlink located in the email and learn more on an applicant’s personal website. The authors monitor whether employers and landlords open the applicants resume (or website) as well as the intensity of information acquisition.

Bartoš et al. (2016) document that employers and landlords allocate their attention to job and rental applicants in line with RI. On a highly selective market, less attention is paid to a negatively-stereotyped group, with the opposite being true on a thin market. On each market, this difference in attention, i.e., “attention discrimination”, further disadvantages the negatively-stereotyped group. The mechanism is explained in Section 3.4. Moreover, attention to applicants on the labor market is responsive to an initial signal about the current employment status of the applicant, too.

The subjects’ behavior thus supports the hypothesis of selective, or adaptive, inattention, and not of a fixed behavioral rule. This is important, because “implicit discrimination” (a fixed rule of less attention to minorities), one of the established theories in psychology, does not hold here. The subjects seem to be well able to adjust attention optimally in the long-run. In other words, employers and landlords do seem to learn what use of information is optimal in their case.

4.4 Lab experiments

Laboratory experiments allow for tests of more refined predictions of the RI model, while evidence using field data typically aims at testing the main qualitative implications only, i.e., those in the example in Section 2.1. These experiments can rigorously test the many predictions of RI regarding the effects of payoff profiles, of stochastic properties of the states, or even of the way information is presented.

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.

Caplin & Dean (2015) then studied the finer details of the behavior implied by RI. They characterize the choice behavior of RI using two conditions of revealed preference. Then, they 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 ones. Subjects observe the balls, and take an incentivized

action. The authors then construct “state-dependent stochastic choice data”, which represent the joint distribution of states, and actions. As is discussed in Section 2.3, this object is very important in the theory of RI as it allows for inference about attention. They find that the results are in line with the qualitative implications of the RI model as subjects’ attention changes with stakes. However, certain choices of subjects would be better explained by a different cost function than the entropy-based one.

Dean & Neligh (2017) use the same design to study further details of attention allocation. They also find that subjects adjust the level of attention, but they do not adjust attention exactly in line with the first order condition (10) given by the entropy-based cost.

Cheremukhin et al. (2011) analyze existing lab data of repeated binary choices, and conclude that the RI model accounts for the distribution of errors subjects make better than other random choice models (with fixed noise). Martin (2017) then experimentally studies the predictions of RI in a strategic setup. Ambuehl et al. (2017) 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.

Finally, Khaw et al. (2016) study dynamic decision-making in the lab. Their subjects are asked to estimate the proportion of red vs. green balls in a hidden box. 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 (Sims, 2003; Woodford, 2008), which is much more discrete than the range of actions (Jung et al., 2015; Matějka, 2016; Stevens, 2017).

4.5 Inference and structural estimation

Finally, we describe papers that do not test RI, but use its insights to infer preferences from agents’ choices. These papers take advantage of the RI-logit formula (11). While in the empirical literature started by McFadden (1974) the logit formula reflects unobserved taste heterogeneity, in the model of Matějka & McKay (2015) it reflects noise in signals. However, the materially important distinction is that the RI-logit includes biases that are driven by prior beliefs, and even the aggregate choices do not reflect utilities only.

The big question now is: how shall an econometrician using the RI-logit disentangle preferences U_i from beliefs $\lambda\alpha_i$ in the exponents of (11)? If RI drives choices in practice, and if one ignores the effect of inattention and uses the approaches in the standard discrete choice, or in the subsequent literature following Berry et al. (1995), then the estimated exponent $U_i + \lambda\alpha_i$ would be misinterpreted for utility only.

One simple way of disentangling the two was proposed in Caplin et al. (2015). They take advantage of the fact that $\lambda\alpha_i$ equals the log of ex ante probability of choosing i . If agents get their prior beliefs about payoffs (e.g., quality) of alternatives by observing what other people bought before them, then the market shares of goods and the ex ante probabilities of choosing the goods must coincide; α_i is thus observable. This is the case once the market is in the steady state. Caplin et al. (2015) show that RI further magnifies market shares of good alternatives through social learning, leading to higher market concentrations.

Joo (2017) uses a more flexible approach. He does not use a strong identifying assumption, instead he assumes that some manipulations change the beliefs, and thus the biases α , but not preferences. He uses A.C.Nielsen supermarket data to estimate preferences for laundry detergents. He observes volumes sold of packages at various prices and of different sizes, but also volumes when the packages were displayed differently, for instance. This variation allowed him to infer that only 40% of quantity surcharge, i.e., puzzling increase of per unit price of large packages, is due to inattention and the rest due to preference for larger packages.

The connection to dynamic structural estimation in the spirit of Rust (1987) is discussed in Steiner et al. (2017). The main implications are in line to those of the static model, and that is the magnifying effect of preferences on frequency of choices. In the dynamic case, an increase in adjustment cost enters preference for inertia, but it also increases the bias for inaction because it is ex ante less likely that the agent chooses to switch. The higher the cost of adjustment, the less likely the switch even if the state is such that the switch would be optimal. This in turn results in higher price elasticity of demand for the switch than under perfect information.

5 Policy implications

RI has non-trivial policy implications for the following main reasons:

1. *Agents make mistakes.* Welfare theorems do not apply.

2. *Biases are adaptive.* The behavior of RI agents is erroneous, but it is not fixed. RI models how agents relocate their cognitive resources when policy is changed. In other words, RI is not subject to Lucas critique;
3. *Cognitive costs enter welfare.* RI considers how the costs of attention drive the choices of attention allocation. If more attention is required under some policies, then these costs enter welfare, too. Because of this, there is no divide between utility driving agents' decisions and welfare as an objective of policy, unlike in many other behavioral approaches.

The main implications of RI described in this text provide some general lessons regarding policy considerations. RI implies that more information is not always helpful (since agents may not pay attention to it), but a subtle policy change, which can induce a qualitative behavioral change, or information of a different form, can be.

(i) RI implies *larger impact of policies that affect agents' beliefs*. Particularly powerful can thus be credible policy actions that change beliefs about stochastic properties of the economy, but also simply a different framing of choice, e.g., via a default (Thaler Richard & Sunstein Cass, 2008), or change in a provision of information.

Naturally, in perfect information models, it is only the true payoffs that matter. Similarly, if agents face uncertainty, and get information of a given form, then the form of prior beliefs matters less. This is because they enter the behavior as one of the signals only. In RI, however, prior beliefs enter as one of the signals, but they also determine the form of all additional information. Agents choose what information to get based on what their prior belief is.

In macroeconomics, Paciello & Wiederholt (2014) show that RI implies a different optimal stabilization monetary policy (more similar to inflation targeting) than New Keynesian models. Monetary policy affects stochastic properties of macroeconomy, and thus it also guides agents' attention. The strategy focusing on stabilization of price level is optimal since then firms pay less attention to changing their price together with their markup.

Regarding information provision, Bartoš et al. (2016) show that even for statistical discrimination (mixing of group- and individual- specific signals) it does not suffice to simply provide more individual specific information, since decision makers may not choose 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, on the other hand, can achieve the same by changing payoffs in the early rounds, and leaving the final choice of who to hire, for instance, completely unconstrained.

(ii) *Effects of relative payoffs and volatility.* Similarly, RI implies that subtle changes in relative payoffs can have large effects, too (see the point (ii) in the example in Section 2.1).

(iii) *Complexity and uncertainty reduce welfare by increasing the cost of contemplation.* It is not only the resulting actions that matter, but the cognitive (attention) costs, too. Complicated tax system could imply that agents do not respond to changes in marginal taxes, because they do not observe them. This could make the system less distortive, but not necessarily more efficient. Higher uncertainty implies higher attention of the RI agents, which decreases welfare, too.

RI would typically suggest that policies should result in less uncertainty than what other models of imperfect information would. A related mechanism is described in Matějka (2015), where sellers choose rigid pricing strategies (i.e., with few price-points only), because they save cognitive effort of consumers. This point is closely related to the idea of scarcity (Mullainathan & Shafir, 2013), which could be formalized by RI.

Concerning policies based on information provision to consumers, households, or voters, RI typically warrants recommendations of what type of information to provide, but not of how much, as in the discrimination example above. If the agents are receptive to information, then a subtle change of beliefs can result in large shifts in behavior.

Is transparency good? Sims (2003) argues that it is, since it warrants less coordination and smoother responses to shocks. However, there can be equilibrium effects which imply the opposite. Gaballo (2016) shows that higher transparency of the central bank may not be the optimal thing since it increases macroeconomic volatility by drawing more attention to aggregate shocks. Matějka & Tabellini (2016) show that higher granularity of information can also increase the power of special interest groups.

(iv) *The effect of choice menus.* RI could allow for the refinement of certain recommendations of behavioral economics (Thaler Richard & Sunstein Cass, 2008). For instance, we know that default choices matter, but for whom, when, and how exactly? Possible regulatory implications range from quality standards to restriction as well as expansion of choices at hand. Reasonable regulatory

interventions include: quality regulation; simplification of the choice environment, for instance by restricting price to be a scalar; advising consumers of their expected costs under each option; or choosing on behalf of consumers.

6 Specific ways of going forward

- Macroeconomics: Did employees at financial institutions have the wrong incentives or did they simply make mistakes? Build a new form of dynamic macro (DSGE). Explore agents' coordination of attention as a driver of aggregate fluctuations. Study optimal communication by central banks;
- Behavioral economics, theory: Look for empirically better valid formulation of the baseline model, i.e., different form of the cost of information acquisition (following the recent work of Woodford, Caplin, Dean). Explore coupling of RI with non-standard biases/preferences. e.g, self-control problems, social preferences;
- Field and lab experiments: What is the effect of a time scale of each decision - if we choose under pressure, if we choose repeatedly? How quantitatively relevant is the assumption of a costless choice of perfect attention strategy? Test various motives for information acquisition (beyond RI), and explore which are the strongest in particular settings;
- Applied work: Given the large popularity of logit in the empirical literature (McFadden, 1974), develop methods for inference and structural estimation with RI;
- Test policy implications and comparative statics, e.g., Bartoš et al (2016) discuss implications of quotas in various stages of hiring process for attention allocation;
- Political economy seems an exciting application of rational inattention, not just since Trump and Brexit - as in Matějka & Tabellini (2016), but much more can be done.

7 Conclusion

To be written.

References

- Ambuehl, S., Ockenfels, A., & Stewart, C. (2017). For they know not what they do: Selection effects of incentives when information is costly.
- Andrade, P. & Le Bihan, H. (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, 60(8), 967–982.
- 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.
- Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, (pp. 841–890).
- 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.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Saliency theory of choice under risk. *The Quarterly journal of economics*, (pp. qjs018).
- 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. & 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. (2016). *Rational Inattention, Optimal Consideration Sets, and Stochastic Choice*. Technical report, NYU working paper.
- Caplin, A., Dean, M., & Leahy, J. (2017). *Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy*. Technical report, National Bureau of Economic Research.
- Caplin, A., Leahy, J., & Matejka, F. (2015). *Learning from Market Share when Consumers are Rationally Inattentive*. Technical report, NYU Working Paper.

- Carpini, M. X. D. & Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.
- Cheremukhin, A., Popova, A., Tutino, A., et al. (2011). Experimental evidence on rational inattention. *Federal Reserve Bank of Dallas Working Paper*, 1112.
- Cheremukhin, A., Restrepo-Echavarria, P., & Tutino, A. (2016). A theory of targeted search.
- Coibion, O. & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1), 116–159.
- Cover, T. M. & Thomas, J. A. (2006). *Elements of Information Theory*. Hoboken, NJ: Wiley.
- Dasgupta, K., Mondria, J., et al. (2014). Inattentive importers. *University of Toronto, Department of Economics*.
- De Oliveira, H., Denti, T., Mihm, M., & Ozbek, K. (2014). Rationally inattentive preferences.
- Dean, M. & Neligh, N. (2017). *Experimental Tests of Rational Inattention*. Technical report.
- Debreu, G. (1960). Review of individual choice behavior by R. D. Luce. *American Economic Review*, 50(1).
- DellaVigna, S. (2007). *Psychology and economics: Evidence from the field*. Technical report, National Bureau of Economic Research.
- Denti, T. (2017). *Unrestricted Information Acquisition*. Technical report.
- Eaton, J. & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5), 1741–1779.
- Ellis, A. (2013). Foundations for optimal inattention.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4), 1661–1710.
- 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.
- Gigerenzer, G., Todd, P. M., ABC Research Group, t., et al. (1999). *Simple heuristics that make us smart*. Oxford University Press.
- Handel, B. & Schwartzstein, J. (2017). *Rational inattention with sequential information sampling*. Technical report.
- Hébert, B., Woodford, M., et al. (2016). *Rational inattention with sequential information sampling*. Technical report.
- Hellwig, C. & Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies*, 76(1), 223–251.
- Joo, J. (2017). Buying a larger package with quantity surcharge: Information friction or preference heterogeneity.
- Jung, J., Kim, J.-h., Matějka, F., & SIMS, C. A. (2015). Discrete actions in information-constrained decision problems.
- 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.
- Kamenica, E. & Gentzkow, M. (2011). Bayesian persuasion. *The American Economic Review*, 101(6), 2590–2615.
- Khaw, M. W., Stevens, L., & Woodford, M. (2016). *Discrete Adjustment to a Changing Environment: Experimental Evidence*. Technical report, National Bureau of Economic Research.
- Kőszegi, B. & Matějka, F. (2017). An attention-based theory of mental accounting.
- Kőszegi, B. & Szeidl, A. (2013). A model of focusing in economic choice. *The Quarterly journal of economics*, 128(1), 53–104.

- Lindbeck, A., Weibull, J., et al. (2017). *Investment, Rational Inattention, and Delegation*. Technical report.
- Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review*, 63(3), 326–334.
- Maćkowiak, B., Matějka, F., Wiederholt, M., et al. (2016). The rational inattention filter.
- 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.
- 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. (2013). Stochastic choice and consideration sets. Forthcoming in *Econometrica*.
- 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. & Koszegi, B. (2017). *An Attention-Based Theory of Mental Accounting*. Technical report.
- Matějka, F. & McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *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.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.
- Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, 145(5).
- 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. (2017). *On the Equivalence of Sequential Sampling and Entropy Cost*. Technical report.
- Mullainathan, S. & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.
- Paciello, L. & Wiederholt, M. (2014). Exogenous information, endogenous information, and optimal monetary policy. *The Review of Economic Studies*, 81(1), 356–388.
- Peng, L. (2005). Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis*, 40(02), 307–329.
- Ravid, D. (2014). *Bargaining with rational inattention*. Technical report, Working paper.
- Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, (pp. 999–1033).

- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27.
- Simon, Herbert A. (1971). Designing organizations for an information-rich world.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3).
- Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *The American Economic Review*, 96(2).
- Steiner, J., Stewart, C., & Matějka, F. (2017). Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica*, 85(2), 521–553.
- Stevens, L. (2017). Coarse pricing policies.
- Thaler Richard, H. & Sunstein Cass, R. (2008). Nudge: Improving decisions about health, wealth, and happiness.
- 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. *Review of Economic Studies*, 77(2).
- Wiederholt, M. (2010). Rational inattention. *The New Palgrave Dictionary of Economics (Online Edition ed.)*.
- Woodford, M. (2002). Imperfect common knowledge and the effects of monetary policy. In P. A. et al. (Ed.), *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. Columbia University working paper.
- Woodford, M. (2014). Stochastic choice: An optimizing neuroeconomic model. *The American Economic Review*, 104(5), 495–500.

Woodford, M. (2015). Inattentive valuation and reference-dependent choice. *Unpublished Manuscript, Columbia University*.

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

Yang, M. (2015b). Optimality of debt under flexible information acquisition.