Behavioral Inattention

Xavier Gabaix*

October 16, 2018

Abstract

Inattention is a central, unifying theme for much of behavioral economics. It permeates such disparate fields as microeconomics, macroeconomics, finance, public economics, and industrial organization. It enables us to think in a rather consistent way about behavioral biases, speculate about their origins, and trace out their implications for market outcomes.

This survey first discusses the most basic models of attention, using a fairly unified framework. Then, it discusses the methods used to measure attention, which present a number of challenges on which a great deal of progress has been achieved, although much more work needs to be done. It then examines the various theories of attention, both behavioral and more Bayesian. It finally discusses some applications. For instance, inattention offers a way to write a behavioral version of basic microeconomics, as in consumer theory and Arrow-Debreu. A last section is devoted to open questions in the attention literature.

This chapter is a pedagogical guide to the literature on attention. Derivations are self-contained.

*xgabaix@fas.harvard.edu. Chapter for the *Handbook of Behavioral Economics*, edited by Douglas Bernheim, Stefano DellaVigna and David Laibson. I thank Vu Chau, Antonio Coppola and Lingxuan Wu for excellent research assistance. For comments and suggestions, I thank the editors of this *Handbook*, Hunt Allcott, Sandro Ambuehl, Pedro Bordalo, Colin Camerer, Cary Frydman, Nicola Gennaioli, Sam Gershman, David Hirshleifer, Filip Matejka, Antonio Rangel, Gautam Rao, Alex Rees-Jones, Michael Thaler, Laura Veldkamp, Patrick Warren and Mirko Wiederholt. For sharing their data, I thank Stefano DellaVigna, Josh Pollet, and Devin Pope. I thank the Sloan Foundation for support.

Contents

1 Introduction

It is clear that our attention is limited. When choosing, say, a bottle of wine for dinner, we think about just a few consideration (the price and the quality of the wine), but not about the myriad of components (for example, future income, the interest rate, the potential learning value from drinking this wine) that are too minor. Traditional rational economics assumes that we process all the information that is freely available to us.

Modifying this classical assumption is empirically desirable and theoretically doable. Moreover, it is necessary in order to attain greater psychological realism in economic modeling, and ultimately to improve our understanding of markets and to design better policies. This chapter is a user-friendly introduction to this research. The style of this chapter is that of a graduate course, with pedagogical, self-contained derivations.[1] We will proceed as follows.

Section 2 is a high-level overview. I use a simple framework to model the behavior of an inattentive consumer. Attention is parameterized by a value m, such that $m = 0$ corresponds to zero attention (hence, to a very behavioral model in which the agent relies on a very crude "default" perception of the world) and $m = 1$ to full attention (hence, to the traditional rational model). At a formal level, this simple framework captures a large number of behavioral phenomena: inattention to prices and to taxes; base rate neglect; inattention to sample size; over- and underreaction to news (which both stem from inattention to the true autocorrelation of a stochastic time series); local inattention to details of the future (also known as "projection bias"); global inattention to the future (also known as hyperbolic discounting). At the same time, the framework is quite tractable. I also use this framework to discuss the psychology of attention.

Once this framework is in place, Section 3 discusses methods used to measure inattention empirically: from observational ones like eye-tracking to some that more closely approach a theoretical ideal.[2] I then survey concrete findings in the empirics of attention. Measuring attention is still a hard task – we still have only a limited number of papers that measure attention in field settings — but it is often rewarded with very good publications.

Figure 1 synthesizes this literature measuring attention. On average, the attention parameter estimated in the literature is 0.44, roughly halfway between no attention and full attention. Sensibly, attention is higher when the incentives to pay attention are stronger.

The survey then takes a more theoretical turn, and explores in greater depth the determinants of attention. In Section 4, I start with the most tractable models, those that yield deterministic predictions (that is, for a given situation, there is a deterministic action).

[1] Other surveys exist. DellaVigna (2009) offers a broad and readable introduction to measurement, in particular of inattention, and Caplin (2016) offers a review, from a more information-theoretic point of view.

[2] I positioned this section early in the survey because many readers are interested in the measurement of attention. While a small fraction of the empirical discussion uses some notions from the later theoretical analysis, I wished to emphasize that most of it simply relies on the basic framework of Section 2.

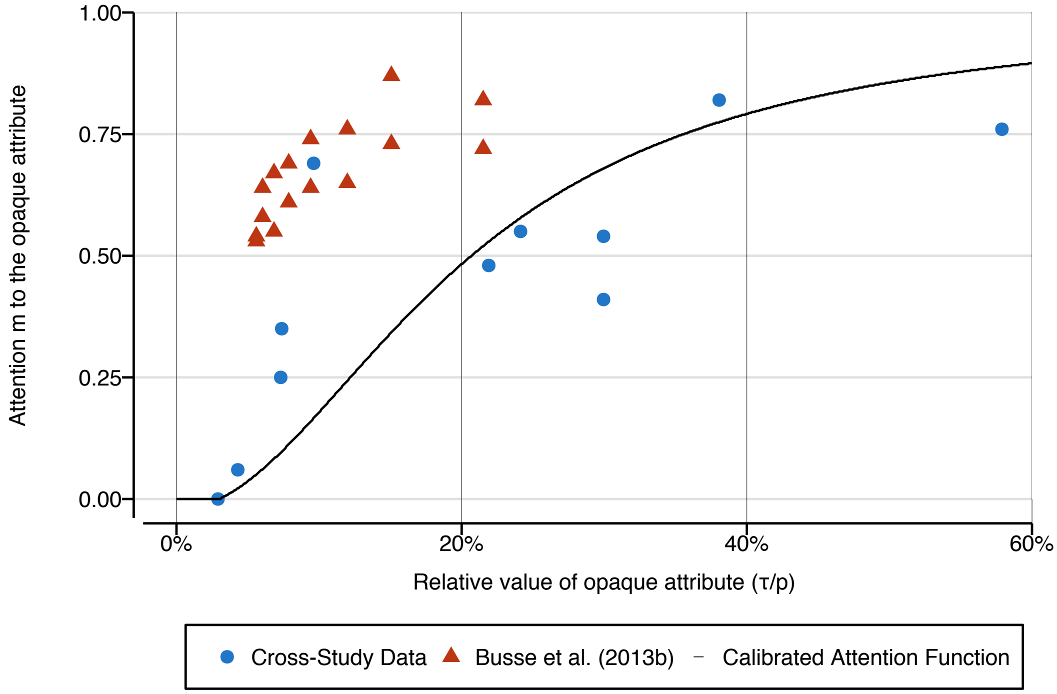

Figure 1: **Attention point estimates (m) vs. relative value of opaque attribute (τ/p), with overlaid calibrated attention function.** Underlying those studies is the following common setup: the full price $q = p + \tau$ of a good is the sum of a based good price p and an opaque price τ. However, an inattentive consumer will perceive $q^s = p + m\tau$, where m captures attention to the opaque attribute. A value $m = 1$ corresponds to full attention, while $m = 0$ implies complete inattention. This figure shows (*circles*) point estimates of the attention parameter m in a cross-section of recent studies (shown in Table 1), against the estimated relative value of the opaque τ add-on attribute relative to the base good or quantity (τ/p). The overlaid curve shows the corresponding calibration of the quadratic-cost attention function in (26), where we impose $\alpha = 1$ and obtain calibrated cost parameters $\bar{\kappa} = 3.0\%$, $q = 22.4$ via nonlinear least squares. Additionally, for comparison, we plot analogous data points (*triangles*) for subsamples from the study of Busse, Lacetera, Pope, Silva-Risso, and Sydnor (2013b), who document inattention to left-digit remainders in the mileage of cars sold at auction, broken down along covariate dimensions. Each data point in the latter series corresponds to a subsample including all cars with mileages within a 10,000 mile-wide bin (e.g., between 15,000 and 25,000 miles, between 25,000 miles and 35,000 miles, and so forth). For each mileage bin, we include data points from both retail and wholesale auctions.

Some models rely on the plain notion that more important dimensions should be given more attention – this is plain, but not actually trivial to capture in a tractable model. Some other models put the accent on proportional thinking rather than absolute thinking: in this view, people pay more attention to relatively more important dimensions.

Section 6 then covers models with stochastic decisions – given an objective situation, the prediction is a probability distribution over the agents' actions. These are more complex models. We will cover random choice models, as well as the strand of the literature in which agents pay to acquire more precise signals. We will then move on to the entropy-based penalty that has found particular favor among macroeconomists.

What are the consequences of introducing behavioral inattention into economic models? This chapter reviews many such implications, in industrial organization, taxation, macroeconomics, and other areas. Section 5 presents something elementary that helps unify all of these strands: a behavioral version of the most basic chapter of the microeconomics textbook à la Varian (1992), including consumer theory, and Arrow-Debreu. As most of rational economics builds on these pillars, it is useful to have a behavioral alternative.

Section 7 moves on to dynamic models. The key pattern there is that of delayed reaction: people react to novel events, but with some variable lag. Sometimes, they do not attend to a decision altogether – we have then a form of "radical inattention". Useful approaches in this domain include models that introduce costs from changing one's action, or costs from changing one's thinking (these are called "sticky action" and "sticky information" models, respectively). We will also discuss models of "habit formation", and models in which agents optimally choose how to acquire information over time. We will understand the benefits and drawbacks of each of these various models.

Finally, Section 8 proposes a list of open questions. The appendices give mathematical complements and additional proofs.

Notation I will typically use subscripts for derivatives, e.g. $f_x(x,y) = \frac{\partial}{\partial x} f(x,y)$, except when the subscript is i or t, in which case it is an index for a dimension i or time t.

I differentiate between the true value of a variable x, and its subjectively perceived value, x^s (the s stands for: "*s*ubjectively perceived value", or sometimes, the value given by *s*alience or *s*parsity).

2 A Simple Framework for Modeling Attention

In this section I discuss a simple framework for thinking about behavioral inattention in economic modeling, and I argue that this simple structure is useful in unifying several themes of behavioral economics, at least in a formal sense. I start from a basic example of prior-anchoring and adjustment toward perceived signals in a model with Gaussian noise, and then move to a more general model structure that captures behavioral inattention in a deterministic fashion.

2.1 An introduction: Anchoring and adjustment via Gaussian signal extraction

Suppose there is a true value x, drawn from a Gaussian distribution $\mathcal{N}\left(x^d, \sigma_x^2\right)$, where x^d is the default value (here, the prior mean) and variance σ_x^2. However, the agent does not know this true value, and instead she receives the signal

$$s = x + \varepsilon \tag{1}$$

where ε is drawn from an independent distribution $\mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. The agent takes the action a. The agent's objective function is $u(a,x) = -\frac{1}{2}(a-x)^2$, so that if she's rational, the agent wants to take the action that solves: $\max_a \mathbb{E}\left[-\frac{1}{2}(a-x)^2 | s\right]$. That is, the agent wants to guess the value of x given the noisy signal s. The first-order condition is

$$0 = \mathbb{E}\left[-(a-x) | s\right] = \mathbb{E}\left[x \mid s\right] - a$$

so that the rational thing to do is to take the action $a(s) = \hat{x}(s)$, where $\hat{x}(s)$ is the expected value of x given s,

$$\hat{x}(s) = \mathbb{E}\left[x \mid s\right] = \lambda s + (1-\lambda) x^d \tag{2}$$

with the dampening factor[3]

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \in [0,1]. \tag{3}$$

Equation (2) says that the agent should anchor at the prior mean x^d, and partially adjust (with a shrinkage factor m) toward the signal s. The average action $\bar{a}(x) := \mathbb{E}\left[a(s) | x\right]$ is then:

$$\bar{a}(x) = mx + (1-m) x^d \tag{4}$$

with $\lambda = m$. In the rest of this paper, I will analogously use the common notation m to index the limited reaction to a stimulus. In this Bayesian subsection, m is microfounded as a Bayesian ratio of variances λ, but in more general contexts (that we shall see shortly) that need not be the case. In the limit case of an agent who is perfectly well-informed, $\sigma_\varepsilon = 0$, and $m = 1$. In the opposite limit of a very confused agent with infinite noise ($\sigma_\varepsilon \to \infty$), $m = 0$ and the agent relies entirely on her default value.

This is akin to the psychology of "anchoring and adjustment". As Tversky and Kahneman (1974, p. 1129) put it: "People make estimates by starting from an initial value that is

[3] The math used here should be familiar, but a refresher is given in Appendix A.

adjusted to yield the final answer [...]. Adjustments are typically insufficient". Here, agents start from the default value x^d and on expectation adjusts it toward the truth x. Adjustments are insufficient, as $m \in [0, 1]$, because signals are generally imprecise.

Most models are variants or generalizations of the model in equation (4), with different weights m (endogenous or not) on the true value. In this review, I discuss a first class of models that eliminates the noise, as not central, at least for the prediction of the average behavior (see Section 4). I then discuss a second class of frameworks in which noise is central – which often leads to more complicated models (see Section 6).

Before discussing these variants and generalizations, I will present a simple formal framework for modeling inattention.

2.2 Models with deterministic attention and action

Most models of inattention have the following common structure. The agent should maximize

$$\max_a u(a, x) \tag{5}$$

where, as before, a is an action (possibly multidimensional), and x is a vector of "attributes", e.g. price innovations, characteristics of goods, additional taxes, deviations from the steady state and so on. So a rational agent will choose $a^r(x) = \operatorname{argmax}_a u(a, x)$.

The behavioral agent replaces this by an "attention-augmented decision utility",

$$\max_a u(a, x, m) \tag{6}$$

where m is a vector that will characterize the degree of attention – i.e. the agent's subjective *m*odel of the world. She takes the action

$$a(x, m) = \operatorname*{argmax}_a u(a, x, m).$$

In inattention models, we will very often take (as in Gabaix and Laibson 2006; Chetty, Looney, and Kroft 2009; DellaVigna 2009; Gabaix 2014)[4]

$$u(a, x, m) = u\left(a, m_1 x_1 + (1 - m_1) x_1^d, \ldots, m_n x_n + (1 - m_n) x_n^d\right). \tag{9}$$

where $m_i \in [0, 1]$ is the attention to variable x_i, and where x_i^d is the "default value" for variable i — it is the value that spontaneously comes to mind with no thinking. This is as if x_i is replaced by the $s$ubjectively perceived x_i^s:

$$x_i^s \coloneqq m_i x_i + (1 - m_i) x_i^d, \tag{10}$$

When $m_i = 0$, the agent "does not think about x_i", i.e. replaces x_i by $x_i^s = x_i^d$.[5] When $m_i = 1$, she perceives the true value ($x_i^s = x_i$). When $0 < m_i < 1$, she perceives partially the true value, though not fully; one microfoundation for that may be the model of Section (2.1), where the agent reacts partially to a noisy signal.[6]

The default x_i^d is typically the prior mean of x_i. However, it can be psychologically richer. For instance, if the mean price of good i is $\mathbb{E}[x_i] = \$10.85$, then the normatively simplest default is $x_i^d = \mathbb{E}[x_i] = \$10.85$. But the default might be a truncated price, e.g. $x_i^d = \$10$ (see Lacetera, Pope, and Sydnor, 2012).

To fix ideas, take the following quadratic example:

$$u(a, x) = -\frac{1}{2}\left(a - \sum_{i=1}^{n} b_i x_i\right)^2. \tag{11}$$

with 0 for the default values ($x_i^d = 0$). Then, the traditional optimal action is

$$a^r(x) = \sum_{i=1}^{n} b_i x_i, \tag{12}$$

where the r superscript is as in the traditional $r$ational actor model. For instance, to choose

[4]Some other models (e.g. Bordalo, Gennaioli, and Shleifer 2013, reviewed below in section 4.2), take the form

$$u(a, x, m) = u(a, m_{a1} x_1, \ldots, m_{an} x_n) \tag{7}$$

where the attention parameters depend on the goods and the action, so that m has dimensions $A \times n$, where A is the number of goods. We keep a simpler form now, as it allows us to use continuous actions (so $A = \infty$) and take derivatives with respect to action a. Also, the attention parameter m is often deployed multiplicatively "outside the utility", as in

$$u(a, x, m) = mu(a, x) + (1 - m) u\left(a, x^d\right) \tag{8}$$

Still, in most cases with continuous actions placing m inside the utility function makes the model more tractable and more expressive.

[5]Responding to the "default" x_i^d (corresponding to $m_i = 0$) is referred to in the psychology literature alternatively as an automatic, habitual or prepotent response.

[6]The linear form is largely for analytical convenience, and is not essential.

a, the decision maker should consider not only innovations x_1 in her wealth, and the deviation of GDP from its trend, x_2, but also the impact of interest rate, x_{10}, demographic trends in China, x_{100}, recent discoveries in the supply of copper, x_{200}, etc. There are an innumerable number of factors that should in principle be taken into account. A sensible agent will "not think" about most of these factors, especially the less important ones. We will formalize this notion.

After attention m is chosen, the behavioral agent optimizes under her simpler representation of the world, i.e. chooses

$$a^s(x,m) = \sum_{i=1}^{n} b_i m_i x_i,$$

so that if $m_i = 0$, she doesn't pay attention to dimension i.

2.3 Unifying behavioral biases: Much of behavioral economics may reflect a form of inattention

Let us see some examples that will show how this form captures – at a formal level at least – many themes of behavioral economics. We shall see that many behavioral biases share a common structure: people anchor on a simple perception of the world, and partially adjusts toward it. Conceptually, there is a "default, simple model" that spontaneously comes to mind,[7] and there is a "true model" that's only imperfectly perceived. Attention m parameterizes the particular convex combination of the default and true models that corresponds to the agent's perception.[8]

This feeling of unity of behavioral economics is tentative — and in part my own speculation, rather than an agreed-upon truth. Still, I find it useful to make a case for it in this chapter. One could imagine doing "attentional interventions" (changing m) to investigate this type of hypothesis experimentally (see e.g. Lombardi and Fehr (2018) for a step in that direction).

Different cognitive functions underlying limited attention Attention involves the allocation of scarce cognitive resources. In the examples below, I will not go much into the detailed exploration of what those scarce resources are. They include limited working memory, limited ability to carry out complex algorithms, or lack of readily-accessible knowledge.[9]

[7]Kahneman and Frederick (2002) call something very close to this idea "attribute substitution" – the use of a simplified model, or question, that is cognitively easier than the original problem.

[8]Gabaix (2014, Online Appendix, Section 9.A) contains an early version of this list, with a fuller treatment some of the biases below, including the endogenization of attention m.

[9]For instance, in the contexts of the examples discussed in this subsection, limited working memory presumably drives the inattention to taxes and the left-digit bias; limited ability to carry out complex algorithms is relevant for the exponential growth bias; and the lack of readily-accessible knowledge is important for the

What drives attention are components of information processing, which are subsumed by the common abstraction m that I use in all these examples.

2.3.1 Inattention to true prices and shrouding of add-on costs

Let us illustrate the misperception of numbers in the context of prices. We start from a default price p^d. The new price is p, while the subjectively price p^s perceived by the agent is

$$p^s = mp + (1-m)\, p^d. \tag{13}$$

The perceived price p^s responds only partially (with a strength m) to the true price p, perhaps as in the noisy-signal microfoundation of Section 2.1.

Take the case without income effects, where the rational demand is $c^r(p)$. Then, the demand of a behavioral agent is $c^s(p) = c^r(p^s(p,m))$. So, the sensitivity of demand to price is $c^s(p)' = mc^r(p^s)'$. The demand sensitivity is muted by a factor m.

We can also reason in logarithmic space, so that the perceived price is:

$$p^s = (p)^m \left(p^d\right)^{1-m}. \tag{14}$$

In general, the psychology of numbers (Dehaene 2011) shows that the latter formulation (in log space) is psychologically more accurate. This Cobb-Douglas formulation is sometimes used in Gabaix (2014) and Khaw, Li, and Woodford (2017) – the latter framework is a stochastic one, and explores how the model's stochasticity is useful to match the empirical evidence.

Similar reasoning applies to the case of goods sold with separate add-ons. Suppose that the price of a base good is p, and the price of an add-on is $\hat{p}$. The consumer might only partially see the add-on, such that she perceives the add-on cost to be $\hat{p}^s = m\hat{p}$. As a result, the myopic consumer perceives total price to be only $p + m\hat{p}$, while the full price is $p + \hat{p}$. Such myopic behavior allows firms to shroud information on add-on costs from consumers in equilibrium (Gabaix and Laibson 2006).

2.3.2 Inattention to taxes

Suppose that the price of a good is p, and the tax on that good is τ. Then, the full price is $q = p + \tau$. But a consumer may pay only partial attention to the tax, so the perceived tax is $\tau^s = m\tau$, and the perceived price is $q^s = p + m\tau$. Chetty, Looney, and Kroft (2009) develop this model, develop the theory of tax incidence, and measure attention to sales taxes in routine consumer purchases. Farhi and Gabaix (2017) provide a systematic theory

inattention to the future, in that costly simulations may need to be performed.

of optimal taxation (encompassing the Ramsey, Pigou and Mirrlees problems) with this type of misperceptions and other biases.

2.3.3 Nominal illusion

The notion of nominal illusion is very related. Suppose that the real price change is q, inflation is π, and the nominal price change is p, so that the real price change is:

$$q = p - \pi.$$

People will often anchor to the nominal price, without removing enough inflation. The real price that they will perceive is:

$$q^s = p - m\pi, \tag{15}$$

where $m = 0$ signifies full nominal illusion. Recent research has shown that nominal anchoring has a surprising impact on stock market analysts (Roger et al. (2018)), and may be even important for concrete outcomes (Shue and Townsend (2018)).

2.3.4 Hyperbolic discounting: Inattention to the future

In an intertemporal choice setting, suppose that true utility is $U_0 = \sum_{t=0}^{\infty} \delta^t u_t$, and call $U_1 = \sum_{t=1}^{\infty} \delta^{t-1} u_t$ the continuation utility, so that

$$U_0 = u_0 + \delta U_1. \tag{16}$$

A present-biased agent (Laibson 1997; O'Donoghue and Rabin 1999) will instead see a perceived utility

$$U_0^s = u_0 + m\delta U_1. \tag{17}$$

The parameter m is equivalent here to the parameter β in the hyperbolic discounting literature.[10]

Still, the normative interpretation is different. If the $m = \beta$ is about misperception, then the favored normative criterion is to maximize over the preferences of the rational agents, i.e maximize $u_0 + \delta U_1$ (Farhi and Gabaix 2017). In contrast, with the multiple selves interpretations usually associated with hyperbolic discounting (Thaler and Shefrin 1981; Fudenberg and Levine 2012) the welfare criterion is not so clear as one needs to trade off the utility of several "selves". Bernheim and Rangel (2009) similarly advocate for an agnostic

[10] Gabaix and Laibson (2017) develop an interpretation of discounting via cognitive myopia much along these lines. Abela et al. (2012) present laboratory evidence that damage to the hippocampus (an area of the brain associated with both working memory and long-term memory) increases impulsive decision-making. This may be because, say, the memory of eating peaches in the past is necessary to simulate the future utility from eating a peach.

welfare criterion for behavioral models that does not privilege the preferences of the rational agent.

2.3.5 When will we see overreaction vs. underreaction?

Suppose that a variable y_{it} follows a process $y_{i,t+1} = \rho_i y_{it} + \varepsilon_{it}$, and ε_{it} is an i.i.d. innovation with mean zero. The decision-maker however has to deal with many such processes, with various autocorrelations, that are ρ^d on average. Hence, for a given process, she may not fully perceive the autocorrelation, and instead use the subjectively perceived autocorrelation ρ_i^s, as in

$$\rho_i^s = m\rho_i + (1-m)\,\rho^d. \tag{18}$$

That is, instead of seeing precisely the fine nuances of each AR(1) process, the agent anchors on a common autocorrelation ρ^d, and then adjusts partially toward the true autocorrelation of variable y_{it}, which is ρ_i. The agent's prediction is $\mathbb{E}_t^s\left[y_{i,t+k}\right] = \left(\rho_i^s\right)^k y_{i,t}$, so that

$$\mathbb{E}_t^s\left[y_{i,t+k}\right] = \left(\frac{\rho_i^s}{\rho_i}\right)^k \mathbb{E}_t\left[y_{i,t+k}\right]$$

where $\mathbb{E}_t^s$ is the subjective expectation, and $\mathbb{E}_t$ is the rational expectation. Hence, the agent exhibits *overreaction* for processes that are less autocorrelated than ρ^d, as $\frac{\rho_i^s}{\rho_i} > 1$, and *underreaction* for processes that are more autocorrelated than ρ^d, as $\frac{\rho_i^s}{\rho_i} < 1$.[11]

For instance, if the growth rate of a stock price is almost not autocorrelated, and the growth rate of earnings has a very small positive autocorrelation, people will overreact to past returns by extrapolating too much (Greenwood and Shleifer 2014). On the other hand, processes that are quite persistent (say, inflation) will be perceived as less autocorrelated than they truly are, and agents will underreact by extrapolating too little (as found by Mankiw, Reis, and Wolfers 2003).[12]

2.3.6 Prospect theory: Inattention to the true probability

There is a literature in psychology (but not widely known by behavioral economists) that finds that probabilities are mentally represented in "log odds space". Indeed, in their survey Zhang and Maloney (2012) assert that this perceptual bias is "ubiquitous" and give a unified account of many phenomena. If $p \in (0,1)$ is the probability of an event, the log odds are $q := \ln \frac{p}{1-p} \in (-\infty, \infty)$. Then, people may misperceive numbers as in (2) and (4), i.e. their

[11] This sort of model is used in Gabaix (2016, 2018).

[12] As of now, this hypothesis for the origin of under/overreaction has not been tested, but it seems plausible and has some indirect support (e.g. from Bouchaud, Krueger, Landier, and Thesmar 2016). A meta-analysis of papers on under/overreaction, perhaps guided by the simple analytics here, would be useful. There are other ways to generate over / underreaction, e.g. Daniel, Hirshleifer, and Subrahmanyam (1998), which relies on investor overconfidence about the accuracy of their beliefs, and biased self-attribution (see also Hirshleifer et al. (2011)).

median perception is[13]

$$q^s = mq + (1-m)\, q^d. \tag{19}$$

Then, people transform their perceived log odds $q^s = \ln \frac{p^s}{1-p^s}$ into a perceived probability $p^s = \frac{1}{1+e^{-q^s}}$, that is $p^s = \pi(p)$ with:

$$\pi(p) = \frac{1}{1 + \left(\frac{1-p}{p}\right)^m \left(\frac{1-p^d}{p^d}\right)^{1-m}}, \tag{20}$$

which is the median perception of a behavioral agent: we have derived a probability weighting function $\pi(p)$. This yields overweighting of small probabilities (and symmetrically underweighting of probabilities close to 1).[14] Psychologically, the intuition is as follows: a probability of 10^{-6} is just too strange and unusual, so the brain "rectifies it" by dilating it toward a more standard probability such as $p^d \simeq 0.36$, and hence overweighting it.[15] This is exactly as in the simple Gaussian updating model of Section 2.1, done in the log odds space, and gives a probability weighing function much like the one in prospect theory (Kahneman and Tversky 1979). This theme is pursued (with a different functional form, not based on the psychology of the log odds space surveyed in Zhang and Maloney 2012) by Steiner and Stewart (2016).[16]

Likewise, for the "diminishing sensitivity" part of prospect theory, one can appeal to the distortions of payoffs as in (14): a payoff is X is perceived as $X^s = X^{m'} \left(X^d\right)^{1-m'}$ (this is done, with experimental evidence, by Khaw et al. 2017). Putting the two themes above (distortions of payoff and distortions of probability) together, we get something much like prospect theory: the perceived value of a gamble offering X with probability p and Y with probability $1-p$ is:

$$V = \left[\pi(p)\, X^{m'} - \pi(1-p)\, Y^{m'}\right] \left(X^d\right)^{1-m'} \tag{21}$$

In the rational model we have $m' = m = 1$, so that $\pi(p) = p$.

How to obtain loss aversion? To get it, we'd need to assume a "pessimistic prior", saying that the typical gamble in life has negative expected value.[17] For instance the default probability for loss events is higher than the default probability in gains events. This might create loss aversion. A complete treatment of this issue is left to future research.

[13] I use the median, because perception contains noise around the mean, and the median is more tractable when doing monotonous non-linear transformations.

[14] This behavioral bias is well-documented empirically: DellaVigna and Pope 2018 conduct a meta-analysis of the relevant experimental (and quasi-experimental) literature – averaging across studies, they estimate a mean probability weight of 6% for a true probability of 1%.

[15] Here I take $p^d \simeq 0.36$ as this is the crossover value where $p = p^s$ in Prelec's (1998) survey.

[16] See also Bordalo, Gennaioli, and Shleifer (2012).

[17] In that hypothesis, people would have "pessimistic prior" about general exchanges in life, and be "overconfident" about their *own* abilities.

2.3.7 Projection bias: Inattention to future circumstances by anchoring on present circumstances

Suppose that I need to forecast x_t, a variable at time t. I might use its time-zero value as an anchor, i.e. $x_t^d = x_0$. Then, my perception at time zero of the future variable is

$$x_t^s = mx_t + (1-m)\,x_0, \tag{22}$$

hence the agent exhibits projection bias. See also Loewenstein, O'Donoghue, and Rabin (2003) for the basic analysis, and Chang, Huang, and Wang (2018) as well as Busse, Knittel, and Zettelmeyer (2013a) for empirical evidence in support of this.

2.3.8 Coarse probabilities and partition dependence

Suppose that there are K disjoint potential outcomes $E_1, ..., E_K$ (which form a partition of the event space). It is hard to know their probability, so a sensible default probability for events $E_1, ..., E_K$ is a uniform prior, $P^d(E) = \frac{1}{K}$ for all $E = E_1, \ldots, E_K$. Then, people may partially adjust towards the truth, and perceive the probability of event E as

$$P^s(E) = mP(E) + (1-m)P^d(E)\,. \tag{23}$$

A correlated notion is that of "partition dependence". When people are asked "what's the probability of dying of cancer" vs "what's the probability of dying of lung cancer, or brain cancer, or breast cancer" etc., their assessed probabilities change, in a way that's partition-dependent (Sonnemann et al. 2013). So, as "cancer" is divided into more categories, people perceive that the likelihood of cancer is higher.

2.3.9 Base-rate neglect: Inattention to the base rate

In base-rate neglect (Tversky and Kahneman 1974) people seem to react a little bit to the base rate, but not enough. A simple way to capture that is is posit that they anchor their perceived based rate on the "uninformative" base rate $P^d(E)$, which is a uniform distribution on the values of E, as in (23). Then, they use Bayes rule with this. This generates base rate neglect.[18]

2.3.10 Correlation neglect

Another way to simplify a situation is to imagine that random variables are uncorrelated, as shown by Enke and Zimmermann (forthcoming). To formalize this, let us say that the true probability of variables $y = (y_1, \ldots, y_n)$ is a joint probability $P(y_1, ..., y_n)$, and the

[18] Grether (1980) uses a logarithmic variant of this equation. One could imagine a deeper model where people do not used Bayes' rule altogether, but that would take use too far afield.

(marginal) distribution of y_i is $P_i(y_i)$. Then the "simpler" default probability is the joint density assuming no correlation $P^d(y) = P_1(y_1)\dots P_n(y_n)$. Correlation neglect is captured by a subjective probability $P^s(y) = mP(y) + (1-m)P^d(y)$.

2.3.11 Insensitivity to sample size

Tversky and Kahneman (1974) show the phenomenon of "insensitivity to sample size". One way to model this is as follows: the true sample size N is replaced by a perceived sample size $N^s = \left(N^d\right)^{1-m} N^m$, and agents update based on that perceived sample size.

2.3.12 Insensitivity to predictability / Misconceptions of regression to the mean / Illusion of validity: Inattention to randomness and noise

Tversky and Kahneman (1974) report that, when people see a fighter pilot's performance, they fail to appreciate the role of mean reversion. Hence, if the pilot does less well the next time, they attribute this to lack of motivation, for instance, rather than reversion to the mean.

Call x the pilot's core ability, and $s_t = x + \varepsilon_t$ the performance on day t, where ε_t is an i.i.d. Gaussian noise term and x is drawn from a $N(0, \sigma_x^2)$ distribution. Given the performance s_t of, say, an airline pilot, an agent predicts next period's performance (Tversky and Kahneman 1974). Rationally, she predicts $\bar{s}_{t+1} := \mathbb{E}[s_{t+1} \mid s_t] = \lambda s_t$ with $\lambda = \frac{1}{1+\sigma_\varepsilon^2/\sigma_x^2}$.

However, a behavioral agent may "forget about the noise", i.e. in her perceived model, $\text{Var}^s(\varepsilon) = m\sigma_\varepsilon^2$. If $m = 0$, they don't think about the existence of the noise, and answer $\overline{y}_{t+1}^s = y_t$. Such agent will predict:

$$\bar{s}_{t+1}^s = \frac{1}{1 + \frac{m\sigma_\varepsilon^2}{\sigma_a^2}} s_t.$$

Hence, behavioral agents with $m = 0$, who fully ignore randomness and noise, will just expect the pilot to do next time as he did last time.

2.3.13 Overconfidence: Inattention to my true ability

If x is my true driving ability, with overoptimism my prior x^d may be a high ability value; perhaps the ability of the top 10% of drivers. There are explanations for this kind of overconfidence: people often have to advocate for themselves (on the job, in the dating scene etc.), and one is a better advocate of one's superiority if one actually believes in it (Mercier and Sperber 2011, see also Bénabou and Tirole 2002). Rosy perceptions come from this high default ability (for myself), coupled with behavioral neglect to make the adjustment. A related bias is that of "overprecision", in which I think that my beliefs are more accurate than they are: then x is the true precision of my signals, and x^d is a high precision.

2.3.14 Cursedness: Inattention to the conditional probability

In a game theoretic setting, Eyster and Rabin (2005) derive the equilibrium implications of *cursedness*, a behavioral bias whereby players underestimate the correlation between their strategies and those of their opponents. The structure is formally similar, with cursedness χ being $1 - m$: the agent forms a belief that is an average of m times to the true probability, and $1 - m$ times a simplified, naïve probability distribution.[19]

2.3.15 Left-digit bias: Inattention to non-leading digits

Suppose that a number, in decimal representation, is $x = a + \frac{b}{10}$, with $a \geq 1$ and $b \in [0, 1)$. An agent's perception of the number might be

$$x^s = a + m\frac{b}{10} \tag{24}$$

where a low value of $m \in [0, 1]$ indicates left-digit bias. Lacetera, Pope, and Sydnor (2012) find compelling evidence of left-digit bias in the perception of the mileage of used cars sold at auction.[20]

2.3.16 Exponential growth bias

Many people appear to have a hard time compounding interest rates, something that Stango and Zinman (2009) call the exponential growth bias. Here, if $x = (1 + r)^t$ is the future value of an asset, then the simpler perceived value is $x^d = 1 + rt$, and the perceived growth is just $x^s = mx + (1 - m)\,x^d$. [21]

2.3.17 Taking stock of these examples

These examples, I submit, illustrate that the simple framework above allows one to think in a unified way about a wide range of behavioral biases, at least in their formal structure. There are four directions in which such baseline examples can be extended. Here I give a brief outline of these four directions, along with a number of examples that are discussed at greater length in later sections of this survey:

1. In the "theoretical economic consequences" direction, economists work out the consequences of partial inattention, e.g. in market equilibrium, or in the indirect effects of

[19] For empirical evidence linking cursedness to limitations in cognitive skill, see also Carroll, Bazerman, and Maury (1988), who show that players in two-person negotiation games generally lack knowledge of how to compute conditional expectations from their opponent's point of view.

[20] A variant it to use exponentially-decreasing weights. If the number is $x = \sum_{i=n}^{N} a_i 10^i$ with $N > n$, its perception is $x^s = \sum_{i=n}^{N} a_i 10^i m^{N-n}$.

[21] Kusev et al. (2018) make further progress in modeling anchoring on linear or other non-exponential approximations in the context of subjective time-series forecasting.

all this.

2. In the "empirical economic measurement" direction, researchers estimate attention m.

3. In the "basic psychology" direction, researchers think more deeply about the "default perception of the world", i.e. what an agent perceives spontaneously. Psychology helps determine this default.[22]

4. In the "endogenization of the psychology" part, attention m is endogenized. This can be helpful, or not, in thinking about the two points above. Typically, endogenous attention is useful to make more refined predictions, though most of those remain to be tested. In the meantime, a simple quasi-fixed parameter like m is useful to have, and allows for parsimonious models – a view forcefully argued by Rabin (2013).

2.4 Psychological underpinnings

Here is a digest of some features of attention from the psychology literature. Pashler (1998) and Nobre and Kastner (2014) handbook offer book-length surveys on the psychology of attention, with primary emphasis on perception. Knudsen (2007) offers a neuroscience-oriented perspective.

2.4.1 Conscious versus unconscious attention

Systems 1 and 2. Recall the terminology for mental operations of Stanovich (1999) and Kahneman (2003), where "system 1" is an intuitive, fast, largely unconscious and parallel system, while "system 2" is a deliberative, slow, relatively conscious and serial system.

System 2, working memory, and conscious attention. It is likely that we do not consciously contemplate thousands of variables when dealing with a specific problem. For instance, research on working memory documents that people handle roughly "seven plus or minus two" items (Miller 1956). At the same time, we do know – in our long term memory – about many variables, x. Hence, we can handle consciously relatively few m_i that are different from 0.[23]

System 1 / Unconscious attention monitoring. At the same time, the mind contemplates unconsciously thousands of variables x_i, and decides which handful it will bring up for conscious examination (that is, whether they should satisfy $m_i > 0$). For instance, my system 1 is currently monitoring if I'm too hot, thirsty, low in blood sugar, but also in the presence of a venomous snake, and so forth. This is not done consciously. But if a variable

[22] It would be nice to have a "meta-model" for defaults, unifying the superficial diversity of default models.

[23] In attentional theories, System 1 chooses the attention (e.g. as in Step 1 in Proposition 4.1), while the decision is done by System 2 (as in Step 2 in the same Proposition).

becomes very alarming (e.g. a snake just appeared), it will be "brought to consciousness" – that is, to the attention of system 2. Those are the variables with an $m_i > 0$.

To summarize, System 1 chooses the m_i's in an unconscious, parallel fashion, while System 2 takes a decision based on the few elements that have been brought to consciousness (i.e. with $m_i > 0$).

This view has a good degree of support in psychology. In a review paper, Dehaene et al. (2017) say:

> William James described attention as "the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought". This definition is close to what we mean by [...] the selection of a single piece of information for entry into the global workspace. There is, however, a clear-cut distinction between this final step, which corresponds to conscious access, and the previous stages of attentional selection, which can operate unconsciously. Many experiments have established the existence of dedicated mechanisms of attention orienting and shown that, like any other processors, they can operate nonconsciously.

2.4.2 Reliance on defaults

What guess does one make when there is no time to think? This is represented by the case $m = 0$: then, variables x are are replaced by their default value, which could be some plain average value, or a crude heuristics. This default model ($m = 0$), and the default action a^d (which is the optimal action under the default model) corresponds to "system 1 under extreme time pressure". The importance of default actions has been shown in a growing literature (e.g. Madrian and Shea 2001; Carroll, Choi, Laibson, Madrian, and Metrick 2009).[24] Here, the default model is very simple (basically, it is "do not think about anything"), but it could be enriched, following other models (e.g. Gennaioli and Shleifer 2010).[25]

2.4.3 Neuroscience: The neural correlates of "mental cost" and "limited attention"

It would be great to know the neural correlates of "limited attention" and things like "mental costs" and "mental fatigue" – what exactly is this scarce resource in the brain? Sadly

[24]This literature shows that default actions matter, not literally that default variables matters. One interpretation is that the action was (quasi-)optimal under some typical circumstances (corresponding to $x = 0$). An agent might not wish to think about extra information (i.e., deviate from $x = 0$), and hence deviate from the default action.

[25]There is no systematic theory of the default yet. This default is a close cousin of what Bayesians call the prior (in fact we might imagine that subjective priors would be very crude, for instance a uniform prior). There is no systematic theory of where the prior comes from either.

neuroscience research has not found it (Section 4 of Kurzban et al. (2013) is a clear discussion of this). A early proposal was glucose in the brain, but it has been discredited. Hence, the attention literature needs to theorize it without clear guidance from the neuro literature.

2.4.4 Other themes

If the choice of attention is largely unconscious, this leads to the curious choice of "attentional blindness". Mack and Rock (1998) studied attentional blindness extensively, and the now canonical experiment for this is the "gorilla" experiment of Simons and Chabris (1999). When asked to perform a task that requires full attentional resources, subjects often didn't see a gorilla in the midst of the experiment.

Another rich field is that of visual attention. Treisman and Gelade (1980) is classic paper in the field, and Carrasco (2011) surveys the literature. One theme – not well integrated by the economics literature, is the "extreme seriality of thought" (see Huang and Pashler 2007): in the context of visual attention, it means that people can process things only one color at a time. In other contexts, like the textbook rabbit / duck visual experiment, it means that one can see a rabbit or a duck in a figure, but not both at the same time.

From an economic point of view, serial models that represent the agent's action step by step tend to be complicated though instructive (see Rubinstein 1998; Gabaix, Laibson, Moloche, and Weinberg 2006; Caplin, Dean, and Martin 2011; Fudenberg, Strack, and Strzalecki 2017, which feature various forms of information search). Because of the limited applicability of process-based models, more "outcome-based models", which directly give the action rather than the intermediary steps, are typically easier to use in economic applications.

3 Measuring Attention: Methods and Findings

I now turn to the literature on the empirical measurement of attention. I first provide a broad taxonomy of the approaches taken in the literature, and then discuss specific empirical findings. The recent empirical literature on inattention has greatly advanced our ability to understand behavioral biases quantitatively, such that we can now begin to form a synthesis of these results. I present such a synthesis at the end of this section.

3.1 Measuring attention: Methods

There are essentially five ways to measure attention:[26]

1. Deviations from an optimal action (this requires to know the optimal action)

[26] This classification builds on DellaVigna's (2009).

2. Deviations from normative cross-partials, e.g. from Slutsky symmetry (this does not require to know the optimal action)

3. Physical measurement, e.g. time on task and eye-tracking.

4. Surveys: eliciting people's beliefs.

5. Imputations from the impact of attentional interventions: impact of reminders, of advice.

As we will see, methods 3-5 can show that attention is not full (hence, help reject the naïve rational and costless-cognition model) and measure correlates of attention (e.g. time spent), and 1 and 2 measure attention as defined in this survey (e.g., measure the parameter m).[27]

3.1.1 Measuring inattention via deviation from an optimal action

Suppose the optimal action function is $a^r(x) := \operatorname{argmax}_a u(a, x)$, and the behavioral action is $a^s(x) = a^r(mx)$. Then, the derivative of the action with respect to x is: $a^s_x(x) = m a^r_x(mx)$. Therefore attention can be measured as[28]

$$m = \frac{a^s_x}{a^r_x}.$$

Hence, the attention parameter m is identified by the ratio of the sensitivities to the signal x of the boundedly-rational action function a^{BR} and of the rational action function a^r. This requires knowing the normatively correct slope, a^r_x. How does one do that?

1. This could be done in a "clear and understood" context, e.g. where all prices are very clear, perhaps with just a simple task (so that in this environment, $m = 1$), which allows us to measure a^r_x. This is the methodology used by Chetty, Looney, and Kroft (2009), Taubinsky and Rees-Jones (2017), and Allcott and Taubinsky (2015).

2. Sometimes, the "normatively correct answer" is the attention of experts. Should one buy generic drugs (e.g. aspirins) or more expensive "branded drugs" – with the same basic molecule? For instance, to find out the normatively correct behavior, Bronnenberg, Dubé, Gentzkow, and Shapiro (2015) look at the behavior of experts – health care professionals – and find that they are less likely to pay extra for premium brands.

[27] Here, I define measuring attention as measuring a parameter m like in the simple model of this chapter, or its multidimensional generalization $m_1, ..., m_n$. However, one could wish to estimate a whole distribution of actions (i.e., $a(x)$ being a random variable, perhaps parametrized by some m). This is the research program in Caplin and Dean (2015); Caplin, Dean, and Leahy (2016). This literature is more conceptual and qualitative at this stage, but hopefully one day it will merge with the more behavioral literature.

[28] To be very precise, $m(x) = \frac{a^s_x(x)}{a^r_x(mx)}$. So, one can get m assuming small deviations x (so that we measure the limit $m(0)$), or the limit of a linearized rational demand $a^r(x)$.

We shall review the practical methods later.[29]

3.1.2 Deviations from Slutsky symmetry

We will see below (Section 5.1.2) that deviations from Slutsky symmetry allow one in principle to measure inattention. Aguiar and Riabov (2016) and Abaluck and Adams (2017) use this idea to measure attention. In particular, Abaluck and Adams (2017) show that Slutsky symmetry should also hold in random demand models. Suppose the utility for good i is $v_i = u_i - \beta p_i$, and the consumer chooses $a = \text{argmax}_i (u_i - \beta p_i + \varepsilon_i)$, where the ε_i are arbitrary noise terms (still, with a non-atomic distribution), which could even be correlated. The probability of choosing i is $c_i(p) = \mathbb{P}(u_i - \beta p_i + \varepsilon_i = \max_j u_j - \beta p_j + \varepsilon_j)$. Define the Slutsky term $S_{ij} = \frac{\partial c_i}{\partial p_j}$. Then, it turns out that we have $S_{ij} = S_{ji}$ again, under the rational model. So, with inattention to prices, and $c^s(p) = c^r\left(Mp + (1-M)p^d\right)$, where $M = diag(m_1, \ldots, m_n)$ is the diagonal matrix of attention, we have

$$S_{ij}^s = S_{ij}^r m_j$$

exactly as in (57). Abaluck and Adams (2017) explore this and similar relations to study the inattention to complex health care plans. They structurally estimate the model described in this section, including in the random choice specification most of the variables that would be available to individuals making Medicare Part D elections, such as premia, deductibles, and so forth. It is nice to see how an a priori abstruse idea (the deviation from Slutsky symmetry in models of limited attention, as in Gabaix 2014) can lead to concrete real-world measurement of the inattention to health-care plan characteristics.

3.1.3 Process tracking: time on task, Mouselab, eye tracking, pupil dilatation, etc.

A popular way to measure activity is with a process-tracing experiment commonly known as Mouselab (Camerer et al. 1993; Payne, Bettman, and Johnson 1993; Johnson et al. 2002; Gabaix, Laibson, Moloche, and Weinberg 2006), or with eye tracking methods. In Mouselab, subjects need to click on boxes to see which information they contain. In eye tracking (Reutskaja, Nagel, Camerer, and Rangel 2011), researchers can follow which part of the screen subjects look at, i.e. track information gathering. There are other physiological methods of measurement as well, such as measuring pupil dilation (which measures effort, Kahneman 1973). See Schulte-Mecklenbeck et al. (2017) for a recent review.

Krajbich, Armel, and Rangel (2010) and Krajbich and Rangel (2011) introduce an class of algorithmic models that link visual perception and simple choices. These attentional drift-

[29] In some cases, the context-appropriate attention parameter m is quite hard to measure. So, people use a "portable already-estimated parameter", e.g. $m = \beta = 0.7$ for hyperbolic discounting.

diffusion models (aDDM) posit that — when choosing among multiple responses — the brain accumulates the evidence in favor of one response relative to others until a decision threshold is reached. These relative choice values evolve according to an exogenous visual attention process. The authors find that aDDM models do very well at explaining empirically observed patterns of eye movement and choice in the lab. Arieli, Ben-Ami, and Rubinstein (2011) use an eye-tracking experiment to trace the decision process of experiment participants in the context of choice over lotteries, and find that that individuals rely on separate evaluations of prizes and probabilities in making their decisions.

Lahey and Oxley (2016), using eye tracking techniques, examine recruiters, and see what information they look at in resumes, in particular from white vs African-American applicants. Bartoš, Bauer, Chytilová, and Matějka (2016) and Ambuehl (2017) study how information acquisition is influenced by incentives.

3.1.4 Surveys

One can also elicit a measurement of attention via surveys. Of course, there is a difficulty. Take an economist. When surveyed, she knows the value of interest rates. But that doesn't mean that she actually takes the interest rate into account when buying a sweater – so as to satisfy her rational Euler equation for consumption. Hence, if people show ignorance in a survey, it is good evidence that they are inattentive. However, when they exhibit knowledge, it does not mean that they actually take into account the variable in their decision. Information, as measured in surveys, is an input of attention, but not the actual attention metric.[30]

For instance, a number of researchers have found that, while people know their average tax rate, they often don't know their marginal one, and often use the average tax rate as a default proxy for the marginal tax rate (De Bartolomé 1995; Liebman and Zeckhauser 2004).[31] Relatedly, Handel and Kolstad (2015) survey the employees of a large firm regarding regarding the health insurance plans available to them and find substantial information frictions.

3.1.5 Impact of reminders, advice

If people don't pay attention, perhaps a reminder will help. In terms of modeling, such a reminder could be a "free signal", or an increase in the default attention m_i^d to a dimension.

A reminder could come, for instance, from the newspaper. Huberman and Regev (2001) show how a New York Times article that re-reports stale news creates a big impact for

[30]In terms of theory, when asked about the "what is the interest rate", I know the interest rate matters a great deal. When asked "what's the best sweater to buy", the interest rate does not matter much (Gabaix 2016).

[31]This is, people perceive the marginal tax rate to me $mx^d + (1-m)\,x$ with x^d the average tax rate and x the marginal tax rate.

one company's stock price. It is not completely clear how that generalizes. There is also evidence that reminders have an impact on savings (Karlan, McConnell, Mullainathan, and Zinman 2016) and medical adherence (Pop-Eleches et al. 2011). The impact of this type of reminders is typically small, possibly because the reminder (which may be for instance a text message) does not shift attention very much, rather than because attention was almost perfect initially.

In a laboratory context, Johnson, Häubl, and Keinan (2007) show that the typical endowment effect (Kahneman et al. 1991) can be reversed by simply asking people to first think about what they could do with the proceeds from selling an object (for example, a coffee mug), and only then listing the reasons for which they want to keep that object. This demonstrates that merely altering the order in which people think about the various aspects of a problem has an impact on their eventual decision (presumably via changing their attention to those various aspects): this is an idea that Johnson, Häubl, and Keinan (2007) call "query theory".

Hanna, Mullainathan, and Schwartzstein (2014) provide summary information to seaweed farmers. This allows the farmers to improve their practice, and achieve higher productivity. This is consistent with a model in which farmers were not optimally using all the information available to them. For instance, this could be described by a model such as Schwartzstein's (2014). In this model, if an agent is pessimistic about the fact that some piece of information is useful, she won't pay attention to it, so that she won't be able to realize that it is useful. Knowledge about the informativeness of the piece of information leads to paying more attention, and better learning.

Again, this type of evidence shows that attention is not full, although it doesn't measure it.

3.2 Measuring attention: Findings

Now that we have reviewed the methods, let us move to specific findings on attention.

3.2.1 Inattention to taxes

People don't fully pay attention to taxes, as the literature has established, using the methodology of Section 3.1.1, and this is important for normative taxation (Mullainathan, Schwartzstein, and Congdon (2012); Farhi and Gabaix (2017)). Chetty, Looney, and Kroft (2009) find a mean attention of between 0.06 (by computing the ratio of the semi-elasticities for sales taxes, which are not included in the sticker price, vs. excise taxes, which are included in the sticker price) and 0.35 (computing the ratio of the semi-elasticities for sales taxes vs. more salient sticker prices).[32][33]

Taubinsky and Rees-Jones (2017) design an online experiment and elicit the maximum tag price that agents would be willing to pay when there are no taxes or when there are standard taxes corresponding to their city of residence. The ratio of these two prices is $1 + m\tau$, where τ is the tax. This allows the estimation of tax salience m. Taubinsky and Rees-Jones (2017) find (in their standard tax treatment)[34] that $\mathbb{E}[m] = 0.25$ and $\mathrm{Var}(m) = 0.13$. So, mean attention is quite small, but the variance is high. The variance of attention is important, because when attention variance is high, optimal taxes are generally lower (Farhi and Gabaix 2017) – roughly, because heterogeneity in attention creates heterogeneity in response, and additional misallocations, which increase the dead-weight cost of the tax. Taubinsky and Rees-Jones (2017) reaffirm this conclusion, and find that accounting for heterogeneity in consumer attention would increase the estimated efficiency losses from realistically-calibrated sales taxes by more than 200%.

3.2.2 Shrouded attributes

It is intuitively clear that many people won't pay attention to "shrouded attributes", such as "surprise" bank fees, minibar fees, shipping charges, and the like (Ellison 2005; Gabaix and Laibson 2006; Ellison and Ellison 2009). Gabaix and Laibson (2006) work out the market equilibrium implication of such attributes with naïve consumers – e.g. consumers who are not paying attention to the existence of shrouded attributes when buying the "base good". In particular, if there are enough naïves there is an inefficient equilibrium where shrouded attributes are priced far above marginal costs. In this equilibrium, naïve consumers are "exploited", to put it crudely: they pay higher prices and subsidize the non-naïves.

There is a growing field literature measuring the effects of such fees and consumers' inattention to them. Using both a field experiment and a natural experiment, Brown, Hossain, and Morgan (2010) find that consumers are inattentive to shrouded shipping costs in eBay online auctions. Grubb (2009) and Grubb and Osborne (2015) show that consumers don't pay attention to sharp marginal charges in three-part tariff pricing schemes,[35] and predict

[32] In an online auction experiment. Greenwood and Hanson (2014) estimate an attention $m = 0.5$ to competitors' reactions and general equilibrium effects.

[33] See also Russo (1977) for earlier field-experimental evidence on the impact of tax salience on consumer purchasing decisions. Relatedly, Finkelstein (2009) studies drivers who pay tolls in cash vs via electronic toll collection (a more automatic and invisible form of payment). The latter were substantially more likely to respond "I don't know" when asked about how much the toll was, and were also more likely to be incorrect if they offered a guess. Abeler and Jäger 2015 also show that attention to tax incentives is lower when they are more complex.

[34] They actually provide a lower bound on variance, and for simplicity we take it here to be a point estimate.

[35] Three-part tariffs are pricing schemes in which a seller offers a good or a service for a fixed fee that comes with a certain usage allowance, as well as a per-unit price that applies to all extra usage in excess of that allowance. One common example is cellphone plans: cellphone carriers commonly offer a certain amount of call minutes and data usage for a fixed price, but charge an extra marginal fee once consumers exceed the allotted quota.

their future demand with excessive ex-ante precision – for example, individuals frequently exhaust their cellular plans' usage allowance, and incur high overage costs. Brown, Camerer, and Lovallo (2012, 2013) show that when film studios withhold movies from critics before their release, moviegoers fail to infer that this may indicate low movie quality – showing that people are indeed behavioral rather than Bayesian (a Bayesian agent should be suspicious of any non-disclosed item, rather than just ignore it like a behavioral agent). Similarly, Jin, Luca, and Martin (2017) use a series of laboratory experiments to show that in general consumers form overly optimistic expectations of product quality when sellers choose not to disclose this information. This literature overlaps with a theoretical literature probing more deeply into firms' incentives to hide these attributes (Heidhues and Kőszegi 2010, 2017), and a related literature modeling competition with boundedly rational agents (Spiegler 2011; Tirole 2009; Piccione and Spiegler 2012; De Clippel, Eliaz, and Rozen 2014). The companion survey on Behavioral Industrial Organization, by Paul Heidhues and Botond Kőszegi, in this volume, details this.

3.2.3 Inattention in health plan choices

There is mounting evidence for the role of confusion and inattention in the choice of health care plans. McFadden (2006) provides an early discussion of consumers' misinformation in health plan choices, particularly in the context of Medicare Part D elections. Abaluck and Gruber (2011) find that people choose Medicare plans less often if premiums are increased by \$100 than if expected out of pocket cost is increased by \$100. Handel and Kolstad (2015) study the choice of health care plans at a large firm. They find that poor information about plan characteristics has a large impact on employees' willingness to pay for the different plans available to them, on average leading them to overvalue plans with more generous coverage and lower deductibles (see also Handel (2013)). This study documents a mistake in an important economic context. Ericson (2014) documents consumer inertia in health plan choices, and Abaluck and Adams (2017) show this inertia is largely attributable to inattention.

3.2.4 Inattention to health consequences

It is intuitively clear that we do not always attend to the health consequences of our choices, e.g. when drinking a sugary soda or smoking a cigarette, we underperceive the future health costs (e.g. cancer) from those enjoyments. Several studies have quantified the magnitude of these underperceived costs, and applied them to the normative study of optimal "sin taxation". Allcott, Lockwood, and Taubinsky (2017) use survey data to measure the price elasticity of demand for sugary beverages, and deliver optimal sin tax formulas in terms of these sufficient statistics. Gruber and Kőszegi (2001) conduct a similar study by postulating hyperbolic discounting, and importing the parameter $m = \beta \simeq 0.7$ into the model.

3.2.5 People use rounded numbers

Lacetera, Pope, and Sydnor (2012) estimate inattention via buyers' "left-digit bias" in evaluating the mileage of used cars sold at auction. Call x the true mileage of a car (i.e., how many miles it has been driven), and x^d the mileage rounded to the leading digit, and let $r = x - x^d$ be the "mileage remainder." For instance, if $x = 12,345$ miles, then $x^d = 10,000$ miles and $r = 2,345$ miles, and the perceived mileage is $x^s = x^d + m\left(x - x^d\right)$. Lacetera, Pope, and Sydnor (2012) estimate a structural model for the perceived value of cars of the form $V = -f\left(x^s\left(x, m\right)\right)$. They find a mean attention parameter of $m = 0.69$ (see also Englmaier et al. (2017)). Busse, Lacetera, Pope, Silva-Risso, and Sydnor (2013b) break down this estimate along covariate dimensions, and find that attention is lower for older and cheaper cars, and lower for lower-income retail buyers.[36]

This is a very nice study, as it offers clean identification and high quality data – hence a more precise estimate of attention than most other papers. It would be nice to see if it matches the quantitative predictions of models discussed in this survey (for example, that in equation (35)).

3.2.6 Do people account for the net present value of future costs and benefits?

When you buy a car, you should pay attention to both the sticker price of the car, and the present value of future gasoline payments. But it is very conceivable that some people will pay less than full attention to the future value of gas payments: the full price of the car $p_{\text{car}}+p_{\text{gas}}$ will be perceived as $m_{\text{car}}p_{\text{car}}+m_{\text{gas}}p_{\text{gas}}$. Two papers explore this, and have somewhat inconsistent findings. Allcott and Wozny (2014) find partial inattention to gas prices: their estimate is $\frac{m_{\text{gas}}}{m_{\text{price}}} = 0.76$. However, Busse, Knittel, and Zettelmeyer (2013a) find that they cannot reject the null hypothesis of equal attention, $\frac{m_{\text{gas}}}{m_{\text{price}}} = 1$. One hopes that similar studies, perhaps with data from other countries, will help settle the issue. One can conjecture that people likewise do not fully pay attention to the cost of car parts – this remains to be seen. A related literature, starting with Hausman (1979), has found that people apply high discount rates to future energy costs when purchasing electricity-consuming durables such as refrigerators or air conditioners (see Frederick, Loewenstein, and O'Donoghue (2002) for a more recent survey of this literature).

3.2.7 Inattention in finance

There is now a large amount of evidence of partial inattention in finance. This is covered in greater depth in the companion chapter on Behavioral Finance, by Nick Barberis. Here are

[36] In addition to the studies already mentioned, Shlain (2018) estimates a structural model of left-digit bias using retail scanner data; mean estimated attention to the left-digit remainder of prices is $m = 0.74$ in the study's main specification. This work was circulated after the final draft of this survey was completed, so that it could not be integrated in the final tables and figures.

some samples from this literature.

When investors have limited attention, the specific format in which accounting statements are presented matters. Hirshleifer and Teoh (2003) is an influential model of that, which relates accounting statements to misvaluations. Relatedly, Hirshleifer, Lim, and Teoh (2009) find that when investors are more distracted (as there are more events that day), inefficiencies are stronger: for instance, the post-earnings announcement drift is stronger. Peng and Xiong (2006) study the endogenous attention to aggregate vs idiosyncratic risk.

DellaVigna and Pollet (2007) find that investors have a limited ability to incorporate some subtle forces (predictable changes in demand because of demographic forces) into their forecasts, especially at long horizons. DellaVigna and Pollet (2009) show that investors are less attentive on Fridays: when companies report their earnings on Fridays, the immediate impact on the price (as a fraction of the total medium run impact) is lower. Hirshleifer, Lim, and Teoh (2009) show how investors are less attentive to a given stock when there are lots of other news in the market. In a similar vein, Barber and Odean (2007) show that retail investors overweight attention-grabbing stocks in their portfolios: for example, the stocks of companies that have recently received high media coverage. In a controlled laboratory context, Frydman and Rangel (2014) find that decreasing attention to a stock's purchase price reduces the extent to which people display a disposition effect (i.e., an abnormally high propensity to sell stocks that have realized capital gains).

Cohen and Frazzini (2008) document that investors are quick at pricing the "direct" impacts of an announcement, but slower at pricing the "indirect" impact (e.g. a new plane by Boeing gets reflected in Boeing's stock price, but less quickly in that of Boeing's supplier network). Giglio and Shue (2014) find that investors underreact to the passage of time after merger announcements, even though lack of news is in fact informative of a higher probability that the deal will be successfully completed: in effect, investors pay limited attention to "no news".

Malmendier and Nagel (2011) find that generations who experienced low stock market returns invest less in the stock market. People seem to put too much weight on their own experience when forming their beliefs about the stock market.

Fedyk (2018) shows that positioning of news on the front page of the Bloomberg terminal platform induces abnormally high trading volumes for the affected companies in the first few minutes after new publication: investors appear to pay a disproportionately low amount of attention to news that is not on the front page.

This literature is growing quickly.[37] It would be nice to have more structural models,

[37] A related study – which may also be linked to models of reference dependence — is that of Baker, Pan, and Wurgler (2012), who find that when thinking about a merger or acquisition price, investors put a lot of weight on recent (trailing 52 weeks) prices. This has real effects: merger waves occur when high returns on the market and likely targets make it easier for bidders to offer a peak price. This shows an intriguing mix of attention to a partially arbitrary price, and its use as an anchor in negotiations and perhaps valuations.

predicting in a quantitative way the speed of diffusion of information.

3.2.8 Evidence of reaction to macro news with a lag

There is much evidence for delayed reaction in macro data. Friedman (1961) talks about "long and variable lags" in the impacts of monetary stimulus. This is also what motivated models of delayed adjustment, e.g. Taylor (1980). Empirical macro research in the past decades has frequently found that a variable (e.g. price) reacts to shocks in other variables (e.g. nominal interest rate) only after a significant delay (e.g., quarters or years).

Delayed reaction is confirmed by the more modern approaches of Romer and Romer (1989) and Romer and Romer (2004), who identify monetary policy shocks using the narrative account of Federal Open Market Committee (FOMC) Meetings[38] and find that the price level would only start falling 25 months after a contractionary monetary policy shock.

This is confirmed also by more formal econometric evidence with identified VARs. Sims (2003) notes that in nearly all Vector Autoregression (VAR) studies, a variable reacts with delay when responding to shocks in other variables, even though a rational theory would predict that it should instantaneously jump. Such finding is robust in VAR specifications of various sizes, variable sets, and identification methods (Leeper, Sims, and Zha 1996; Christiano, Eichenbaum, and Evans 2005). While it is feasible to generate delayed response using adjustment costs, large adjustment costs would imply that a variable's reactions to *all* shocks are smooth, contradicting the VAR evidence that responses to own shocks tend to be large. A model of inattention, however, can account for both phenomena simultaneously.

Finally, micro survey data suggest that macro sluggishness is not just the result of delayed action, but rather the result of infrequent observation as well. Alvarez, Guiso, and Lippi (2012) and Alvarez, Lippi, and Paciello (2017) provide evidence of infrequent reviewing of portfolio choice and price setting, respectively, with clean analytics (see also Abel, Eberly, and Panageas 2013 for a sophisticated model along those lines). A median investor reviews her portfolio 12 times and makes changes only twice annually, while a median firm in many countries reviews price only 2-4 times a year.

3.2.9 Evidence on level-k thinking in games

An impactful literature on level-k thinking in games has generated a large amount of empirical evidence for the lack of strategic sophistication in games (Nagel 1995; Ho et al. 1998). Level-k models (Stahl and Wilson 1995; Camerer et al. 2004) allow players to vary in their levels of strategic sophistication: level-0 players randomize their strategies; level-1 players best-respond under the assumption that all other players are level-0; level-2 players best-respond under the assumption that everyone else is a level-1 thinker, and so forth. Within

[38] The intended interest rate changes identified in accounts of FOMC Meetings are further orthogonalized by relevant variables in Fed's information set (Greenbook Forecasts), making it plausibly exogenous.

the framework laid out in this survey, level-k thinking can be understood as lack of attention to higher-order strategical issues, perhaps because of cognitive limitations. Level-0 players corresponds to fully inattentive agents ($m = 0$), while level-∞ players are the traditional rational agents who make full use of all available information ($m = 1$).

Camerer, Ho, and Chong (2004) estimate that an average parameter value $k = 1.5$ fits well empirical data from a wide variety of games, which implies a rather low value of m. A related empirical literature corroborates this finding and contributed to many of the advances in experimental process-tracking methods discussed in subsection 3.1.3: see for example Costa-Gomes et al. (2001), Wang, Spezio, and Camerer (2010), and Brocas et al. (2014).[39]

3.3 Attention across stakes and studies

Attention over many studies Figure 1 (in the introduction) and Table 1 and contain a synthesis of eleven measurements of attention – I selected all the studies I could find that measured attention (i.e., gave an estimate of the parameter m). They are a tribute to the hard work of many behavioral economists. I am sure that this table will be enriched over time.

Table 1 shows point estimates of the attention parameter m in the literature discussed in this survey. For each distinct study or experimental setting, I report the most aggregated available estimates. In each of these studies, m is measured as the degree to which individuals underperceive the value of an opaque add-on attribute τ to a quantity or price p, such that the subjectively perceived total value of the quantity is $q^s(m) = p + m\tau$.

Correspondingly, for each economic setting I show the estimated ratio of the values p and τ, which is a measure of the relative significance of the add-on attribute τ. Appendix B outlines the details of the methodology used to compile this data. Figure 1 plots the point estimates of m against the estimated value of τ/p.[40] In addition to this cross-study data, Figure 1 plots a second set of intra-study data points from Busse, Lacetera, Pope, Silva-Risso, and Sydnor (2013b), who offer very precise estimates of attention broken down along covariate dimensions. By looking at subsamples of Busse, Lacetera, Pope, Silva-Risso, and Sydnor's (2013b) dataset of more than 22 million of used car transactions, we are able to effectively highlight the co-movement between m and the relative importance of the add-on attribute. Attention is high for car mileage presumably because the absolute dollar stakes (rather than just the relative stakes $\frac{\tau}{p}$) are particularly high.

[39]See Avoyan and Schotter (2018) for a related exploration of limited attention in games.

[40]I use the relative value as a determinant of attention. This arises if, for a given decision, there is a fixed attention budget between different attributes of the problem, and their relative importance determine attention. In the model of Section 4.1.1, this is microfounded by endogenizing the κ and using the "scale-free" κ, as in Gabaix (2014), section V.C.

Table 1: **Attention estimates in a cross-section of studies.** This table shows point estimates of the attention parameter m in a cross-section of recent studies, alongside the estimated relative value of the opaque add-on attribute with respect to the relevant good or quantity (τ/p). I report the most aggregated available estimates for each distinct study or experimental setting. The quantity τ is the estimated mean value of the opaque good or quantity against which m is measured; the quantity p is the estimated mean value of the good or quantity itself, exclusive of the opaque attribute. Appendix B describes the construction methodology and details. Studies are arranged by their τ/p value, in descending order.

Study	**Good or Quantity**	**Opaque Attribute**	**Attribute Importance** (τ/p)	**Attention Estimate** (m)
Allcott and Wozny (2014)	Expense associated with car purchase	Present value of future gasoline costs	0.58	0.76
Hossain and Morgan (2006)	Price of CDs sold at auction on eBay	Shipping costs	0.38	0.82
DellaVigna and Pollet (2009)	Public company equity value	Value innovation due to earnings announcements	0.30	0.54
DellaVigna and Pollet (2009)	Public company equity value	Value innovation due to earnings announcements *that occur on Fridays*	0.30	0.41
Hossain and Morgan (2006)	Price of CDs sold at auction on eBay	Shipping costs	0.24	0.55
Taubinsky and Rees-Jones (2017)	Price of products purchased in laboratory experiment	Sales tax, *tripled relative to standard tax*	0.22	0.48
Lacetera, Pope, and Sydnor (2012)	Mileage of used cars sold at auction	Mileage left-digit remainder	0.10	0.69
Chetty, Looney, and Kroft (2009)	Price of grocery store items	Sales tax	0.07	0.35
Taubinsky and Rees-Jones (2017)	Price of products purchased in laboratory experiment	Sales tax	0.07	0.25
Chetty, Looney, and Kroft (2009)	Price of retail beer cases	Sales tax	0.04	0.06
Brown, Hossain, and Morgan (2010)	Price of iPods sold at auction on eBay	Shipping costs	0.03	0.00
Mean	—	—	0.21	0.44
Standard Deviation	—	—	0.18	0.28

As noted above, mean attention is 0.44 – roughly in the middle of the two poles of full rationality and complete inattention. We see that, sensibly, attention does increase with incentives. [41] We can actually propose a model-based calibration of this attention.

Calibrating the attention function Figure 1 additionally shows a calibration of an attention model in which estimated attention $\hat{m}$ is a function of the attribute's relative importance τ/p:[42]

$$\hat{m} = \mathcal{A}_\alpha \left(\left[\frac{\tau/p}{\bar{\kappa}} \right]^2 \right), \tag{25}$$

where $\mathcal{A}_\alpha$ is an attention function, which will be derived in Section 4.1.1. For now, the reader can think of the attention function A_α as the solution to a problem in which an agent chooses optimal attention m subject to the tradeoff between the penalty resulting from inattention and the cost of paying attention. For this calibration, I allow the attention cost function to depend quadratically on m, to a degree parameterized by the scalar parameter $q \geq 0$, such that the attention function is given by the following variant of (34),

$$\mathcal{A}_\alpha(\sigma^2) \coloneqq \arg \min_{m \in [0,1]} \frac{1}{2} (1-m)^2 \sigma^2 + \left(m + qm^2\right)^\alpha, \tag{26}$$

where the parameter q captures the curvature of the attention function. In order to retain both continuity and sparsity of the attention function (26), I impose the sparsity-inducing restriction $\alpha = 1$, which results in

$$\mathcal{A}_1(\sigma^2) = \max \left(\frac{\sigma^2 - 1}{\sigma^2 + 2q}, 0 \right). \tag{27}$$

I estimate the cost parameters $\bar{\kappa}$ and q on the cross-study data via nonlinear least squares, according to the model in (25). This yields calibrated parameters $\bar{\kappa} = 3.0\%$ and $q = 22.4$.

3.4 Different meanings of "attention"

The concept of "attention" has several meanings. The primary meaning in this survey is that of "the extent to which an agent's cognitive process is able to make use of all available

[41] It is reassuring to see the attention covarying positively with incentives in Figure 1. In part, this is because the problems collected in the figure happen to have arguably roughly similar levels of complexity. If problem complexity ("κ" in later notations) varied a lot across studies, a future update of Figure 1 could potentially show a less clean positive relation between stakes and attention. Likewise, in situations where stakes are very important in dollar terms (e.g. for car purchases), attention is likely to be higher.

[42] Things are expressed in terms of the "scale free cost" $\bar{\kappa}$ (see Gabaix 2016, Sections 4.2 and 10.2, which conjectures that "a reasonable parameter might be $\bar{\kappa}$ =5%"), which is unitless, so potentially portable across contexts. It means that agents don't consider attributes τ whose relative importance $\left|\frac{\tau}{p}\right|$ is less than $\bar{\kappa}$. It also justifies the scaling $\frac{\tau}{p}$, where the "natural scale" of the decision is p. This assumes that the attention budget has a fixed amount for the task at hand. Understanding the degree of fungibility of attention across tasks is an interesting research frontier.

data" (relative to a normative, rational benchmark, so that $m = 1$ means full attention), and is captured by this summary measure m — something we might call "effective attention". Measuring it requires some normative model. In much of psychology, however, attention means "observable sensory inputs devoted to the task" – let us denote this by T, as in time spent on the information, and call T the "observable attention". For instance, a bored student may look at a whole lecture (which lasts for $T = 80$ minutes), but still not really exert effort (let us call M that mental effort) so that the total amount learned (indexed by m) is very low. We have a relation of the type:

$$m = f(T, M),$$

e.g. $m = 1 - e^{-\alpha MT}$. I note that T is directly measurable, m can also be measured, though more indirectly, and M is quite nebulous at this stage. More generally, we can think of attentional inputs as "information gathering" (T in our example) and "information processing" (M).

In this survey I have emphasized m because it's the summary quantity we need to predict behavior, and because in many studies (e.g. on the attention to the tax) we don't have access to the time spent T. Also, even the whole time spent T is not a sufficient statistics for the effective attention m – as active mental effort is hard to measure.

Still, the "observable attention" T is very precious – as it can be directly measured, and much progress of science comes from focusing on directly measurable quantities. Hence, the process tracking studies reviewed in Section 3.1.3 represent great advances. Sometimes, observable attention T yields an excellent prediction of choice (for example, in context of simple choices among two or three alternatives, as in Krajbich and Rangel (2011)).

One can hope that the understanding of observable attention T, and effective attention m will continue to increase, with a healthy interplay between the two.

4 Models of Endogenous Attention: Deterministic Action

We have seen that attention can be modeled in a simple way and that it can be measured. In this section, we will study some deterministic models that endogenize attention.

4.1 Paying more attention to more important variables: The sparsity model

The model in Gabaix (2014) aims at a high degree of applicability – to do so, it generalizes the max operator in economics, by assuming that agents can be less than fully attentive. This provides a foundation for a behavioral version of basic textbook microeconomics (Section 5), of the basic theory of taxation (Farhi and Gabaix (2017)), of basic dynamic macroeconomics (Gabaix (2016)), and macroeconomic fiscal and monetary policy (Gabaix (2018)).[43]

The agent faces a maximization problem which is, in its traditional version, $\max_a u(a,x)$ subject to $b(a,x) \geq 0$, where u is a utility function, and b is a constraint. In this section I present a way to define the "sparse max" operator defined and analyzed in Gabaix (2014):[44]

$$\operatorname*{smax}_a u(a,x) \text{ subject to } b(a,x) \geq 0, \tag{28}$$

which is a less than fully attentive version of the "max" operator. Variables a, x and function b have arbitrary dimensions.[45]

The case $x = 0$, will sometimes be called the "default parameter." We define the default action as the optimal action under the default parameter: $a^d := \arg\max_a u(a,0)$ subject to $b(a,0) \geq 0$. We assume that u and b are concave in a (and at least one of them is strictly concave) and twice continuously differentiable around $\left(a^d, 0\right)$. We will typically evaluate the derivatives at the default action and parameter, $(a,x) = \left(a^d, 0\right)$.

4.1.1 The sparse max without constraints

For clarity, we shall first define the sparse max without constraints, i.e. study $\operatorname{smax}_a u(a,x)$.

Motivation for optimization problem The agent maximizing (6) will take the action

$$a(x,m) := \arg\max_a u(a,x,m) \tag{29}$$

and she will experience utility $v(x,m) = u(a(x,m),x)$. Let us posit that attention creates a psychic cost, parametrized by

$$\mathcal{C}(m) = \kappa \sum_i m_i^\alpha$$

with $\alpha \geq 0$. The case $\alpha = 0$ corresponds to a fixed cost κ paid each time m_i is non-zero. The parameter $\kappa \geq 0$ is a penalty for lack of sparsity. If $\kappa = 0$, the agent is the traditional, rational agent model with costless cognition.

It follows that agent would allocate attention m as:

$$\max_m \mathbb{E}\left[u(a(x,m),x)\right] - \mathcal{C}(m). \tag{30}$$

However, ever since Simon (1955), many researchers have seen that problem (30) is very com-

[43] This subsection and Section 5 draw extensively from Gabaix (2014).

[44] I draw on fairly recent literature on statistics and image processing to use a notion of "sparsity" that still entails well-behaved, convex maximization problems (Tibshirani 1996, Candes and Tao 2006).

[45] We shall see that parameters will be added in the definition of sparse max.

plicated – more complex than the original problem (we are threatened by "infinite regress" problem). The key step of the sparse max is that the agent will solve a version of this problem.

Definition 4.1 (Sparse max – abstract definition). *In the sparse max, the agent optimizes in two steps. In Step 1, she selects the optimal attention m^* under a simplified version of the optimal problem* (30)*: (i) she replaces her utility by a linear-quadratic approximation, and (ii) imagines that the vector x is drawn from a mean 0 distribution, with no correlations, but the accurate variances. In Step 2, she picks the best action* (29) *under the exact utility function, modulated by the attention vector m^* selected in Step 1.*

To to see this analytically, we introduce some notation. The expected size of x_i is $\sigma_i = \mathbb{E}\left[x_i^2\right]^{1/2}$, in the "ex ante" version of attention. In the "ex post allocation of attention" version, we set $\sigma_i := |x_i|$. We define $a_{x_i} := \frac{\partial a}{\partial x_i}$, which indicates how much a change in x_i should change the action, for the traditional agent.[46]

The agent entertaining the simplified problem of Definition 4.1 will want to solve:[47]

$$m^* = \arg\min_{m\in[0,1]^n} \frac{1}{2}\sum_{i=1}^{n}\left(1-m_i\right)^2 \Lambda_{ii} + \kappa \sum_{i=1}^{n} m_i^{\alpha}, \tag{32}$$

where $\frac{1}{2}\Lambda_{ii} := -\frac{1}{2}\mathbb{E}\left[x_i^2\right] a_{x_i} u_{aa} a_{x_i}$ is the gain that the consumer enjoys when he goes from zero attention to full attention in dimension i (up to third order terms in the Taylor expansion). Hence, (32) shows how the agent trades off the benefits of more attention (captured by $\left(1-m_i\right)^2\Lambda_{ii}$) with the costs of attention (captured by κm_i^{α}).

The attention function To build at least some intuition, let us start with the case with just one variable, $x_1 = x$ and call $\sigma^2 = \Lambda_{11}$. Then, problem (32) becomes:

$$\min_m \frac{1}{2}\left(1-m\right)^2\sigma^2 + \kappa m^{\alpha}. \tag{33}$$

[46]This implies $a_{x_i} = -u_{aa}^{-1} u_{a x_i}$. Derivatives are evaluated at the default action and parameter, i.e. at $(a,x) = \left(a^d, 0\right)$.

[47]The justification is as follows. We call $V(m) = \mathbb{E}\left[u\left(a\left(x,m\right),x\right)\right]$ the expected consumption utility. Then, a Taylor expansion shows that we have, for small x (call $\iota = (1,\ldots,1)$ the vector corresponding to full attention, like the traditional agent):

$$V(m) - V(\iota) = -\frac{1}{2}\sum_{i,j}\left(1-m_i\right)\Lambda_{ij}\left(1-m_j\right) + o\left(\sigma^2\right), \tag{31}$$

defining $\Lambda_{ij} := -\sigma_{ij} a_{x_i} u_{aa} a_{x_j}$, $\sigma_{ij} := \mathbb{E}\left[x_i x_j\right]$ and $\sigma^2 = \left\|\left(\sigma_i^2\right)_{i=1\ldots n}\right\|$. The Taylor expansions is for small noises in x, rather than for m close to 1. The agent drops the non-diagonal terms (this is an optional, but useful, assumption of sparse max).

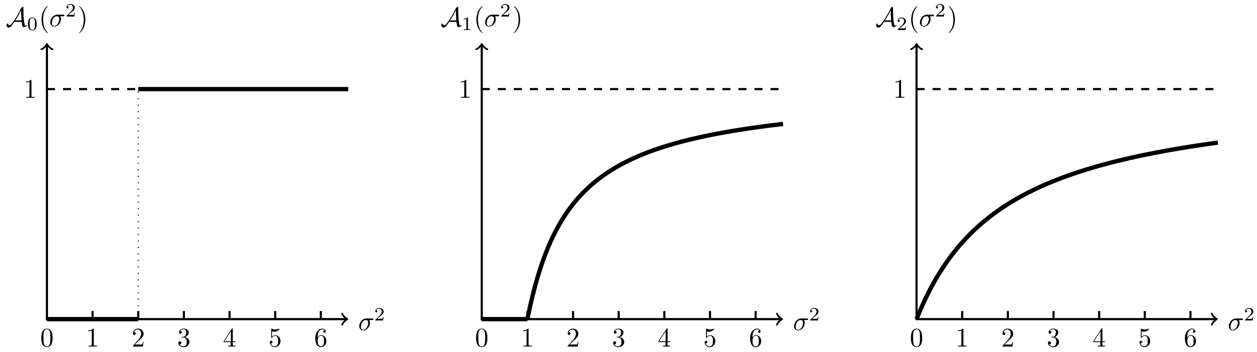

Figure 2: Three attention functions $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$, corresponding to fixed cost, linear cost and quadratic cost respectively. We see that $\mathcal{A}_0$ and $\mathcal{A}_1$ induce sparsity – i.e. a range where attention is exactly 0. $\mathcal{A}_1$ and $\mathcal{A}_2$ induce a continuous reaction function. $\mathcal{A}_1$ alone induces sparsity and continuity.

Optimal attention is $m = \mathcal{A}_\alpha\left(\frac{\sigma^2}{\kappa}\right)$, where the "attention function" $\mathcal{A}_\alpha$ is defined as[48]

$$\mathcal{A}_\alpha\left(\sigma^2\right) := \arg\min_{m\in[0,1]} \frac{1}{2}\left(1-m\right)^2\sigma^2 + m^\alpha. \tag{34}$$

Figure 2 plots how attention varies with the variance σ^2 for fixed, linear and quadratic cost: $\mathcal{A}_0\left(\sigma^2\right) = 1_{\sigma^2\geq 2}$, $\mathcal{A}_1\left(\sigma^2\right) = \max\left(1-\frac{1}{\sigma^2}, 0\right)$, $\mathcal{A}_2\left(\sigma^2\right) = \frac{\sigma^2}{2+\sigma^2}$. In particular, let us examine $\mathcal{A}_1\left(\sigma^2\right)$. When the stakes are small, attention is 0 ($\mathcal{A}_1\left(\sigma^2\right) = 0$ for $\sigma^2 \leq 1$). As stakes increase, attention becomes non-0, and as stakes becomes very large, attention becomes full ($\lim_{\sigma^2\to\infty}\mathcal{A}_1\left(\sigma^2\right) = 1$).

We now explore the case in which a^s indeed induces no attention to certain variables.[49]

Lemma 4.1 (Special status of linear costs). *When $\alpha \leq 1$ (and only then) the attention function $\mathcal{A}_\alpha\left(\sigma^2\right)$ induces sparsity: when the variable is not very important, then the attention weight is 0 ($m = 0$). When $\alpha \geq 1$ (and only then) the attention function is continuous. Hence, only for $\alpha = 1$ do we obtain both sparsity and continuity.*

For this reason $\alpha = 1$ is recommended for most applications. Below I state most results in their general form, making clear when $\alpha = 1$ is required.[50]

The sparse max: Values of attention The abstract definition of the sparse max, and the functional form assumptions made above, lead to the following concrete procedure that

[48]If there are multiple minimizers m, we take the the highest one.

[49]Lemma 4.1 has direct antecedents in statistics: the pseudo norm $\|m\|_\alpha = \left(\sum_i |m_i|^\alpha\right)^{1/\alpha}$ is convex and sparsity-inducing iff $\alpha = 1$ (Tibshirani 1996).

[50]The sparse max is, properly speaking, sparse (in the narrow sense of inducing zero attention) only when $\alpha \leq 1$. When $\alpha > 1$, the abuse of language seems minor, as the smax still offers a way to economize on attention (as it generally shrinks attention towards 0). Perhaps smax should be called a "bmax" or behavioral / boundedly rational max.

describes that behavioral agent.

Proposition 4.1 *The sparse max of Definition 4.1 is solved in two steps.*

Step 1: Choose the attention vector m^**, which is optimally equal to:*

$$m_i^* = \mathcal{A}_\alpha \left(\sigma_i^2 \left| a_{x_i} u_{aa} a_{x_i} \right| / \kappa \right), \tag{35}$$

where $\mathcal{A}_\alpha : \mathbb{R} \to [0,1]$ *is the attention function expressed in* (34)*,* σ_i^2 *is the perceived variance of* x_i^2*,* $a_{x_i} = -u_{aa}^{-1} u_{ai}$ *is the traditional marginal impact of a small change in* x_i*, evaluated at* $x = 0$*, and* κ *is the cost of cognition.*

Step 2: Choose the action

$$a^s = \arg\max_a u(a, x, m^*). \tag{36}$$

Hence more attention is paid to variable x_i if it is more variable (high σ_i^2), if it should matter more for the action (high $|a_{x_i}|$), if an imperfect action leads to greater losses (high $|u_{aa}|$), and if the cost parameter κ is low.

The sparse max procedure in (35) entails (for $\alpha \leq 1$): "Eliminate each feature of the world that would change the action by only a small amount".[51] This is how a sparse agent sails through life: for a given problem, out of the thousands of variables that might be relevant, he takes into account only a few that are important enough to significantly change his decision.[52] He also devotes "some" attention to those important variables, not necessarily paying full attention to them.[53]

Let us revisit the initial example.[54]

Example 1 *In the quadratic loss problem,* (11)*, the traditional action is* $a^r = \sum_{i=1}^n b_i x_i$*, and the behavioral action is:*

$$a^s = \sum_{i=1}^n m_i^* b_i x_i, \qquad m_i^* = \mathcal{A}_\alpha \left(b_i^2 \sigma_i^2 / \kappa \right). \tag{37}$$

Discrete goods All this can be extended to discrete actions. For instance, suppose that the agent must choose one good among A goods. Good $a \in \{1...A\}$ has value $u(a, x) =$

[51] For instance, when $\alpha = 1$, eliminate the x_i such that $\left| \sigma_i \cdot \frac{\partial a}{\partial x_i} \right| \leq \sqrt{\frac{\kappa}{|u_{aa}|}}$.

[52] To see this formally (with $\alpha = 1$), note that m has at most $\sum_i b_i^2 \sigma_i^2 / \kappa$ non-zero components (because $m_i \neq 0$ implies $b_i^2 \sigma_i^2 \geq \kappa$). Hence, when κ increases, the number of non-zero components becomes arbitrarily small. When x has infinite dimension, m has a finite number of non-zero components, and is therefore sparse (assuming $\mathbb{E}\left[(a^r)^2 \right] < \infty$).

[53] There is anchoring with partial adjustment, i.e. dampening. This dampening is pervasive, and indeed optimal, in "signal plus noise" models (more on this later).

[54] The proof is as follows. We have $a_{x_i} = b_i$, $u_{aa} = -1$, so (35) gives $m_i = \mathcal{A}_\alpha \left(b_i^2 \sigma_i^2 / \kappa \right)$.

$\sum_{i=1}^n b_i x_{ia}$. Then, the boundedly rational perception of the utility from good a is $u(a,x,m) = \sum_{i=1}^n m_i b_i x_{ia}$, and the optimal attention is again:

$$m_i^* = \mathcal{A}_\alpha \left(b_i^2 \sigma_i^2 / \kappa \right) \tag{38}$$

where σ_i^2 is the variance of x_{ia} across goods. The behavioral action is then $a^s = \arg\max_a u(a,x,m^*)$.

4.1.2 Sparse max allowing for constraints

Let us now extend the sparse max so that it can handle maximization under $K(=\dim b)$ constraints, which is problem (28). As a motivation, consider the canonical consumer problem:

$$\max_{c_1,\ldots,c_n} u(c_1,\ldots,c_n) \text{ subject to } p_1 c_1 + \ldots + p_n c_n \leq w. \tag{39}$$

We start from a default price $\boldsymbol{p}^d$. The new price is $p_i = p_i^d + x_i$, while the price perceived by the agent is $p_i^s(m) = p_i^d + m_i x_i$, i.e.[55]

$$p_i^s(p_i, m) = m_i p_i + (1 - m_i) p_i^d.$$

How to satisfy the budget constraint? An agent who underperceives prices will tend to spend too much – but he's not allowed to do so. Many solutions are possible, but the following makes psychological sense and has good analytical properties. In the traditional model, the ratio of marginal utilities optimally equals the ratio of prices: $\frac{\partial u/\partial c_1}{\partial u/\partial c_2} = \frac{p_1}{p_2}$. We will preserve that idea, but in the space of perceived prices. Hence, the ratio of marginal utilities equals the ratio of *perceived* prices:[56]

$$\frac{\partial u/\partial c_1}{\partial u/\partial c_2} = \frac{p_1^s}{p_2^s}, \tag{40}$$

i.e. $u'(\boldsymbol{c}) = \lambda \boldsymbol{p}^s$, for some scalar λ.[57] The agent will tune λ so that the constraint binds, i.e. the value of $\boldsymbol{c}(\lambda) = u'^{-1}(\lambda \boldsymbol{p}^s)$ satisfies $\boldsymbol{p} \cdot \boldsymbol{c}(\lambda) = w$.[58] Hence, in step 2, the agent "hears clearly" whether the budget constraint binds.[59] This agent is behavioral, but smart enough to exhaust his budget.

[55] The constraint is $0 \leq b(\boldsymbol{c}, \boldsymbol{x}) := w - \left(\boldsymbol{p}^d + \boldsymbol{x}\right) \cdot \boldsymbol{c}$.

[56] Otherwise, as usual, if we had $\frac{\partial u/\partial c_1}{\partial u/\partial c_2} > \frac{p_1^s}{p_2^s}$, the consumer could consume a bit more of good 1 and less of good 2, and project to be better off.

[57] This model, with a general objective function and K constraints, delivers, as a special case, the third adjustment rule discussed in Chetty, Looney, and Kroft (2007) in the context of consumption with two goods and one tax.

[58] If there are several λ, the agent takes the smallest value, which is the utility-maximizing one.

[59] See footnote 69 for additional intuitive justification.

Consequences for consumption Section 5.1.1 develops consumer demand from the above procedure, and contains many examples. For instance, Proposition 5.1 finds that the Marshallian demand of a behavioral agent is

$$\boldsymbol{c}^s(\boldsymbol{p}, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w'), \tag{41}$$

where the as-if budget w' solves $\boldsymbol{p} \cdot \boldsymbol{c}^r(\boldsymbol{p}^s, w') = w$, i.e. ensures that the budget constraint is satisfied under the true price.

Determination of the attention to prices, m^*. The exact value of attention, m, is not essential for many issues, and this subsection might be skipped in a first reading. Call λ^d the Lagrange multiplier at the default price.[60]

Proposition 4.2 (Attention to prices). *The sparse agent's attention to price i is: $m_i^* = \mathcal{A}_\alpha\left(\left(\frac{\sigma_{p_i}}{p_i^d}\right)^2 \psi_i \lambda^d p_i^d c_i^d / \kappa\right)$, where ψ_i is the price elasticity of demand for good i.*

Hence attention to prices is greater for goods (i) with more volatile prices $(\frac{\sigma_{p_i}}{p_i^d})$, (ii) with higher price elasticity ψ_i (i.e. for goods whose price is more important in the purchase decision), and (iii) with higher expenditure share $(p_i^d c_i^d)$. These predictions seem sensible, though not extremely surprising. What is important is that we have some procedure to pick the m, so that the model is closed. Still, it would be interesting to investigate empirically the prediction of Proposition 4.2.

Generalization to arbitrary problems of maximization under constraints We next generalize this approach to satisfying the budget constraint to arbitrary problems. The reader may wish to skip to the next section, as this material is more notationally dense. We define the Lagrangian $L(a, x) := u(a, x) + \lambda^d \cdot b(a, x)$, with $\lambda^d \in \mathbb{R}_+^K$ the Lagrange multiplier associated with problem (28) when $x = 0$ (the optimal action in the default model is $a^d = \arg\max_a L(a, 0)$). The marginal action is: $a_x = -L_{aa}^{-1} L_{ax}$. This is quite natural: to turn a problem with constraints into an unconstrained problem, we add the "price" of the constraints to the utility. Applying the Definition 4.1 to this case this the following characterization.[61]

Proposition 4.3 (Sparse max operator with constraints). *The sparse max,* $\text{smax}_{a|\kappa,\sigma}\, u(a, x)$ *subject to $b(a, x) \geq 0$, is solved in two steps.*

[60]This is, $u'(\boldsymbol{c}^d) = \lambda^d \boldsymbol{p}^d$, where $\boldsymbol{p}^d$ is the exogenous default price, and $\boldsymbol{c}^d$ is the (endogenous) optimal consumption as the default.

[61]For instance, in a consumption problem (39), λ^d is the "marginal utility of a dollar", at the default prices. This way we can use Lagrangian L to encode the importance of the constraints and maximize it without constraints, so that the basic sparse max can be applied.

Step 1: Choose the attention m^* *as in* (32)*, using* $\Lambda_{ij} := -\sigma_{ij} a_{x_i} L_{aa} a_{x_j}$*, with* $a_{x_i} = -L_{aa}^{-1} L_{ax_i}$*. Define* $x_i^s = m_i^* x_i$ *the associated sparse representation of* x*.*

Step 2: Choose the action. *Form a function* $a(\lambda) := \arg\max_a u(a, x^s) + \lambda b(a, x^s)$*. Then, maximize utility under the true constraint:* $\lambda^* = \arg\max_{\lambda \in \mathbb{R}_+^K} u(a(\lambda), x^s)$ *subject to* $b(a(\lambda), x) \geq 0$*. (With just one binding constraint this is equivalent to choosing* λ^* *such that* $b(a(\lambda^*), x) = 0$*; in case of ties, we take the lowest* λ^**.) The resulting sparse action is* $a^s := a(\lambda^*)$*. Utility is* $u^s := u(a^s, x)$*.*

Step 2 of Proposition 4.3 allows generally for the translation of a boundedly rational maximum without constraints into a boundedly rational maximum with constraints. To obtain further intuition on the constrained maximum, we turn to consumer theory.

4.2 Proportional thinking: The salience model of Bordalo, Gennaioli, Shleifer

In a series of papers, Bordalo, Gennaioli, and Shleifer (2012; 2013; 2016a) introduce a model of context-dependent choice in which attention is drawn toward those attributes of a good that are *salient* – that is, attributes that are particularly unusual with respect to a given reference frame.

4.2.1 The salience framework in the absence of uncertainty

The theory of salience in the context of choice over goods is developed in Bordalo, Gennaioli, and Shleifer (2013). In a general version of the model, the decision-maker chooses a good from a set $\mathcal{C} = \{\mathbf{x}_a\}_{a=1,\ldots,A}$ of $A > 1$ goods. Each good in the choice set $\mathcal{C}$ is a vector $\mathbf{x_a} = (x_{a1}, \ldots, x_{an})$ of attributes x_{ai} which characterize the utility obtained by the agent along a particular dimension of consumption. In the baseline case without behavioral distortions, the utility of good a is separable across consumption dimensions, with relative weights $(b_i)_{i=1,\ldots,n}$ attached to each dimension, such that $u(a) = \sum_{i=1}^n b_i x_{ai}$. Each weight b_i captures the relative significance of a dimension of consumption, absent any salience distortions. In the boundedly rational case, the agent's valuation of good a instead gets the subjective (or salience-weighted) utility:

$$u^s(a) = \sum_{i=1}^n b_i m_{ai} x_{ai} \tag{42}$$

where m_{ai} is a weight capturing the extent of the behavioral distortion, which is determined independently for each of the good's attributes. The distortion m_{ai} of the decision weight b_i is taken to be an increasing function of the *salience* of attribute i for good a with respect to a *reference point* $\bar{x}_i$. Bordalo, Gennaioli, and Shleifer (2013) propose using the average value of the attribute among goods in the choice set as a natural reference point, that is

$\bar{x}_i = \frac{1}{A}\sum_{a=1}^{A} x_{ai}$. Valuation is comparative in that what is salient about an option depends on what you compare it to.

Formally, the salience of x_{ai} with respect to $\bar{x}_i$ is given by $\sigma(x_{ai}, \bar{x}_i)$, where the salience function σ satisfies the following conditions:

Definition 4.2 *The salience function $\sigma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfies the following properties:*[62]

1. *Ordering. If $[x, y] \subset [x', y'] \in \mathbb{R}$, then $\sigma(x, y) < \sigma(x', y')$.*
2. *Diminishing Sensitivity. If $x, y \in \mathbb{R}_{>0}$, then for all $\epsilon > 0$, $\sigma(x + \epsilon, y + \epsilon) < \sigma(x, y)$.*
3. *Reflection.*[63] *If $x, y, x', y' \in \mathbb{R}_{>0}$, then $\sigma(x, y) < \sigma(x', y')$ if and only if $\sigma(-x, -y) < \sigma(-x', -y')$.*

According to these axioms, the salience of an attribute increases in its distance to that attribute's reference value, and decreases in the absolute magnitude of the reference point. The agent focuses her attention on those attributes that depart from the usual, but any given difference in an attribute's value is perceived with less intensity when the magnitude of values is uniformly higher. The reflection property guarantees a degree of symmetry between gains and losses. A tractable functional form that satisfies the properties in Definition 4.2 is

$$\sigma(x, y) = \frac{|x - y|}{|x| + |y| + \theta} \tag{43}$$

with $\theta \geq 0$ is an application-dependent parameter. This functional form additionally is symmetric in the two arguments x, y. The model is completed by specifying how the salience values σ translate into the distortion weights m_{ai}. Letting $(r_{ai})_{i=1,\ldots,n}$ be the ranking of good a's attributes according to their salience (where rank 1 corresponds to the most salient attribute), Bordalo, Gennaioli, and Shleifer (2013) propose the functional form

$$m_{ai} = Z_a \delta^{r_{ai}} \tag{44}$$

where the parameter $\delta \in (0, 1]$ measures the strength of the salience distortion, and Z_a is an application-dependent normalization.[64] In the case $\delta = 1$ we recover the fully rational agent, while in the limit case $\delta \to 0$ the agent only attends to the attribute that is most salient.[65]

[62]In Bordalo, Gennaioli, and Shleifer (2013), the additional axiom that σ is symmetric is introduced. Since the assumption of symmetry is relaxed in the case of choice among multiple goods, with multiple attributes, for expository purposes I omit it in Definition 4.2.

[63]This property is only relevant if σ admits both negative and positive arguments. This is discussed in further depth in Bordalo, Gennaioli, and Shleifer (2012).

[64]For the choice over goods, the authors propose $Z_a = \frac{n}{\sum_i \delta^{r_{ai}}}$. For probabilities, Z_a ensures that total probability is 1, $Z_a = 1/\left(\sum_i \pi_i \delta^{r_{ai}}\right)$.

[65]The distortion function (44) can exhibit discontinuous jumps. An alternative specification introduced in Bordalo, Gennaioli, and Shleifer (2016a) that allows for continuous salience distortions is $m_{ai} = Z_a e^{(1-\delta)\sigma(x_{ai}, \bar{x}_i)}$.

To see this model of salience in action, consider the case of a consumer choosing between two bottles of wine, with high (H) and low (L) quality, in a store or at a restaurant. The two relevant attributes for each good $a \in \{H, L\}$ are quality q_a and price p_a. Suppose that utility in the absence of salience distortions is $U_a = q_a - p_a$. The quality of bottle H, $q_H = 30$, is 50 percent higher than the quality of bottle L, $q_L = 20$. At the store, bottle H retails for \$20, while bottle L retails for \$10. At the restaurant, each bottle is marked up and the prices are \$60 and \$50, respectively. When is the consumer likely to choose the more expensive bottle?

While in the absence of salience distortions the agent is always indifferent between the two bottles, salience will tilt the choice in one or the other direction depending on the choice context. Taking the reference point for each attribute to be its average in the choice set, at the store we have a "reference good" $(\bar{q}_{\text{s}}, \bar{p}_s) = (25, 15)$, while at the restaurant we have a reference good $(\bar{q}_r, \bar{p}_r) = (25, 55)$. Under the functional form in (43), we can readily verify that in the store price is the more salient attribute for each wine, while at the restaurant quality is. Hence in the store the consumer focuses her attention on price and chooses the cheaper wine, while at the restaurant the markup drives attention away from prices and toward quality, leading her to choose the higher-end wine. Dertwinkel-Kalt et al. (2017) provide evidence for this effect.

Bordalo, Gennaioli, and Shleifer (2016a) further embed the salience-distorted preference structure over price and quality into a standard model of market competition. This yields a set of predictions that depart from the rational benchmark, as firms strategically make price and quality choices so as to tilt the salience of these attributes in their favor.

4.2.2 Salience and choice over lotteries

Bordalo, Gennaioli, and Shleifer (2012) develop the salience model in the context of choice over lotteries. The framework is very similar to the one discussed for the case in which we have no uncertainty. The decision-maker must choose among a set $\mathcal{C}$ of $A > 1$ lotteries. We let S be the minimal state space associated with $\mathcal{C}$, defined as the set of distinct payoff combinations that occur with positive probability. The state space S is assumed to be discrete, such that each state of the world $i \in S$ occurs with known probability π_i. The payoff of lottery a in state of the world i is x_{ai}. Absent any salience distortions, the value of lottery a is $u(a) = \sum_{i \in S} \pi_i v(x_{ai})$. Under salient thinking, the agent distorts the true state probabilities and correspondingly assigns utility

$$u^s(a) = \sum_{i \in S} \pi_i m_{ai} v(x_{ai}) \tag{45}$$

to lottery a, where the distortion weights m_{ai} are increasing in the salience of state i. Bordalo, Gennaioli, and Shleifer (2012) propose evaluating the salience of state i in lottery a by weighing its payoff against the average payoff yielded by the other lotteries in the same state of the world, meaning that the salience is given by $\sigma(x_{ai}, \bar{x}_i)$, where $\bar{x}_i = \frac{1}{A-1}\sum_{\tilde{a}\in C:\, \tilde{a}\neq a} x_{\tilde{a}i}$.

The salience model of choice under uncertainty presented in this section accounts for several empirical puzzles, including the Allais paradoxes, yielding tight quantitative predictions for the circumstances under which such choice patterns are expected to occur. For a concrete example, we consider the "common-consequence" Allais paradox as presented in Bordalo, Gennaioli, and Shleifer (2012).[66] In this version of the common-consequence Allais paradox, originally due to Kahneman and Tversky (1979), experimental participants are asked to choose between the two lotteries

$$L_1(z) = (2500,\ 0.33;\ \ 0,\ 0.01;\ \ z,\ 0.66)$$
$$L_2(z) = (2400,\ 0.34;\ \ z,\ 0.66)$$

for varying levels of the common consequence z. In a laboratory setting, when the common consequence z is high ($z = 2400$), participants tend to exhibit risk-averse behavior, preferring $L_2(2400)$ to $L_1(2400)$. However, when $z = 0$ most participants shift to risk-seeking behavior, preferring $L_1(0)$ to $L_2(0)$. This empirical pattern is not readily accounted for by the standard theory of choice under uncertainty, as it violates the axiom of independence.

In order to see how the salience model accounts for the Allais paradox, we need only derive the conditions that determine the preference ranking over lotteries in the two cases $z = 2400$ and $z = 0$. For this example, we assume the linear value function $v(x) = x$ and we take σ to be symmetric in its arguments, such that for all states $i \in S$ we have homogeneous salience rankings in the case of choice between two lotteries $a, \tilde{a}$ – that is, $r_{ai} = r_{\tilde{a}i} := r_i$. We further assume the distortion function is defined analogously to (44). These conditions yield the following necessary and sufficient criterion for lottery a to be preferred in a choice between a and $\tilde{a}$:

$$\sum_{i\in S} \delta^{r_i}\pi_i[v(x_{ai}) - v(x_{\tilde{a}i})] > 0. \tag{46}$$

When $z = 2400$, the minimum state space for the lotteries in the choice set is $S = \{(2500, 2400),\ (0, 2400),\ (2400, 2400)\}$ which from the ordering and diminishing sensitivity properties of σ yields the salience rankings

$$\sigma(0, 2400) > \sigma(2500, 2400) > \sigma(2400, 2400).$$

By criterion (46), in order to account for the preference relation $L_2(2400) \succ L_1(2400)$ it must then hold be that

$$.01\,(2400) - .33\delta\,(100) > 0,$$

[66]For experimental support of salience theory, see also Mormann and Frydman (2016).

which is true whenever $\delta < .73$. Intuitively, for low enough δ the agent focuses her attention on the salient downside of 0 in $L_1(0)$, which lowers her valuation of it. By an analogous argument, when $z = 0$ a necessary and sufficient condition for $L_1(0) \succ L_2(0)$ is that $\delta \geq 0$. Hence the Allais paradox is resolved for $\delta \in [0, .73)$, when the decision-maker exhibits salience bias of great enough significance.

4.3 Other themes

4.3.1 Attention to various time dimensions: "Focusing"

The model of focusing of Kőszegi and Szeidl (2013) expresses a shrinkage assumption similar to that of sparsity (Section 4.1), but with a different emphasis in applications, and an assumption of additivity. The model assumes that the decision-maker gives higher attention on those dimensions of the choice problem that are of primary order – which Kőszegi and Szeidl (2013) take to be the attributes along which her options vary by the largest amount. Given a choice set $\mathcal{C} = \{\mathbf{x}_a\}_{a=1\ldots A}$ of $A > 1$ actions that yield utilities $(x_{ai})_{i=1\ldots n}$ along n dimensions, the decision-maker departs from the rational benchmark $u(a) = \sum_{i=1}^{n} x_{ai}$ by distorting the importance of each consumption dimension to a degree that is increasing in the latitude of the options available to her in that dimension. Formally, we capture the range σ_i of dimension i as the range (one could imagine another way, e.g. the standard deviation of the x_{ai} across actions a)

$$\sigma_i = \max_a x_{ai} - \min_a x_{ai}. \tag{47}$$

Subjectively perceived utility is:

$$u^s(a) = \sum_{i=1}^{n} m_i x_{ai}, \tag{48}$$

where the attention weight is

$$m_i = \mathcal{A}(\sigma_i) \tag{49}$$

and the attention function $\mathcal{A}$ is increasing in the range of outcomes σ_i. Intuitively, the decision-maker attends to those dimensions of the problem in which her choice is most consequential. Hence, we obtain a formulation related to sparsity, though it does not use its general apparatus, e.g. the nonlinear framework and microfoundation for attention.

In the context of consumer finance, the focusing model explains why consumers occasionally choose expensive financing options even in the absence of liquidity constraints. Suppose an agent is buying a laptop, and has the option of either paying \$1000 upfront, or enrolling in the vendor's financing plan, which requires 12 future monthly payments of \$100. For simplicity, we assume no time-discounting and linear consumption disutility from monetary payments. We also take consumption in each period of life to be a separate dimension of the choice problem. The agent therefore chooses between two actions a_1, a_2 yielding payoff vectors $x_1 = (-1000, 0, \ldots, 0)$ and $x_2 = (0, -100, \ldots, -100)$ respectively. The vector of utility ranges is therefore $\sigma = (1000, 100, \ldots, 100)$, such that the prospect of a large upfront payment attracts the agent's attention more than the repeated but small subsequent payments. The choice-relevant comparison is between $u^s(a_1) = -\mathcal{A}(1000) \cdot 1000$ and $u^s(a_2) = -\mathcal{A}(100) \cdot 1200$. As long as $\frac{\mathcal{A}(1000)}{\mathcal{A}(100)} > 1.2$, the agent will choose the more expensive monthly payment plan, even though she does not discount the future or face liquidity constraints. Relatedly, Kőszegi and Szeidl (2013) also demonstrate how the model explains present-bias and time-inconsistency in preferences in a generalized intertemporal choice context.

Bushong, Rabin, and Schwartzstein (2016) develop a related model, where however $\mathcal{A}(\sigma)$ is decreasing in σ, though $\sigma\mathcal{A}(\sigma)$ is increasing in σ. This tends to make the agent relatively insensitive to the absolute importance of a dimension. Interestingly, it tends to make predictions opposite to those of Kőszegi and Szeidl (2013), Bordalo, Gennaioli, and Shleifer (2013) and Gabaix (2014). The authors propose that this is useful to understand present bias, if "the future" is lumped in one large dimension in the decision-making process.

4.3.2 Motivated attention

The models discussed in this section do not feature motivated attention (a close cousin of motivated reasoning) – e.g. the fact that I might pay more attention to things that favor or are pleasing to me (a self-serving bias), and avoid depressing thoughts. There is empirical evidence for this; for instance Olafsson and Pagel (2017) find that people are more likely to look at their banking account when it is flush than when it is low, an "ostrich effect" (Karlsson et al. 2009; Sicherman et al. 2016). The evidence is complex: in loss aversion, people pay more attention to losses than gains, something prima facie opposite to a self-serving attention bias. Hopefully future research will clarify this.

Let me propose the following simple model of motivated attention. Paying attention m_i to dimension i gives a psychic utility $u_{x_i} x_i \phi(m_i)$, for some increasing function ϕ. Note that x_i is "good news" iff $u_{x_i} x_i \geq 0$ (as a small innovation x_i creates a change in utility $u_{x_i} x_i$).

Using a simple variant of Step 1 in the sparsity model yields an optimal attention:[67]

$$m_i^* = \mathcal{A}_\alpha \left(\frac{1}{\kappa} \max \left(\sigma_i^2 \left| a_{x_i} u_{aa} a_{x_i} \right| + \mu u_{x_i} x_i, 0 \right) \right), \tag{51}$$

This is as in (35), but with the extra term for motivated cognition, $\mu u_{x_i} x_i$. It implies that the agent pays more attention to "good news" (i.e. news with $u_{x_i} x_i \geq 0$), with a particular strength μ of motivated cognition that might be empirically evaluated. For instance, in the basic quadratic problem the traditional action is $a^r = \sum_{i=1}^n b_i x_i$, and the behavioral action augmented by motivated reasoning becomes:

$$a^s = \sum_{i=1}^n m_i^* b_i x_i, \qquad m_i^* = \mathcal{A}_\alpha \left(\frac{\max \left(b_i^2 \sigma_i^2 + \mu b_i x_i, 0 \right)}{\kappa} \right). \tag{52}$$

Yet another model is that people might be monitoring information but are mindful of their loss aversion, i.e. avoid "bad news", along the lines of Kőszegi and Rabin (2009), Olafsson and Pagel (2017) and Andries and Haddad (2017).

4.3.3 Other decision-theoretic models of bounded rationality

In the spirit of "model substitution", interesting work of the "bounded rationality" tradition include Jehiel's (2005) analogy-based equilibrium (which has generated a sizable literature), and work of Compte and Postlewaite (2017).

4.4 Limitation of these models

These models are of course limited. They do not feature a refined "cost" of attention: for example, why it's harder to pay attention to tax changes than a funny story remains unmodeled.

Attention can be controlled, but not fully. For instance, consider someone who had a bad breakup, and can't help thinking about it during an exam. That doesn't seem fully optimal, but (in the same way that paying attention to pain is generally useful, but one would like to be able to stop paying attention to pain once under torture) this may be optimal given some constraints on the design of attention.

[67] When including psychic utility, the problem (32) becomes

$$m^* = \arg \max_{m \in [0,1]^n} \sum_{i=1}^n \left[\frac{-1}{2} (1 - m_i)^2 \Lambda_{ii} + u_{x_i} x_i \phi(m_i) - \kappa m_i^\alpha \right], \tag{50}$$

On the right-hand side, the first term is the utility loss from inattention ($\frac{-1}{2} (1 - m_i)^2 \Lambda_{ii}$), the new term is the psychic utility from the news ($u_{x_i} x_i \phi(m_i)$), and the last term is the cost of attention ($-\kappa m_i^\alpha$). If we use, for tractability, $\phi(m) = \frac{\mu}{2} \left(1 - (1 - m)^2 \right)$, where $\mu \geq 0$ parametrizes the importance of motivated cognition, then we obtain (51).

This notion of "bottom-up attention" may be modeled in a way similar to the "top-down attention" that is the center of this paper. Instead of the attention responding as $m_i^* = \mathcal{A}_\alpha (I_i)$ to the "deep" usefulness I_i of a piece of information (as captured by (35) with $I_i = \sigma_i^2 |a_{x_i} u_{aa} a_{x_i}| / \kappa$), the agent might rely on the "surface" usefulness of that information, e.g. it's written in red, and in a large font, or with some "emotional" characteristics (that is, she might use a default attention $m_i^* = \mathcal{A}_\alpha \left(I_i^d\right)$, where $I_i^d = \gamma' Y_i$, for some vector Y_i related to the characteristics of information i). Itti and Koch (2000) is an influential empirical model of such visual characteristics.

Rather than seeing those objections as fatal flaws, we shall see them as interesting research challenges.

5 A Behavioral Update of Basic Microeconomics: Consumer Theory, Arrow-Debreu

Here I present a behavioral version of basic microeconomics, based on limited attention. It is based on Gabaix (2014). Its structure does not, however, depends on the details of the endogenization of attention (i.e. from sparsity or some other procedure). Hence, the analysis applies to a large set of behavioral models, provided they incorporate some inattention to prices.

5.1 Textbook consumer theory

5.1.1 Basic consumer theory: Marshallian demand

We are now ready to see how textbook consumer theory changes for this partially inattentive agent. The rational consumer's Marshallian demand is:

$$\boldsymbol{c}(\boldsymbol{p}, w) := \arg\max_{\boldsymbol{c} \in \mathbb{R}^n} u(\boldsymbol{c}) \text{ subject to } \boldsymbol{p} \cdot \boldsymbol{c} \leq w \tag{53}$$

where $\boldsymbol{c}$ and $\boldsymbol{p}$ are the consumption vector and price vector. We denote by $\boldsymbol{c}^r(\boldsymbol{p}, w)$ the demand under the traditional $r$ational model, and by $\boldsymbol{c}^s(\boldsymbol{p}, w)$ the demand of a behavioral agent (the s stands for: demand given "subjectively perceived prices").

The price of good i is $p_i = p_i^d + x_i$, where p_i^d is the default price (e.g., the average price) and x_i is the deviation between the default price and the true price, as in Section 2.3.1. The price perceived by a behavioral agent is $p_i^s = p_i^d + m_i x_i$, i.e.:

$$p_i^s(m) = m_i p_i + (1 - m_i) p_i^d. \tag{54}$$

When $m_i = 1$, the agent fully perceives price p_i, while when $m_i = 0$, he replaces it by the default price.[68]

Proposition 5.1 (Marshallian demand). *Given the true price vector* $\boldsymbol{p}$ *and the perceived price vector* $\boldsymbol{p}^s$, *the Marshallian demand of a behavioral agent is*

$$\boldsymbol{c}^s(\boldsymbol{p}, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w'), \tag{55}$$

where the as-if budget w' *solves* $\boldsymbol{p} \cdot \boldsymbol{c}^r(\boldsymbol{p}^s, w') = w$, *i.e. ensures that the budget constraint is exactly satisfied under the true price (if there are several such* w', *take the largest one).*

To obtain intuition, we start with an example.

Example 2 (Demand by a behavioral agent with quasi-linear utility). *Take* $u(\boldsymbol{c}) = v(c_1, ..., c_{n-1}) + c_n$, *with* v *strictly concave, and assume that the price of good* n *is correctly perceived. Demand for good* $i < n$ *is independent of wealth and is:* $c_i^s(\boldsymbol{p}) = c_i^r(\boldsymbol{p}^s)$.

In this example, the demand of the behavioral agent is the rational demand given the perceived price (for all goods but the last one). The residual good n is the "shock absorber" that adjusts to the budget constraint. In a dynamic context, this good n could be "savings". Here it is a polar opposite to quasilinear demand.

Example 3 (Demand proportional to wealth). *When rational demand is proportional to wealth, the demand of a behavioral agent is:* $c_i^s(\boldsymbol{p}, w) = \frac{c_i^r(\boldsymbol{p}^s, w)}{\boldsymbol{p} \cdot \boldsymbol{c}^r(\boldsymbol{p}^s, 1)}$.

Example 4 (Demand by behavioral Cobb-Douglas and CES agents). *When* $u(\boldsymbol{c}) = \sum_{i=1}^n \alpha_i \ln c_i$, *with* $\alpha_i \geq 0$, *demand is:* $c_i^s(\boldsymbol{p}, w) = \frac{\alpha_i}{p_i^s} \frac{w}{\sum_j \alpha_j \frac{p_j}{p_j^s}}$. *When instead* $u(\boldsymbol{c}) = \sum_{i=1}^n c_i^{1-1/\eta} / (1 - 1/\eta)$, *with* $\eta > 0$, *demand is:* $c_i^s(\boldsymbol{p}, w) = (p_i^s)^{-\eta} \frac{w}{\sum_j p_j (p_j^s)^{-\eta}}$.

More generally, say that the consumer goes to the supermarket, with a budget of $w =$ \$100. Because of the lack of full attention to prices, the value of the basket in the cart is actually \$101. When demand is linear in wealth, the consumer buys 1% less of all the goods, to hit the budget constraint, and spends exactly \$100 (this is the adjustment factor $1/\boldsymbol{p} \cdot \boldsymbol{c}^r(\boldsymbol{p}^s, 1) = \frac{100}{101}$). When demand is not necessarily linear in wealth, the adjustment is (to the leading order) proportional to the income effect, $\frac{\partial \boldsymbol{c}^r}{\partial w}$, rather than to the current basket, $\boldsymbol{c}^r$. The behavioral agent cuts "luxury goods", not "necessities".[69]

[68] More general functions $p_i^s(m)$ could be devised. For instance, perceptions can be in percentage terms, i.e. in logs, $\ln p_i^s(m) = m_i \ln p_i + (1 - m_i) \ln p_i^d$. The main results go through with this log-linear formulation, because in both cases, $\frac{\partial p_i^s}{\partial p_i}_{|\boldsymbol{p}=\boldsymbol{p}^d} = m_i$.

[69] For instance, the consumer at the supermarket might come to the cashier, who'd tell him that he is over budget by \$1. Then, the consumer removes items from the cart (e.g. lowering the as-if budget w' by \$1), and presents the new cart to the cashier, who might now say that he's \$0.10 under budget. The consumers now will adjust his consumption a bit (increase w' by \$0.10). This demand here is the convergence point of this "tatonnement" process. In computer science language, the agent has access to an "oracle" (like the cashier) telling him if he's over- or under budget.

5.1.2 Asymmetric Slutsky matrix, and inferring attention from choice data, and nominal illusion,

The Slutsky matrix The Slutsky matrix is an important object, as it encodes both elasticities of substitution and welfare losses from distorted prices. Its element S_{ij} is the (compensated) change in consumption of c_i as price p_j changes:

$$S_{ij}(\boldsymbol{p}, w) := \frac{\partial c_i(\boldsymbol{p}, w)}{\partial p_j} + \frac{\partial c_i(\boldsymbol{p}, w)}{\partial w} c_j(\boldsymbol{p}, w). \tag{56}$$

With the traditional agent, the most surprising fact about it is that it is symmetric: $S^r_{ij} = S^r_{ji}$. Mas-Colell, Whinston, and Green (1995, p.70) comment: "Symmetry is not easy to interpret in plain economic terms. As emphasized by Samuelson (1947), it is a property just beyond what one would derive without the help of mathematics."

Now, if a prediction is non-intuitive to Mas-Colell, Whinston, and Green, it might require too much sophistication from the average consumer. We now present a less rational, and psychologically more intuitive, prediction.

Proposition 5.2 (Slutsky matrix). *Evaluated at the default price, the Slutsky matrix S^s is, compared to the traditional matrix S^r:*

$$S^s_{ij} = S^r_{ij} m_j, \tag{57}$$

i.e. the behavioral demand sensitivity to price j is the rational one, times m_j, the salience of price j. As a result the behavioral Slutsky matrix is not symmetric in general. Sensitivities corresponding to "non-salient" price changes (low m_j) are dampened.

Proof sketch. For simplicity, here I only show the proof in the the quasi-linear case of Example 2, when the price of the last good is correctly perceived at 1. We call $\boldsymbol{C} = (c_1, \ldots, c_{n-1})$, so that $u(\boldsymbol{C}, c_n) = v(\boldsymbol{C}) + c_n$, and set $\boldsymbol{P} = (p_1, \ldots, p_{n-1})$. We restrict our attention to good $i, j < n$. The consumer's first order condition is $v_{c_i} = p^s_i$, i.e. $v'(\boldsymbol{C}) = \boldsymbol{P}^s$, so the demand is $\boldsymbol{C}^s(\boldsymbol{P}^s) = \boldsymbol{C}^r(\boldsymbol{P}^s)$, where $\boldsymbol{C}(\boldsymbol{P}) = v'^{-1}(\boldsymbol{P})$ is the rational demand. Then, $S_{ij} = \frac{\partial c^s_i(\boldsymbol{p}, w)}{\partial p_j}$, as there is no income effect (by the way, that expression allows to verify that the Slutsky matrix is symmetric in the rational case). So, evaluating the derivatives at $\boldsymbol{p}^s = \boldsymbol{p}^d$,

$$S^s_{ij} = \frac{\partial c^s_i(\boldsymbol{p})}{\partial p_j} = \frac{dc^r_i\left(m_j p_j + (1 - m_j) p^d_j\right)}{dp_j} = m_j \frac{\partial c^r_i(q_j)}{\partial q_j} = m_j S^r_{ij}$$

This gives the result. □

Instead of looking at the full price change, the consumer just reacts to a fraction m_j of it. Hence, he's typically less responsive than the rational agent. For instance, say that $m_i > m_j$, so that the price of i is more salient than price of good j. The model predicts that $\left|S^s_{ij}\right|$ is lower than $\left|S^s_{ji}\right|$: as good j's price isn't very salient, quantities don't react much to it. When $m_j = 0$, the consumer does not react at all to price p_j, hence the substitution effect is zero.

The asymmetry of the Slutsky matrix indicates that, in general, *a behavioral consumer cannot be represented by a rational consumer who simply has different tastes or some adjustment costs.* Such a consumer would have a symmetric Slutsky matrix.

To the best of my knowledge, this is the first derivation of an asymmetric Slutsky matrix in a model of bounded rationality.[70]

Equation (57) makes tight testable predictions. It allows us to infer attention from choice data, as we shall now see.[71]

Proposition 5.3 (Estimation of limited attention). *Choice data allows one to recover the attention vector m, up to a multiplicative factor $\overline{m}$. Indeed, suppose that an empirical Slutsky matrix S^s_{ij} is available. Then, m can be recovered as $m_j = \overline{m}\prod_{i=1}^n \left(\frac{S^s_{ij}}{S^s_{ji}}\right)^{\gamma_i}$, for any $(\gamma_i)_{i=1...n}$ such that $\sum_i \gamma_i = 1$.*

Proof: We have $\frac{S^s_{ij}}{S^s_{ji}} = \frac{m_j}{m_i}$, so $\prod_{i=1}^n \left(\frac{S^s_{ij}}{S^s_{ji}}\right)^{\gamma_i} = \prod_{i=1}^n \left(\frac{m_j}{m_i}\right)^{\gamma_i} = \frac{m_j}{\overline{m}}$, for $\overline{m} := \prod_{i=1}^n m_i^{\gamma_i}$.$\square$

The underlying "rational" matrix can be recovered as $S^r_{ij} := S^s_{ij}/m_j$, and it should be symmetric, a testable implication.[72] There is a old literature estimating Slutsky matrices – but it had not explored the role of non-salient prices (a recent exception is the nascent literature mentioned in Section 3.1.2).

It would be interesting to test Proposition 5.2 directly. The extant evidence is qualitatively encouraging, via the literature on tax salience and shrouded attributes (Sections 3.2.1-3.2.2), and the recent literature exploring this Slutsky asymmetry (Section 3.1.2).

Marginal demand

Proposition 5.4 *The Marshallian demand $\boldsymbol{c}^s(\boldsymbol{p}, w)$ has the marginals (evaluated at $\boldsymbol{p} = \boldsymbol{p}^d$): $\frac{\partial \boldsymbol{c}^s}{\partial w} = \frac{\partial \boldsymbol{c}^r}{\partial w}$ and*

$$\frac{\partial c^s_i}{\partial p_j} = \frac{\partial c^r_i}{\partial p_j} \times m_j - \frac{\partial c^r_i}{\partial w} c^r_j \times (1 - m_j). \tag{58}$$

This means that, though substitution effects are dampened, income effects ($\frac{\partial \boldsymbol{c}}{\partial w}$) are preserved (as w needs to be spent in this one-shot model).

Nominal illusion Recall that the consumer "sees" only a part m_j of the price change (eq. (54)). One consequence is nominal illusion.

[70] Browning and Chiappori (1998) have in mind a very different phenomenon: intra-household bargaining, with full rationality.

[71] The Slutsky matrix does not allow one to recover $\overline{m}$: for any $\overline{m}$, S^s admits a dilated factorization $S^s_{ij} = \left(\overline{m}^{-1} S^r_{ij}\right)(\overline{m} m_j))$. To recover $\overline{m}$, one needs to see how the demand changes as $\boldsymbol{p}^d$ varies. Aguiar and Serrano (2017) explore further the link between Slutsky matrix and bounded rationality.

[72] Here, we find again a less intuitive aspect of the Slutsky matrix.

Proposition 5.5 (Nominal illusion) *Suppose that the agent pays more attention to some goods than others (i.e. the m_i are not all equal). Then, the agent exhibits nominal illusion, i.e. the Marshallian demand $\boldsymbol{c}(\boldsymbol{p}, w)$ is (generically) not homogeneous of degree 0.*

To gain intuition, suppose that prices and the budget all increase by 10%. For a rational consumer, nothing really changes and he picks the same consumption. However, consider a behavioral consumer who pays more attention to good 1 ($m_1 > m_2$). He perceives that the price of good 1 has increased more than the price of good 2 has (he perceives that they have respectively increased by $m_1 \cdot 10\%$ vs $m_2 \cdot 10\%$). So, he perceives that the *relative* price of good 1 has increased ($\boldsymbol{p}^d$ is kept constant). Hence, he consumes less of good 1, and more of good 2. His demand has shifted. In abstract terms, $\boldsymbol{c}^s(\chi \boldsymbol{p}, \chi w) \neq \boldsymbol{c}^s(\boldsymbol{p}, w)$ for $\chi = 1.1$, i.e. the Marshallian demand is not homogeneous of degree 0. The agent exhibits nominal illusion.[73]

5.2 Textbook competitive equilibrium theory

We next revisit the textbook chapter on competitive equilibrium, with less than fully rational agents. We will use the following notation. Agent $a \in \{1, ..., A\}$ has endowment $\boldsymbol{\omega}^a \in \mathbb{R}^n$ (i.e. he is endowed with ω_i^a units of good i), with $n > 1$. If the price is $\boldsymbol{p}$, his wealth is $\boldsymbol{p} \cdot \boldsymbol{\omega}^a$, so his demand is $\boldsymbol{D}^a(\boldsymbol{p}) := \boldsymbol{c}^a(\boldsymbol{p}, \boldsymbol{p} \cdot \boldsymbol{\omega}^a)$. The economy's excess demand function is $\boldsymbol{Z}(\boldsymbol{p}) := \sum_{a=1}^{A} \boldsymbol{D}^a(\boldsymbol{p}) - \boldsymbol{\omega}^a$. The set of equilibrium prices is $\mathcal{P}^* := \left\{\boldsymbol{p} \in \mathbb{R}_{++}^n : \boldsymbol{Z}(\boldsymbol{p}) = 0\right\}$. The set of equilibrium allocations for a consumer a is $\mathcal{C}^a := \{\boldsymbol{D}^a(\boldsymbol{p}) : \boldsymbol{p} \in \mathcal{P}^*\}$. The equilibrium exists under weak conditions laid out in Debreu (1970).

5.2.1 First and second welfare theorems: (In)efficiency of equilibrium

We start with the efficiency of Arrow-Debreu competitive equilibrium, i.e. the first fundamental theorem of welfare economics.[74] We assume that competitive equilibria are interior, and consumers are locally non-satiated.

Proposition 5.6 (First fundamental theorem of welfare economics revisited: (In)efficiency of competitive equilibrium). *An equilibrium is Pareto efficient if and only if the perception of relative prices is identical across agents. In that sense, the first welfare theorem generally fails.*

Hence, typically the equilibrium is not Pareto efficient when we are not at the default price. The intuitive argument is very simple (the appendix has a rigorous proof): recall

[73] See Eyster et al. (2017) for progress on the impact of nominal illusion from inattention.

[74] This chapter does not provide the producer's problem, which is quite similar and is left for a companion paper (and is available upon request). Still, the two negative results in Propositions 5.6 and 5.7 apply to exchange economies, hence apply a fortiori to production economies.

that given two goods i and j, each agent equalizes relative marginal utilities and relative perceived prices (see equation (40)):

$$\frac{u^a_{c_i}}{u^a_{c_j}} = \left(\frac{p^s_i}{p^s_j}\right)^a, \qquad \frac{u^b_{c_i}}{u^b_{c_j}} = \left(\frac{p^s_i}{p^s_j}\right)^b, \tag{59}$$

where $\left(\frac{p^s_i}{p^s_j}\right)^a$ is the relative price perceived by consumer a. Furthermore, the equilibrium is efficient if and only if the ratio of marginal utilities is equalized across agents, i.e. there are no extra gains from trade, i.e.

$$\frac{u^a_{c_i}}{u^a_{c_j}} = \frac{u^b_{c_i}}{u^b_{c_j}}. \tag{60}$$

Hence, the equilibrium if efficient if and only if any consumers a and b have the same perceptions of relative prices ($\left(\frac{p^s_i}{p^s_j}\right)^a = \left(\frac{p^s_i}{p^s_j}\right)^b$).

The second welfare theorem asserts that any desired Pareto efficient allocation $(\boldsymbol{c}^a)_{a=1\ldots A}$ can be reached, after appropriate budget transfers (for a formal statement, see e.g., Mas-Colell, Whinston, and Green 1995, section 16.D). The next Proposition asserts that it generally fails in this behavioral economy. The intuition is as follows: typically, if the first welfare theorem fails, then a fortiori the second welfare theorem fails, as an equilibrium is typically not efficient.

Proposition 5.7 (Second theorem of welfare economics revisited). *The second welfare theorem generically fails, when there are strictly more than two consumers or two goods.*

5.2.2 Excess volatility of prices in an behavioral economy

To tractably analyze prices, we follow the macro tradition, and assume in this section that there is just one representative agent. A core effect is the following.

Bounded rationality leads to excess volatility of equilibrium prices. Suppose that there are two dates, and that there is a supply shock: the endowment $\boldsymbol{\omega}(t)$ changes between $t = 0$ and $t = 1$. Let $d\boldsymbol{p} = \boldsymbol{p}(1) - \boldsymbol{p}(0)$ be the price change caused by the supply shock, and consider the case of infinitesimally small changes (to deal with the arbitrariness of the price level, assume that $p_1 = p^d_1$ at $t = 1$). We assume $m_i > 0$ (and will derive it soon).

Proposition 5.8 (Bounded rationality leads to excess volatility of prices). *Let $d\boldsymbol{p}^{[r]}$ and $d\boldsymbol{p}^{[s]}$ be the change in equilibrium price in the rational and behavioral economies, respectively. Then:*

$$dp^{[s]}_i = \frac{dp^{[r]}_i}{m_i}, \tag{61}$$

i.e., after a supply shock, the movements of price i in the behavioral economy are like the movements in the rational economy, but amplified by a factor $\frac{1}{m_i} \geq 1$. Hence, ceteris paribus, *the prices of non-salient goods are more volatile. Denoting by* σ_i^k *the price volatility in the rational* ($k = r$) *or behavioral* ($k = s$) *economy, we have* $\sigma_i^s = \frac{\sigma_i^r}{m_i}$.

Hence, non-salient prices need to be more volatile to clear the market. This might explain the high price volatility of many goods, such as commodities. Consumers are quite price inelastic, because they are inattentive. *In a behavioral world, demand underreacts to shocks; but the market needs to clear, so prices have to overreact to supply shocks.*[75]

5.3 What is robust in basic microeconomics?

I distinguish what appears to be robust and not robust in the basic microeconomic theory of consumer behavior and competitive equilibrium – when the specific deviation is a sparsity-seeking agent. I use the sparsity benchmark not as "the truth," of course, but as a plausible extension of the traditional model, when agents are less than fully rational. I contrast the traditional (or "classical") model to a behavioral model with inattention.

Propositions that are not robust

Tradition: There is no money illusion. *Behavioral* model: There is money illusion: when the budget and prices are both increased by 5%, the agent consumes less of goods with a salient price (which he perceives to be relatively more expensive); Marshallian demand $\boldsymbol{c}(\boldsymbol{p}, w)$ is not homogeneous of degree 0.

Tradition: The Slutsky matrix is symmetric. *Behavioral* model: It is asymmetric, as elasticities to non-salient prices are attenuated by inattention.

Tradition: The competitive equilibrium allocation is independent of the price level. *Behavioral* model: Different aggregate price levels lead to materially different equilibrium allocations, as implied by a Phillips curve.

Greater robustness: Objects are very close around the default price, up to second order terms

Tradition: People maximize their "objective" welfare. *Behavioral* model: people maximize in default situations, but there are losses away from it.

Tradition: Competitive equilibrium is efficient, and the two Arrow-Debreu welfare theorems hold. *Behavioral* model: Competitive equilibrium is efficient if it happens at the default price. Away from the default price, competitive equilibrium has inefficiencies, unless all agents have the same misperceptions. As a result, the two welfare theorems do not hold in general.

Traditional economics gets the signs right – or, more prudently put, the signs predicted by the rational model (e.g. Becker-style price theory) are robust under a sparsity variant. Those predictions are of the type "if the price of good 1 goes down, demand for it goes

[75] Gul, Pesendorfer, and Strzalecki (2017) offer a very different model leading to volatile prices, with a different mechanism linked to endogenous heterogeneity between agents.

up", or more generally "if there's a good incentive to do X, people will indeed tend to do X."[76,77] Those sign predictions make intuitive sense, and, not coincidentally, they hold in the behavioral model:[78] those sign predictions (unlike quantitative predictions) remain unchanged even when the agent has a limited, qualitative understanding of his situation. Indeed, when economists think about the world, or in much applied microeconomic work, it is often the sign predictions that are used and trusted, rather than the detailed quantitative predictions.

6 Models with Stochastic Attention and Choice of Precision

We now move on to models with noisy signals. They are more complex to handle, as they provide a stochastic prediction, not a deterministic one. There are pros and cons to that. One pro is that economists can stick to optimal information processing. In addition, the amount of noise may actually be a guide to the thought process, hence might be a help rather than a hindrance: see Glimcher (2011) and Caplin (2016). The drawback is basically the complexity of this approach – these models become quickly intractable.

Interestingly, much of the neuroeconomics (Glimcher and Fehr 2013) and cognitive psychology (Gershman, Horvitz, and Tenenbaum 2015; Griffiths, Lieder, and Goodman 2015) literatures sees the brain as an optimal information processor. Indeed, for low-level processes (e.g. vision), the brain may well be optimal, though for high-level processes (e.g. dealing with the stock market) it is not directly optimal.

6.1 Bayesian models with choice of information

There are many Bayesian models in which agents pay to get more precise signals. An early example is Verrecchia (1982): agents pay to receive more precise signals in a financial market. In Geanakoplos and Milgrom (1991), managers pay for more information. They essentially all work with linear-quadratic settings – otherwise the task is intractable. In the basic problem of Section 2.1, the expected loss is

$$\mathbb{E}\left[\max_{a}\mathbb{E}\left[-\frac{1}{2}\left(a-x\right)^{2}|s\right]\right]=-\frac{1}{2}\left(1-m\right)\sigma_{x}^{2}$$

[76]Those predictions need not be boring. For instance, when divorce laws are relaxed, spouses kill each other less (Stevenson and Wolfers 2006).

[77]This is true for "direct" effects, though not necessarily once indirect effects are taken into account. For instance, this is true for compensated demand (see the part on the Slutsky matrix), and in partial equilibrium. This is not necessarily true for uncompensated demand (where income effects arise) or in general equilibrium – though in many situations those "second round" effects are small.

[78]The closely related notion of strategic complements and substitutes (Bulow, Geanakoplos, and Klemperer 1985) is also robust to a sparsity deviation.

so that the agent's problem is:

$$\max_{\tau} -\frac{1}{2}(1-m)\sigma_x^2 - \kappa G(\tau) \text{ subject to } m = \frac{\tau}{1+\tau}$$

where $\tau = \frac{\sigma_x^2}{\sigma_\varepsilon^2}$ is the relative precision of the signal, and G is the cost of precision, which is increasing. This can be equivalently reformulated as:

$$\max_{m} -\frac{1}{2}(1-m)\sigma_x^2 - \kappa g(m)$$

by defining $g(m)$ appropriately $\left(g(m) := G\left(\frac{m}{1-m}\right)\right)$. So, we have a problem very much like (33).

This allows us to think about the optimal choice of information. When actions are strategic complements, you can get multiple equilibria in information gathering (Hellwig and Veldkamp 2009). When actions are strategic substitutes, you often obtain specialization in information (Van Nieuwerburgh and Veldkamp 2010). More generally, rational information acquisition models do seem to predict qualitatively relevant features of real markets (Kacperczyk, Van Nieuwerburgh, and Veldkamp 2016).

6.2 Entropy-based inattention: "Rational inattention"

Sims (1998, 2003) extends the ideas to allow for larger choice sets in which agents freely choose the properties of their signals. He uses the entropy penalty to handle non-Gaussian variables.

6.2.1 Information theory: A crash course

Here is a brief introduction to information theory, as developed by Shannon (1948). The basic textbook for this is Cover and Thomas (2006).

Discrete variables Take a random variable X with probability p_i of a value x_i. Throughout, we will use the notation f to refer to the probability mass function of a given random variable (when discrete), or to its probability density function (when continuous). Then the entropy of X for the discrete case is defined as

$$H(X) = -\mathbb{E}[\log f(X)] = -\sum_i p_i \log p_i$$

so that $H \geq 0$ (for a discrete variable; it won't be true for a continuous variable). In the case where uncertainty between outcomes is greatest, X can take n equally probable values,

$p_i = \frac{1}{n}$. This distribution gives the maximum entropy,

$$H(X) = \log n$$

which illustrates that higher uncertainty yields higher entropy.

This measure of "complexity" is really a measure of the complexity of communication, not of finding or processing information. For instance, the entropy of a coin flip is $\log 2$ – one bit if we use the base 2 logarithm. But also, suppose that you have to communicate the value of the 1000th figure in the binary expansion of $\sqrt{17}$. Then, the entropy of that is again simply $\log 2$. This is the not the cost of actually processing information (which is a harder thing to model), just the cost of transmitting the information.

Suppose we have two independent random variables, with $X = (Y, Z)$. Then, $f^X(y, z) = f^Y(y) f^Z(z)$ so

$$\begin{aligned} H(X) &= -\mathbb{E}\left[\log f^X(X)\right] = -\mathbb{E}\left[\log\left(f^Y(Y) f^Z(Z)\right)\right] \\ &= -\mathbb{E}\left[\log f^Y(Y) + \log f^Z(Z)\right] \\ H(X) &= H(Y) + H(Z). \end{aligned} \tag{62}$$

This shows that the information of independent outcomes is additive. The next concept is that of mutual information, for which we drop the assumption that the two variables of interest are independent. It is defined by the reduction of entropy of X when you know Y:

$$\begin{aligned} I(X,Y) &= H(X) - H(X|Y) \\ &= -\mathbb{E}\left[\log f^X(X)\right] + \mathbb{E}\left[\log f^{X|Y}(X|Y)\right] = -\mathbb{E}\left[\log f^X(X)\right] + \mathbb{E}\left[\log \frac{f(X,Y)}{f^Y(Y)}\right] \\ &= -\mathbb{E}\left[\log f^X(X) + \log f^Y(Y)\right] + \mathbb{E}\left[\log f(X,Y)\right] \\ &= H(X) + H(Y) - H(X,Y) = I(Y,X), \end{aligned}$$

and so it follows that mutual information is symmetric. The next concept is the Kullback-Leibler divergence between two distributions p, q,

$$D(p\|q) = \mathbb{E}^P\left[\log \frac{p(X)}{q(X)}\right] = \sum_i p_i \log \frac{p_i}{q_i}. \tag{63}$$

Note that the Kullback-Leibler divergence is not actually a proper distance, since $D(p\|q) \neq D(q\|p)$, but it is similar to a distance – it is nonnegative, and equal to 0 when $p = q$.

Hence, we have:

$$I(X,Y) = D\left(f(x,y) \,\|\, f^X(x) f^Y(y)\right) = \sum_{x,y} f(x,y) \log \frac{f(x,y)}{f^X(x) f^Y(y)}, \tag{64}$$

or, in other words, mutual information $I(X,Y)$ is the Kullback-Leibler divergence between the full joint probability $f(x,y)$ and its "decoupled" approximation $f^X(x)\, f^Y(y)$.

Continuous variables With continuous variables with density $f(x)$, entropy is defined to be:

$$H(X) = -\mathbb{E}[\log f(X)] = -\int f(x) \log f \, dx$$

with the convention that $f(x) \log f(x) = 0$ if $f(x) = 0$. For instance, if X is a uniform $[a,b]$, then $f(x) = \frac{1}{b-a} 1_{x \in [a,b]}$ and

$$H(X) = \log(b-a) \tag{65}$$

which shows that we can have a negative entropy, $(H(X) < 0)$ with continuous variables.

If $Y = a + \sigma X$, then because $f^Y(y)\, dy = f^X(x)\, dx$, i.e. $f^Y(y) = \frac{1}{\sigma} f^X(x)$, we have

$$H(Y) = -E\left[\log f^Y(Y)\right] = -E\left[\log f^X(X)\right] + \log \sigma$$

$$H(Y) = H(X) + \log \sigma \tag{66}$$

so with continuous variables, multiplying a variable by σ increases its entropy by $\log \sigma$.

The entropy of a Gaussian $N(\mu, \sigma^2)$ variable is, as shown in Appendix A,

$$H(X) = \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log(2\pi e) \tag{67}$$

and for a multi-dimensional Gaussian with variance-covariance matrix V, the entropy is

$$H(X) = \frac{1}{2} \log(\det V) + \frac{n}{2} \log(2\pi e) \tag{68}$$

which is analogous to the one-dimensional formula, but σ^2 is replaced by $\det V$.

Mutual information in a Gaussian case Suppose X, Y are jointly Gaussian with correlation ρ. Then, their mutual information is

$$I(X,Y) = \frac{1}{2} \log \frac{1}{1-\rho^2} \tag{69}$$

so that the mutual information is increasing in the correlation.

6.2.2 Using Shannon entropy as a measure of cost

Sims (2003) proposed the following problem, which had two innovations: the use of entropy, and a reformulation of the choice of the signal structure (both of which can be generalized). Consider an agent that has no information or attention costs and makes choices by maximizing $u(a,x)$. In the Sims version, the agent will pick a stochastic action A drawn from an endogenously chosen density $q(a|x)$ – i.e., the probability density of a given the true state is x – where q is chosen by the optimization problem

$$\max_{q(a|x)} \int u(a,x)\, q(a|x)\, f(x)\, da\, dx \text{ s.t. } I(A,X) \leq K, \tag{70}$$

that is, the agent instructs some black box to give him a stochastic action: the box sees the true x, and then returns a noisy prescription $q(a|x)$ for his action.[79] Of course, the nature of this black box is a bit unclear, but may be treated as some thought process.[80]

A simple example To get a feel for this problem, revisit the "targeting problem" seen above, where $x \sim N(0, \sigma^2)$, and $u(a,x) = -\frac{1}{2}(a-x)^2$. The solution is close to that in Section 2.1: the agent receives a noisy signal $s = x + \varepsilon$, and takes the optimal action $a(s) = \mathbb{E}[x|s] = mx + m\varepsilon$ with $m = \frac{\sigma^2}{\sigma^2+\sigma_\varepsilon^2}$. The loss utility achieved is:

$$U = \mathbb{E}[u(a(s), x)] = -\frac{1}{2}(mx + m\varepsilon - x)^2 = -\frac{1}{2}(1-m)\sigma^2 \tag{71}$$

The analytics in (92) shows that $\rho^2 = \text{corr}(a(s), x)^2 = m$, and using equation (69), the mutual information is $I(a(s), x) = \frac{1}{2}\log\frac{1}{1-\rho^2} = \frac{1}{2}\log\frac{1}{1-m}$. Hence, the decision problem (70) boils down to:

$$\max_m -\frac{1}{2}(1-m)\sigma^2 \text{ s.t. } \frac{1}{2}\log\frac{1}{1-m} \leq K$$

This gives:

$$m = 1 - e^{-2K} \tag{72}$$

and the action has the form:

$$a^{\text{Sims}} = mx + \eta$$

with $\eta = m\varepsilon$. So, we get a solution as in the basic problem of Section 4.1, with a cost function $\mathcal{C}(m) = \frac{1}{2}\log\frac{1}{1-m}$.

If we wish to calibrate an attention $m \leq 0.9$, equation (72) implies that we need $K \leq \frac{-\log(1-m)}{2}$, which gives $K \leq 1.15$ "natural units" or a capacity of at most 1.7 bits. This is a very small capacity.

Extensions to multiple dimensions Now, let us see the multidimensional version, with the basic quadratic problem (11): $u(a,x) = -\frac{1}{2}(a - \sum_{i=1}^n b_i x_i)^2$. We assume that the x_i are uncorrelated, jointly Gaussian, that $\text{Var}(x_i) = \sigma_i^2$. Then, one can show that the solution

[79]The density chosen is non-parametric, and does not involve explicitly sending intermediary signals as in the prior literature – which is an innovation by Sims (2003).

[80]In the Shannon theory, this nature is clear. Originally, the Shannon theory is a theory of communication. Someone has information x at one end of the line, and needs to communicate information a at the other end of the line. Under the setup of the Shannon theory, the cost is captured by the mutual information $I(A,X)$.

takes the form

$$a^{\text{Sims}} = m \sum_i b_i x_i + \eta \tag{73}$$

with a η orthogonal to the x.

This may be a good place to contrast the Sims approach and sparsity. For the same quadratic problem, Equation (37) yields

$$a^s = \sum_{i=1}^{n} m_i b_i x_i \tag{74}$$

so that the agent can pay more attention to source 1 than to source 10 (if $m_1 > m_{10}$). Hence, with the global entropy constraint of Sims we obtain uniform dampening across all variables (i.e. $m_i = m$ for all i in equation (73)) – not source-specific dampening as in (37)-(74).[81]

Now, a drawback of sparsity (and related models like Bordalo et al. (2013)) is that framing (in the sense of partitioning of attributes into dimensions) matters in this model, whereas it does not in the Sims approach. This is important to its ability to generate non-uniform dampening across dimensions. On the other hand, this seems realistic: in the base example of the price with a tax (section 2.3.2), it does matter empirically whether the price is presented as two items (price, tax), or as one composite information price plus tax.

Discussion One advantage of the entropy-based approach is that we have a universally applicable measure of the cost of information. In simple linear-quadratic-Gaussian cases, the models are similar (except for the important fact that Sims generates uniform, rather than source-dependent, dampening). When going away from this linear-quadratic-Gaussian case, the modeling complexity of sparsity remains roughly constant (and attention still obtains in closed form). In Sims, this is not the case, and quickly problems become extremely hard to solve. For instance, in Jung, Kim, Matějka, and Sims (2015), the solution has atoms – it is non-smooth. One needs a computer to solve it.[82] All in all, it is healthy for economics that different approaches are explored in parallel.

Sims called this modelling approach "rational inattention". This name may be overly broad, as Sims really proposed a particular modelling approach, not at all the general notion that attention allocation responds to incentives. That notion comes from the 1960s and information economics–dating back at least to Stigler (1961), where agents maximize utility subject to the cost of acquiring information, so that information and attention responds to costs and benefits. There are many papers under that vein, e.g. Verrecchia (1982); Geanakoplos and Milgrom (1991). Hence, a term such as "entropy-based inattention" seems

[81] There is a "water-filling" result in information theory that generates source-dependent attention, but it requires different channels, not the Sims unitary attention channel.

[82] This is can be seen as a drawback, but Matějka (2016) proposes that this can be used to model pricing with a discrete set of prices.

like a proper name for the specific literature initiated by Sims.[83]

Still, a great virtue of the entropy-based approach is that it has attracted the energy of many economists, especially in macro (Maćkowiak and Wiederholt 2009, 2015; Veldkamp 2011; Khaw, Stevens, and Woodford 2016), e.g. to study the inattention to the fine determinants of pricing by firms. Many depart from the "global entropy penalty," which allows one to have source-specific inattention. But then, there is no real reason to stick to the entropy penalty in the first place–other cost functions will work similarly. Hence researchers keep generalizing the Shannon entropy, for instance Caplin, Dean, and Leahy (2017); Ellis (2018).

6.3 Random choice via limited attention

6.3.1 Limited attention as noise in perception: Classic perspective

A basic model is the random choice model. The consumer must pick one of n goods. Utility is v_i, drawn from a distribution $f(v)$. In the basic random utility model à la Luce-McFadden (Manski and McFadden 1981), the probability of choosing v_i is

$$p_i = \frac{e^{v_i/\sigma}}{\sum_j e^{v_j/\sigma}}, \tag{75}$$

with the following classic microfoundation: agents receive a signal

$$s_i = v_i + \sigma \varepsilon_i \tag{76}$$

where the ε_i are i.i.d. with a Gumbel distribution, $\mathbb{P}(\varepsilon_i \leq x) = e^{-e^{-x}}$. The idea is that ε_i is noise in perception, and perhaps it could be decreased actively by agents, or increased by firms.

Agents have diffuse priors on v_i. Hence, they choose the good j with the highest signal s_t, $p_i = \mathbb{P}\left(i \in \operatorname{argmax}_j s_j\right)$. With Gumbel noise, this leads (after some calculations as in e.g. Anderson, De Palma, and Thisse 1992) to (75). When the noise scaling parameter σ is higher, there is more uncertainty; when $\sigma \to \infty$, then $p_i \to \frac{1}{n}$. The choice is completely random.

This is a useful model, because it captures in a simple way "noisy perceptions". It has proven very useful in industrial organization (e.g. Anderson, De Palma, and Thisse 1992) – where the typical interpretation is "rational differences in tastes", rather than "noise in the perception". It can be generalized in a number of ways, including with correlated noises, and non-Gumbel noise (Gabaix et al. 2016). This framework can be used to analyze equilibrium prices when consumers are confused and/or when firms create noise to confuse consumers. Then, the equilibrium price markup (defined as price minus cost) is generally proportional

[83] Maćkowiak et al. (2018) is a recent survey.

to σ, the amount of noise. For a related model with two types of agents, see Carlin (2009).

6.3.2 Random choice via entropy penalty

Matějka and McKay (2015) derive an entropy-based foundation for the logit model. In its simplest form, the idea is as follows. The consumer must pick one of n goods. Utility is v_i, drawn from a distribution $f(v)$. The endogenous probability of choosing i is p_i. The problem is to maximize utility subject to a penalty for having an inaccurate probability:

$$\max_{(p_i(v))_{i=1\ldots n}} \mathbb{E}\left[\sum_i p_i(v)\, v_i\right] - \kappa D\left(P \| P^d\right)$$

where the expectation is taken over the value of v, and $D\left(P\|P^d\right)$ is the Kullback-Leibler distance between the probability and a default probability, P^d. Hence, we have a penalty for a "sophisticated" probability distribution that differs from the default probability.[84] So, the Lagrangian is

$$L = \int \sum_i p_i(v)\, v_i f(v)\, dv - \kappa \int \left[\sum_i p_i(v) \log \frac{p_i(v)}{p_i^d}\right] f(v)\, dv - \int \mu(v) \left(\sum p_i(v) - 1\right) f(v) dv.$$

Differentiation with respect to $p_i(v)$ gives $0 = v_i - \kappa(1 + \log \frac{p_i(v)}{p_i^d}) - \mu(v)$, i.e.$p_i(v) = p_i^d e^{v_i/\kappa} K(v)$ for a value $K(v)$. Ensuring that $\sum_i p_i(v) = 1$ gives

$$p_i(v) = \frac{p_i^d e^{v_i/\kappa}}{\sum_j p_j^d e^{v_j/\kappa}}. \tag{77}$$

so this is like (75), with σ replaced by κ, and with default probabilities being uniform. When the cost κ is 0, then the agent is the classical rational agent.

Matějka and McKay's setup (2015) actually gives the default: $\max_{\left(p_i^d\right)} \mathbb{E}\left[\log\left(\sum_i p_i^d e^{v_i/\kappa}\right)\right]$ s.t. $\sum_i p_i^d = 1$. So, when the v_i are drawn from the same distribution, $p_i^d = 1/n$.

In some cases, some options will not even be looked at, so $p_i^d = 0$. This gives a theory of consideration sets. For related work, see Masatlioglu, Nakajima, and Ozbay (2012); Manzini and Mariotti (2014); Caplin, Dean, and Leahy (2016). This in turn helps explore dynamic problems, as in Steiner, Stewart, and Matějka (2017).

[84] The math is analogous to the basic derivation of the "Boltzmann distribution" familiar to statistical mechanics. Maximizing the entropy $H(P)$ subject to a given energy constraint $\sum_i p_i v_i = V$ yields a distribution $p_i = \frac{e^{-\beta v_i}}{\sum_j e^{-\beta v_j}}$ for some β.

7 Allocation of Attention over Time

The models discussed so far were static. We now move on to models that incorporate the allocation of attention over time. One important theme is that people are initially inattentive to a new piece of information, but over time they adjust to the news – a form of "sluggishness". I cover different ways to generate sluggishness, particularly over time. They are largely substitutes for inattention from a modeling standpoint, but they generate sometimes different predictions, as we shall see.

7.1 Generating sluggishness: Sticky action, sticky information, and habits

7.1.1 Sticky action and sticky information

The most common models are those of sticky action and sticky information. In the sticky action model, agents need to pay a cost to change their action. In the sticky information model, agents need to pay a cost to change their information. Sticky action has been advocated in macroeconomics by Calvo (1983) and Caballero (1995), and in a finance context by Duffie and Sun (1990) and Lynch (1996). Sticky information has been advocated in macro by Gabaix and Laibson (2002), Carroll (2003), and then by Mankiw and Reis (2002), Reis (2006a), Bacchetta and Van Wincoop (2010) and numerous authors since. Coibion and Gorodnichenko (2015) finds evidence for slow adjustment to information. Intuitively, this generates sluggishness in the aggregate action. To see this, consider the following tracking problem. The agent should maximize

$$V = \sum_{t=0}^{\infty} \beta^t u(a_t, x_t) \tag{78}$$

$$u(a, x) = -\frac{1}{2}(a - x)^2 \tag{79}$$

where a_t is a decision variable, and x_t an exogenous variable satisfying:

$$x_{t+1} = \rho x_t + \varepsilon_{t+1} \tag{80}$$

with $|\rho| \leq 1$. In the frictionless version, the optimal action at date t is:

$$a_t^r = x_t.$$

Simple case: Random walk To keep the math simple, take $\rho = 1$ at first. Consider first the "sticky action" case. We will consider two benchmarks. In the "Calvo" model (as in the pricing model due to Calvo (1983)) the agent changes her action only with a Poisson probability $1-\theta$ at each period. In the "fixed delay D" model (as in Gabaix and Laibson 2002; Reis 2006a), the agent changes her action every D periods. Both models imply that the action is changed with a lag.

Call $a_{t,s}^{A}$ (respectively $a_{t,s}^{I}$) the action of an agent at time t, who re-optimized her action (respectively, who refreshed her information) s periods ago, in the sticky Action model (respectively, Information). Then

$$a_{t,s}^{A} = a_{t-s}^{r} = x_{t-s}$$

and

$$a_{t,s}^{I} = \mathbb{E}_{t-s}\left[a_t^r\right] = x_{t-s}.$$

Hence, in the random walk case, sticky action and sticky information make the same prediction. However, when we go beyond the random walk, predictions are different (see Section A.2).

So, consider the impact of a change in ε_t in x_t, on the aggregate action

$$\bar{a}_t = \sum_{s=0}^{\infty} f\left(s\right) a_{t,s}.$$

In the Calvo model, $f\left(s\right) = \left(1-\theta\right)\theta^s$. In the "fixed delay D" model, $f\left(s\right) = \frac{1}{D}1_{0\leq s<D}$. Look at $\overline{a}_t\left(\varepsilon_t, \varepsilon_{t-1}, ...\right)$, and $\mathbb{E}_t \frac{d\bar{a}_{t+T}}{d\varepsilon_t}$. Then:

$$\mathbb{E}_t\left[\frac{d\bar{a}_{t+T}}{d\varepsilon_t}\right] = \sum_{s=0}^{T} f\left(s\right) =: F\left(T\right)$$

with

$$F\left(T\right) = 1-\theta^{T+1} \tag{81}$$

in the Calvo model; and

$$F\left(T\right) = \min\left(\frac{T+1}{D}, 1\right)$$

in the updating-every-D periods model. Hence, we have a delayed reaction. This is the first lesson. Models with sticky action, and sticky reaction, generate a sluggish, delayed response in the aggregate action.

Put another way, in the Calvo model, aggregate dynamics are:

$$\bar{a}_t{}^{A} = \theta\bar{a}_{t-1}^{A} + \left(1-\theta\right)x_t \tag{82}$$

and they are the same (in the random walk case that we are presently considering) in the sticky information case.

7.1.2 Habit formation generates inertia

Macroeconomists who want to generate inertia often use habits. That is, instead of a utility function $u(a_t, x_t)$, one uses a utility function

$$v(a_{t,}, a_{t-1}, x_t) := u\left(\frac{a_t - h a_{t-1}}{1-h}, x_t\right) \tag{83}$$

where $h \in [0, 1)$ is a habit parameter. This is done in order to generate stickiness. To see how, consider again the targeting problem (78), but with no frictions except for habit:

$$\max_{a_t} - \sum_{t=0}^{\infty} \beta^t \left(\frac{a_t - h a_{t-1}}{1-h} - x_t\right)^2.$$

The first best can be achieved simply by setting the square term to 0 at each date, e.g. $\frac{a_t - h a_{t-1}}{1-h} - x_t = 0$. That is,

$$a_t = h a_{t-1} + (1-h) x_t \tag{84}$$

which is exactly an AR(1) process, like (82), replacing θ by h. This is a sense in which a habit model can generate the same behavior as a sticky action / information model. In more general setups, the correspondence is not perfect, but it qualitatively carries over.

Macroeconomists have used this habit model to generate inertia in consumption, and even in investment – see Christiano, Eichenbaum, and Evans (2005). Havranek, Rusnak, and Sokolova's (2017) meta-analysis finds a median estimate of $h = 0.5$ for macro studies, and $h = 0.1$ for micro studies. The discrepancy is probably due to the fact that at the micro level there is so much volatility of consumption, that this is only consistent with a small degree of habit formation. In macro studies, aggregate consumption is much smoother, so aggregate sluggishness of the reaction to information results in a higher measured h.

Of course, for normative purposes the analysis is completely different. In the habit model above, the agent achieves the first best utility. However, in the sticky information model, if the agent could remove her friction (e.g. lower the stickiness θ to 0), she would do it. In a more complex macro model, the same holds. Likewise for optimal retirement savings policy, the specific reason for people's sensitivity to the default matters a great deal (Bernheim, Fradkin, and Popov 2015).

Which is true? Most macroeconomists acknowledge that habits are basically just a device to generate stickiness. Carroll, Crawley, Slacalek, Tokuoka, and White (2017) argue that stickiness is indeed about inattention, rather than habits.

7.1.3 Adjustment costs generate inertia

Adjustment costs also generate inertia. Suppose that the problem is

$$\max_{a_t} -\sum_{t=0}^{\infty} \beta^t [(a_t - x_t)^2 + \kappa (a_t - a_{t-1})^2]$$

such that the first order condition with respect to a_t is

$$a_t - x_t + \kappa (a_t - a_{t-1}) - \beta\kappa (\mathbb{E}_t a_{t+1} - a_t) = 0 \tag{85}$$

so we obtain a second order difference equation. When x_t is a random walk, we have

$$a_t = \theta a_{t-1} + (1-\theta) x_t \tag{86}$$

where θ solves $\theta = \frac{\kappa}{\kappa+1+\beta\kappa(1-\theta)}$.[85] So, θ is 0 when $\kappa = 0$, and $\theta = 1$ as $\kappa \to \infty$.

Hence again, adjustment costs yield an isomorphic behavior, but with a more complex mathematical result, as θ has to be solved for.

7.1.4 Observable difference between inattention vs. habits / adjustment costs: Source-specific inattention

Both inattention and habits / adjustment costs create delayed reaction. Let us see this in a one-period model.

In an adjustment cost model, the agent solves $\max_a - (a - \sum_i b_i x_i)^2 - \kappa (a - a_{-1})^2$, which yields an action:

$$a = m a^r + (1-m) a_{-1} \tag{87}$$

with $m = \frac{1}{1+\kappa}$. Likewise, a habit

$$\max_a u\left(\frac{a - h a_{-1}}{1-h}, x\right) \tag{88}$$

creates the same expression (87) for the action, this time with $m = 1-h$.

[85] Proof: we use (86) at $t+1$, which gives

$$\mathbb{E} a_{t+1} - a_t = (\theta - 1) a_t + (1-\theta) \mathbb{E}_t x_{t+1} = (1-\theta)(x_t - a_t).$$

Then, we plug it into (85),

$$\begin{aligned} 0 &= a_t - x_t + \kappa (a_t - a_{t-1}) - \beta\kappa (\mathbb{E} a_{t+1} - a_t) \\ &= (1 + \beta\kappa (1-\theta)) (a_t - x_t) + \kappa (a_t - a_{t-1}) \end{aligned}$$

Solving for a_t, this implies (86) at time t, provided that $\theta = \frac{\kappa}{\kappa+1+\beta\kappa(1-\theta)}$.

In contrast, inattention creates an action:

$$a = \sum_i m_i b_i x_i.$$

Hence we can differentiate between them as follows. The presence of adjustment costs (or the sticky action model) creates *uniform* under-reaction ($m_i = m$ for all i's), while inattention (of the sticky information kind) creates *source-specific* under-reaction (the m_i in general differ across i's).

7.1.5 Dynamic default value

Within behavioral models, a simple way to model dynamic attention is via the default value. For instance, the default value could jump to the optimal default value, with some Poisson probability, much as in the sticky information model. In a Bayesian context, the "prior" could be updated with some Poisson probability.

7.2 Optimal dynamic inattention

How to optimize the allocation of attention? The agent minimizes the following objective function over the information acquisition policy, in which a denotes a state-contingent policy:

$$V(a, \beta) = -\mathbb{E}\left[\sum_{t \geq 0} (1-\beta) \beta^t \left(\frac{1}{2}(a_t - x_t)^2 + \kappa C_t\right)\right]$$

where $C_t = 1$ if a cost is paid, 0 otherwise. Here, to simplify calculations and concentrate on the economics, we take the "timeless perspective", and take the limit $\beta \to 1$. That is, the agent maximizes, over the adjustment policy, $V(a) = \lim_{\beta \to 1} V(a, \beta)$, that is

$$V(a) = -\mathbb{E}\left[\frac{1}{2}(a_t - x_t)^2 + \kappa C_t\right]$$

which is the average consumption loss plus a penalty for the average cost of looking up information.[86]

If the information is s periods old, then $a_{t,s} - x_t = \sum_{u=1}^{s} \varepsilon_{t-u}$, so

$$\mathbb{E}\left[(a_{t,s} - x_t)^2\right] = s\sigma^2$$

hence, the losses from misoptimization are:

$$\mathbb{E}\left[(a_t - x_t)^2\right] = \sigma^2 \mathbb{E}[T] = \sigma^2 \frac{D-1}{2} \text{ for the } D\text{-period model,}$$

[86]Indeed, as $\sum_{t \geq 0} (1-\beta) \beta^t X_t \to \mathbb{E}[X_t]$ as $\beta \to 1$ if X is an "ergodic" process.

$$\mathbb{E}\left[(a_t - x_t)^2\right] = \sigma^2 \frac{\theta}{1-\theta} \text{ for the Calvo model.}$$

Now, we calculate[87]

$$\begin{aligned}
\mathbb{E}[C_t] &= \frac{1}{D} \text{ for the } D\text{-period model,} \\
\mathbb{E}[C_t] &= 1-\theta \text{ for the Calvo model,}
\end{aligned}$$

so that the optimal reset time solves, in the D period model:

$$\min_D \frac{1}{2}\sigma^2 \frac{D-1}{2} + \kappa \frac{1}{D}$$

i.e. a frequency of price adjustments

$$\frac{1}{D} = \frac{\sigma}{2\sqrt{\kappa}} \tag{89}$$

as in Gabaix and Laibson 2002; Reis 2006a; Alvarez, Lippi, and Paciello 2011; Reis 2006b. Likewise, in the Calvo model, the optimal frequency θ is

$$\min_\theta \frac{1}{2}\sigma^2 \frac{\theta}{1-\theta} + \kappa(1-\theta)$$

i.e. the frequency of price adjustments is

$$1-\theta = \min\left(\frac{\sigma}{\sqrt{2\kappa}}, 1\right). \tag{90}$$

The same generalizes to the case where the signal has n components. Suppose that

$$\begin{aligned}
x_t &= \sum_i x_{it} \\
x_{it} &= x_{i,t-1} + \varepsilon_{it}
\end{aligned}$$

and reset costs are κ_i. Then the average per-period loss is

$$\sum_i \left[\frac{1}{2}\sigma_i^2 \frac{D_i - 1}{2} + \kappa_i \frac{1}{D_i}\right]$$

so that the frequency at which agents look up source i is

$$\frac{1}{D_i} = \frac{\sigma_i}{2\sqrt{\kappa_i}}. \tag{91}$$

[87] In the D model, the information is looked up every D periods. In the Calvo model, the probability of looking up the information next period is $1-\theta$.

I do know not of systematic evidence on this, although the research on this topic is progressing vigorously (e.g. Alvarez, Lippi, and Paciello 2011; Alvarez, Gonzalez-Rozada, Neumeyer, and Beraja 2016).

7.3 Other ways to generate dynamic adjustment

7.3.1 Procrastination

Another way to generate sluggishness is to use procrastination, as in Carroll, Choi, Laibson, Madrian, and Metrick (2009). In this view, agents hope to act, but procrastinate for a long time – because they are optimistic about their future behavior (O'Donoghue and Rabin (1999)). A related issues is forgetting and lapsed attention. For instance, Ericson (2017) finds that an important factor is that people overestimate the likelihood that they will remember that they have to make a decision, which amplifies sluggishness (see also Ericson 2011).

7.3.2 Unintentional inattention

Most models are about fairly "intentional" attention – agents choose to pay attention (though, given attention is dictated more so by System 1 than by System 2, in the language of Kahneman (2003), the distinction isn't completely clear cut). If unintentional inattention is the first-order issue, how do we model that? A simple way would be to say that the agent has the wrong "priors" over the importance of variable x_i. That is, in truth σ_i is high, but the agent thinks that σ_i is low – for instance, at the allocation of attention stage the agent thinks that an employer's retirement savings match rate is small. Concretely, at Step 1 in Proposition 4.1, the agent might have too low a perception of σ_i. One could imagine an iterated allocation problem, where the agent also optimizes over his perception of the costs and benefits.

7.3.3 Slow accumulation of information with entropy-based cost

Sims (1998) was motivated by evidence for sluggish adjustment. Maćkowiak and Wiederholt (2009, 2015) pursue that idea in macroeconomics – while breaking the unitary entropy of Sims, such that agents are allowed to have heterogeneous attention to different news sources. The dynamics are much more complex to derive, but are not unrealistic.

7.4 Behavioral macroeconomics

There has been a recent interest in behavioral macroeconomics. It is too early to present a comprehensive survey of this literature. Themes includes rules of thumb (Campbell and Mankiw 1989), limited information updating (Caballero 1995, Gabaix and Laibson 2002, Mankiw and Reis 2002, Reis 2006a, Alvarez et al. (2011)), and noisy signals (Sims 2003, Maćkowiak and Wiederholt 2015). A small but growing literature in theoretical macroeconomics draws consequences for general equilibrium and policy from features like inattention and imperfect information (Woodford 2013; García-Schmidt and Woodford 2015; Angeletos and Lian 2017, 2016; Farhi and Werning 2017; Bordalo, Gennaioli, and Shleifer 2016b). For instance, Gabaix (2018) presents a behavioral version of the textbook New Keynesian model, which gives a way to model monetary and fiscal policy with behavioral agents. There is also a budding literature on limited attention by firms beyond the issues of price stickiness (Goldfarb and Xiao (2016)). We can expect this literature to grow in the future.

8 Open Questions and Conclusion

The field of inattention has become extremely active. Here are some important open issues.

We need more measures of inattention, and go beyond rejecting the "full attention" null hypothesis Currently, to produce one good measure of attention m, we need a full paper. It would be nice to scale up production – in particular, to always attempt to provide a quantitative measure of attention, rather than a demonstration that it is not full.

Investigating Varian in the lab Consider the difference between a physics textbook and a microeconomics textbook. In a physics textbook, assertions and results (e.g. force = mass × acceleration) have been verified exquisitely in the lab. Not so in economics. When one opens a textbook such as Mas-Colell, Whinston, and Green (1995) or Varian (1992), one is confronted with a few chapters that have been extensively investigated: for example, expected utility (with prospect theory as a behavioral benchmark), or basic game theory, with some behavioral models as an alternative (Camerer 2003). But other chapters, such as basic microeconomics of the consumer-theory and Arrow-Debreu styles, have been investigated very little.[88] One sees many assertions and predictions, with very few experimental counterparts – and indeed, one suspects that the assertions will actually be wrong if they are to be tested.

This is a result – I believe – of the lack of a systematic behavioral alternative. The material in Section 5 fills this gap by proposing a behavioral counterpart of the major parts of basic microeconomics, including directions in which inattention will modify the rational predictions. It would be a great advance to implement procedures to investigate its predictions empirically. Such a study would be very useful for economics.

[88]There is a literature estimating GARP and Afriat's theorem, but it is generally not guided by a specific behavioral alternative, so that "rejection of rationality" usually gives little guidance to a behavioral alternative. See Aguiar and Serrano (2017) for progress on this, and the references therein to this strand of literature.

The challenges are: (i) to implement a notion of "clearly perceived" and "more opaque" prices, (ii) measure attention m, and (iii) implement in a roughly natural way the basic problem (53). The outcome of this would be very valuable, as we would then have a worked-out and tested counterpart of basic microeconomics.

We need more experimental evidence on the determinants of attention There are now several theories of attention, but measurement is somewhat lagging in refinement (the reason is that it is already hard to measure attention in the first place, so that the study of the determinants of attention is even harder). What's the cost of inattention? Could we get some sense of the shape of the cost, and of the attention function (e.g. that in Figure 1)? At a more basic level, the global-entropy constraint à la Sims predicts a unitary shadow value of attention, as in equation (73), without source-dependent inattention. Other models, e.g. behavioral models and older models where people pay for precision (Verrecchia 1982; Veldkamp 2011), predict source-dependent inattention, as in (37). Other theories emphasize the fact that attention is commodity- and action-dependent (Bordalo, Gennaioli, and Shleifer 2013). Empirical guidance would be useful.

More structural estimation The early papers found evidence for imperfect attention, with large economic effects. A more recent wave of papers has estimated inattention – its mean, variance, and how it varies with income, education and the like. A third generation of papers might estimate more structural models of inattention, to see if the predictions do fit, and perhaps suggest newer models.

Using this to do better policy: generating attention All this work may lead to progress in how to generate attention, e.g. for policy. Making consumers more rational is difficult even when the right incentives are in place – for example, consumers overwhelmingly fail to minimize fees in allocating their portfolios (Choi, Laibson, and Madrian 2009). The work on nudges (Thaler and Sunstein 2008) is based on psychological intuition rather than quantified principles. Also, knowing better "best practices" for disclosure would be helpful. Firms are good at screening for consumer biases (Ru and Schoar 2016), but public institutions less so, and debiasing is quite hard.

A Appendix: Further Derivations and Mathematical Complements

A.1 Further Derivations

Basic signal-extraction problem (Section 2.1) We have $s = x + \varepsilon$. So $\mathbb{E}[x|s] = ms$, with $m = \frac{\mathrm{Cov}(x,s)}{\mathrm{Var}(s)} = \frac{v_x}{v_s}$, with $v_x = \sigma_x^2$ and $v_s = \sigma_x^2 + \sigma_\varepsilon^2$. Hence, the optimal $a = \mathbb{E}[x|s]$ is $a = ms = mx + m\varepsilon$. A little bit of algebra gives $v_\varepsilon = v_s - v_x = v_x\left(\frac{1}{m} - 1\right)$ and

$$\mathrm{Var}(m\varepsilon) = m^2 v_\varepsilon = m(1-m)\,v_x$$

so a is distributed as:

$$a = mx + \sqrt{m(1-m)}\,\eta_x \tag{92}$$

where η_x is another draw from the distribution of x. This implies $\mathrm{Var}(a) = m\mathrm{Var}(x)$, and $\mathbb{E}\left[(a-x)^2\right] = (1-m)\,\sigma_x^2$.

Derivation of the losses from inattention (equation (31)) Let us start with a 1-dimensional action, with a utility function $u(a)$. Call a^* the optimum. But the agent does $a = a^* + \hat{a}$, where $\hat{a}$ is a deviation (perhaps coming from inattention). Then utility losses are

$$L(\hat{a}) := u(a^* + \hat{a}) - u(a^*).$$

Let's do a Taylor expansion,

$$L_a(\hat{a}) = u'(a^* + \hat{a}),\ L_{aa}(\hat{a}) = u''(a^* + \hat{a})$$
$$L(\hat{a}) = L(0) + L_a(0)\,\hat{a} + \frac{1}{2}L_{aa}(0)\,\hat{a}^2 + o\left(\hat{a}^2\right)$$

which implies $L(0) = L_a(0) = 0$. Hence:

$$L(\hat{a}) = \frac{1}{2}u_{aa}(0)\,\hat{a}^2 + o\left(\hat{a}^2\right).$$

Next, for a small x, the deviation is

$$\hat{a} = a^*(x^s) - a^*(x) = a_x(x^s - x) + o(x) = a_x(m-1)\,x + o(x)$$

hence, for a one-dimensional x, the loss is:

$$\begin{aligned} 2L(x) &= u_{aa}(a^*(x))\,\hat{a}^2 + o\left(\hat{a}^2\right) = u_{aa}(a^*(0))\,\hat{a}^2 + o\left(\hat{a}^2\right) \\ &= u_{aa}a_x^2x^2(1-m)^2 + o\left(|x|^2\right). \end{aligned}$$

With an $n-$dimensional x, the math is similar, with matrices:

$$\hat{a} = a^* (x^s) - a^* (x) = a_x (x^s - x) = a_x (M - I) x + o (x)$$

with $M = diag (m_1, ..., m_n)$, I the identity matrix of dimension n. So, neglecting $o\left(\|\hat{a}\|^2\right)$ terms,

$$\begin{aligned}
2L &= \hat{a}' u_{aa} (0) \hat{a} + o\left(\|\hat{a}\|^2\right) = x' (I - M)' a_x' u_{aa} (0) a_x (I - M) x \\
&= \sum_{i,j} (1 - m_i) x_i a_{x_i}' u_{aa} (0) a_{x_j} x_j (1 - m_j) \\
&= -\sum_{i,j} (1 - m_i) \tilde{\Lambda}_{ij} (1 - m_j) = -(\iota - m) \tilde{\Lambda} (\iota - m)' \\
\tilde{\Lambda}_{ij} &= -x_i a_{x_i}' u_{aa} (0) a_{x_j} x_j, \ \iota := (1, ..., 1) .
\end{aligned}$$

We then obtain (31) by taking expectations.

Derivation of the entropy of Gaussian variables (Section 6.2.1) The entropy doesn't depend on the mean, so we normalized it to 0.

One dimension. The density is $f(x) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$, so

$$\begin{aligned}
H(X) &= -\mathbb{E}[\log f(X)] = -\mathbb{E}\left[-\frac{x^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right] \\
&= \frac{1}{2} + \frac{1}{2}\log\left(2\pi\sigma^2\right) = \frac{1}{2}\log\sigma^2 + \frac{1}{2}\log(2\pi e) .
\end{aligned}$$

Higher dimensions. The density is $f(x) = \frac{e^{-\frac{1}{2}x'V^{-1}x}}{(2\pi)^{n/2}(\det V)^{1/2}}$, where $V = \mathbb{E}[XX']$ is the variance covariance matrix. Using the notation $|V| = \det V$, and Tr for the trace, we first note

$$\begin{aligned}
\mathbb{E}\left[x'V^{-1}x\right] &= \mathbb{E}\left[\mathrm{Tr}\left(x'V^{-1}x\right)\right] = \mathbb{E}\left[\mathrm{Tr}\left(xx'V^{-1}\right)\right] \\
&= \mathrm{Tr}\,\mathbb{E}\left[xx'V^{-1}\right] = \mathrm{Tr}\,\mathbb{E}\left[VV^{-1}\right] = \mathrm{Tr}\, I_n = n.
\end{aligned}$$

Then, the entropy is

$$\begin{aligned}
H(X) &= -\mathbb{E}[\log f(X)] = -\mathbb{E}\left[-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|V| - \frac{1}{2}x'V^{-1}x\right] \\
&= \frac{1}{2}\log\left((2\pi)^n |V|\right) + \frac{n}{2} = \frac{1}{2}\log\left((2\pi e)^n |V|\right) .
\end{aligned}$$

Mutual information of two Gaussian variables (Section 6.2.1) Suppose X, Y are jointly Gaussian, with variance-covariance matrix $V = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$, where $\rho = \operatorname{corr}(X, Y)$. Then, $\det V = \sigma_X^2\sigma_Y^2(1-\rho^2)$, so

$$H(X, Y) = \frac{1}{2}\log(\det V) + \log(2\pi e)$$

and using (67) gives

$$I(X, Y) = H(X) + H(Y) - H(X, Y) = -\frac{1}{2}\log\left(1-\rho^2\right).$$

Proof of Proposition 5.1 From Definition 4.3, the optimum satisfies: $u'(c) = \lambda p^s$ for some λ. Hence, this consumption is the consumption of a rational agent facing prices p^s, and wealth $w' = p^s \cdot c$.

Proof of Proposition 5.2 Here I show only the proof in the most transparent case – see the original paper for the general case. Utility is $u(c) = U(C) + c_n$, where $C = (c_1, ..., c_{n-1})$, and the price of good n is 1 and correctly perceived. Then, demand satisfies $u'(c) = \lambda p^s$. Applying this to the last good gives $1 = \lambda$. So, demand for the other goods satisfies $U'(C) = P^s$, where $P = (p_1, \ldots, p_n)$. Differentiating w.r.t. P, $U''(C)C_P^s = M$, where $M = diag(m_1, \ldots, m_{n-1})$ is the vector of attention to prices. Now, the Slutsky matrix (for the goods $1, \ldots, n-1$) is $S^s = C_P^s = U''^{-1}(C)M$, as all the income effects are absorbed by the last good ($\frac{\partial c_i}{\partial w} = 0$ for $i < n$). As a particular case where $M = I$, the rational Slutsky matrix is $S^r = U''^{-1}(C)$. So, we have $S^s = S^r M$.

Proof of Proposition 5.4 The part $\frac{\partial \boldsymbol{c}^s}{\partial w} = \frac{\partial \boldsymbol{c}^r}{\partial w}$ follows from Proposition 5.1: at the default prices $\boldsymbol{p} = \boldsymbol{p}^s$, so $\boldsymbol{c}^s\left(\boldsymbol{p}^d, w\right) = \boldsymbol{c}^r\left(\boldsymbol{p}^d, w\right)$, which implies $\frac{\partial \boldsymbol{c}^s}{\partial w} = \frac{\partial \boldsymbol{c}^r}{\partial w}$. Then, the definition of the Slutsky matrix and Proposition 5.2 imply (58).

Proof of Proposition 5.8 In an endowment economy, equilibrium consumption is equal to the endowment, $\boldsymbol{c}(t) = \boldsymbol{\omega}(t)$. We have $\frac{u_i(\boldsymbol{c}(t))}{u_1(\boldsymbol{c}(t))} = \frac{p_i^s(t)}{p_1^s(t)}$ for $t = 0, 1$: the ratio of marginal utilities is equal to the ratio of perceived prices – both in the rational economy (where perceived prices are true prices) and in the behavioral economy (where they're not). Using $p_1^s(t) = p_1^r(t) = p_1(0)$, that implies that the perceived price needs to be the same in the behavioral and rational economy: $\left(p_i^{[s]}(t)\right)^{\text{perceived}} = p_i^{[r]}(t)$. Thus, we have $m_i dp_i^{[s]} = d\left[\left(p_i^{[s]}\right)^{\text{perceived}}\right] = dp_i^{[r]}$, i.e. $dp_i^{[s]} = \frac{1}{m_i}dp_i^{[r]}$.

A.2 Mathematical Complements

Here I provide some mathematical complements.

Dynamic attention: Beyond the random walk case Here I expand on Section 7.1, beyond the random cases which made the analytics very transparent. I consider the case (80) with ρ not necessarily equal to 1. The sticky action is a bit more delicate to compute. Consider an agent who can change her action at time t. At period $t+s$, she will still have to perform action $a_{t,s}^A = a_{t,0}^A$ with probability θ^s (we use the Calvo formulation here). Hence, the optimal action at t satisfies

$$\max_a -\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \theta^s \left(a - x_{t+s}\right)^2 .$$

The first order condition is

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \theta^s \left(a - x_{t+s}\right) = 0$$

i.e. $\frac{1}{1-\beta\theta} a - \sum_{s=0}^{\infty} \beta^s \theta^s \mathbb{E}_t \left[x_{t+s}\right] = 0$, i.e. $a = a_{t,0}^A$ with

$$a_{t,0}^A = \left(1-\beta\theta\right) \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \theta^s \mathbb{E}_t \left[x_{t+s}\right]. \tag{93}$$

In the AR(1) case, $\mathbb{E}_t \left[x_{t+s}\right] = \rho^s x_t$, and

$$a_{t,0}^A = \frac{1-\beta\theta}{1-\beta\theta\rho} x_t. \tag{94}$$

In the sticky information model, the problem is, for each period t,

$$\max_{a_{t,s}^I} -\mathbb{E}_{t-s} \left(a_{t,s}^I - x_t\right)^2$$

which yields

$$a_{t,s}^I = \mathbb{E}_{t-s} \left[x_t\right]. \tag{95}$$

Hence, we see that the two models are generally different – even though they generate the same predictions in the random walk case.

B Appendix: Data Methodology

This appendix outlines the details of the methodology used to compile the data in Table 1 and Figure 1, which present point estimates of the attention parameter m in a cross-section of recent studies, alongside the estimated relative value of the opaque add-on attribute with respect to the relevant good or quantity (τ/p).

- In the study of Allcott and Wozny (2014), we take τ to be the standard deviation of the present discounted value of future gasoline costs in the authors' sample; p is correspondingly the standard deviation of vehicle price, such that $\tau = \$4,147$ and $p = \$9,845$. The point estimate for m is as reported by the authors.

- Hossain and Morgan (2006) and Brown, Hossain, and Morgan (2010) both conduct a series of paired experiments by selling various goods on eBay and varying the shrouded shipping costs. This setup allows us to deduce the implied degree of inattention, following the same methodology as in DellaVigna (2009). We consider auction pairs in which the auction setup and the sum of reserve price are held constant, while the shipping cost is altered. As in DellaVigna (2009), we assume buyers are bidding their true willingness to pay in eBay's second price auctions, such that their bid is $b = p+m\tau$, where p is the buyer's valuation of the object and τ is the shipping cost. Seller's revenue is $p+(1-m)c$. Under this model, the ratio of the difference in revenues to the difference in shipping costs across the two auction conditions corresponds to the quantity $1 - m$.

  The estimates for the attention parameter m in the experiments of Hossain and Morgan (2006) are as reported in DellaVigna (2009). We use the same methodology to derive the analogous estimate for the eBay Taiwan field experiment of Brown, Hossain, and Morgan (2010). The raw implied estimate for the latter experimental setting is negative ($m = -0.43$), as the mean revenue difference between the two auction conditions is greater than the difference in shipping costs. For consistency with the definition of m and in order to account for measurement error, we constrain the final implied estimate of m to the interval $[0, 1]$.

  Given that each estimate of m is inferred from a set of two paired auctions, the value p of the good under auction is defined as average revenue minus shipping costs across the two auction conditions. The value τ of the opaque attribute is analogously defined as the average shipping cost across the two auction conditions.

- For the study of DellaVigna and Pollet (2009) we take τ/p to be the ratio of the standard deviation of abnormal returns at earnings announcement to abnormal returns for the quarter, pooled across all weekdays and computed following the methodology in DellaVigna and Pollet (2009). The quarterly cadence is chosen to match the frequency of earnings announcements in the authors' sample. The return at earnings announcement is for two trading days from the close of the market on the trading day before the earnings announcement to the close of the trading day after the earnings announcement. The standard deviation of the abnormal returns at earnings announcement is 0.0794. The standard deviation of the abnormal returns for the quarter, starting from the close of the market on the trading day before the earnings announcement and continuing to the close of the market on trading day 60 after the announcement, is 0.2651. The estimates for the attention parameter m are as in DellaVigna (2009).

- In the case of Lacetera, Pope, and Sydnor (2012), τ is taken to be the average mileage remainder in the sample, which is approximately $5,000$, per correspondence with the authors. The quantity p is obtained by subtracting $\tau = 5,000$ from the mileage of the median car in the sample, which is $56,997$. Hence $p = 51,997$. The estimate for m is as reported by the authors in the full-sample specification that includes all car transactions, pooled across fleet/lease and dealer categories.

- For the field experiment of Chetty, Looney, and Kroft (2009), we take τ/p to be the relevant sales tax rate of 7.38%. Correspondingly, for the natural experiment of Chetty, Looney, and Kroft (2009) we take τ/p to be 4.30%, which is the mean sales tax rate for alcoholic products across U.S. states as reported by the authors. The estimates for the attention parameter m are as reported by the authors.

- For the study of Taubinsky and Rees-Jones (2017), we analogously let τ/p be the sales tax rate applied in the laboratory experiment, which is 7.31% in the standard-tax treatment arm, and triple that value in the triple-tax treatment arm. The estimate for the attention parameter m are as reported by the authors for the two treatment arms.

- Figure 1 additionally shows data points from Busse, Lacetera, Pope, Silva-Risso, and Sydnor (2013b), who measure inattention to left-digit remainders in the mileage of used cars in auctions along several covariate dimensions. Each data point corresponds to a subsample of cars with mileages within a 10,000 mile-wide bin (e.g., between 15,000 and 25,000 miles, between 25,000 and 35,000 miles, and so forth). Data is available for two data sets, one including retail auctions and one including wholesale auctions. For each mileage bin, we include data points from both of these data sets. The estimates of m are as reported by the authors. The metric τ/p is the average ratio of mileage remainder to true mileage net of mileage remainder in the subsamples. As this ratio is most readily available for the data set of wholesale car auctions, we compute the τ/p estimates on subsamples of the wholesale data set only, under the assumption that the mileage distribution is not systematically different across the two data sets. We do not expect substantive impact on our results from this assumption.

References

Abaluck, Jason, Adams, Abi, 2017. What do consumers consider before they choose? Identification from asymmetric demand responses. NBER Working Paper No. 23566.

Abaluck, Jason, Gruber, Jonathan, 2011. Heterogeneity in choice inconsistencies among the elderly: Evidence from prescription drug plan choice. The American Economic Review Papers and Proceedings 101 (3), 377–381.

Abel, Andrew B., Eberly, Janice C., Panageas, Stavros, 2013. Optimal inattention to the stock market with information costs and transactions costs. Econometrica 81 (4), 1455–1481.

Abela, Andrew R, Dougherty, Stephen D, Fagen, Erin D, Hill, Carolyn JR, Chudasama, Y, 2012. Inhibitory control deficits in rats with ventral hippocampal lesions. Cerebral Cortex 23 (6), 1396–1409.

Abeler, Johannes, Jäger, Simon, 2015. Complex tax incentives. American Economic Journal: Economic Policy 7 (3), 1–28.

Aguiar, Victor H., Riabov, Nickolai, 2016. Estimating high dimensional demand under bounded rationality: The ESMAX demand system. Working Paper.

Aguiar, Victor H., Serrano, Roberto, 2017. Slutsky matrix norms: The size, classification, and comparative statics of bounded rationality. Journal of Economic Theory 172, 163 – 201.

Allcott, Hunt, Lockwood, Benjamin B., Taubinsky, Dmitry, 2017. A theory of regressive sin taxes, with an application to the optimal soda tax. Working Paper.

Allcott, Hunt, Taubinsky, Dmitry, 2015. Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. The American Economic Review 105 (8), 2501–2538.

Allcott, Hunt, Wozny, Nathan, 2014. Gasoline prices, fuel economy, and the energy paradox. Review of Economics and Statistics 96 (5), 779–795.

Alvarez, Fernando, Gonzalez-Rozada, Martin, Neumeyer, Andy, Beraja, Martin, 2016. From hyperinflation to stable prices: Argentina's evidence on menu cost models. Forthcoming at the Quarterly Journal of Economics.

Alvarez, Fernando, Guiso, Luigi, Lippi, Francesco, 2012. Durable consumption and asset management with transaction and observation costs. The American Economic Review 102 (5), 2272–2300.

Alvarez, Fernando, Lippi, Francesco, Paciello, Luigi, 2017. Monetary shocks in models with observation and menu costs. Forthcoming at the Journal of the European Economic Association.

Alvarez, Fernando E., Lippi, Francesco, Paciello, Luigi, 2011. Optimal price setting with observation and menu costs. Quarterly Journal of Economics 126 (4), 1909–1960.

Ambuehl, Sandro, 2017. An offer you can't refuse. Working Paper.

Anderson, Simon P., De Palma, Andre, Thisse, Jacques François, 1992. Discrete choice theory of product differentiation. MIT Press.

Andries, Marianne, Haddad, Valentin, 2017. Information aversion. NBER Working Paper No. 23958.

Angeletos, George-Marios, Lian, Chen, 2016. Forward guidance without common knowledge. NBER Working Paper No. 23379.

Angeletos, George-Marios, Lian, Chen, 2017. Dampening general equilibrium: From micro to macro. NBER Working Paper No. 22785.

Arieli, Amos, Ben-Ami, Yaniv, Rubinstein, Ariel, 2011. Tracking decision makers under uncertainty. American Economic Journal: Microeconomics 3 (4), 68–76.

Avoyan, Ala, Schotter, Andrew, 2018. Attention in games: An experimental study. Working Paper.

Bacchetta, Philippe, Van Wincoop, Eric, 2010. Infrequent portfolio decisions: A solution to the forward discount puzzle. American Economic Review 100 (3), 870–904.

Baker, Malcolm, Pan, Xin, Wurgler, Jeffrey, 2012. The effect of reference point prices on mergers and acquisitions. Journal of Financial Economics 106 (1), 49–71.

Barber, Brad M, Odean, Terrance, 2007. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. The Review of Financial Studies 21 (2), 785–818.

Bartoš, Vojtěch, Bauer, Michal, Chytilová, Julie, Matějka, Filip, 2016. Attention discrimination: Theory and field experiments with monitoring information acquisition. The American Economic Review 106 (6), 1437–1475.

Bénabou, Roland, Tirole, Jean, 2002. Self-confidence and personal motivation. Quarterly Journal of Economics 117 (3), 871–915.

Bernheim, B. Douglas, Fradkin, Andrey, Popov, Igor, 2015. The welfare economics of default options in 401(k) plans. The American Economic Review 105 (9), 2798–2837.

Bernheim, B. Douglas, Rangel, Antonio, 2009. Beyond revealed preference: Choice theoretic foundations for behavioral welfare economics. Quarterly Journal of Economics 124 (1), 51–104.

Bordalo, Pedro, Gennaioli, Nicola, Shleifer, Andrei, 2012. Salience theory of choice under risk. Quarterly Journal of Economics 127 (3), 1243–1285.

Bordalo, Pedro, Gennaioli, Nicola, Shleifer, Andrei, 2013. Salience and consumer choice. Journal of Political Economy 121 (5), 803–843.

Bordalo, Pedro, Gennaioli, Nicola, Shleifer, Andrei, 2016a. Competition for attention. The Review of Economic Studies 83 (2), 481–513.

Bordalo, Pedro, Gennaioli, Nicola, Shleifer, Andrei, 2016b. Diagnostic expectations and credit cycles. NBER Working Paper No. 22266.

Bouchaud, Jean-Philippe, Krueger, Philipp, Landier, Augustin, Thesmar, David, 2016. Sticky expectations and the profitability anomaly. HEC Paris Research Paper No. FIN-2016-1136.

Brocas, Isabelle, Carrillo, Juan D, Wang, Stephanie W, Camerer, Colin F, 2014. Imperfect choice or imperfect attention? understanding strategic thinking in private information games. Review of Economic Studies 81 (3), 944–970.

Bronnenberg, Bart J., Dubé, Jean-Pierre, Gentzkow, Matthew, Shapiro, Jesse M., 2015. Do pharmacists buy Bayer? Informed shoppers and the brand premium. Quarterly Journal of Economics 130 (4), 1669–1726.

Brown, Alexander L, Camerer, Colin F, Lovallo, Dan, 2012. To review or not to review? limited strategic thinking at the movie box office. American Economic Journal: Microeconomics 4 (2), 1–26.

Brown, Alexander L, Camerer, Colin F, Lovallo, Dan, 2013. Estimating structural models of equilibrium and cognitive hierarchy thinking in the field: The case of withheld movie critic reviews. Management Science 59 (3), 733–747.

Brown, Jennifer, Hossain, Tanjim, Morgan, John, 2010. Shrouded attributes and information suppression: Evidence from the field. Quarterly Journal of Economics 125 (2), 859–876.

Browning, Martin, Chiappori, Pierre-Andre, 1998. Efficient intra-household allocations: A general characterization and empirical tests. Econometrica 66 (6), 1241–1278.

Bulow, Jeremy I., Geanakoplos, John D., Klemperer, Paul D., 1985. Multimarket oligopoly: Strategic substitutes and complements. Journal of Political Economy 93 (3), 488–511.

Bushong, Benjamin, Rabin, Matthew, Schwartzstein, Joshua, 2016. A model of relative thinking. Working Paper.

Busse, Meghan R., Knittel, Christopher R., Zettelmeyer, Florian, 2013a. Are consumers myopic? Evidence from new and used car purchases. The American Economic Review 103 (1), 220–256.

Busse, Meghan R., Lacetera, Nicola, Pope, Devin G., Silva-Risso, Jorge, Sydnor, Justin R., 2013b. Estimating the effect of salience in wholesale and retail car markets. The American Economic Review Papers and Proceedings 103 (3), 575–579.

Caballero, Ricardo J., February 1995. Near-rationality, heterogeneity, and aggregate consumption. Journal of Money, Credit and Banking 27 (1), 29–48.

Calvo, Guillermo A., 1983. Staggered prices in a utility-maximizing framework. Journal of Monetary Economics 12 (3), 383–398.

Camerer, Colin, 2003. Behavioral game theory: Experiments in strategic interaction. Princeton University Press.

Camerer, Colin F, Ho, Teck-Hua, Chong, Juin-Kuan, 2004. A cognitive hierarchy model of games. The Quarterly Journal of Economics 119 (3), 861–898.

Camerer, Colin F, Johnson, Eric, Rymon, Talia, Sen, Sankar, 1993. Cognition and framing in sequential bargaining for gains and losses. Frontiers of game theory 104, 27–47.

Campbell, John Y., Mankiw, N. Gregory, 1989. Consumption, income, and interest rates: Reinterpreting the time series evidence. NBER Macroeconomics Annual 4, 185–216.

Candes, Emmanuel J., Tao, Terence, 2006. Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory 52 (12), 5406–5425.

Caplin, Andrew, 2016. Measuring and modeling attention. Annual Review of Economics 8, 379–403.

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

Caplin, Andrew, Dean, Mark, Leahy, John, 2016. Rational inattention, optimal consideration sets and stochastic choice. Working Paper.

Caplin, Andrew, Dean, Mark, Leahy, John, 2017. Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. NBER Working Paper No. 23652.

Caplin, Andrew, Dean, Mark, Martin, Daniel, 2011. Search and satisficing. The American Economic Review 101 (7), 2899–2922.

Carlin, Bruce I., 2009. Strategic price complexity in retail financial markets. Journal of Financial Economics 91 (3), 278–287.

Carrasco, Marisa, 2011. Visual attention: The past 25 years. Vision research 51 (13), 1484–1525.

Carroll, Christopher D., 2003. Macroeconomic expectations of households and professional forecasters. Quarterly Journal of Economics 118 (1), 269–298.

Carroll, Christopher D., Crawley, Edmund, Slacalek, Jiri, Tokuoka, Kiichi, White, Matthew N., 2017. Sticky expectations and consumption dynamics. Working Paper.

Carroll, Gabriel D., Choi, James J., Laibson, David, Madrian, Brigitte C., Metrick, Andrew, 2009. Optimal defaults and active decisions. Quarterly Journal of Economics 124 (4), 1639–1674.

Carroll, John S, Bazerman, Max H, Maury, Robin, 1988. Negotiator cognitions: A descriptive approach to negotiators' understanding of their opponents. Organizational Behavior and Human Decision Processes 41 (3), 352–370.

Chang, Tom Y, Huang, Wei, Wang, Yongxiang, 2018. Something in the air: Pollution and the demand for health insurance. Forthcoming at the Review of Economic Studies.

Chetty, Raj, Looney, Adam, Kroft, Kory, 2007. Salience and taxation: Theory and evidence. NBER Working Paper No. 13330.

Chetty, Raj, Looney, Adam, Kroft, Kory, 2009. Salience and taxation: Theory and evidence. The American Economic Review 99 (4), 1145–1177.

Choi, James J., Laibson, David, Madrian, Brigitte C., 2009. Why does the law of one price fail? An experiment on index mutual funds. The Review of Financial Studies 23 (4), 1405–1432.

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

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

Coibion, Olivier, Gorodnichenko, Yuriy, 2015. Information rigidity and the expectations formation process: A simple framework and new facts. The American Economic Review 105 (8), 2644–2678.

Compte, Olivier, Postlewaite, Andrew, 2017. Ignorance and Uncertainty. Unpublished Manuscript.

Costa-Gomes, Miguel, Crawford, Vincent P, Broseta, Bruno, 2001. Cognition and behavior in normal-form games: An experimental study. Econometrica 69 (5), 1193–1235.

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

Daniel, Kent, Hirshleifer, David, Subrahmanyam, Avanidhar, 1998. Investor psychology and security market under- and overreactions. The Journal of Finance 53 (6), 1839–1885.

De Bartolomé, Charles A. M., 1995. Which tax rate do people use: Average or marginal? Journal of Public Economics 56 (1), 79–96.

De Clippel, Geoffroy, Eliaz, Kfir, Rozen, Kareen, 2014. Competing for consumer inattention. Journal of Political Economy 122 (6), 1203–1234.

Debreu, Gerard, 1970. Economies with a finite set of equilibria. Econometrica 38 (3), 387–392.

Dehaene, Stanislas, 2011. The Number Sense: How the Mind Creates Mathematics. Oxford University Press.

Dehaene, Stanislas, Lau, Hakwan, Kouider, Sid, 2017. What is consciousness, and could machines have it? Science 358 (6362), 486–492.

DellaVigna, Stefano, 2009. Psychology and economics: Evidence from the field. Journal of Economic Literature 47 (2), 315–372.

DellaVigna, Stefano, Pollet, Joshua M., 2007. Demographics and industry returns. The American Economic Review 97 (5), 1667–1702.

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

DellaVigna, Stefano, Pope, Devin, 2018. What motivates effort? evidence and expert forecasts. The Review of Economic Studies 85 (2), 1029–1069.

Dertwinkel-Kalt, Markus, Köhler, Katrin, Lange, Mirjam RJ, Wenzel, Tobias, 2017. Demand shifts due to salience effects: Experimental evidence. Journal of the European Economic Association 15 (3), 626–653.

Duffie, Darrell, Sun, Tong-sheng, 1990. Transactions costs and portfolio choice in a discrete-continuous-time setting. Journal of Economic Dynamics and Control 14 (1), 35–51.

Ellis, Andrew, 2018. Foundations for optimal inattention. Journal of Economic Theory 173, 56–94.

Ellison, Glenn, 2005. A model of add-on pricing. Quarterly Journal of Economics 120 (2), 585–637.

Ellison, Glenn, Ellison, Sara Fisher, 2009. Search, obfuscation, and price elasticities on the internet. Econometrica 77 (2), 427–452.

Englmaier, Florian, Schmöller, Arno, Stowasser, Till, 2017. Price discontinuities in an online market for used cars. Management Science 64 (6), 2754–2766.

Enke, Benjamin, Zimmermann, Florian, forthcoming. Correlation neglect in belief formation. Review of Economic Studies.

Ericson, Keith M, 2014. Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange. American Economic Journal: Economic Policy 6 (1), 38–64.

Ericson, Keith M. Marzilli, 2011. Forgetting we forget: Overconfidence and memory. Journal of the European Economic Association 9 (1), 43–60.

Ericson, Keith M. Marzilli, 2017. On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. Journal of the European Economic Association 15 (3), 692–719.

Eyster, Erik, Madarasz, Kristof, Michaillat, Pascal, 2017. Pricing when Customers Care about Fairness but Misinfer Markups. NBER Working Paper No. 23778.

Eyster, Erik, Rabin, Matthew, 2005. Cursed equilibrium. Econometrica 73 (5), 1623–1672.

Farhi, Emmanuel, Gabaix, Xavier, 2017. Optimal Taxation with Behavioral Agents. NBER Working Paper No. 21524.

Farhi, Emmanuel, Werning, Iván, 2017. Monetary Policy, Bounded Rationality, and Incomplete Markets. NBER Working Paper No. 23281.

Fedyk, Anastassia, 2018. Front page news: The effect of news positioning on financial markets. Working Paper.

Finkelstein, Amy, 2009. E-ztax: Tax Salience and Tax Rates. Quarterly Journal of Economics 124 (3), 969–1010.

Frederick, Shane, Loewenstein, George, O'Donoghue, Ted, 2002. Time discounting and time preference: A critical review. Journal of economic literature 40 (2), 351–401.

Friedman, Milton, 1961. The lag in effect of monetary policy. Journal of Political Economy 69 (5), 447–466.

Frydman, Cary, Rangel, Antonio, 2014. Debiasing the disposition effect by reducing the saliency of information about a stock's purchase price. Journal of economic behavior & organization 107, 541–552.

Fudenberg, Drew, Levine, David K., 2012. Timing and self-control. Econometrica 80 (1), 1–42.

Fudenberg, Drew, Strack, Philipp, Strzalecki, Tomasz, 2017. Speed, accuracy, and the optimal timing of choices. Working Paper.

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

Gabaix, Xavier, 2016. Behavioral macroeconomics via sparse dynamic programming. NBER Working Paper No. 21848.

Gabaix, Xavier, 2018. A behavioral new keynesian model. NBER Working Paper No. 22954.

Gabaix, Xavier, Laibson, David, 2002. The 6D bias and the equity-premium puzzle. NBER Macroeconomics Annual 16, 257–312.

Gabaix, Xavier, Laibson, David, 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. Quarterly Journal of Economics 121 (2), 505–540.

Gabaix, Xavier, Laibson, David, 2017. Myopia and discounting. NBER Working Paper No. 23254.

Gabaix, Xavier, Laibson, David, Li, Deyuan, Li, Hongyi, Resnick, Sidney, Vries, Casper Gde, 2016. The impact of competition on prices with numerous firms. Journal of Economic Theory 165, 1–24.

Gabaix, Xavier, Laibson, David, Moloche, Guillermo, Weinberg, Stephen, September 2006. Costly information acquisition: Experimental analysis of a boundedly rational model. The American Economic Review 96 (4), 1043–1068.

García-Schmidt, Mariana, Woodford, Michael, 2015. Are low interest rates deflationary? A paradox of perfect-foresight analysis. NBER Working Paper 21614.

Geanakoplos, John, Milgrom, Paul, 1991. A theory of hierarchies based on limited managerial attention. Journal of the Japanese and International Economies 5 (3), 205–225.

Gennaioli, Nicola, Shleifer, Andrei, 2010. What comes to mind. Quarterly Journal of Economics 125 (4), 1399–1433.

Gershman, Samuel J., Horvitz, Eric J., Tenenbaum, Joshua B., 2015. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. Science 349 (6245), 273–278.

Giglio, Stefano, Shue, Kelly, 2014. No news is news: do markets underreact to nothing? The Review of Financial Studies 27 (12), 3389–3440.

Glimcher, Paul W., 2011. Foundations of Neuroeconomic Analysis. Oxford University Press.

Glimcher, Paul W., Fehr, Ernst, 2013. Neuroeconomics: Decision Making and the Brain. Academic Press.

Goldfarb, Avi, Xiao, Mo, 2016. Transitory shocks, limited attention, and a firm's decision to exit. Working paper.

Greenwood, Robin, Hanson, Samuel G., 2014. Waves in ship prices and investment. Quarterly Journal of Economics 130 (1), 55–109.

Greenwood, Robin, Shleifer, Andrei, 2014. Expectations of returns and expected returns. The Review of Financial Studies 27 (3), 714–746.

Grether, David M, 1980. Bayes rule as a descriptive model: The representativeness heuristic. The Quarterly Journal of Economics 95 (3), 537–557.

Griffiths, Thomas L, Lieder, Falk, Goodman, Noah D, 2015. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. Topics in cognitive science 7 (2), 217–229.

Grubb, Michael D., 2009. Selling to overconfident consumers. The American Economic Review 99 (5), 1770–1807.

Grubb, Michael D., Osborne, Matthew, 2015. Cellular service demand: Biased beliefs, learning, and bill shock. The American Economic Review 105 (1), 234–71.

Gruber, Jonathan, Kőszegi, Botond, 2001. Is addiction "rational"? Theory and evidence. Quarterly Journal of Economics 116 (4), 1261–1303.

Gul, Faruk, Pesendorfer, Wolfgang, Strzalecki, Tomasz, 2017. Coarse competitive equilibrium and extreme prices. The American Economic Review 107 (1), 109–137.

Handel, Benjamin R., 2013. Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. The American Economic Review 103 (7), 2643–2682.

Handel, Benjamin R., Kolstad, Jonathan T., 2015. Health insurance for "humans": Information frictions, plan choice, and consumer welfare. The American Economic Review 105 (8), 2449–2500.

Hanna, Rema, Mullainathan, Sendhil, Schwartzstein, Joshua, 2014. Learning through noticing: Theory and evidence from a field experiment. Quarterly Journal of Economics 129 (3), 1311–1353.

Hausman, Jerry A, 1979. Individual discount rates and the purchase and utilization of energy-using durables. The Bell Journal of Economics, 33–54.

Havranek, Tomas, Rusnak, Marek, Sokolova, Anna, 2017. Habit formation in consumption: A meta-analysis. European Economic Review 95, 142–167.

Heidhues, Paul, Kőszegi, Boton, 2010. Exploiting naivete about self-control in the credit market. The American Economic Review 100 (5), 2279–2303.

Heidhues, Paul, Kőszegi, Botond, 2017. Naivete-based discrimination. Quarterly Journal of Economics 132 (2), 1019–1054.

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

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

Hirshleifer, David, Lim, Sonya S, Teoh, Siew Hong, 2011. Limited investor attention and stock market misreactions to accounting information. The Review of Asset Pricing Studies 1 (1), 35–73.

Hirshleifer, David, Teoh, Siew Hong, 2003. Limited attention, information disclosure, and financial reporting. Journal of accounting and economics 36 (1), 337–386.

Ho, Teck-Hua, Camerer, Colin, Weigelt, Keith, 1998. Iterated dominance and iterated best response in experimental" p-beauty contests". The American Economic Review 88 (4), 947–969.

Hossain, Tanjim, Morgan, John, 2006. ... Plus shipping and handling: Revenue (non) equivalence in field experiments on eBay. Advances in Economic Analysis & Policy 5 (2).

Huang, Liqiang, Pashler, Harold, 2007. A boolean map theory of visual attention. Psychological Review 114 (3), 599.

Huberman, Gur, Regev, Tomer, 2001. Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. The Journal of Finance 56 (1), 387–396.

Itti, Laurent, Koch, Christof, 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision research 40 (10-12), 1489–1506.

Jehiel, Philippe, 2005. Analogy-based expectation equilibrium. Journal of Economic Theory 123 (2), 81–104.

Jin, Ginger Zhe, Luca, Michael, Martin, Daniel, March 2017. Is no news (perceived as) bad news? An experimental investigation of information disclosure. NBER Working Paper No. 21099.

Johnson, Eric J., Camerer, Colin, Sen, Sankar, Rymon, Talia, 2002. Detecting failures of backward induction: Monitoring information search in sequential bargaining. Journal of Economic Theory 104 (1), 16–47.

Johnson, Eric J, Häubl, Gerald, Keinan, Anat, 2007. Aspects of endowment: a query theory of value construction. Journal of experimental psychology: Learning, memory, and cognition 33 (3), 461.

Jung, Junehyuk, Kim, Jeong-Ho, Matějka, Filip, Sims, Christopher A., 2015. Discrete actions in information-constrained tracking problems. Working Paper.

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

Kahneman, Daniel, 1973. Attention and effort. Vol. 1063. Prentice-Hall.

Kahneman, Daniel, 2003. Maps of bounded rationality: Psychology for behavioral economics. The American Economic Review 93 (5), 1449–1475.

Kahneman, Daniel, Frederick, Shane, 2002. Representativeness revisited: Attribute substitution in intuitive judgment. Heuristics and biases: The psychology of intuitive judgment 49, 81.

Kahneman, Daniel, Knetsch, Jack L, Thaler, Richard H, 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. Journal of Economic perspectives 5 (1), 193–206.

Kahneman, Daniel, Tversky, Amos, 1979. Prospect theory: An analysis of decision under risk. Econometrica, 263–291.

Karlan, Dean, McConnell, Margaret, Mullainathan, Sendhil, Zinman, Jonathan, 2016. Getting to the top of mind: How reminders increase saving. Management Science 62 (12), 3393–3411.

Karlsson, Niklas, Loewenstein, George, Seppi, Duane, 2009. The ostrich effect: Selective attention to information. Journal of Risk and uncertainty 38 (2), 95–115.

Khaw, Mel Win, Li, Ziang, Woodford, Michael, 2017. Risk aversion as a perceptual bias. NBER Working Paper No. 23294.

Khaw, Mel Win, Stevens, Luminita, Woodford, Michael, 2016. Discrete adjustment to a changing environment: Experimental evidence. NBER Working Paper No. 22978.

Knudsen, Eric I, 2007. Fundamental components of attention. Annu. Rev. Neurosci. 30, 57–78.

Kőszegi, Botond, Rabin, Matthew, 2009. Reference-dependent consumption plans. The American Economic Review 99 (3), 909–36.

Kőszegi, Botond, Szeidl, Adam, 2013. A model of focusing in economic choice. Quarterly Journal of Economics 128 (1), 53–104.

Krajbich, Ian, Armel, Carrie, Rangel, Antonio, 2010. Visual fixations and the computation and comparison of value in simple choice. Nature neuroscience 13 (10), 1292.

Krajbich, Ian, Rangel, Antonio, 2011. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. Proceedings of the National Academy of Sciences 108 (33), 13852–13857.

Kurzban, Robert, Duckworth, Angela, Kable, Joseph W, Myers, Justus, 2013. An opportunity cost model of subjective effort and task performance. Behavioral and Brain Sciences 36 (6), 661–679.

Kusev, Petko, Schaik, Paulvan, Tsaneva-Atanasova, Krasimira, Juliusson, Asgeir, Chater, Nick, 2018. Adaptive anchoring model: How static and dynamic presentations of time series influence judgments and predictions. Cognitive science 42 (1), 77–102.

Lacetera, Nicola, Pope, Devin G., Sydnor, Justin R., 2012. Heuristic thinking and limited attention in the car market. The American Economic Review 102 (5), 2206–2236.

Lahey, Joanna N, Oxley, Douglas, 2016. The power of eye tracking in economics experiments. The American Economic Review Papers and Proceedings 106 (5), 309–313.

Laibson, David, 1997. Golden eggs and hyperbolic discounting. Quarterly Journal of Economics 112 (2), 443–478.

Leeper, Eric M., Sims, Christopher A., Zha, Tao, 1996. What does monetary policy do? Brookings Papers on Economic Activity 27 (2), 1–78.

Liebman, Jeffrey B., Zeckhauser, Richard J., 2004. Schmeduling. Working Paper.

Loewenstein, George, O'Donoghue, Ted, Rabin, Matthew, 2003. Projection bias in predicting future utility. Quarterly Journal of Economics 118 (4), 1209–1248.

Lombardi, Gaia, Fehr, Ernst, 2018. The attentional foundations of framing effects. Working Paper.

Lynch, Anthony W, 1996. Decision frequency and synchronization across agents: Implications for aggregate consumption and equity return. The Journal of Finance 51 (4), 1479–1497.

Mack, Arien, Rock, Irvine, 1998. Inattentional blindness. MIT Press.

Maćkowiak, Bartosz, Matějka, Filip, Wiederholt, Mirko, 2018. Rational Inattention: A Disciplined Behavioral Model.

Maćkowiak, Bartosz, Wiederholt, Mirko, 2009. Optimal sticky prices under rational inattention. The American Economic Review 99 (3), 769–803.

Maćkowiak, Bartosz, Wiederholt, Mirko, 2015. Business cycle dynamics under rational inattention. Review of Economic Studies 82 (4), 1502–1532.

Madrian, Brigitte C., Shea, Dennis F., 2001. The power of suggestion: Inertia in 401(k) participation and savings behavior. Quarterly Journal of Economics 116 (4), 1149–1187.

Malmendier, Ulrike, Nagel, Stefan, 2011. Depression babies: Do macroeconomic experiences affect risk taking? Quarterly Journal of Economics 126 (1), 373–416.

Mankiw, N. Gregory, Reis, Ricardo, 2002. Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. Quarterly Journal of Economics 117 (4), 1295–1328.

Mankiw, N. Gregory, Reis, Ricardo, Wolfers, Justin, 2003. Disagreement about inflation expectations. NBER Macroeconomics Annual 18, 209–248.

Manski, Charles F, McFadden, Daniel (Eds.), 1981. Structural analysis of discrete data with econometric applications. MIT Press.

Manzini, Paola, Mariotti, Marco, 2014. Stochastic choice and consideration sets. Econometrica 82 (3), 1153–1176.

Mas-Colell, Andreu, Whinston, Michael Dennis, Green, Jerry R, 1995. Microeconomic Theory. Oxford University Press, New York.

Masatlioglu, Yusufcan, Nakajima, Daisuke, Ozbay, Erkut Y, 2012. Revealed attention. American Economic Review 102 (5), 2183–2205.

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

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

McFadden, Daniel, 2006. Free markets and fettered consumers. The American Economic Review 96 (1), 3–29.

Mercier, Hugo, Sperber, Dan, 2011. Why do humans reason? arguments for an argumentative theory. Behavioral and brain sciences 34 (2), 57–74.

Miller, George A., 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63 (2), 81.

Mormann, Milica Milosavljevic, Frydman, Cary, 2016. The role of salience and attention in choice under risk: An experimental investigation. Working Paper.

Mullainathan, Sendhil, Schwartzstein, Joshua, Congdon, William J., 2012. A reduced-form approach to behavioral public finance. Annual Review of Economics 4 (1), 511–540.

Nagel, Rosemarie, 1995. Unraveling in Guessing Games: An Experimental Study. The American Economic Review 85 (5), 1313–1326.

Nobre, Anna C. (Kia), Kastner, Sabine, January 2014. The Oxford Handbook of Attention, 1st Edition. Oxford University Press.

O'Donoghue, Ted, Rabin, Matthew, 1999. Doing it now or later. The American Economic Review 89 (1), 103–124.

Olafsson, Arna, Pagel, Michaela, 2017. The ostrich in us: Selective attention to financial accounts, income, spending, and liquidity. NBER Working Paper No. 23945.

Pashler, Harold E., 1998. The Psychology of Attention. MIT Press.

Payne, John W., Bettman, James R., Johnson, Eric J., 1993. The Adaptive Decision Maker. Cambridge University Press.

Peng, Lin, Xiong, Wei, 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80 (3), 563–602.

Piccione, Michele, Spiegler, Ran, 2012. Price competition under limited comparability. Quarterly Journal of Economics 127 (1), 97–135.

Pop-Eleches, Cristian, Thirumurthy, Harsha, Habyarimana, James P., Zivin, Joshua G., Goldstein, Markus P., De Walque, Damien, Mackeen, Leslie, Haberer, Jessica, Kimaiyo, Sylvester, Sidle, John, et al., 2011. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 25 (6), 825.

Prelec, Drazen, 1998. The probability weighting function. Econometrica 66 (3), 497–527.

Rabin, Matthew, 2013. Incorporating limited rationality into economics. Journal of Economic Literature 51 (2), 528–543.

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

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

Reutskaja, Elena, Nagel, Rosemarie, Camerer, Colin F., Rangel, Antonio, 2011. Search dynamics in consumer choice under time pressure: An eye-tracking study. The American Economic Review 101 (2), 900–926.

Roger, Tristan, Roger, Patrick, Schatt, Alain, 2018. Behavioral bias in number processing: Evidence from analysts' expectations. Journal of Economic Behavior & Organization 149, 315–331.

Romer, Christina D., Romer, David H., 1989. Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. NBER Macroeconomics Annual 4, 121–170.

Romer, Christina D., Romer, David H., September 2004. A new measure of monetary shocks: Derivation and implications. The American Economic Review 94 (4), 1055–1084.

Ru, Hong, Schoar, Antoinette, 2016. Do credit card companies screen for behavioral biases? NBER Working Paper No. 22360.

Rubinstein, Ariel, 1998. Modeling Bounded Rationality. MIT Press.

Russo, J Edward, 1977. The value of unit price information. Journal of Marketing Research, 193–201.

Samuelson, Paul A., 1947. Foundations of Economic Analysis. Harvard University Press.

Schulte-Mecklenbeck, Michael, Johnson, Joseph G., Böckenholt, Ulf, Goldstein, Daniel G., Russo, J. Edward, Sullivan, Nicolette J., Willemsen, Martijn C., 2017. Process-tracing methods in decision making: On growing up in the 70s. Current Directions in Psychological Science 26 (5), 442–450.

Schwartzstein, Joshua, 2014. Selective attention and learning. Journal of the European Economic Association 12 (6), 1423–1452.

Shannon, Claude E., 1948. A mathematical theory of communication. Bell System Technical Journal 27 (4), 623–656.

Shlain, Avner S., 2018. More than a penny's worth: Left-digit bias and firm pricing. Working Paper.

Shue, Kelly, Townsend, Richard, 2018. Can the market multiply and divide? non-proportional thinking in financial markets. Working Paper.

Sicherman, Nachum, Loewenstein, George, Seppi, Duane J., Utkus, Stephen P., 2016. Financial attention. The Review of Financial Studies 29 (4), 863–897.

Simon, Herbert A., 1955. A behavioral model of rational choice. Quarterly Journal of Economics 69 (1), 99–118.

Simons, Daniel J., Chabris, Christopher F., 1999. Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception 28 (9), 1059–1074.

Sims, Christopher A., December 1998. Stickiness. Carnegie-Rochester Conference Series on Public Policy 49 (1), 317–356.

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

Sonnemann, Ulrich, Camerer, Colin F, Fox, Craig R, Langer, Thomas, 2013. How psychological framing affects economic market prices in the lab and field. Proceedings of the National academy of Sciences 110 (29), 11779–11784.

Spiegler, Ran, 2011. Bounded Rationality and Industrial Organization. Oxford University Press.

Stahl, Dale O, Wilson, Paul W, 1995. On players' models of other players: Theory and experimental evidence. Games and Economic Behavior 10 (1), 218–254.

Stango, Victor, Zinman, Jonathan, 2009. Exponential growth bias and household finance. The Journal of Finance 64 (6), 2807–2849.

Stanovich, Keith E, 1999. Who is rational?: Studies of individual differences in reasoning. Psychology Press.

Steiner, Jakub, Stewart, Colin, 2016. Perceiving prospects properly. The American Economic Review 106 (7), 1601–1631.

Steiner, Jakub, Stewart, Colin, Matějka, Filip, 2017. Rational inattention dynamics: Inertia and delay in decision-making. Econometrica 85 (2), 521–553.

Stevenson, Betsey, Wolfers, Justin, 2006. Bargaining in the shadow of the law: Divorce laws and family distress. Quarterly Journal of Economics 121 (1), 267–288.

Stigler, George J., 1961. The economics of information. Journal of Political Economy 69 (3), 213–225.

Taubinsky, Dmitry, Rees-Jones, Alex, 2017. Attention variation and welfare: Theory and evidence from a tax salience experiment. Forthcoming at the Review of Economic Studies.

Taylor, John B., 1980. Aggregate dynamics and staggered contracts. Journal of Political Economy 88 (1), 1–23.

Thaler, Richard H, Shefrin, Hersh M, 1981. An economic theory of self-control. Journal of Political Economy 89 (2), 392–406.

Thaler, Richard H., Sunstein, Cass R., 2008. Nudge. Yale University Press.

Tibshirani, Robert, 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.

Tirole, Jean, 2009. Cognition and incomplete contracts. The American Economic Review 99 (1), 265–294.

Treisman, Anne M, Gelade, Garry, 1980. A feature-integration theory of attention. Cognitive psychology 12 (1), 97–136.

Tversky, Amos, Kahneman, Daniel, 1974. Judgment under uncertainty: Heuristics and biases. Science 185, 1124–30.

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

Varian, Hal R., 1992. Microeconomic Analysis. WW Norton.

Veldkamp, Laura L., 2011. Information Choice in Macroeconomics and Finance. Princeton University Press.

Verrecchia, Robert E., 1982. Information acquisition in a noisy rational expectations economy. Econometrica 50 (6), 1415–1430.

Wang, Joseph T, Spezio, Michael, Camerer, Colin F, 2010. Pinocchio's pupil: Using eye-tracking and pupil dilation to understand truth telling and deception in sender-receiver games. American Economic Review 100 (3), 984–1007.

Woodford, Michael, 2013. Macroeconomic analysis without the rational expectations hypothesis. Annual Reviews of Economics 5, 303–346.

Zhang, Hang, Maloney, Laurence T., 2012. Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. Frontiers in Neuroscience 6.