Behavioral Inattention

Xavier Gabaix*

December 4, 2017

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 much progress has been 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, producer 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. Draft chapter for the *Handbook of Behavioral Economics*, edited by Douglas Bernheim, Stefano DellaVigna and David Laibson. *Comments solicited.* Please email me if you see an important reference that's missing (including your own papers). I thank Vu Chau and Antonio Coppola for excellent research assistance. For comments and suggestions, I thank the editors of this Handbook, Hunt Allcott, and Gautam Rao. For sharing their with data, I thank Stefano DellaVigna, Josh Pollet, and Devin Pope. I thank the Sloan Foundation for support.

Contents

1 Motivation

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. Still, traditional rational economics assumes that we process all the information that is freely available to us.

Modifying this assumption is empirically relevant, theoretically doable, and has great consequences in making economics more psychologically realistic, understanding markets, and designing 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 level overview. I use a simple framework to model the behavior of an inattentive consumer. Attention is parametrized by a value m, such that $m = 0$ corresponds to zero attention (hence, to a very behavioral model) 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 – roughly, a new good measure of attention is rewarded with a publication in the *American Economic Review*. I synthesize this literature. On average, the attention parameter estimated in the literature is 0.44, roughly halfway between the very behavioral and the very rational. Sensibly, attention is higher when the incentives to pay attention are stronger, as shown in Figure 1.

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). Some models rely on the plain notion that more important dimensions should be more given more attention – this is plain, but not actually trivial to capture in a clean model. Some

[1] Other surveys exist. DellaVigna (2009) offers a broad and readable introduction to measurement, in particular in 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 much of it simply relies on the basic framework of Section 2.

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 5 then covers models with stochastic decision – given an objective situation, the prediction is a probability distribution over the agents' actions. These are inherently 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 6 presents something elementary that yet helps unify all this: 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 version of them.

Much until now was static. 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: demand given "$s$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 unifying 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 $u\left(a, x\right) = -\frac{1}{2}\left(a - x\right)^2$, so that if she's rational, the agent wants to take the action that solves: $\max_a \mathbb{E}\left[-\frac{1}{2}\left(a - x\right)^2 |s\right]$. That is, the agent wants to guess the value of x given the noisy signal s. The first-order condition is

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

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

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

with the dampening factor[3]

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

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 $\bar{a}\left(x\right) := \mathbb{E}\left[a\left(s\right)|x\right]$ is then:

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

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 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 vast generalizations of the model in equation (4), with different weights m (endogenous or not) on the true value. A first class of models eliminates the noise, as not central, at least for the prediction of the average behavior (see Section 4). A second keeps the noise as central – which often leads to more complicated models (see Section 5).

Before discussing these variants and generalizations, I will present a simple formal frame-

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

work 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. 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]

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

This is as if x_i is replaced by the subjectively perceived x_i :

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

with an attention parameter $m_i \in [0, 1]$, and where x_i^d is the "default value" of variable i. When $m_i = 0$, the agent "does not think about x_i", i.e. replaces x_i by $x_i^s = 0$; when $m_i = 1$, she perceives the true value ($x_i^s = x_i$). We call $m = (m_i)_{i=1\ldots n}$ the attention vector.

The default x_i^d is typically the prior mean of x_i. However, it can be psychologically

[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 x. Also, the attention parameter m is often placed "outside the utility", as in

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

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

more sophisticated. 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:

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

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 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, say, $n > 10,000$ 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. choose

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

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

2.3 Unifying behavioral biases: Much of behavioral economics reflects 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. This exercise illustrates 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 "true model", and there is a "default, simple model" that spontaneously comes to mind. Attention m parameterizes the particular convex combination of the default and true models that corresponds to the agent's perception.[5]

[5]This feeling of unity of behavioral economics is not universal, but I find it useful to make a case for it in this chapter. 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.

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 price perceived by the agent is

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

Hence, take the case without income effect (see Section 6.1.1 for the case with 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 formulation is sometimes used in Gabaix (2014) and Khaw, Li, and Woodford (2017) – the latter formulation 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. Mullainathan, Schwartzstein, and Congdon (2012) offer an overview of behavioral public finance, and Farhi and Gabaix (2017) provide a systematic theory of optimal taxation (encompassing the Ramsey, Pigou and Mirrlees problems) with misperceptions and other biases.

2.3.3 Neglected risks

Suppose that the probability of a bad state of the world happening is p. The perceived probability is $p^s = mp$, if the default probability is $p^d = 0$. This generates underreaction to neglected risks (Gennaioli, Shleifer, and Vishny 2012).

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

$$U_0 = u_0 + \delta U_1. \tag{15}$$

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

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

The parameter m is equivalent here to the parameter β in the hyperbolic discounting literature.[6]

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 $u_0 + \delta U_1$ (Bernheim and Rangel 2009; Farhi and Gabaix 2017). In contrast, with hyperbolic discounting or a planner-doer model (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".

2.3.5 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 gives 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 median perception is[7]

$$q^s = mq + (1 - m)\, q^d. \tag{17}$$

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

$$p^s = \pi (p) = \frac{1}{1 + \left(\frac{1-p}{p}\right)^m \left(\frac{1-p^d}{p^d}\right)^{1-m}}, \tag{18}$$

which is the median perception of a behavioral agent: we have derived a probability weighting

[6] Gabaix and Laibson (2017) develop an interpretation of discounting via cognitive myopia much along these lines.

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

function $\pi(p)$. This yields overweighting of small probabilities (and symmetrically underweighting of probabilities close to 1). 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.[8] This is exactly as in the simple Gaussian updating model of Section 2.1, done in the log odds space. This gives a probability weighing function much like 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).[9]

Putting the two themes above (distortions of payoff and distortions of probability) together, we get something much like prospect theory (see also Woodford 2012 for the distortions of payoffs). 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. For instance the default probability for loss events is higher than the default probability in gains events. This will create loss aversion.

This thinking is all somewhat ex-post. Still, with this perspective, we can at least imagine why nature made people prospect-theoretic – and indeed, why we *should* make robots prospect-theoretic if their perceptions were noisy: they would misperceive payoffs and probabilities (because of the inherent noisiness of the intuitive treatment of numbers in the mind), and they would do so in an environment where gambles have in general negative expected values. Optimal correction of such features creates respectively: diminishing sensitivity, distortion of probabilities, and loss aversion. This generates a set of predictions much akin to those of prospect theory.

There is a tension. There is a tendency to *neglect* lots of small probability events in the "editing phase" of Kahneman and Tversky (1979), where agents decides which states of the world to take into account at all; and to *overestimate* them in the decision phase. This is the kind of tension that irks non-behavioral economists, and embarrasses behavioral economists. Using endogenous attention and sparsity, though, this tension can be solved.[10]

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

[9]We will see Bordalo, Gennaioli, and Shleifer's (2012) theory, which is more distantly related, in Section 4.2.

[10]This editing phase could be accounted for in the following way, along the lines of material of section 4.1.1. In the "first cut editing phase", the agent does as in Step 1 of the sparse max (Proposition 4.1), but with $p_i^d = 0$, and with $p_i^s = m_i p_i$. This yields the set of events i to which the agent does pay attention ($m_i > 0$); those are the events that are "important" for the decision at hand. Hence, as the end of this first Step 1, the agent has an "edited model", where non-surviving probabilities set to 0. Then, in the second phase, which could be called "behavioral decision making in a pre-simplified world" the agents does a regular sparse max of Proposition 4.1 (including a new Step 1, on the edited model, with a positive default probability, as in 18). This description seems to account reasonably well for the "editing phase" of prospect theory, of which Kahneman and Tversky (1979) gave a verbal rather than algorithmic description. A systematic exploration of that model would be interesting.

2.3.6 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. 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. There are other explanations for overconfidence and overprecision, e.g. motivation or signaling (Bénabou and Tirole 2002).

2.3.7 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.

2.3.8 Projection bias: Inattention to future circumstances by anchoring too much 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

$$x_t^s = mx_t + (1-m)\,x_0, \tag{19}$$

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

2.3.9 Base-rate neglect: Inattention to the base rate

In base-rate neglect (Tversky and Kahneman 1974) the true base probability P is replaced by $P^s(y) = mP(y) + (1-m)P^d(y)$, where $P^d(y)$ is a uniform distribution on the values of y.

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 (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 $N^s = (N^d)^{1-m} N^m$, 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 the stochasticity of the world

Tversky and Kahneman (1974) report that, when people see a fighter pilot's performance, they fail to appreciate reversion to the mean. 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 $\bar{x}_{t+1} := \mathbb{E}[x_{t+1} \mid x_t] = \lambda x_t$ 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, $\mathrm{Var}^s(\varepsilon) = m\sigma_\varepsilon^2$. 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:

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

Hence, very behavioral agents (with $m = 0$), who fully ignore the stochasticity of the world, will just expect the pilot to do next time as he did last time.

2.3.13 When will see 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} 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

$$\rho_i^s = m\rho_i + (1-m)\rho^d. \tag{20}$$

That is, instead of seeing precisely the fine nuances of many AR(1) processes, 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 $\mathbb{E}_t^s[y_{i,t+k}] = (\rho_i^s)^k y_{i,t}$, 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]

2.3.14 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

$$x^s=a+m\frac{b}{10} \tag{21}$$

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. I review this in greater detail in Section 3.2.5.

2.3.15 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 $x^s=mx+(1-m)\,x^d$.

2.3.16 Taking stocks of all these examples

All these examples, I submit, illustrate that the very simple framework above allows one to think in a relatively 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

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

[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 over 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.

extended, all of them worthwhile. 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 that partial inattention, e.g. in market equilibrium, or in the indirect effects of all this.

2. In the "empirical economic measurement" direction, researchers estimate attention m: see if it is full or not, and, even better, measure it.

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.[13]

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 Styles (2006) offer book-length surveys on the psychology of attention, in particular in perception.

2.4.1 Conscious versus unconscious attention

Systems 1 and 2. Recall the terminology for mental operations of Kahneman (2003), where "system 1" is the intuitive, fast, largely unconscious and parallel system, while "system 2" is the analytical, slow, conscious system.

System 2, working memory, and conscious attention. It is clear that we do not handle 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.[14]

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

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

[14] 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).

conscious examination (that is, whether they should satisfy $m_i > 0$). For instance, my system 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 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 $m_i > 0$.

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 (the Bayesian analogue of the default is the "prior"). 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).[15] 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).

2.4.3 Other themes

If the choice of attention is largely unconscious, this leads to the curious choice of "attentional blindness". The now canonical experiment for this is the gorilla experiment of Simons and Chabris (1999). When asked to perform a time-consuming task, subject often didn't see a gorilla in the midst of the experiment.

Another 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 the 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 but instructive (Rubinstein 1998; Gabaix, Laibson, Moloche, and Weinberg 2006; Caplin, Dean, and Martin 2011; Fudenberg, Strack, and Strzalecki 2017), so that more "outcome based models", that directly give the action rather than the intermediary steps, can be useful.

[15] 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.

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 in 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:[16]

1. Deviations from an optimal action.
2. Deviations from normative cross-partials, e.g. from Slutsky symmetry.
3. Physical measurement, e.g. eye-tracking.
4. Surveys: eliciting people's beliefs.
5. Qualitative measures: impact of reminders, of advice.

As we shall see, methods 3-5 can show that attention is not full (hence, help reject the naïve rational model), and 1 and 2 truly measure attention (i.e., measure the parameter m).[17]

3.1.1 Measuring inattention via deviation from an optimal action

Suppose the optimal action function is $a^{BR}(x) = a^r(mx)$, so the derivative with respect to x is $a_x^{BR}(x) = m a_x^r(mx)$. Therefore attention can be measured as[18]

$$m = \frac{a_x^{BR}}{a_x^r}.$$

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_x^r. How does one do that?

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

[17] 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.

[18] To be very precise, $m(x) = \frac{a_x^{BR}(x)}{a_x^r(mx)}$. 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 $a^r(x)$.

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_x^r. This is the methodology used by Chetty, Looney, and Kroft (2009), Taubinsky and Rees-Jones (2017), and Allcott and Taubinsky (2015), that we will review in Section 3.2.1.

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.

We shall review the practical methods later.[19]

3.1.2 Deviations from Slutsky symmetry

We will see below (Section 6.1.2) 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 $S_{it} = \frac{\partial c_i}{\partial p_j}$. 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 $M = diag(m_1, \dots, m_n)$ is the diagonal matrix of attention, we have

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

exactly like in the basic model. Abaluck and Adams (2017) explore this and similar relations to study the inattention to complex health care plans. It is nice to see how an a priori abstruse idea (the deviation from Slutsky symmetry in limited attention models , as in Gabaix 2014) can lead to concrete real-world measurement of the inattention to health-care plans characteristics.

3.1.3 Process tracking: Mouselab, eye tracking, pupil dilatation, etc.

A popular way to measure activity is with a process-tracing experiment commonly known as Mouselab (Payne, Bettman, and Johnson 1993; Gabaix, Laibson, Moloche, and Weinberg

[19] 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.

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. There are other physiological methods of measurement as well, such as measuring pupil dilation (Kahneman 1973). See Schulte-Mecklenbeck et al. (2017) for a recent review.

Those measures are useful, but they are not entirely ideal, as they measure attentional inputs, not attention itself.[20] To see this, conceptually, call T the time spend on dimension i – time here is a stand in for other measures, e.g. time gazing at the dimension, fMRI intensity, pupil dilatation, and so forth. Then, we could model the "attention processing function" as a function of time:

$$m = f(T).$$

Hence, time spent is an input in the attention production function, but it is not attention per se. Also, given attention is limited on a scale $[0, 1]$, and time T is unbounded, the function f cannot be linear. In addition, the function must be modulated by some "mental effort", let's call it M, as in:

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

For instance, a student may look at a whole lecture ($T = 80$ minutes), but still not really exert effort (low M), so that the total amount learned (indexed by m) is very low. It would be great to measure the production function of attention, $f(T, M)$.

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. Krajbich and Rangel (2011) find that the drift-diffusion model is a good predictor of choice and reaction times when subjects are faced with choices over two or three alternatives. 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) provide theory and evidence on how statistical discrimination guides information acquisition.

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 level of interest rate. But that doesn't mean that she actually takes the interest rate into account when buying a sweater – so as to

[20] There is another difficulty. The lab evidence is not necessarily directly analogous to the way people make big decisions, like deciding which college to attend. I may attend to information about college A on one day, and to information about college B on another, but over time I attend to a lot of diverse information before I make my choice, and I may even keep notes of what I considered salient while I was attending. That is different than the one-shot multifaceted information arrival typically studied in the lab.

satisfy her rational Euler equation for sweaters. Hence, if people show ignorance in a survey, it is good evidence that they are inattentive. However, when they show a good knowledge, it does not mean that they actually take into account the variable in their decision. Information, as measured in surveys, is an input into attention, but not the actual attention metric.[21]

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).

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 creates a big impact for 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).

Hanna, Mullainathan, and Schwartzstein (2014) provide summary information to seaweed farmers. This allows them 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 ($\sigma_{x_i}^2 a_{x_i}^2$ in equation 31) leads to paying more attention, and better learning.

Again, this type of evidence clearly 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. The first experimental measure of attention to taxes was (to the best of my knowledge) in Chetty, Looney, and Kroft (2009) using a field experiment. Chetty,

[21]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 2016a).

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).

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)[22] that $\mathbb{E}[m] = 0.25$ and $\text{Var}(m) = 0.13$. 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.

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 their existence when buying the "base good" product. In particular, if there are enough naïves there is an inefficient equilibrium where shrouded attributes are priced much 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,[23] and predict their future demand with excessive ex-ante precision – for example, individuals frequently exhaust their cellular plans' usage allowance, and incur high overage costs. 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 – showing that people are indeed behavioral rather than Bayesian (a Bayesian

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

[23] 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.

agent should be suspicious of any non-disclosed item, rather than just ignore it like a behavioral agent). This literature works in a healthy interplay with a theoretical literature probing deeper 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) contains 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 more 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. This study shows very clearly a mistake in an important economic context. Abaluck and Adams (2017) show that consumers' inertia in health plan choices 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. But how big is this effect? One approach is postulate hyperbolic discounting, and import the parameter $m = \beta \simeq 0.7$ into the model (e.g. Gruber and Kőszegi 2001). Allcott, Lockwood, and Taubinsky (2017) make further progress by using a specially designed survey measuring temptation and health knowledge.

3.2.5 People use rounded numbers when thinking about the mileage of used cars

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 already drove), 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 $m = 0.69$. 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.

This is a very nice study, as it offers high quality data. It would be nice to see if it matches the quantitative predictions of models discussed in this survey (for example, that in equation 31).

3.2.6 When people buy cars, do they pay full attention to the present value of gasoline expenses?

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 indeed 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.

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 some samples from this literature.

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.

DellaVigna and Pollet (2007) find that investors have a limited ability to incorporate some subtle forces (predictable change 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.

Cohen and Frazzini (2008) find that investors are quick at pricing the "direct" impacts on 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).

Baker, Pan, and Wurgler (2012) find that when thinking about a merger or acquisition price, investors put a lot of attention 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.

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.

This literature is growing fast. It would be nice to have more structural models, 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.

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[24] 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 smoothly and with delay when responding to shocks in other variables, but contemporaneously and significantly different from zero when responding to its own shocks. Such finding is robust in VAR specifications of various sizes, variable sets, and identification method (Leeper, Sims, and Zha 1996; Christiano, Eichenbaum, and Evans 2005). While it is feasible to generate delayed response using adjustment costs, large adjustment cost 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

[24] 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.

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.

3.3 Attention across stakes and studies

Attention over many studies Table 1 and Figure 1 contain a synthesis of ten studies 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 they 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 $p^s(m) = p + m\tau$

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. 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.

Calibrating the attention function Figure 1 additionally shows a calibration of an attention model in which estimated attention $\hat{m}$ as a function of the attribute's relative importance τ/p is[25]

$$\hat{m} = \mathcal{A}_\alpha\left(\left[\frac{\tau/p}{\bar{\kappa}}\right]^2\right) \tag{22}$$

where $\mathcal{A}_\alpha$ is an attention function, which will be derived in detail 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

[25] Things are expressed in terms of the "scale free cost" $\bar{\kappa}$ (see Gabaix 2016a, 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 think 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.

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

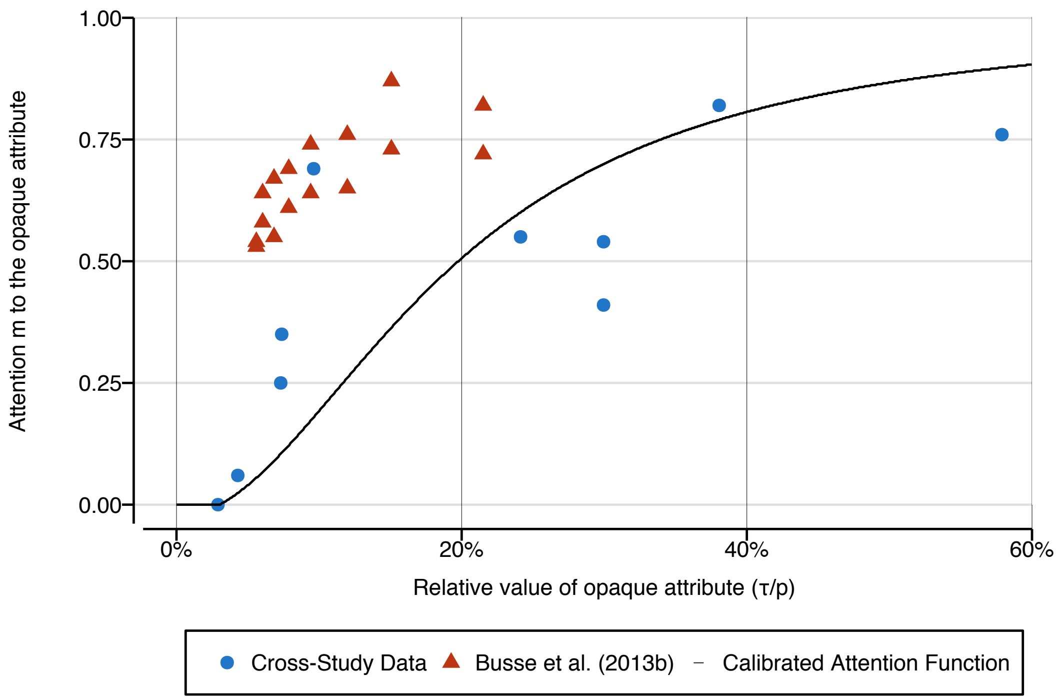

Figure 1: **Attention point estimates (m) vs. relative value of opaque attribute (τ/p), with overlaid calibrated attention function.** This figure shows (*circles*) point estimates of the attention parameter m in a cross-section of recent studies (shown in Table 1), against the estimated relative value of the opaque add-on attribute with respect to the relevant good or quantity (τ/p). A value $m = 1$ corresponds to full attention, while $m = 0$ implies complete inattention. The overlaid curve (*black curve*) shows the corresponding calibration of the quadratic-cost attention function in (23), where we impose $\alpha = 1$ and obtain calibrated cost parameters $\bar{\kappa} = 3.0\%$, $q = 20.4$ 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.

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 (30),

$$\mathcal{A}_\alpha(\sigma^2) \coloneqq \sup \left[\arg \min_{m \in [0,1]} \frac{1}{2} (1-m)^2 \sigma^2 + \left(m + qm^2\right)^\alpha \right] \tag{23}$$

where the parameter q was useful to capture the curvature of the attention function. In order to retain both continuity and sparsity of the attention function (23), I impose the restriction $\alpha = 1$, and estimate the cost parameters $\bar{\kappa}$ and q on the cross-study data via nonlinear least squares, according to the model in (22), which yields calibrated parameters $\bar{\kappa} = 3.0\%$, $q = 20.4$.

4 Models of Endogenous Attention: Deterministic Action

We have seen how attention can be modeled in a simple way, and that it can be measured. In this section, we will study some models that endogenize attention. They are deterministic, and differ in emphasis. The sparsity model of Section 4.1 emphasizes the absolute importance of effects. The salience model of section 4.2 is instead mostly interested in relative importance.

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 presents a generalization of the max operator used in economics, where agents can be less than fully attentive. This helps write a behavioral version of basic textbook microeconomics (Section 6), of the basic theory of taxation (Farhi and Gabaix (2017), which also uses more general models), of basic dynamic macroeconomics (Gabaix (2016a)), and macroeconomic fiscal and monetary policy (Gabaix (2016b)).

The agent 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):[26]

$$\operatorname*{smax}_a u(a,x) \text{ subject to } b(a,x) \geq 0, \tag{24}$$

which is a less than fully attentive version of the "max" operator. Variables a, x and function

[26] 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).

b have arbitrary dimensions.[27]

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 strictly concave) and twice continuously differentiable around $(a^d, 0)$. We will typically evaluate the derivatives at the default action and parameter, $(a, x) = (a^d, 0)$.

4.1.1 The sparse max: First, without constraints

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

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

$$a(x, m) := \arg\max_a u(a, x, m) \tag{25}$$

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.

So it would be sensible to allocate attention m as:

$$\max_m \mathbb{E}[u(a(x, m), x)] - \mathcal{C}(m). \tag{26}$$

However, ever since Simon (1955), many researchers have seen that problem (26) is very complicated – more complex than the original problem (that's the "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 agents does two things. In Step 1, she solves the optimal problem* (26)*, but in a simplified version: (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,* (25)*.*

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

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

the default action and parameter, i.e. at $(a,x)=\left(a^d,0\right)$. 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):

$$V(m)-V(\iota)=-\frac{1}{2}\sum_{i,j}\left(1-m_i\right)\Lambda_{ij}\left(1-m_j\right)+o\left(\sigma^2\right), \tag{27}$$

defining $\Lambda_{ij}:=-\sigma_{ij}a_{x_i}u_{aa}a_{x_j}$, $\sigma_{ij}:=\mathbb{E}\left[x_ix_j\right]$ and $\sigma^2=\|(\sigma_i^2)_{i=1\ldots n}\|$.[28] The agent drops the non-diagonal terms (this is an optional, but useful, feature of the sparse max). The agent entertaining the simplified problem of Definition 4.1 will want to solve:

$$m^*=\arg\min_{m\in\mathbb{R}^n}\frac{1}{2}\sum_{i=1}^n\left(1-m_i\right)^2\Lambda_{ii}+\kappa\sum_{i=1}^n m_i^\alpha. \tag{28}$$

The attention function To build some intuition, let us start with the case with just one variable, $x_1=x$. Then, problem (28) becomes:

$$\min_m\frac{1}{2}\left(1-m\right)^2\sigma^2+\kappa\left|m\right|^\alpha. \tag{29}$$

Attention is $m=\mathcal{A}_\alpha\left(\frac{\sigma^2}{\kappa}\right)$, where the "attention function" $\mathcal{A}_\alpha$ is defined as[29]

$$\mathcal{A}_\alpha\left(\sigma^2\right):=\sup\left[\arg\min_{m\in[0,1]}\frac{1}{2}\left(1-m\right)^2\sigma^2+m^\alpha\right]. \tag{30}$$

Figure 2 plots how attention varies with the variance σ^2 for fixed, linear and quadratic cost: $\mathcal{A}_0\left(\sigma^2\right)=1_{\sigma^2\geq 2}$, $\mathcal{A}_1\left(\sigma^2\right)=\max\left(1-\frac{1}{\sigma^2},0\right)$, $\mathcal{A}_2\left(\sigma^2\right)=\frac{\sigma^2}{2+\sigma^2}$.

We now explore when a^s indeed induces no attention to many variables.[30]

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.[31]

---

[28] The Taylor expansions is for small noises in x, rather than for m close to 1.

[29] That is: $\mathcal{A}_\alpha\left(\sigma^2\right)$ is the value of $m\in[0,1]$ that minimizes $\frac{1}{2}\left(1-m\right)^2\sigma^2+m^\alpha$ (as conveyed by the arg min), taking the highest m if there are multiple minimizers (as conveyed by the sup).

[30] 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).

[31] The sparse max is, properly speaking, sparse 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. Perhaps smax should be called a "bmax" or behavioral / boundedly rational max.

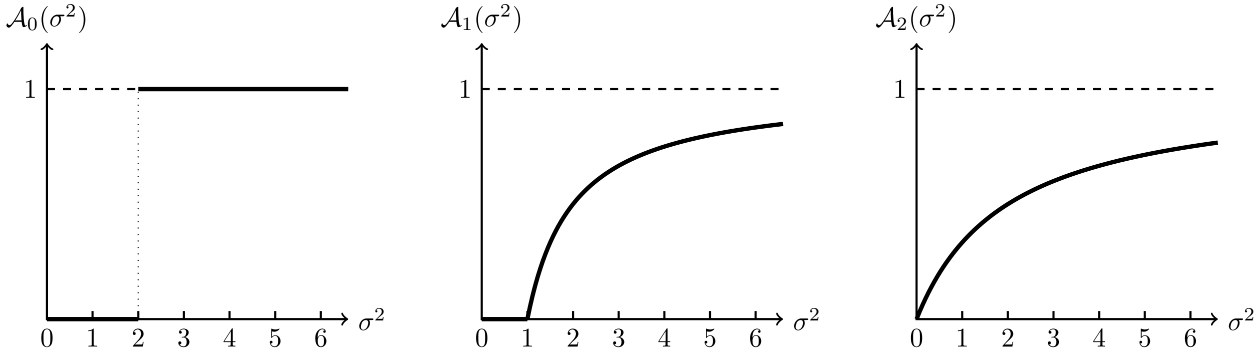

Figure 2: Three attention functions $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$, corresponding to fixed cost, linear cost and quadratic cost respectively. We see that $\mathcal{A}_0$ and $\mathcal{A}_1$ induce sparsity – i.e. a range where attention is exactly 0. $\mathcal{A}_1$ and $\mathcal{A}_2$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

The sparse max: Values of attention

Proposition 4.1 *The sparse max is done in two steps.*

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

$$m_i^* = \mathcal{A}_\alpha \left(\sigma_i^2 \left|a_{x_i} u_{aa} a_{x_i}\right| / \kappa\right), \tag{31}$$

where $\mathcal{A} : \mathbb{R} \to [0,1]$ *is the attention function expressed in* (30)*,* 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

$$a^s = \arg\max_a u\left(a, x, m^*\right). \tag{32}$$

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 great losses (high $|u_{aa}|$), and if the cost parameter κ is low.

The sparse max procedure in (31) entails (for $\alpha \leq 1$): "Eliminate each feature of the world that would change the action by only a small amount" (i.e., when $\alpha = 1$, eliminate the x_i such that $\left|\sigma_i \cdot \frac{\partial a}{\partial x_i}\right| \leq \sqrt{\frac{\kappa}{|u_{aa}|}}$). 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.[32] He also devotes "some" attention to those important variables, not necessarily paying full attention

[32]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 $\mathbb{E}\left[(a^r)^2\right] < \infty$).

to them.[33]

Let us revisit the initial example.

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

$$a^s = \sum_{i=1}^n m_i b_i x_i, \qquad m_i = \mathcal{A}_\alpha \left(b_i^2 \sigma_i^2 / \kappa\right). \tag{33}$$

Proof: We have $a_{x_i} = b_i$, $u_{aa} = -1$, so (31) gives $m_i = \mathcal{A}_\alpha \left(b_i^2 \sigma_i^2 / \kappa\right)$. □

4.1.2 Sparse max: Full version, allowing for constraints

Let us now extend the sparse max so that it can handle maximization under $K(= \dim b)$ constraints, problem (24). As a motivation, consider problem

$$\max_{c_1,\ldots,c_n} u\left(c_1, \ldots, c_n\right) \text{ subject to } p_1 c_1 + \ldots + p_n c_n \leq w. \tag{34}$$

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.[34]

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

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:[35]

$$\frac{\partial u/\partial c_1}{\partial u/\partial c_2} = \frac{p_1^s}{p_2^s}, \tag{35}$$

i.e. $u'(\boldsymbol{c}) = \lambda \boldsymbol{p}^s$, for some scalar λ.[36] The agent will tune λ so that the constraint binds, i.e. the value of $\boldsymbol{c}(\lambda) = u'^{-1}\left(\lambda \boldsymbol{p}^s\right)$ satisfies $\boldsymbol{p} \cdot \boldsymbol{c}(\lambda) = w$.[37] Hence, in step 2, the agent "hears

[33] 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).

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

[35] 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.

[36] 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.

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

clearly" whether the budget constraint binds.[38] This agent is boundedly rational, but smart enough to exhaust his budget.

We next generalize this idea to arbitrary problems. This is heavier to read, so the reader may wish to skip to the next section. 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 (24) 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.[39]

Definition 4.2 (Sparse max operator with constraints). *The sparse max,* $\text{smax}_{a|\kappa,\sigma}\, u(a,x)$ *subject to* $b(a,x) \geq 0$*, is defined as follows.*

Step 1: Choose the attention m^* *as in (28), 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 Definition 4.2 allows quite generally for the translation of a boundedly rational maximum without constraints, into a boundedly maximum with constraints. To obtain further intuition on the constrained maximum, we turn to consumer theory.

Consequences for consumption Section 6.1.1 develops consumer demand from the above procedure, and contains many examples. For instance, Proposition 6.1 finds that the Marshallian demand of a behavioral agent is

$$\boldsymbol{c}^s(\boldsymbol{p}, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w'), \tag{36}$$

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 hit 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. Recall that λ^d is the Lagrange multiplier at the default price.[40]

[38] See footnote 52 for additional intuitive justification.

[39] For instance, in a consumption problem (34), 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.

[40] λ^d is endogenous, and characterized by $u'(\boldsymbol{c}^d) = \lambda^d \boldsymbol{p}^d$, where $\boldsymbol{p}^d$ is the exogenous default price, and $\boldsymbol{c}^d$ is the (endogenous) optimal consumption as the default. The comparative statics hold, keeping λ^d constant.

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.

Many more consequences will emerge in Section 6.

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

In a series of papers, Bordalo, Gennaioli, and Shleifer (2012; 2013; 2015) 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:

$$u^s(a) = \sum_{i=1}^n b_i m_{ai} x_{ai} \quad (37)$$

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 $\bar{x}_i = \frac{1}{A}\sum_{a=1}^A x_{ai}$.

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 4.3 *The salience function $\sigma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfies the following properties:*[41]

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.*[42] *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 4.3 is

$$\sigma(x, y) = \frac{|x - y|}{|x| + |y| + \theta} \tag{38}$$

with $\theta \geq 0$. 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 $(r_{ai})_{i=1,\dots,n}$ 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

$$m_{ai} = \frac{\delta^{r_{ai}}}{\sum_{j=1}^{n} \delta^{r_{aj}}} \tag{39}$$

where the parameter $\delta \in (0, 1]$ measures the strength of the salience distortion. 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.[43]

[41] 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 4.3.

[42] 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).

[43] The distortion function (39) can exhibit discontinuous jumps. An alternative specification introduced in Bordalo, Gennaioli, and Shleifer (2015) that allows for continuous salience distortions is

$$m_{ai} = \frac{\exp\{(1 - \delta)\,\sigma(x_{ai}, \bar{x}_i)\}}{\sum_{j=1}^{n} \exp\{(1 - \delta)\,\sigma(x_{aj}, \bar{x}_j)\}}$$

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 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" $(\bar{q}_{\mathrm{s}}, \bar{p}_s) = (25, 15)$, while at the restaurant we have a reference good $(\bar{q}_r, \bar{p}_r) = (25, 55)$. Under the functional form in (38), 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.

Bordalo, Gennaioli, and Shleifer (2015) 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 is to 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

$$u^{BR}(a) = \sum_{i \in S} \pi_i m_{ai} v(x_{ai}) \tag{40}$$

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).[44] In this version of the common-consequence Allais paradox, originally due to Kahneman and Tversky (1979), experiment 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 (2). These conditions yield the following necessary and sufficient criterion for lottery a to be preferred in a choice between a and $\tilde{a}$:

$$\sum_{i \in S} \delta^{r_i} \pi_i [v(x_{ai}) - v(x_{\tilde{a}i})] > 0. \tag{41}$$

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

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

which is true whenever $\delta < .73$. Intuitively, for low enough δ the agent focuses her attention

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

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 a focus on additive problems. Assuming that the decision-maker focuses her 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 $\mathcal{C} = \{\mathbf{x}_a\}_{a=1...A}$ of $A > 1$ actions that yield utilities $(x_{ai})_{i=1...n}$ 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 in 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)

$$\sigma_i = \max_a x_{ai} - \min_a x_{ai}. \tag{42}$$

Subjectively perceived utility is:

$$u^s(a) = \sum_{i=1}^{n} m_i x_{ai}, \tag{43}$$

where the attention weight is

$$m_i = \mathcal{A}(\sigma_i) \tag{44}$$

and is increasing in the range of outcomes σ_i, and $\mathcal{A}$ is an attention function. Intuitively, the decision-maker attends to those dimensions of the problem in which her choice is most consequential, and we have $\mathcal{A}'(\sigma) > 0$. Hence, we obtained the 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 choose between two actions a_1, a_2 yielding payoff vectors $x_1 = (-1000, 0, \dots, 0)$ and $x_2 = (0, -100, \dots, -100)$ respectively. The vector of utility ranges is therefore $\sigma = (1000, 100, \dots, 100)$, 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 me (a self-serving bias), and avoid thinking about depressing thoughts. There is empirical evidence on 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". The evidence is complex: in loss aversion, people pay more attention to losses than gains, something prima facie opposite to a self-serving bias. Hopefully future research will clarify this.

A simple model of that would be the following. Call $v(x) = \max_a u(a, x)$. Then, paying attention m_i to dimension i has a psychic cost $\kappa_{\text{mot}} v_{x_i} x_i$. Another, related model, is that attention to variable i is

$$m_i = \mathcal{A}\left(\frac{v_{x_i} x_i}{\kappa_{\text{mot}}}\right) \tag{45}$$

such that I pay more attention to things that favor more (increase my utility). 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. Why do we pay attention to funny stories, to cherished photos? Presumably, because of the enjoyment value. This is not captured above.[45]

Likewise, these models do not feature a refined "cost" of attention – e.g. why it's harder to pay attention to tax changes than a funny story. This is not modeled.

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.

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

5 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) 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.

[45] Enjoying a story give me pleasure, but not the sort of utility captured in the "motivated attention" above – which rests on paying attention to variables (such as my driving ability) that make me seem better to myself or others.

5.1 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

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

so that the agent's problem is:

$$\max_{\tau_{\varepsilon}}-\frac{1}{2}(1-m)\,\sigma_{x}^{2}-\kappa G(\tau_{\varepsilon})\ \text{ subject to } m=\frac{\tau}{1+\tau}$$

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 (29).

This allows us to think the optimal choice of information. When actions are strategic complements, you can get multiple equilibria in information gathering (Hellwig and Veldkamp 2009). When action 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).

5.2 Entropy-based inattention

How to handle non-Gaussian variables? Sims (1998, 2003) proposed the entropy penalty, kindling the interest of many macroeconomists and in this way helping make macroeconomics more realistic. An advantage of that entropy-based penalty is that it applies to all distributions. This generality comes at a cost, but first let's review some information theory.

Sims called this technique "rational inattention". This name has proved confusing, as Sims really proposed an entropy-based penalty, not at all the general notion that attention allocation responds to incentives, which vastly predates the 2000s. "Rational inattention" ought to refer to a much more ancient idea 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 like a proper name for the literature initiated by Sims.

5.2.1 Information theory: A crash course

Here is a brief introduction to information theory, as done 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 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 calculate the value of the 1000th decimal 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

$$\begin{aligned} H(X) &= -\mathbb{E}[\log f(X)] = -\mathbb{E}\left[\log\left(f^Y(Y) f^Z(X)\right)\right] \\ &= -\mathbb{E}\left[\log f^Y(Y) + \log f^Z(X)\right] \\ H(X) &= H(Y) + H(Z). \end{aligned} \tag{46}$$

This shows that the information of independent outcomes is additive. The next concept is that of mutual information. It is defined by the reduction of entropy of X when you know

Y:

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

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

$$D(p\|q) = \mathbb{E}^P\left[\log \frac{p(X)}{q(X)}\right] = \sum_i p_i \log \frac{p_i}{q_i}. \tag{47}$$

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:

$$I(X,Y) = D\left(p(x,y)\|p^X(x)\,p^Y(y)\right) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,q(y)}. \tag{48}$$

In other terms, mutual information $I(X,Y)$ is the Kullback-Leibler divergence between the full joint probability $p(x,y)$ and its "decoupled" approximation $p(x)\,q(y)$.

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

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

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

$$H(X) = \log(b-a) \tag{49}$$

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

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

$$H(Y) = H(X) + \log\sigma \tag{50}$$

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,

$$H(X) = \frac{1}{2}\log\sigma^2 + \frac{1}{2}\log(2\pi e) \tag{51}$$

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

$$H(X) = \frac{1}{2}\log(\det V) + n\log(2\pi e) \tag{52}$$

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

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

$$I(X,Y) = \frac{1}{2}\log\frac{1}{1-\rho^2} \tag{53}$$

so that the mutual information is increasing in the correlation.

5.2.2 Using Shannon entropy as a measure of cost

Sims (2003) proposed the following problem. Suppose the agent does $\max_a u(a,x)$ in the traditional version. 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

$$\max_{q(a|x)} \mathbb{E}\int u(a,x)\, q(a|x)\, f(x)\, da\, dx \text{ s.t. } I(A,X) \leq K. \tag{54}$$

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. Of course, the nature of this black box is a bit unclear, but may be treated as some thought process.[46]

To get a feel for this problem, consider the case 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 $a = \mathbb{E}[x|s] = mx + m\varepsilon$. The analytics in (82) shows that $\rho^2 = \operatorname{corr}(a,x)^2 = m$, and using equation (53), the mutual information is $I(S,A) = \frac{1}{2}\log\frac{1}{1-m}$. Saturating the constraint in (54), $I(S,A) = K$, which gives:

$$m = 1 - e^{-2K}$$

[46]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 some consideration developed in the Shannon theory, the cost is captured by the mutual information $I(A,X)$.

So, we get a solution like the basic problem of Section Section 2.1, with a special cost function.

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 $var(x_i) = \sigma_i^2$. Then, one can show that the solution takes the form:[47]

$$a^{\text{Sims}} = m \sum_i b_i x_i + \eta \tag{55}$$

with a η orthogonal to the x. Let us contrast this to the answer of sparsity, (33) gives

$$a^s = \sum_{i=1}^{n} m_i b_i x_i \tag{56}$$

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 obtain uniform dampening across all variables (i.e. $m_i = m$ for all i in equation (55)) – not source-specific dampening as in (33)-(56).[48]

One advantage is that we have a universally applicable measure of the cost of information. However, the cost is that it's not particularly psychologically founded. Relatedly, the procedure is generically quite intractable. For instance, in Jung, Kim, Matejka, and Sims (2015), the solution has atoms – it is non-smooth. One needs a computer to solve it. Those problems can be mathematically fascinating, but they seem to lead us away from a behavioral account of what people actually do.[49]

Still, at a minimum 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). 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).

5.3 Random choice via limited attention

5.3.1 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,

[47] The solution is left as an exercise for the reader.

[48] 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.

[49] 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.

the probability of choosing v_i is

$$p_i = \frac{e^{\sigma v_i}}{\sum_j e^{\sigma v_j}}. \tag{57}$$

The classic microfoundation is the following. Agents receive a signal

$$s_i = v_i + \sigma \varepsilon_i \tag{58}$$

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, $p_i = \mathbb{P}\left(i \in \text{argmax}_j\, s_j\right)$. With Gumbel noise, this leads (after some calculations as in e.g. Anderson, De Palma, and Thisse 1992) to (57). When the noise size σ is higher, there is more uncertainty; when $\sigma \to \infty$, then $p_i \to \frac{1}{n}$. The choice is completely random. When the cost κ is 0, then the agent is the traditional, rational agent.

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 is useful to see thing like: what's the equilibrium price, when consumers are confused? Then, the equilibrium price markup (defined as price minus cost) is generally proportional to σ, the amount of noise. For a related model with two types of agents, see Carlin (2009).

5.3.2 Random choice via entropy penalty

Matějka and McKay (2015) derive a 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.[50] So,

[50] 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.

the Lagrangian is

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

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

$$p_i(v) = \frac{p_i^d e^{v_i/\kappa}}{\sum_j p_j^d e^{v_j/\kappa}}. \tag{59}$$

Matějka and McKay's setup (2015) actually gives the default: $\max_{(p_i^d)} \mathbb{E}\left[\log\left(\sum_i p_i^d e^{v_i/\kappa}\right)\right]$ s.t. $\sum_i p_i^d = 1$. So, when the v_i are drawn 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. See Caplin, Dean, and Leahy (2016). This in turn helps explore dynamic problems, as in Steiner, Stewart, and Matějka (2017).

6 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 effect works for a host of behavioral models, provided they generate some inattention to prices.

6.1 Textbook consumer theory: A behavioral update

6.1.1 Basic consumer theory: Marshallian demand

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

$$c(\boldsymbol{p}, w) := \arg\max_{\boldsymbol{c} \in \mathbb{R}^n} u(\boldsymbol{c}) \text{ subject to } \boldsymbol{p} \cdot \boldsymbol{c} \leq w \tag{60}$$

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 stand 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 an innovation. The price perceived by a behavioral agent is $p_i^s = p_i^d + m_i x_i$, i.e.:

$$p_i^s(m) = m_i p_i + (1 - m_i) p_i^d. \quad (61)$$

When $m_i = 1$, the agent fully perceives price p_i, while when $m_i = 0$, he replaces it by the default price.[51]

Proposition 6.1 (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*

$$\boldsymbol{c}^s(\boldsymbol{p}, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w'), \quad (62)$$

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 hit 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. 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.

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

[51] 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, $\frac{\partial p_i^s}{\partial p_i}_{|\boldsymbol{p} = \boldsymbol{p}^d} = m_i$.

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".[52]

6.1.2 Nominal illusion, asymmetric Slutsky matrix, and inferring attention from choice data

Recall that the consumer "sees" only a part m_j of the price change (eq. 61). One consequence is nominal illusion.

Proposition 6.2 (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 the 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.

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:

$$S_{ij}(\boldsymbol{p}, w) := \frac{\partial c_i(\boldsymbol{p}, w)}{\partial p_j} + \frac{\partial c_i(\boldsymbol{p}, w)}{\partial w} c_j(\boldsymbol{p}, w). \tag{63}$$

With the traditional agent, the most surprising fact about it is that it is symmetric: $S^r_{ij} = S^r_{ji}$. Kreps(2012, Chapter 11.6) comments: "The fact that the partial derivatives are identical and not just similarly signed is quite amazing. Why is it that whenever a \$0.01 rise in the price of good i means a fall in (compensated) demand for j of, say, 4.3 units, then a \$0.01 rise in the price of good j means a fall in (compensated) demand for i by [...] 4.3 units? [...] I am unable to give a good intuitive explanation." Varian (1992, p.123) concurs:

[52] 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.

"This is a rather nonintuitive result." Mas-Colell, Whinston, and Green (1995, p.70) add: "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 6.3 (Slutsky matrix). *Evaluated at the default price, the Slutsky matrix S^s is, compared to the traditional matrix S^r:*

$$S_{ij}^s = S_{ij}^r m_j, \tag{64}$$

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.

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 $\left|S_{ij}^s\right|$ is lower than $\left|S_{ji}^s\right|$: 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.[53]

Equation (64) makes tight testable predictions. It allows us to infer attention from choice data, as we shall now see.[54]

Proposition 6.4 (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_{ij}^s is available. Then, m can be recovered as $m_j = \overline{m} \prod_{i=1}^n \left(\frac{S_{ij}^s}{S_{ji}^s}\right)^{\gamma_i}$, for any $(\gamma_i)_{i=1...n}$ such that $\sum_i \gamma_i = 1$.*

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

[53] Browning and Chiappori (1998) have in mind a very different phenomenon: intra-household bargaining, with full rationality. Their model adds $2n + O(1)$ degrees of freedom, while sparsity adds $n + O(1)$ degrees of freedom.

[54] The Slutsky matrix does not allow one to recover $\overline{m}$: for any $\overline{m}$, S^s admits a dilated factorization $S_{ij}^s = \left(\overline{m}^{-1} S_{ij}^r\right)\left(\overline{m} m_j\right))$. 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.

The underlying "rational" matrix can be recovered as $S^r_{ij} := S^s_{ij}/m_j$, and it should be symmetric, a testable implication.[55] There is a literature estimating Slutsky matrices, which does not yet seem to have explored the role of non-salient prices.

It would be interesting to test Proposition 6.3 directly. The extant evidence is qualitatively encouraging, via the literature on obfuscation and shrouded attributes (Gabaix and Laibson 2006, Ellison and Ellison 2009) and tax salience.[56] Those papers find field evidence that some prices are partially neglected by consumers.

Marginal demand

Proposition 6.5 *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*

$$\frac{\partial c^s_i}{\partial p_j} = \frac{\partial c^r_i}{\partial p_j} \times m_j - \frac{\partial c^r_i}{\partial w} c^r_j \times (1 - m_j)\,. \tag{65}$$

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).

6.2 Textbook competitive equilibrium theory: A behavioral update

We next revisit the textbook chapter on competitive equilibrium, with a less than fully rational agent. 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^a_i 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).

6.2.1 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.[57] We assume that competitive equilibria are interior,

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

[56] Chetty, Looney, and Kroft (2009) show that a \$1 increase in tax that is included in the posted prices reduces demand more than when it is not included. Anagol and Kim (2012) found that many firms sold closed-end mutual funds because they can charge more fees by 'initial issue expense' (which can be amortized, so is not visible to customers) than by 'entry load' (a more obvious one time charge). In an online auction experiment. Greenwood and Hanson (2014) estimate an attention $m = 0.5$ to competitors' reactions and general equilibrium effects.

[57] 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 6.6 and 6.7 apply to exchange economies, hence apply a fortiori to production economies.

and consumers are locally non-satiated.

Proposition 6.6 (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 that given two goods i and j, each agent equalizes relative marginal utilities and relative perceived prices (see equation 35):

$$\frac{u^a_{c_i}}{u^a_{c_j}} = \left(\frac{p^s_i}{p^s_j}\right)^a, \qquad \frac{u^b_{c_i}}{u^b_{c_j}} = \left(\frac{p^s_i}{p^s_j}\right)^b, \tag{66}$$

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.

$$\frac{u^a_{c_i}}{u^a_{c_j}} = \frac{u^b_{c_i}}{u^b_{c_j}}. \tag{67}$$

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

6.2.2 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).

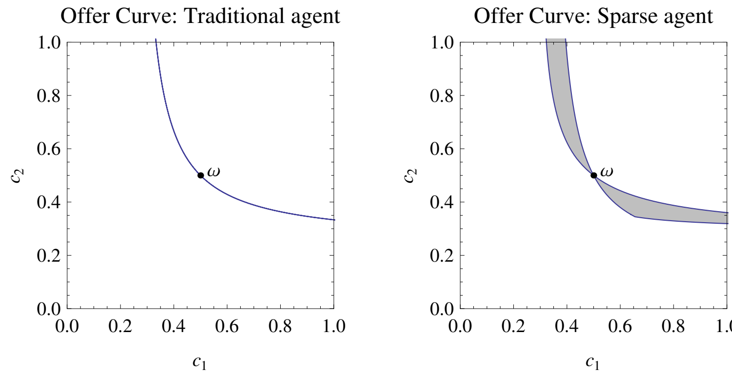

Figure 3: This Figure shows the agent's offer curve: the set of demanded consumptions $\boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p} \cdot \boldsymbol{\omega})$, as the price vector $\boldsymbol{p}$ varies. The left panel is the traditional (rational) agent's offer curve. The right panel is the behavioral agent's offer curve (in gray): it is a 2-dimensional surface.

Proposition 6.8 (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:*

$$dp_i^{[s]} = \frac{dp_i^{[r]}}{m_i}, \tag{68}$$

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.*[58]

6.2.3 Behavioral Edgeworth box: Extra-dimensional offer curve

We move on to the Edgeworth box. Take a consumer with endowment $\boldsymbol{\omega} \in \mathbb{R}^n$. Given a price vector $\boldsymbol{p}$, his wealth is $\boldsymbol{p} \cdot \boldsymbol{\omega}$, and so his demand is $\boldsymbol{D}(\boldsymbol{p}) := \boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p} \cdot \boldsymbol{\omega}) \in \mathbb{R}^n$. The offer curve OC is defined as the set of demands, as prices vary: $OC := \left\{\boldsymbol{D}(\boldsymbol{p}) : \boldsymbol{p} \in \mathbb{R}^n_{++}\right\}$.[59]

Let us start with two goods ($n = 2$). The left panel of Figure 3 is the offer curve of the rational consumer: it has the traditional shape. The right panel plots the offer curve

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

[59] One can imagine in the background a sequence of i.i.d. economies with a stochastic aggregate endowment, as in section 6.2.4. That would generate the average price (hence a default price), and a variability of prices (which will lead to the allocation of attention).

of a behavioral consumer with the same basic preferences: the offer curve is the gray area. The offer curve has acquired an extra dimension, compared to the one-dimensional curve of the rational consumer. The OC is a now two-dimensional “ribbon”, with a pinch at the endowment; if mistakes are unbounded, the OC is the union of quadrants north-west or south-east of $\boldsymbol{\omega}$.[60]

What is going on here? In the traditional model, the offer curve is one-dimensional: as demand $\boldsymbol{D}(\boldsymbol{p}) = \boldsymbol{c}(\boldsymbol{p}, \boldsymbol{p} \cdot \boldsymbol{\omega})$ is homogeneous of degree 0 in $\boldsymbol{p} = (p_1, p_2)$, only the relative price p_1/p_2 matters. However, in the behavioral model, demand $\boldsymbol{D}(\boldsymbol{p})$ is not homogeneous of degree 0 in $\boldsymbol{p}$ any more: this is the nominal illusion of Proposition 6.2. Hence, the offer curve is effectively described by two parameters (p_1, p_2) (rather than just their ratio), so it is 2-dimensional (the online appendix has a formal proof in section XII).[61] Note that this holds even though the Marshallian demand is a nice, single-valued function.

In the traditional model, equilibria are the intersection of offer curves. However, this is typically not the case here, as we shall now see.

6.2.4 A Phillips curve in the Edgeworth box

In the traditional model with one equilibrium allocation, the set of equilibrium prices $\mathcal{P}^*$ is one-dimensional ($\mathcal{P}^* = \{\chi\overline{\boldsymbol{p}} : \chi \in \mathbb{R}_{++}\}$), and $\mathcal{C}^a$ is just a point, $\boldsymbol{D}^a(\overline{\boldsymbol{p}})$.[62]

In the behavioral setup, $\mathcal{P}^*$ is still one-dimensional.[63] However, *to each equilibrium price level corresponds a different real equilibrium.* This is analogous to a “Phillips curve”: $\mathcal{C}^a$ has dimension 1.

To fix ideas, it is useful to consider the case of one rational consumer and one behavioral consumer.

Proposition 6.9 *Suppose agent a is rational, and the other agent is behavioral with $m_1 = 1, m_2 = 0$, and two goods. The set $\mathcal{C}^a$ of a's equilibrium allocations is one-dimensional: it is equal to a's offer curve.*

Suppose we start at a middle point of the curve in Figure 4, right panel. Suppose for concreteness that consumer b is a worker, good 2 is food, and good 1 is “leisure,” so that when he consumes less of good 1, he works more. Let us say that $m_1 > m_2$; he pays keen attention to his nominal wage, p_1, and less to the price of food, p_2. Suppose now that

[60] A point $\boldsymbol{c}$ in the OC must be in the two quadrants north-west or south-east of $\boldsymbol{\omega}$ (otherwise, we would have $\boldsymbol{c} \ll \boldsymbol{\omega}$ or $\boldsymbol{c} \gg \boldsymbol{\omega}$; however, there is a $\boldsymbol{p}$ s.t. $\boldsymbol{p} \cdot \boldsymbol{c} = \boldsymbol{p} \cdot \boldsymbol{\omega}$: a contradiction).

[61] This “2-dimensional offer curve” is distinct from the previously-known “thick indifference curve”. The latter arises when the consumer violates strict monotonicity (i.e. likes equally 5.3 and 5.4 bananas), is not associated to any endowment or prices, and has no pinch. The behavioral offer curve, in contrast, arises from nominal illusion, needs an endowment and prices, and has a pinch at the endowment.

[62] More generally, equilibria consist of a finite union of such sets, under weak conditions given in Debreu (1970).

[63] By Walras' law, $\mathcal{P}^* = \{\boldsymbol{p} : \boldsymbol{Z}_{-n}(\boldsymbol{p}) = 0\}$, where $\boldsymbol{Z}_{-n} = (\boldsymbol{Z}_i)_{1 \le i < n}$. As $\boldsymbol{Z}_{-n}$ is a function $\mathbb{R}^n_{++} \to \mathbb{R}^{n-1}$, $\mathcal{P}^*$ is generically a one-dimensional manifold.

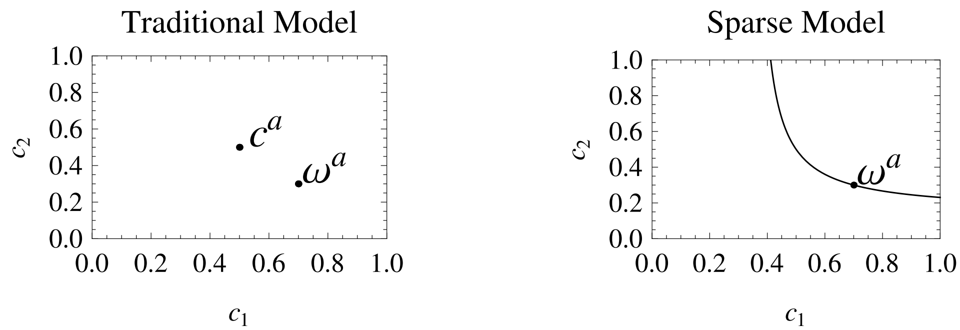

Figure 4: These Edgeworth boxes show competitive equilibria when both agents have Cobb-Douglas preferences. The left panel illustrates the traditional model with rational agents: there is just one equilibrium, $\boldsymbol{c}^a$. The right panel illustrates the situation when type a is rational, and type b is boundedly rational: there is a one-dimensional continuum of competitive equilibria (one for each price level) – a "Phillips curve." Agent a's share of the total endowment (ω^a) is the same in both cases.

the central bank raises the price level. Then, consumer b sees that his nominal wage has increased, and sees less clearly the increase in the price of good 2. So he perceives that his real wage ($\frac{p_1}{p_2}$) has increased. Hence (under weak assumptions) he supplies more labor: i.e., he consumes less of good 1 (leisure) and more of good 2. Hence, the central bank, by raising the price level, has shifted the equilibrium to a different point.

Is this Phillips curve something real and important? This question is debated in macroeconomics, with an affirmative answer from New Keynesian analyses (Galí 2011). Standard macro deals with one equilibrium, conditioning on the price level (and its expectations). To some extent, this is what we have here. Given a price level, there is (locally) only one equilibrium (as in Debreu 1970), but changes in the price level change the equilibrium (when there are some frictions in the perception or posting of prices). This is akin to a (temporary) Phillips curve: when the price level goes up, the perceived wage goes up, and people supply more labor. Hence, we observe here the price-level dependent equilibria long theorized in macro, but in the pristine and general universe of basic microeconomics. One criticism of the influential Lucas (1972) view is that inflation numbers are in practice very easy to obtain, contrary to Lucas' postulate. This criticism does not apply here: behavioral agents actively neglect inflation numbers, which means the Phillips curve effect is valid even when information is readily obtainable.

6.3 What is robust in basic microeconomics?

I gather 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.

Propositions that are not robust

Tradition: There is no money illusion. *Behavioral* model: There is money illusion: when the budget and prices are 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 offer curve is one-dimensional in the Edgeworth box. *Behavioral* model: It is typically a two-dimensional pinched ribbon.[64]

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

Tradition: The Slutsky matrix is the second derivative of the expenditure function. *Behavioral* model: They are linked in a different way.

Tradition: The Slutsky matrix is negative semi-definite. The weak axiom of revealed preference holds. *Behavioral* model: These properties generally fail in a psychologically interpretable way.

Small robustness: Propositions that hold at the default price, but not away from it, to the first order

Marshallian and Hicksian demands, Shephard's lemma and Roy's identity: the values of the underlying objects are the same in the *traditional* and *behavioral* model at the default price,[65] but differ (to the first order in $\boldsymbol{p} - \boldsymbol{p}^d$) away from the default price. This leads to a U-shape of errors in welfare assessment (in an analysis that does not take into account bounded rationality) as a function of consumer sophistication, because the econometrician would mistake a low elasticity due to inattention for a fundamentally low elasticity.

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.

[64] When the prices of the two goods change, in the traditional model only their ratio matters. So there is only one free parameter. However, as a behavioral agent exhibits some nominal illusion, both prices matter, not just their ratio, and we have a two-dimensional curve.

[65] The default price is the price expected by a fully inattentive agent.

The values of the expenditure function $e(\boldsymbol{p}, u)$ and indirect utility function $v(\boldsymbol{p}, w)$ are the same, under the *traditional* and *behavioral* models, up to second order terms in the price deviation from the default $(\boldsymbol{p} - \boldsymbol{p}^d)$.[66]

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 does down, demand for it goes up", or more generally "if there's a good incentive to do X, people will indeed tend to do X,"[67,68] Those sign predictions make intuitive sense, and, not coincidentally, they hold in the behavioral model:[69] those sign predictions (unlike quantitative predictions) remained 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.

7 Allocation of Attention over Time

The models so far were static. We now move on to models that discuss the allocation of attention over time. One important theme is that on impact with a new piece of information, people are quite inattentive, but over time people 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

[66] The above points about second-order losses are well-known (Akerlof and Yellen 1985), and are just a consequence of the envelope theorem. I mention them here for completeness.

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

[68] 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.

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

by Duffie and Sun (1990). Sticky information has been advocated in macro by Gabaix and Laibson (2002), Carroll (2003), and then by Mankiw and Reis (2002), Reis (2006a), 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

$$V = \sum_{t=0}^{\infty} \beta^t u(a_t, x_t) \tag{69}$$

$$u(a, x) = -\frac{1}{2}(a - x)^2 \tag{70}$$

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

$$x_{t+1} = \rho x_t + \varepsilon_{t+1} \tag{71}$$

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 (like 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 capture 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

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

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

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

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

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

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

with

$$F(T) = 1-\theta^{T+1} \tag{72}$$

in the Calvo model; and

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

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:

$$\bar{a}_t^{\,A} = \theta\bar{a}_{t-1}^{A} + (1-\theta)x_t \tag{73}$$

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

$$v(a_{t,}, a_{t-1}, x_t) := u\left(\frac{a_t - ha_{t-1}}{1-h}, x_t\right) \tag{74}$$

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 (69), but with no frictions except for habit:

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

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

$$a_t = ha_{t-1} + (1-h)x_t \tag{75}$$

which is exactly an AR(1) process, like (73), 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 as 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 to 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 default matters a great deal (Bernheim, Fradkin, and Popov 2015).

Which is true? Most macroeconomists, privately, acknowledge that habits are basically just a device to generate stickiness. Still, scientific evidence would be nice. 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

$$a_t - x_t + \kappa (a_t - a_{t-1}) - \beta\kappa (\mathbb{E}_t a_{t+1} - a_t) = 0 \tag{76}$$

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

$$a_t = \theta a_{t-1} + (1-\theta) x_t \tag{77}$$

where θ solves $\theta = \frac{\kappa}{\kappa+1+\beta\kappa(1-\theta)}$.[70]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.

[70]Proof: we use (77), which gives $\mathbb{E}a_{t+1} - a_t = (1-\theta)(x_t - a_t)$. Then, we plug it into 76, using

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

which gives the expression for θ.

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

Both inattention and habits / adjustment costs create delayed reaction. But we can differentiate between them as follows. Inattention (of the sticky information kind) creates *source-specific* under-reaction, whereas adjustment costs (or the sticky action model) create *uniform* under-reaction. Let us see this in a one-period model. Inattention creates an action:

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

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

$$a = m a^r + (1 - m)\, a_{-1} \tag{78}$$

with $m = \frac{1}{1+\kappa}$. This is a uniform dampening, across all dimensions i. Likewise, a habit

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

creates the same expression (78) for the action, this time with $m = 1 - h$.

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:

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

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

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

which is the average consumption loss plus a penalty for the average cost of looking up information.[71]

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:

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

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

Now, we calculate[72]

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

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

$$\frac{1}{D} = \frac{\sigma}{2\sqrt{\kappa}} \tag{79}$$

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

$$1 - \theta = \min\left(\frac{\sigma}{\sqrt{2\kappa}}, 1\right). \tag{80}$$

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

[72] 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 $1 - \theta$.

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

$$\frac{1}{D_i} = \frac{\sigma_i}{2\sqrt{\kappa_i}}. \tag{81}$$

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. 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 at all 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 Slows 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), 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 2016). For instance, Gabaix (2016b) presents a behavioral version of the textbook New Keynesian model, which gives a way to model monetary and fiscal policy with behavioral agents. We can expect this literature to grow in the future.

8 Open Questions and Conclusion

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

We need more measures of inattention This survey showed a number of measure of attention (Section 3.3). 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. In particular, can we relate the "physical measures of inputs to attention" (e.g. eye-tracking) to attention itself (see Section 3.1.3)?

Investigating Varian in the lab Let us ponder the difference between a physics textbook and a microeconomics textbook. In physics textbooks, assertions and results (e.g. force = mass times acceleration) have been verified exquisitely in the lab. Not so in economics. You open, say, Varian (1992) or Mas-Colell, Whinston, and Green (1995), and see 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. It would be great to make economics more like physics. To do so, it seems important to experimentally investigate basic microeconomics (à la Varian 1992) in the lab. The material in Section 6 gives 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 a procedure to investigate its predictions empirically.

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 naturalistic way the basic problem (60). The rewards would be very nice, as we'd have a worked-out and tested counterpart of basic microeconomics.

When you look at Mas-Colell, Whinston, and Green (1995), you see that a few chapters have been extensively investigated (for example, expected utility, with prospect theory as a 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 / Arrow-Debreu style, have been investigated very little – as a result, I think, of the lack of a clean behavioral alternative, a gap that is now filled (see Section 6). Such a study would be drier than, say, work on discrimination or fairness, but useful for economics.[73] Hopefully that imbalance will be corrected.

We need more experimental evidence on the determinants of attention There are now several theories of attention, but measurement is somewhat lagging in refinement. 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 attention, as in equation (55), 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 (33). 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 structurally models of inattention, to see if the predictions do fit, and perhaps suggest newer models.

[73]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.

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.

More work on the consequences of inattention Work on the consequences of inattention for markets outcomes and public policy will continue.

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 $m = \frac{\text{Cov}(x,s)}{\text{Var}(s)} = \frac{v_x}{v_s}$, with $v_x = \sigma_x^2$ and $v_\varepsilon = \sigma_\varepsilon^2$. 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

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

so a is distributed as:

$$a = mx + \sqrt{m(1-m)}\eta_x \tag{82}$$

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

Derivation of the losses from inattention (equation 27) 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:

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

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 identify 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 (27) by taking expectations.

Derivation of the entropy of Gaussian variables (Section 5.2.1) 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 5.2.1) 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

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

and using (51) gives

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

Proof of Proposition 6.1 From Definition 4.2, 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 6.3 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 6.5 The part $\frac{\partial \boldsymbol{c}^s}{\partial w} = \frac{\partial \boldsymbol{c}^r}{\partial w}$ follows from Proposition 6.1: 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 6.3 imply (65).

Proof of Proposition 6.8 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 (71) 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

$$a_{t,0}^A = \left(1 - \beta\theta\right) \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \theta^s \mathbb{E}_t \left[x_{t+s}\right] . \tag{83}$$

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

$$a_{t,0}^A = \frac{1-\beta\theta}{1-\beta\theta\rho} x_t . \tag{84}$$

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

$$a_{t,s}^I = \mathbb{E}_{t-s} \left[x_t\right] . \tag{85}$$

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%. The estimate for the attention parameter m is as reported by the authors for the standard-tax sample.

- 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.

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.

Akerlof, George A., Yellen, Janet L., 1985. Can small deviations from rationality make significant differences to economic equilibria? The American Economic Review 75 (4), 708–720.

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.

Anagol, Santosh, Kim, Hugh Hoikwang, 2012. The impact of shrouded fees: Evidence from a natural experiment in the Indian mutual funds market. The American Economic Review 102 (1), 576–593.

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.

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.

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, 2015. Competition for attention. The Review of Economic Studies 83 (2), 481–513.

Bordalo, Pedro, Gennaioli, Nicola, Shleifer, Andrei, 2016. 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.

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, 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.

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.

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.

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.

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.

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.

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.

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.

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

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, 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.

Friedman, Milton, 1961. The lag in effect of monetary policy. Journal of Political Economy 69 (5), 447–466.

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, 2016a. Behavioral macroeconomics via sparse dynamic programming. NBER Working Paper No. 21848.

Gabaix, Xavier, 2016b. 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.

Galí, Jordi, 2011. The return of the wage phillips curve. Journal of the European Economic Association 9 (3), 436–461.

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.

Gennaioli, Nicola, Shleifer, Andrei, Vishny, Robert, 2012. Neglected risks, financial innovation, and financial fragility. Journal of Financial Economics 104 (3), 452–468.

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.

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.

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.

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., 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.

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.

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.

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.

Jung, Junehyuk, Kim, Jeong-Ho, Matejka, 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, 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.

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.

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, 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.

Kreps, David M., 2012. Microeconomic Foundations I: Choice and Competitive Markets. Princeton University Press.

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.

Lucas, Robert E., 1972. Expectations and the neutrality of money. Journal of Economic Theory 4 (2), 103–124.

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.

Mas-Colell, Andreu, Whinston, Michael Dennis, Green, Jerry R, 1995. Microeconomic Theory. Oxford University Press, New York.

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.

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.

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.

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.

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.

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.

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.

Spiegler, Ran, 2011. Bounded Rationality and Industrial Organization. Oxford University Press.

Stango, Victor, Zinman, Jonathan, 2009. Exponential growth bias and household finance. The Journal of Finance 64 (6), 2807–2849.

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.

Styles, Elizabeth, 2006. The Psychology of Attention. Psychology Press.

Taubinsky, Dmitry, Rees-Jones, Alex, 2017. Attention variation and welfare: Theory and evidence from a tax salience experiment. NBER Working Paper No. 22545.

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.

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.

Woodford, Michael, 2012. Prospect theory as efficient perceptual distortion. The American Economic Review Papers and Proceedings 102 (3), 41–46.

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.