

A Sparsity-Based Model of Bounded Rationality*

Xavier Gabaix

May 8, 2014

Abstract

This paper defines and analyzes a “sparse max” operator, which is a less than fully attentive and rational version of the traditional max operator. The agent builds (as economists do) a simplified model of the world which is sparse, considering only the variables of first-order importance. His stylized model and his resulting choices both derive from constrained optimization. Still, the sparse max remains tractable to compute. Moreover, the induced outcomes reflect basic psychological forces governing limited attention.

The sparse max yields a behavioral version of two basic chapters of the microeconomics textbook: consumer demand and competitive equilibrium. We obtain a behavioral version of Marshallian and Hicksian demand, the Slutsky matrix, the Edgeworth box, Roy’s identity etc. The Slutsky matrix is no longer symmetric: non-salient prices are associated with anomalously small demand elasticities. Because the consumer exhibits nominal illusion, in the Edgeworth box, the offer curve is a two-dimensional surface rather than a one-dimensional curve. As a result, different aggregate price levels correspond to materially distinct competitive equilibria, in a similar spirit to a Phillips curve. This framework provides a way to assess which parts of basic microeconomics are robust, and which are not, to the assumption of perfect maximization.

*xgabaix@stern.nyu.edu. I thank David Laibson for a great many enlightening conversations about behavioral economics over the years. For very helpful comments, I thank the editor and the referees, and Andrew Abel, Nick Barberis, Daniel Benjamin, Douglas Bernheim, Andrew Caplin, Pierre-André Chiappori, Vincent Crawford, Stefano DellaVigna, Alex Edmans, Ed Glaeser, Oliver Hart, David Hirshleifer, Harrison Hong, Daniel Kahneman, Paul Klemperer, Botond Kőszegi, Sendhil Mullainathan, Matthew Rabin, Antonio Rangel, Larry Samuelson, Yuliy Sannikov, Thomas Sargent, Josh Schwartzstein, and participants at various seminars and conferences. I am grateful to Jonathan Libgober, Elliot Lipnowski, Farzad Saidi, and Jerome Williams for very good research assistance, and to the NYU CGEB, INET, the NSF (grant SES-1325181) for financial support.

1 Introduction

This paper proposes a tractable model of some dimensions of bounded rationality (BR). It develops a “sparse max” operator, which is a behavioral version of the traditional “max” operator, and applies to general problems of maximization under constraint.¹ In the sparse max, the agent pays less or no attention to some features of the problem, in a way that is psychologically founded. I use the sparse max to propose a behavioral version of two basic chapters of the economic textbooks: consumer theory (problem 1) and basic equilibrium theory.

The principles behind the sparse max are the following. First, the agent builds a simplified model of the world, somewhat like economists do, and thinks about the world through this simplified model. Second, this representation is “sparse,” i.e., uses few parameters that are non-zero or differ from the usual state of affairs. These choices are controlled by an optimization of his representation of the world that depends on the problem at hand. I draw from fairly recent literature on statistics and image processing to use a notion of “sparsity” that still entails well-behaved, convex maximization problems (Tibshirani (1996), Candès and Tao (2006)). The idea is to think of “sparsity” (having lots of zeroes in a vector) instead of “simplicity” (which is an amorphous notion), and measure the lack of “sparsity” by the sum of absolute values. This paper follows this lead to use sparsity notions in economic modelling, and to the best of my knowledge is the first to do so.²

“Sparsity” is also a psychologically realistic feature of life. For any decision, in principle, thousands of considerations are relevant to the agent: his income, but also GDP growth in his country, the interest rate, recent progress in the construction of plastics, interest rates in Hungary, the state of the Amazonian forest, etc. Since it would be too burdensome to take all of these variables into account, he is going to discard most of them. The traditional modelling for this is to postulate a fixed cost for each variable. However, that often leads to discontinuous reactions and intractable problems (fixed costs, with their non-convexity, are notoriously ill-behaved). In contrast, the notion of sparsity I use leads to continuous reactions and problems that are easy to solve.

The model rests on very robust psychological notions. It incorporates limited attention, of course. To supply the missing elements due to limited attention, people rely on defaults – which are typically the expected values of variables. At the same time, attention is allocated purposefully, towards features that seem important. When taking into account some information, agents anchor on the default and do a limited adjustment towards the truth, as in Tversky and Kahneman’s (1974) “anchoring and adjustment”.³

¹The meaning of “sparse” is that of a sparse vector or matrix. For instance, a vector $m \in \mathbb{R}^{100,000}$ with only a few non-zero elements is sparse. In this paper, the vector of things the agent considers is (endogenously) sparse.

²Econometricians have already successfully used sparsity (e.g. Belloni and Chernozhukov 2011).

³In models with noisy perception, an agent optimally responds by shading his noisy signal, so that he optimally underreacts (conditionnally on the true signal). Hence, he behaves on average as he misperceives the truth – indeed, perceives only a fraction of it. The sparsity model displays this “partial adjustment” behavior even though it is deterministic (see Proposition 15). The sparse agent is in part a deterministic “representative agent” idealization of such an agent with noisy perception.

If the agent is confused about prices, how is the budget constraint still satisfied? I propose a way to incorporate maximization under constraint (building on Chetty, Kroft and Looney (2007)), in a way that keep the model plausible and tractable.

After the sparse max has been defined, I apply it to write a behavioral version of textbook consumer theory and competitive equilibrium theory. By consumer theory, I mean the optimal choice of a consumption bundle subject to a budget constraint:

$$\max_{c_1, \dots, c_n} u(c_1, \dots, c_n) \text{ subject to } p_1 c_1 + \dots + p_n c_n \leq w. \quad (1)$$

There does not appear to be any systematic treatment of this building block with a limited rationality model other than sparsity in the literature to date.⁴

One might think that there is little to add to such an old and basic topic. However, it turns out that (sparsity-based) limited rationality leads to enrichments that may be both realistic and intellectually intriguing. I assume that agents do not fully pay attention to all prices. The sparse max determines how much attention they pay to each price, and how they adjust their budget constraint.

The agent exhibits a form of nominal illusion. If all prices and his budget increase by 10%, say, the consumer does not react in the traditional model. However, a sparse agent might perceive the price of bread did not change, but that his nominal wage went up. Hence, he supplies more labor. In a macroeconomic context, this leads to a “Phillips curve”.

The Slutsky matrix is no longer symmetric: non-salient prices will lead to small terms in the matrix, breaking symmetry. I argue below that indeed, the extant evidence seems to favor the effects theorized here. In addition, the model offers a way to recover quantitatively the extent of limited attention.

We can also revisit the venerable Edgeworth box, and meet its younger cousin, the “behavioral Edgeworth box”. In the traditional Edgeworth box, the offer curve is, well, a curve: a one-dimensional object.⁵ However, in the sparsity model, it becomes a two-dimensional object (see Figure 3).⁶ This is again because of nominal illusion displayed by a sparse agent.

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.⁷ 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.

⁴Dufwenberg et al. (2011) analyze competitive equilibrium with other-regarding, but rational, preferences.

⁵Recall that the “offer curve” of an agent is the set of consumption bundles he chooses as prices change (those price changes also affecting the value of his endowment).

⁶This notion is very different from the idea of a “thick indifference curve”, in which the consumer is indifferent between dominated bundles. A sparse consumer has only a thin indifference curve.

⁷The paper discusses the empirical relevance and underlying conditions for the deviations expressed here.

Propositions that are not robust

Tradition: There is no money illusion. *Sparse* 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 $c(\mathbf{p}, w)$ is not homogeneous of degree 0.

Tradition: The Slutsky matrix is symmetric. *Sparse* 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. *Sparse* model: It is typically a two-dimensional pinched ribbon.⁸

Tradition: The competitive equilibrium allocation is independent of the price level. *Sparse* 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. *Sparse* model: They are linked in a different way.

Tradition: The Slutsky matrix is negative semi-definite. The weak axiom of revealed preference holds. *Sparse* 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 *sparse* model at the default price,⁹ but differ (to the first order in $\mathbf{p} - \mathbf{p}^d$) away from the default price. That leads to a U-shape of errors in welfare assessment (in an analysis that would 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. *Sparse* model: people maximize in default situations, but there are losses away from it.

Tradition: Competitive equilibrium is efficient. *Sparse* model: it is efficient if it happens at the default price. Away from the default price, competitive equilibrium has inefficiencies, unless people have the same misperceptions.

The values of the expenditure function $e(\mathbf{p}, u)$ and indirect utility function $v(\mathbf{p}, w)$ are the same, under the *traditional* and *sparse* models, up to second order terms in the price deviation from the default $(\mathbf{p} - \mathbf{p}^d)$.¹⁰

⁸When 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 sparse agent exhibits some nominal illusion, both prices matter, not just their ratio, and we have a two-dimensional curve.

⁹The default price is the price expected by a fully inattentive agent.

¹⁰The above points about second-order losses are well-known (Akerlof and Yellen 1985), and are just a consequence

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,”¹¹ Those sign predictions make intuitive sense, and, not coincidentally, they hold in the sparse model:¹² 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.

In addition, I work out one consequence of inattention: “fiscal illusion”. A sparse employee prefers a tax increase to be levied on the employer, rather than on himself – contradicting a basic result of public finance that the division does not matter. This is due to his (endogenous) neglect of the general equilibrium effect of a labor tax. More generally, agents will perceive direct effects more readily than indirect, general equilibrium effects. Buchanan (1967) argues that this fiscal illusion is a cause of dysfunction in political decisions. The online appendix sketches other applications, in particular to behavioral biases. Those applications might be best expanded in future research.

This research builds on prior insights on the modelling of costly attention, including reference points, salience, and costly information: they will be extensively reviewed below. The main methodological contribution here is to provide a tractable model that applies quite generally, so that hitherto too difficult problems (including maximization under smooth constraints) can be handled.

The limitations of sparse max will be clear below (and remedies suggested). One point that should be kept in mind:

*The sparse max is, for now, the only available modelling technology that is able to handle the basic consumption problem (1) – and a fortiori to handle general problems of constrained maximization. Other modelling technologies fail to apply, or are too complex to apply to (1).*¹³

Some modelling technologies *fail to apply*. For instance, the “near rational” approach says that agents will lose at most ε utils: it is often useful (Akerlof and Yellen 1985, Chetty 2012), but it does not offer a precise model of which actions people will take. Another approach says that information is updated slowly (e.g. Gabaix and Laibson 2002, Mankiw and Reis 2002). But it relies on the crutches of time, so it does not apply when all actions are taken in one period.

of the envelope theorem. I mention them here for completeness.

¹¹This 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.

¹²The closely related notion of strategic complements and substitutes (Bulow, Geanakoplos and Klemperer 1985) is also robust to a sparsity deviation.

¹³Echenique, Golovin and Wierman (2013) analyze consumer demand with indivisible goods. They show that a boundedly rational model is equivalent to a rational model with a different utility – which is not the case here (Proposition 6). A key reason is that indivisible goods prevent the existence of a Slutsky matrix.

Other technologies appear to be *too complicated* to handle the consumption problem tractably. For instance, “thinking as rational payment of fixed costs” leads to intractable calculations when applied to general problems¹⁴, and doesn’t allow for partial inattention. “Bayesian inference based on noisy signals” (Sims 2003, Veldkamp 2011) leads to a variety of nice insights, but is quite intractable in most cases, and doesn’t allow for source-independent inattention. Again, a plain problem like (1), with its general utility function would lead to formidable computations – and indeed has never been attacked by this strand of literature.¹⁵ There are also differences of substance, discussed in section 6.2.

The plan of the paper is as follows. Section 2 defines the sparse max and analyzes it. It also discusses its psychological underpinnings. Section 3 develops consumer theory, and section 4 analyzes competitive equilibrium theory. Section 5 provide additional information on the sparse max, e.g. how it respects min-max duality and is invariant to rescaling. Section 6 discusses links with existing themes in behavioral and information economics. Section 7 presents concluding remarks. Many proofs are in the appendix or the online appendix, which contains extensions and other applications.

2 The Sparse Max Operator

The agent faces a maximization problem which is, in its traditional version, $\max_a u(a, x)$ subject to $b(a, x) \geq 0$, where u is a utility function, and b is a constraint. I want to define the “sparse max” operator:

$$\operatorname{smax}_a u(a, x) \text{ subject to } b(a, x) \geq 0, \tag{2}$$

which is a less than fully attentive version of the “max” operator. Variables a , x and function b have arbitrary dimensions.¹⁶

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, $(a, x) = (a^d, 0)$.

¹⁴They are “NP-complete” problems. To get an intuitive sense of that, suppose that each of the n prices can be examined by paying a fixed cost. There are 2^n ways to allocated those fixed costs.

¹⁵However, if that study could be performed, I suspect that it would find many insights similar to those offered by the present analysis. To generate broad forces, the modelling specifics do not matter, though those specifics do matter a lot in terms of tractability.

¹⁶We shall see that parameters will be added in the definition of sparse max.

2.1 The Sparse Max: First, Without Constraints

For clarity, we shall first define the sparse max without constraints, i.e. study $\text{smax}_a u(a, x)$. To fix ideas, take the following quadratic example:

$$u(a, x) = -\frac{1}{2}\left(a - \sum_{i=1}^n \mu_i x_i\right)^2. \quad (3)$$

Then, the traditional optimal action is

$$a^r(x) = \sum_{i=1}^n \mu_i x_i, \quad (4)$$

(r like in the traditional *r*ational actor model). For instance, to choose a , the decision maker should consider not only innovations x_1 in his 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 $n > 10,000$ (say) factors that should in principle be taken into account. A sensible agent will “not think” about most of factors, especially the small ones. We will formalize that notion.

We define the perceived representation of x_i as:

$$x_i^s := m_i x_i, \quad (5)$$

where $m_i \in [0, 1]$ is the attention to x_i . When $m_i = 0$, the agent “does not think about x_i ”, i.e. replaces x_i by $x_i^s = 0$; when $m_i = 1$, he perceives the true value ($x_i^s = x_i$). We call $m = (m_i)_{i=1\dots n}$ the attention vector.

After attention m is chosen, the sparse agent optimizes under his simpler representation of the world, i.e. choose $a^s = \arg \max_a u(a, x^s) = \sum_{i=1}^n \mu_i x_i^s$.

Attention creates a psychic cost, parametrized as $g(m) = \kappa m^\alpha$ for $\alpha \geq 0$. The case $\alpha = 0$ corresponds to a fixed cost κ paid each time m is non-zero. Parameter $\kappa \geq 0$ is a penalty for lack of sparsity. If $\kappa = 0$, the agent is the traditional, rational agent model.

The agent takes the x to be drawn from a distribution where $\sigma_{ij} = \mathbb{E}[x_i x_j]$ and $\mathbb{E}[x_i] = 0$.¹⁷ The expected size of x_i is $\sigma_i = \mathbb{E}[x_i^2]^{1/2}$. 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 the default action and parameter, i.e. at $(a, x) = (a^d, 0)$. I next define the sparse max.

Definition 1 (Sparse max operator without constraints). *The sparse max, $\text{smax}_{a|\kappa, \sigma} u(a, x)$, is defined by the following procedure.*

¹⁷This perceived covariance could be the objective one, or, in some applications, an (endogenously) “sparsified” covariance, where most correlations are 0.

Step 1: Choose the attention vector m^* :

$$m^* = \arg \min_{m \in [0,1]^n} \frac{1}{2} \sum_{i,j=1 \dots n} (1 - m_i) \Lambda_{ij} (1 - m_j) + \kappa \sum_{i=1 \dots n} m_i^\alpha, \quad (6)$$

with the cost-of-inattention factors $\Lambda_{ij} := -\sigma_{ij} a_{x_i} u_{aa} a_{x_j}$. Define $x_i^s = m_i^* x_i$, the sparse representation of x .

Step 2: Choose the action

$$a^s = \arg \max_a u(a, x^s), \quad (7)$$

and set the resulting utility to be $u^s = u(a^s, x)$.

The Appendix describes a microfoundation for sparse max, via costs and benefit of thinking for m . Here are the highlights. In (6), the agent solves for the attention m^* that trades off a proxy for the utility losses (the first term in the right-hand side, which is the leading term in the Taylor expansion of utility losses from imperfect attention) and a psychological penalty for deviations from a sparse model (the second term on the left-hand side of 6). Then, in (7), the agent maximizes over the action a , taking the perceived parameter x^s at face value. The problem may appear complex, but we shall see that the sparse max is actually quite simple to use.

The attention function To build some intuition, let us start with the case with just one variable, $x_1 = x$. Then, problem (6) becomes: $\min_m \frac{1}{2} (m - 1)^2 \sigma^2 + \kappa |m|^\alpha$. Attention is $m = \mathcal{A}_\alpha \left(\frac{\sigma^2}{\kappa} \right)$, where the ‘‘attention function’’ \mathcal{A}_α is defined as¹⁸

$$\mathcal{A}_\alpha (\sigma^2) := \inf \left[\arg \min_m \frac{1}{2} (m - 1)^2 \sigma^2 + |m|^\alpha \right].$$

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

We now state sparse max in a leading special case.

Proposition 1 *Suppose that agent views the x_i ’s as uncorrelated with standard deviation σ_i . Then, the perceived x_i^s is:*

$$x_i^s = x_i \mathcal{A}_\alpha \left(\frac{\sigma_i^2 |a_{x_i} u_{aa} a_{x_i}|}{\kappa} \right), \quad (8)$$

where $a_{x_i} = -u_{aa}^{-1} u_{ai}$ is the traditional marginal impact of a small change in x_i , evaluated at $x = 0$. The action is $a^s = \arg \max_a u(a, x^s)$.

¹⁸That is: $\mathcal{A}_\alpha (\sigma^2)$ is the value of m that minimizes $\frac{1}{2} (m - 1)^2 \sigma^2 + |m|^\alpha$ (as conveyed by the $\arg \min$), taking the lowest $m \geq 0$ if there are multiple minimizers (as conveyed by the \inf).

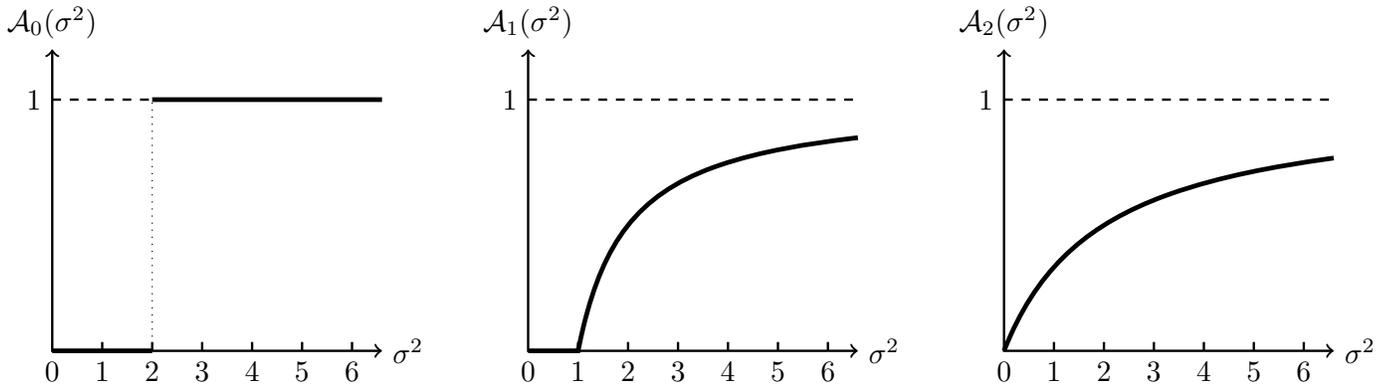


Figure 1: Three attention functions $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$, corresponding to fixed cost, linear cost and quadratic cost respectively. We see that \mathcal{A}_0 and \mathcal{A}_1 induce sparsity – i.e. a range where attention is exactly 0. \mathcal{A}_1 and \mathcal{A}_2 induce a continuous reaction function. \mathcal{A}_1 alone induces sparsity and continuity.

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 (8) 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 $|\sigma_i \cdot \frac{\partial a}{\partial x_i}| \leq \sqrt{\frac{\kappa}{|u_{aa}|}}$). 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. He also devotes “some” attention to those important variables, not necessarily paying full attention to them.¹⁹

Let us revisit the initial example.²⁰

Example 1 *In the quadratic loss problem, (3), the traditional and the sparse actions are: $a^r = \sum_{i=1}^n \mu_i x_i$, and*

$$a^s = \sum_{i=1}^n \mathcal{A}_\alpha(\mu_i^2 \sigma_i^2 / \kappa) \mu_i x_i. \quad (9)$$

Proof: We have $a_{x_i} = \mu_i$, $u_{aa} = -1$, so (8) gives $x_i^s = x_i \mathcal{A}_\alpha(\mu_i^2 \sigma_i^2 / \kappa)$. \square

We now explore when a^s indeed induces no attention to many variables.²¹

Lemma 1 (Special status of linear costs). *When $\alpha \leq 1$ (and only then) the attention function $\mathcal{A}_\alpha(\sigma^2)$ induces sparsity: when the variable is not very important, then the attention weight is 0*

¹⁹There is anchoring with partial adjustment, i.e. dampening. This dampening is pervasive, and indeed optimal, in “signal plus noise” models (more on this later).

²⁰Also (with $\alpha = 1$), m has at most $\sum \mu_i^2 \sigma_i^2 / \kappa$ non-zero components (because $m_i \neq 0$ implies $\mu_i^2 \sigma_i^2 \geq \kappa$). Hence, even when x has infinite dimension, m has a finite number of non-zero components, and is therefore sparse (assuming $\mathbb{E}[(a^r)^2] < \infty$).

²¹Lemma 1 has direct antecedents in statistics: the pseudo norm $\|m\|_\alpha = (\sum_i |m_i|^\alpha)^{1/\alpha}$ is convex and sparsity-inducing iff $\alpha = 1$ (Tibshirani (1996)). Hassan and Mertens (2011) also use $\alpha = 1$.

($m = 0$). 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.²²

Let us examine one simple application.

Application to Fiscal illusion: Agents underperceiving general equilibrium effects

In theory, whether a wage tax is paid by the employee or the employer does not matter: only the total tax matters. Economists know this, and so does a rational agent. However, this is at first very counterintuitive to many non-economists: if taxes are to be raised, most employees prefer the employer tax to be raised, rather than the employee tax – “of course”. This leads to host of political controversies about “who should pay the tax” (Kerschbamer and Kirchsteiger 2000). Buchanan (1967) bemoans this misunderstanding of the indirect effect taxes, and calls it “fiscal illusion,” attributing the original theme to John Stuart Mill.

We shall see that a sparse agent behaves like a non-economist in his (lack of) understanding of the theory of tax equivalence (Sausgruber and Tyran (2005) provide supporting experimental evidence). To see this, suppose a tax t^e is paid by the employee, and t^f by the firm, so that the total tax increase is $t = t^e + t^f$. As a result, the (pre-tax) wage will change by an endogenous quantity Δw , and the net (i.e., after-tax) wage received by the employees will change by $\Delta n = -t^e + \Delta w$. The true change in the net wage is $\Delta n = -\beta t$, for a constant β , which implies that only the total tax matters, not the specific part paid by the employer and the employee.²³

The employee is asked to predict his net wage change, which we can write $\Delta n(x_1, x_2) = -x_1 + x_2$, where $x_1 = t^e$ is the direct effect, while $x_2 = \Delta w$ is the indirect effect. Formally, we apply the sparse model to $u(a, x_1, x_2) = -\frac{1}{2}(a - \Delta n(x_1, x_2))^2$. The next statement is proved in the appendix.

Example 2 (“Fiscal illusion”: a sparse agent does not understand employer/employee tax neutrality). Consider a tax increase t , with a share $s \in [0, 1]$ paid by the employee, and the rest paid by the firm. A sparse employee prefers to pay a low share $s < \beta$ rather than a high share $s' > \beta$. The preference is strict for some parameters. In contrast, a rational agent is indifferent between the two shares, as he understands that only the total tax increase matters.

This theme could be developed greatly.²⁴ In general, any policy change will have both a direct

²²The 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.

²³This is detailed in the derivation of Example 2, and implies $\Delta w = -\beta t + t^e$.

²⁴This derivation of a lack of understanding of the employer / employee tax neutrality appears to be new. A few papers provide evidence for a lack of understanding of general equilibrium effects (in their theory part, they posit rather than derive it): see Camerer and Lovallo (1999), Greenwood and Hanson (2013), Dal Bó, Dal Bó and Eyster (2013); see also the more distant literature on strategic interactions discussed in section 6.2.

effect and an indirect (general equilibrium) effect, as agents are induced to change their equilibrium behavior. It is plausible that a less than fully rational agent will find it hard to foresee the indirect effect. This way, he will put too little weight on indirect, general equilibrium effects. For instance, disliking free trade or favoring rent control, often comes from a failure to understand indirect effects, such as cheaper goods, or harder-to-find rentals. The sparse model offers one way to model the naive agent, who is untutored in economics, and endogenously sees the direct effects more easily than the indirect effects.

2.2 Psychological Underpinnings

The model is based on the following very robust psychological facts.

Limited 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 . The model roughly represents that selective use of information. In step 1, the mind contemplates thousands of x_i , and decides which handful it will bring up for conscious examination. Those are the variables with a non-zero m_i . We simplify problems, and can attend to only a few things – this is what sparsity represents.

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. One could say that the choice of “what comes to mind” in Step 1 is a system 1 operation, that (operating in the unconscious background) selects what to bring up to the conscious mind (the attention m). Step 2 is more like a system 2 operation, determining what to choose, given a restricted set of variables actively considered.

Reliance on defaults What guess does one make with no time to think? This is represented by $x = 0$: the variables x are not taken into account when we have no time to think (the Bayesian analogue of the default is the “prior”). This default model ($x = 0$), 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. Carroll et al. 2009).²⁵ 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).

Anchoring and adjustment The mind, in the model, anchors on the default model. Then, it does a full or partial adjustment towards the truth. This is akin to the psychology of “anchoring

²⁵This 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 $x = 0$), hence deviate from the default action.

and adjustment”. There is anchoring on a default value and partial adjustment towards the truth: “People make estimates by starting from an initial value that is adjusted to yield the final answer [...]. Adjustments are typically insufficient” (Tversky and Kahneman, 1974, p. 1129).

The sparse max exhibits anchoring on the default model, and partial adjustment towards the truth, with the attention function \mathcal{A} . It would be interesting to experimentally investigate the \mathcal{A} function – perhaps to refine it. The comparative statics make sense (less important variables are used less). Hence, even though there is no specific experimental evidence regarding the exact value of this function, the extensive psychological evidence qualitatively supports its basic elements.

2.3 Sparse Max: Full Version, Allowing For Constraints

Let us now extend sparse max so that it can handle maximization under $K (= \dim b)$ constraints, problem (2). As a motivation, consider problem (1), $\max_{\mathbf{c}} u(\mathbf{c})$ s.t. $\mathbf{p} \cdot \mathbf{c} \leq w$. We start from a default price \mathbf{p}^d . The new price is $p_i = p_i^d + x_i$, while the price perceived by the agent is $p_i^s(m) = p_i^d + m_i x_i$.²⁶

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 (see section 6.1), 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:²⁷

$$\frac{\partial u / \partial c_1}{\partial u / \partial c_2} = \frac{p_1^s}{p_2^s}, \quad (10)$$

i.e. $u'(\mathbf{c}) = \lambda \mathbf{p}^s$, for some scalar λ .²⁸ The agent will tune λ so that the constraint binds, i.e. the value of $\mathbf{c}(\lambda) = u'^{-1}(\lambda \mathbf{p}^s)$ satisfies $\mathbf{p} \cdot \mathbf{c}(\lambda) = w$.²⁹ Hence, in step 2, the agent “hears clearly” whether the budget constraint binds.³⁰ 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 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 (2) 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

²⁶The constraint is $0 \leq b(\mathbf{c}, \mathbf{x}) := w - (\mathbf{p}^d + \mathbf{x}) \cdot \mathbf{c}$.

²⁷Otherwise, 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.

²⁸This 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.

²⁹If there are several λ , the agent takes the smallest value, which is the utility-maximizing one.

³⁰See footnote 33 for additional intuitive justification.

the constraints to the utility.³¹

Definition 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 (6), 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: $\lambda^ = \arg \max_{\lambda \in \mathbb{R}_+^K} u(a(\lambda), x^s)$ s.t. $b(a(\lambda), x) \geq 0$. (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 non-negative λ^* .) The resulting sparse action is $a^s := a(\lambda^*)$. Utility is $u^s := u(a^s, x)$.*

Step 2 of Definition 2 allows quite generally for the translation of a BR maximum without constraints, into a BR maximum with constraints. It could be reused in other contexts. To obtain further intuition on the constrained maximum, we turn to consumer theory.

3 Textbook Consumer Theory: A Behavioral Update

3.1 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: $\mathbf{c}(\mathbf{p}, w) := \arg \max_{\mathbf{c} \in \mathbb{R}^n} u(\mathbf{c})$ subject to $\mathbf{p} \cdot \mathbf{c} \leq w$, where \mathbf{c} and \mathbf{p} are the consumption vector and price vector. We denote by $\mathbf{c}^r(\mathbf{p}, w)$ the demand under the traditional *rational* model, and by $\mathbf{c}^s(\mathbf{p}, w)$ the demand of a sparse agent.

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 sparse agent is $p_i^s = p_i^d + m_i x_i$, i.e.:

$$p_i^s(m) = m_i p_i + (1 - m_i) p_i^d. \quad (11)$$

When $m_i = 1$, the agent fully perceives price p_i , while when $m_i = 0$, he replaces it by the default price.³²

³¹For instance, in a consumption problem (1), 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.

³²More 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, $\frac{\partial p_i^s}{\partial p_i} \Big|_{p=p^d} = m_i$ (see online appendix).

Proposition 2 (Marshallian demand). *Given the true price vector \mathbf{p} and the perceived price vector \mathbf{p}^s , the Marshallian demand of a sparse agent is*

$$\mathbf{c}^s(\mathbf{p}, w) = \mathbf{c}^r(\mathbf{p}^s, w'), \quad (12)$$

where the as-if budget w' solves $\mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, w') = w$, 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 3 (Demand by a sparse agent with quasi-linear utility). *Take $u(\mathbf{c}) = v(c_1, \dots, c_{n-1}) + c_n$, with v strictly concave. Demand for good $i < n$ is independent of wealth and is: $c_i^s(\mathbf{p}) = c_i^r(\mathbf{p}^s)$.*

In this example, the demand of the sparse 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 is a polar opposite.

Example 4 (Demand proportional to wealth). *When rational demand is proportional to wealth, the demand of a sparse agent is: $c_i^s(\mathbf{p}, w) = \frac{c_i^r(\mathbf{p}^s, w)}{\mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, 1)}$.*

Example 5 (Demand by a sparse Cobb-Douglas agent). *Take $u(\mathbf{c}) = \sum_{i=1}^n \alpha_i \ln c_i$, with $\alpha_i \geq 0$. Demand is: $c_i^s(\mathbf{p}, w) = \frac{\alpha_i}{p_i^s} \frac{w}{\sum_j \alpha_j \frac{p_j}{p_j^s}}$.*

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 $1/\mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, 1) = \frac{100}{101}$). When demand is not necessarily linear in wealth, the adjustment is (to the leading order) proportional to the marginal demand, $\frac{\partial \mathbf{c}^r}{\partial w}$, rather than the average demand, \mathbf{c}^r . The sparse agent cuts “luxury goods”, not “necessities”.³³ Figure 2 illustrates the resulting consumption.³⁴

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.³⁵

³³For 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 a bit his consumption (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.

³⁴It is analogous to a tariff in international trade, where the price distortion is rebated to consumers.

³⁵ λ^d is endogenous, and characterized by $u'(\mathbf{c}^d) = \lambda^d \mathbf{p}^d$, where \mathbf{p}^d is the exogenous default price, and \mathbf{c}^d is the (endogenous) optimal consumption as the default. The comparative statics hold, keeping λ^d constant.

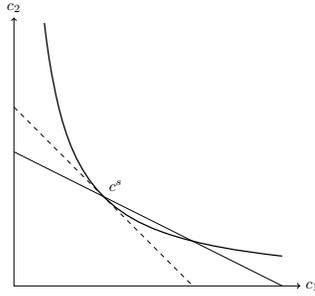


Figure 2: The indifference curve is tangent to the perceived budget set (dashed line) at the chosen consumption \mathbf{c}^s , which also lies on the true budget set (solid line). Parameters: $\ln c_1 + \ln c_2$, $\mathbf{p} = (1, 2)$, $\mathbf{p}^s = (1, 1)$, $w = 3$, $\mathbf{c}^s = (1, 1)$.

Proposition 3 (Attention to prices). *In the basic consumption problem, assuming that price shocks are perceived as uncorrelated, 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 $\left(\frac{\sigma_{p_i}}{p_i^d} \right)$, (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. This allows us to derive the “indirect” consequences of limited attention to prices. More surprises happen here, as we shall now see.

3.2 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. 11).

Proposition 4 *The Marshallian demand $\mathbf{c}^s(\mathbf{p}, w)$ has the marginals (evaluated at $\mathbf{p} = \mathbf{p}^d$): $\frac{\partial \mathbf{c}^s}{\partial w} = \frac{\partial \mathbf{c}^r}{\partial w}$ and*

$$\frac{\partial c_i^s}{\partial p_j} = \frac{\partial c_i^r}{\partial p_j} \times m_j - \frac{\partial c_i^r}{\partial w} c_j^r \times (1 - m_j). \quad (13)$$

This means, as we detail shortly, that income effects $\left(\frac{\partial \mathbf{c}}{\partial w} \right)$ are preserved (as w needs to be spent in this one-shot model), but substitution effects are dampened. One consequence is nominal illusion.

Proposition 5 (Nominal illusion) *Suppose that the agent pays more attention to some goods than others (i.e. the m_i are not all equal). Then, the agent exhibits nominal illusion, i.e. the Marshallian demand $\mathbf{c}(\mathbf{p}, w)$ 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 sparse

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 (\mathbf{p}^d is kept constant). Hence, he consumes less of good 1, and more of good 2. His demand has shifted. In abstract terms, the $\mathbf{c}^s(\chi\mathbf{p}, \chi w) \neq \mathbf{c}^s(\mathbf{p}, w)$ for $\chi = 1.1$, i.e. the Marshallian demand is not homogeneous of degree 0. He 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:

$$S_{ij}(\mathbf{p}, w) := \frac{\partial c_i(\mathbf{p}, w)}{\partial p_j} + \frac{\partial c_i(\mathbf{p}, w)}{\partial w} c_j(\mathbf{p}, w). \quad (14)$$

With the traditional agent, the most surprising fact about it is that it is symmetric: $S_{ij}^r = S_{ji}^r$. 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: “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 et al., it might require too much sophistication from the average consumer. We now present a less rational, and psychologically more intuitive, prediction.

Proposition 6 (Slutsky matrix). *Evaluated at the default price, the Slutsky matrix S^s is, compared to the traditional matrix S^r :*

$$S_{ij}^s = S_{ij}^r m_j, \quad (15)$$

i.e. the sparse demand sensitivity to price j is the rational one, times m_j , the salience of price j . As a result the sparse 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 $|S_{ij}^s|$ is lower than $|S_{ji}^s|$: 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 sparse 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.³⁶

Equation (15) makes tight testable predictions. It allows us to infer attention from choice data, as we shall now see.³⁷

Proposition 7 (Estimation of limited attention). *Choice data allow one to recover the attention vector m , up to a multiplicative factor \bar{m} . Indeed, suppose that an empirical Slutsky matrix S_{ij}^s is available. Then, m can be recovered as $m_j = \bar{m} \prod_i \left(\frac{S_{ij}^s}{S_{ji}^s} \right)^{\gamma_i}$, for any $(\gamma_i)_{i=1\dots n}$ s.t. $\sum_i \gamma_i = 1$.*

Proof: We have $\frac{S_{ij}^s}{S_{ji}^s} = \frac{m_j}{m_i}$, so $\prod_i \left(\frac{S_{ij}^s}{S_{ji}^s} \right)^{\gamma_i} = \prod_i \left(\frac{m_j}{m_i} \right)^{\gamma_i} = \frac{m_j}{\bar{m}}$, for $\bar{m} := \prod_i m_i^{\gamma_i}$. \square

The underlying “rational” matrix can be recovered as $S_{ij}^r := S_{ij}^s/m_j$, and it should be symmetric, a testable implication. 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 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.³⁸ Those papers find field evidence that some prices are partially neglected by consumers.

4 Textbook Competitive Equilibrium Theory: A Behavioral Update

4.1 (In)efficiency of Equilibrium

We next revisit the textbook chapter on competitive equilibrium, with a less than fully rational agent. We will use the following notation. Agent $a \in \{1, \dots, A\}$ has endowment $\omega^a \in \mathbb{R}^n$ (i.e.

³⁶Browning 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.

³⁷The Slutsky matrix does not allow one to recover \bar{m} : for any \bar{m} , S^s admits a dilated factorization $S_{ij}^s = (\bar{m}^{-1} S_{ij}^r) (\bar{m} m_j)$. To recover \bar{m} , one needs to see how the demand changes as \mathbf{p}^d varies. Aguiar and Serrano (2014) explore further the link between Slutsky matrix and BR.

³⁸Chetty, 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. Abaluck and Gruber (2009) 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. 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, Brown, Hossain and Morgan (2010) showed that the seller increases revenue by increasing his shipping charge and lowering his opening price by an equal amount.

he is endowed with ω_i^a units of good i). If the price is \mathbf{p} , his wealth is $\mathbf{p} \cdot \boldsymbol{\omega}^a$, so his demand is $\mathbf{D}^a(\mathbf{p}) := \mathbf{c}^a(\mathbf{p}, \mathbf{p} \cdot \boldsymbol{\omega}^a)$. The economy's excess demand function is $\mathbf{Z}(\mathbf{p}) := \sum_{a=1}^A \mathbf{D}^a(\mathbf{p}) - \boldsymbol{\omega}^a$. The set of equilibrium prices is $\mathcal{P}^* := \{\mathbf{p} \in \mathbb{R}_{++}^n : \mathbf{Z}(\mathbf{p}) = 0\}$. The set of equilibrium allocations for a consumer a is $\mathcal{C}^a := \{\mathbf{D}^a(\mathbf{p}) : \mathbf{p} \in \mathcal{P}^*\}$. The equilibrium exists under weak conditions laid out in Debreu (1970). We start with the efficiency of competitive equilibrium.

Proposition 8 ((In)efficiency of competitive equilibrium). *Assume that competitive equilibria are interior. An equilibrium is Pareto efficient if and only if the perception of relative prices is identical across agents.*

Hence, typically the equilibrium is not Pareto efficient when we are not at the default price. The argument is very simple: if consumers a and b have the same perceptions of prices, then for two goods i and j , $\frac{u_{c_i}^a}{u_{c_j}^a} = \frac{p_i^s}{p_j^s} = \frac{u_{c_i}^b}{u_{c_j}^b}$, so that the ratio of marginal utilities is equalized across agents; there are no extra gains from trade.³⁹

4.2 Excess Volatility of Prices in a Sparse 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 $d\mathbf{p} = \mathbf{p}(1) - \mathbf{p}(0)$ 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 $p_1 = p_1^d$ at $t = 1$). We assume $m_i > 0$ (and will derive it soon).

Proposition 9 (Bounded rationality leads to excess volatility of prices). *Let $d\mathbf{p}^{[r]}$ and $d\mathbf{p}^{[s]}$ be the change in equilibrium price in the rational and sparse economies, respectively. Then:*

$$dp_i^{[s]} = \frac{dp_i^{[r]}}{m_i}, \quad (16)$$

i.e., after a supply shock, the movements of price i in the sparse 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 sparse ($k = s$) economy, we have $\sigma_i^s = \frac{\sigma_i^r}{m_i}$.

³⁹Conversely, if the perceptions of relative prices differ between a and b , then $\frac{u_{c_i}^a}{u_{c_j}^a} = \left(\frac{p_i^s}{p_j^s}\right)_a \neq \left(\frac{p_i^s}{p_j^s}\right)_b = \frac{u_{c_i}^b}{u_{c_j}^b}$; as the ratio of marginal utilities is not equalized across agents, the equilibrium is not efficient. The assumption of an “interior” equilibrium ensures that $\frac{u_{c_i}^a}{u_{c_j}^a} = \frac{p_i^s}{p_j^s}$.

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 sparse world, demand underreacts to shocks; but the market needs to clear, so prices have to overreact to supply shocks.*⁴⁰

Hence, higher volatility leads to higher attention (Proposition 3), and higher attention leads to lower price volatility (Proposition 9). The next proposition describes the resulting fixed point – which ensures that, even with sparse agents, we have $m_i > 0$ endogenously.

Proposition 10 (Endogenous attention and price volatility in an endowment economy). *Assume the linear cost version ($\alpha = 1$) of the sparse max, and that the agents perceive price shocks as uncorrelated. Attention to the price of good i is $m_i = \frac{-J_i + \sqrt{J_i^2 + 4J_i}}{2}$, with $J_i = (\sigma_i^r)^2 \frac{\psi_i c_i^d p_i^d}{\kappa}$. Price volatility is: $\sigma_i^s = \frac{\sigma_i^r}{m_i}$, and is increasing in fundamental volatility σ_i^r (i.e., the volatility in the benchmark, non-sparse economy).*

Consumers need to be attentive ($m_i > 0$), otherwise price volatility would be infinite.⁴¹ Here, endogenously, the actual price volatility of each good is high enough to motivate consumers to pay attention to the price.

4.3 Behavioral Edgeworth Box: Extra-dimensional Offer Curve

We move on to the Edgeworth box. Take a consumer with endowment $\omega \in \mathbb{R}^n$. Given a price vector \mathbf{p} , his wealth is $\mathbf{p} \cdot \omega$, and so his demand is $\mathbf{D}(\mathbf{p}) := \mathbf{c}(\mathbf{p}, \mathbf{p} \cdot \omega) \in \mathbb{R}^n$. The offer curve OC is defined as the set of demands, as prices vary: $OC := \{\mathbf{D}(\mathbf{p}) : \mathbf{p} \in \mathbb{R}_{++}^n\}$.⁴²

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 of a sparse consumer with the same basic preferences: the offer curve is the gray area. It is a two-dimensional “ribbon”, with a pinch at the endowment, rather than the one-dimensional curve of the rational consumer.⁴³ The offer curve has acquired an extra dimension.⁴⁴

⁴⁰Gul, Pesendorfer and Strzalecki (2014) offer a very different behavioral model leading to volatile prices.

⁴¹Things would change in an economy with heterogeneous agents, who might specialize: only some agents might attend to the price of good i (e.g., heavy users of it).

⁴²One can imagine in the background a sequence of i.i.d. economies with a stochastic aggregate endowment, as in section 4.4. That would generate the average price (hence a default price), and a variability of prices (which will lead to the allocation of attention). Note that the default comes from the default price, not from a default action that might be “no trade”.

⁴³A point \mathbf{c} in the OC must be in the two quadrants north-west or south-east of ω (otherwise, we would have $\mathbf{c} \ll \omega$ or $\mathbf{c} \gg \omega$; however, there is a \mathbf{p} s.t. $\mathbf{p} \cdot \mathbf{c} = \mathbf{p} \cdot \omega$: a contradiction). If mistakes are unbounded, the OC is the union of those two quadrants.

⁴⁴To see this directly, take $u(\mathbf{c}) = \ln c_1 + \ln c_2$, $\mathbf{p}^d = (1, 1)$, and $\mathbf{m} = (1, 0)$. Then, $p_1^s = p_1$, $p_2^s = 1$. The OC is the set of (c_1, c_2) for which there are (p_1, p_2) such that: $\frac{u_{c_1}}{u_{c_2}} = \frac{p_1^s}{p_2^s}$ and $\mathbf{p} \cdot (\mathbf{c} - \omega) = 0$, i.e.: $\frac{c_2}{c_1} = p_1$ and $\frac{p_1}{p_2}(c_1 - \omega_1) + c_2 - \omega_2 = 0$. The OC is described by two parameters: $\frac{p_1}{p_2}$ and p_1 , so is two-dimensional.

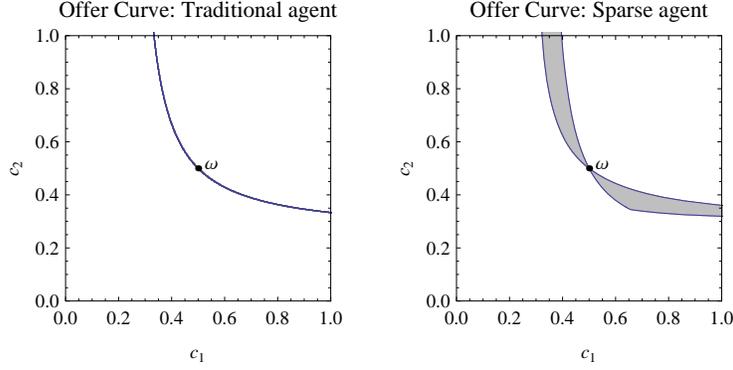


Figure 3: This Figure shows the agent’s offer curve: the set of demanded consumptions $\mathbf{c}(\mathbf{p}, \mathbf{p} \cdot \boldsymbol{\omega})$, as the price vector \mathbf{p} varies. The left panel is the traditional (rational) agent’s offer curve. The right panel is the sparse agent’s offer curve (in gray): it is a 2-dimensional surface. Parameters: $u(\mathbf{c}) = \ln c_1 + \ln c_2$, $\mathbf{p}^d = (1, 1)$, $\mathbf{p} \in [1/5, 5]^2$, $\mathbf{m} = (1, 0.7)$.

What is going on here? In the traditional model, the offer curve is one-dimensional: as demand $\mathbf{D}(\mathbf{p}) = \mathbf{c}(\mathbf{p}, \mathbf{p} \cdot \boldsymbol{\omega})$ is homogeneous of degree 0 in $\mathbf{p} = (p_1, p_2)$, only the relative price p_1/p_2 matters. However, in the sparse model, demand $\mathbf{D}(\mathbf{p})$ is not homogeneous of degree 0 in \mathbf{p} any more: this is the nominal illusion of Proposition 5. 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).⁴⁵ Note that this holds even though the Marshallian demand is a nice, single-valued function.⁴⁶

4.4 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 ($\mathcal{P}^* = \{\chi \bar{\mathbf{p}} : \chi \in \mathbb{R}_{++}\}$), and \mathcal{C}^a is just a point, $\mathbf{D}^a(\bar{\mathbf{p}})$.⁴⁷

In the sparse setup, \mathcal{P}^* is still one-dimensional.⁴⁸ However, *to each equilibrium price level corresponds a different real equilibrium*. This is analogous to a “Phillips curve”: \mathcal{C}^a has dimension 1.⁴⁹ To fix ideas, it is useful to consider the case of one rational consumer and one sparse consumer (the online appendix generalizes).

⁴⁵This “2-dimensional offer curve” appears to be new. It is distinct from the previously-known “thick indifference curve”. The latter arises when the consumer violates strict monotonicity (i.e. likes equally 7 and 7.1 bananas), is not associated to any endowment or prices, and has no pinch. The sparse offer curve, in contrast, arises from nominal illusion, needs an endowment and prices, and has a pinch at the endowment.

⁴⁶I thank Peter Diamond for the following: view the Edgeworth box as 3-dimensional, the third dimension being the price level, say p_1 . For each p_1 , OC is 1-dimensional. However, when all OCs (indexed by p_1) are projected down onto one graph (as in Figure 3), they lead to a 2-dimensional OC.

⁴⁷More generally, equilibria consist of a finite union of such sets, under weak conditions given in Debreu (1970).

⁴⁸By Walras’ law, $\mathcal{P}^* = \{\mathbf{p} : \mathbf{Z}_{-n}(\mathbf{p}) = 0\}$, where $\mathbf{Z}_{-n} = (\mathbf{Z}_i)_{1 \leq i < n}$. As \mathbf{Z}_{-n} is a function $\mathbb{R}_{++}^n \rightarrow \mathbb{R}^{n-1}$, \mathcal{P}^* is generically a one-dimensional manifold.

⁴⁹In the traditional model, equilibria are the intersection of offer curves. However, this is typically not the case

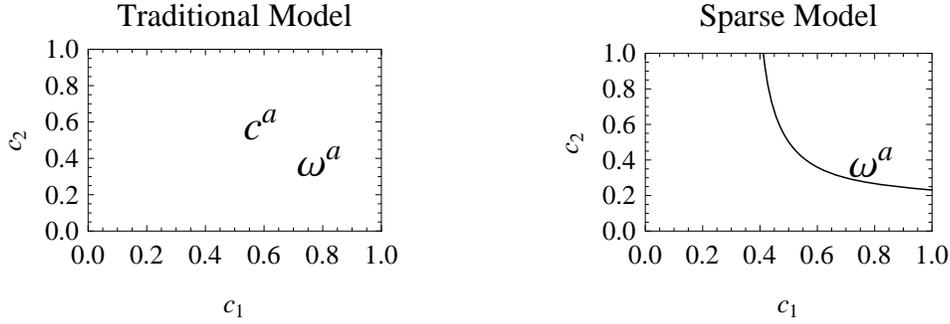


Figure 4: These Edgeworth boxes show competitive equilibria when both agents have Cobb-Douglas preferences. The left panel illustrates the traditional model with rational agents: there is just one equilibrium, c^a . The right panel illustrates the situation when type a is rational, and type b is boundedly 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.

Proposition 11 *Suppose agent a is rational, and the other agent is sparse with $m_1 = 1, m_2 = 0$, and two goods. The set 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.⁵⁰ 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 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 heavily debated in macroeconomics. 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 (Galí 2011): 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

here. The reason is left as an exercise for the reader.

⁵⁰This result linking bounded rationality to a price-dependent real equilibrium appears to be new. The most closely related may be Geanakoplos and Mas-Colell (1989), who analyze a two-period asset-market model. They study incomplete markets with full rationality, here I study complete markets with bounded rationality.

(e.g. Lucas 1972), but in the pristine and general universe of basic microeconomics. One criticism of the Lucas view is that inflation numbers are in practice very easy to obtain, contrary to Lucas’ postulate. This criticism does not apply here: sparse agents actively neglect inflation numbers, which means the Phillips curve effect is valid even when information is readily obtainable.⁵¹

5 Complements to the Sparse Max

5.1 Ex-post Allocation of Attention

What happens when attention is chosen after seeing the x_i ? To capture this, say that the agent uses the actual magnitude of the variable, rather than its expected magnitude: set $\sigma_i = |x_i|$, and $\sigma_{ij} = x_i^2 1_{i=j}$.⁵² This way, the model can be applied to deterministic settings. Everything else is the same: for instance, the sparsified x_i is: $x_i^s = x_i \mathcal{A}_\alpha(x_i^2 (\partial a / \partial x_i)^2 |u_{aa}| / \kappa)$.

5.2 Scale-free κ

The parameter κ has units of utility. Hence, arguably, when the units in which utility is measured double, so should κ . Here is a way to ensure that.

Scale-free κ : Use the unitless parameter $\bar{\kappa} \geq 0$ as a primitive, and set:

$$\kappa := \bar{\kappa} \sum_{i,j} \Lambda_{ij}. \tag{17}$$

Here, we take the average utility gain from thinking as the “scale” of κ .⁵³ In the quadratic problem, that gives: $\kappa = \bar{\kappa} \sum_j \mu_j^2 \sigma_j^2$, i.e. $a^s = \sum_{i=1}^n \mathcal{A}_\alpha \left(\frac{\mu_i^2 \sigma_i^2}{\bar{\kappa} \sum_{j=1}^n \mu_j^2 \sigma_j^2} \right) x_i$. What matters is the relative importance of variable i , compared to the other variables j . Bordalo, Gennaioli and Shleifer (2012, 2013) have emphasized the importance of this proportional thinking. As $\bar{\kappa}$ is unitless, it might be portable from one context to the next.

On the other hand, it is useful keep the regular sparse max (without scaled κ) when we want to capture “this is a small decision, so agents will think little about it”, where the κ might come from some other, prior maximization problem. Also, the scale-free κ is a bit more complex to use than the plain κ .

⁵¹How important sparsity is compared to other explanations (e.g. sticky wages) would be an interesting topic for future research.

⁵²To use a simple problem: “Calculate $a^r = 20 - 5 + 600 + 12 - 232 + 3 - 10000 + 454 - 2000$.” The psychology is that the agent will consider a few large items, e.g. the 10000, 2000, and 600, mentally (provisionally) eliminate the others, and do the addition. The agent “eliminates the signs” at first, to detect what to pay attention to (step 1 of sparse max), then puts them back in the simplified problem (step 2).

⁵³A justification is the following. To have κ proportional to u , we might have $\kappa := 2\bar{\kappa}\mathbb{E}[v(\iota) - v(0)]$, for some unitless $\bar{\kappa}$. Lemma 2 implies $\kappa = \bar{\kappa} \sum_{i,j} \Lambda_{ij} + o(\|x\|^2)$.

5.3 Min-Max Duality

The sparse max has the following nice duality property, analogous to the one of the regular max. Other ways to handle the budget constraints typically lead to a violation of duality.

Proposition 12 (How min-max duality holds for the sparse max). *Suppose $u, -w$ are concave in a , at least one of them strictly so, and let \hat{u}, \hat{w} be two real numbers. Consider the dual problems: (i) $\bar{u}(\hat{w}) := \text{smax}_a u(a, x)$ s.t. $w(a, x) \leq \hat{w}$, (ii) $\bar{w}(\hat{u}) := \text{smin}_a w(a, x)$ s.t. $u(a, x) \geq \hat{u}$. Assume that the constraint binds for problem (i) at \hat{w} . Then, for a given attention m^* (i.e., applying just Step 2) the two problems are duals of each other, i.e. $\bar{w}(\bar{u}(\hat{w})) = \hat{w}$ and $\bar{u}(\bar{w}(\hat{u})) = \hat{u}$. If we assume the “scale-free” version of κ , they also yield the same attention m^* .*

5.4 When Sparse Max is Ordinal Rather Than Simply Cardinal

We say that the sparse max is *ordinal* or “reparametrization invariant” when the action it generates depends on the preferences and the constraints, but not on the specific functions (u, b) representing them.⁵⁴ For instance, the static maximization operator is ordinal, but expected utility is simply cardinal, not ordinal. Ordinality is a nice formal property, though it is not psychologically necessary: people’s attention might depend on their risk aversion, an effect ordinality would eliminate.

A slight reformulation of sparse max is useful here. Define compensated action $\bar{a}(x) := \arg \max_a u(a, x)$ s.t. $b(a, x) \geq b(a^d, x)$, and its derivative at $x = 0$, $\bar{a}_x := -(I + a_y b_a) L_{aa}^{-1} L_{ax}$. We shall call “*compensated sparse max*”: the sparse max of Definition 2, replacing a_x by \bar{a}_x . The justification for this definition is detailed in the online appendix.⁵⁵ The situation is summarized by the following Proposition.

Proposition 13 (Is sparse max ordinal or simply cardinal?) *Given an exogenous attention m , (i.e., just applying Step 2), the sparse max is ordinal. With an endogenous attention m (i.e., applying Steps 1 and 2), assume the scaled version of κ (17): with unconstrained maximization problems, the sparse max is ordinal; with general maximization problems, the “compensated” sparse max is also ordinal.*

The online appendix discusses the pros and cons of the compensated vs plain sparse max. As they are very close, the plain sparse max is generally recommended, as it is easier to use.

⁵⁴E.g. it returns the same answer when $u(a, x)$ is transformed into $f(u(a, x))$ for a arbitrary increasing function f .

⁵⁵ $\bar{a}(x)$ is the extension to general problems of the “compensated demand” of consumption theory. It is useful as welfare losses from inattention are $-\frac{1}{2}(x^s - x)' \bar{a}'_x L_{aa} \bar{a}_x (x^s - x)$. The a_y is the derivative at $(x, y) = 0$ of: $\bar{a}(x, y) := \arg \max_a u(a, x)$ s.t. $b(a, x) + y \geq b(a^d, x)$.

6 Discussion

6.1 Discussion of the Sparse Max

Any departure from the standard rational model involves making particular modelling decisions. The sparsity-based model is, of course, not the only way to model boundedly rational behavior of the partial-inattention type. The main advantages of the sparsity-based model relative to similar approaches are the following key points: (i) it predicts actions that are deterministic (in contrast with “noisy signal” models, say); (ii) it predicts actions that are continuous as a function of the parameters (in contrast with models with fixed costs of attention, say); and (iii) it can be applied in a wide variety of contexts, and in particular, to any problem which can be expressed as in problem (2). I address some potential questions about the model below.

Doesn't sparse maximization complicate the agent's problem? One could object that it is easier to optimize on a , as in the traditional model, than on a and m , as in the sparse model. However, we can interpret the situation in the following way: at time 0, so to speak, the agent chooses an “attentional policy”, i.e. the vector m^* . He is then prepared to react to many situations, with a precompiled sparse attention vector that allows him to focus on just a few variables. Hence, it is economical for the agent to use sparse maximization. In addition, as shown in Proposition 1, in many situations the sparse max leads to a procedure for the agent that is computationally much simpler than the traditional model.

If the agent knows x , why simplify it? One interpretation is that it is system 1 (Kahneman 2003) that, at some level, knows x , and chooses not to bring it to the attention of system 2 for a more thorough analysis. System 1 chooses the representation x^s , while system 2 takes care of the actual maximization, with a simpler problem.

How does the agent know u_{aa} and a_x ? Again drawing on Kahneman (2003), this can be interpreted as system 1 having a sense of which variables are important and which are not, in the default model. It seems intuitive that, for many problems at least, agents do have a sense of which variables are important or not. To keep the model simple, this is represented by the agent's knowledge of u_{aa} and a_x .

Why isn't attention “all or nothing”? (i.e. why don't we have $m \in \{0, 1\}$?) First, the model does allow for all-or-nothing attention, with the choice of a particular attention function, \mathcal{A}_0 . Secondly, in many inattention models, the aggregate behavior is equivalent to partial inattention $m \in (0, 1)$ (see section 7.2). In addition, in many applications the all-or-nothing approach generates discontinuous reaction curves (e.g. demand curves) that are empirically implausible.

“Framing” matters here; is that good or bad? The framing of the problem affects the agent's decision here. For instance, suppose we ask an agent to predict real wage growth, under two different “frames”, i.e. bases of x . In the “nominal” frame, inputs are nominal wage growth (x_1) and inflation (x_2). In the “real” frame, inputs are real wage growth (x'_1) and inflation (x'_2). (The basis is $(x'_1, x'_2) = (x_1 - x_2, x_2)$). So the correct prediction for real wage growth is $a = x_1 - x_2 = x'_1$.

However, a sparse agent will make different predictions in the different frames. In the nominal frame, $m_1 > m_2$ (see Lemma 4 for a justification), so he will exhibit nominal illusion: inflation leads to an overestimation of real wage growth (as in Shafir, Diamond and Tversky 1997). In the real frame, however, there is no nominal illusion; thus, it is clear that the framing of the problem matters. This is arguably a desirable feature of the model, however. In contrast, in entropy-based models of rational inattention (e.g., Sims 2003), the agent would not exhibit any nominal illusion in either frame: he will dampen his prediction of the real wage by the same amount in both frames, i.e. predict $\lambda x'_1$, with some dampening $\lambda \in [0, 1]$.

Why evaluate the derivatives at the default model rather than at the true model? Lemma 2 justifies this mathematically: it evaluates derivatives at the default model. The agent needs to know approximately what to do at the default, but not elsewhere. This simplifies the agent’s decision-making problem.

What “cost” is preventing the agent from using the traditional model? One could interpret the model as assuming that it is costly for the agent to reduce the noise in his perception of each x_i (see section (7.2), in which the sparse max corresponds to the average behavior of agents with noisy signals). One could also interpret the model as incorporating a mental cost of processing the data. Research in neuroscience has not yet converged on a definitive characterization of what the source of these costs might be. (Possibilities include “working memory”, “mental effort” and “fatigue”).⁵⁶

Why the specific choice procedure that requires the constraint to be satisfied? The model assumes a choice procedure where the multiplier λ is adjusted to satisfy the budget constraint (see Section 2.3). There are certainly other choice procedures which could be considered instead (see Chetty et al. 2007). One such procedure is: “choose the optimal action under the perceived x^s , but adjust the ‘last’ action to satisfy the constraint”. This is an appropriate procedure when there is a clear “last” action (e.g., the choice of savings, as in Gabaix (2013a)), but in many cases no such “last” action exists. There are also many procedures that could be appropriate for consumption problems, but don’t have any counterpart in the general problem. Examples are: “decide how much money to spend on each good”, or “multiply all components of your action by a parameter”). The choice procedure used in the model applies to general decision problems and has useful properties, in particular: (i) invariance to reparametrization of the action (e.g., outcomes are the same for the choice of consumption and the choice of log consumption); and (ii) min-max duality. None of the alternative procedures above satisfy both these properties.

Can’t the same results be obtained with existing models of inattention? Yes, other models of inattention would likely yield similar results, *if* they could be applied and solved.⁵⁷ However, I consider this an advantage of the model. We are interested here in the general impact of inattention,

⁵⁶I thank, without implication, neuroscientists I queried about this.

⁵⁷For instance, Slutsky asymmetry could presumably also be derived for other models of inattention. Note, however, that relative inattention is necessary for the asymmetry of the Slutsky matrix. J.P. Bouchaud (personal communication) has shown that with pure noise in the demand function, Slutsky symmetry is preserved.

so it is desirable that the predictions of the model match those of related models. The contribution of the sparse max is its tractability and generalizability, which allow inattention to be applied to the basic chapters of microeconomics for the first time, and thus allows many new properties to be derived.

Is it a problem to present a model without axioms? It is conceivable that axioms could be formulated for the sparse max.⁵⁸ We note that many of the useful innovations in basic modelling have started without any axiomatic basis: prospect theory, hyperbolic discounting, learning in games, fairness models, Calvo pricing etc. Sometimes the axioms came, but later.

6.2 Links with Themes of the Literature

Sparsity is another line of attack on the polymorphous problem of confusion, inattention, simplification, and bounded rationality. It is a complement rather than a substitute for existing models. For instance, one could join sparsity to research on learning (Sargent 1993, Fudenberg and Levine 1998, Fuster, Laibson, and Mendel 2010), and study “sparse learning.” Some of the most active themes are the following.

Behavioral economics. This research complements a recent surge of interest in behavioral modelling, especially of the “differential attention” type. In Bordalo, Gennaioli and Shleifer (2012, 2013), agents choosing between two goods (or gambles) pay more attention to dimensions (or states) where the two choices are most different. In Kőszegi and Szeidl (2013), people focus more on features that differ most in the choice set. Sparsity is another way to express these features. Much of this behavioral work wants to derive rich implications from psychology, so it develops basic, additive setups. Here the sparse max is designed to be able to tackle quite general problems (equation 2). This greater generality allows it to revisit chapters of the microeconomics textbook – at a level of generality hitherto not accessible.⁵⁹

Interestingly, many classic behavioral biases are of the “inattention” and “simplification” type: for instance, inattention to sample size, base-rate neglect, insensitivity to predictability, anchoring and partial adjustment (Tversky and Kahneman 1974), and projection bias (which is neglect of mean-reversion – Loewenstein, O’Donoghue and Rabin (2003)). Sparsity might be a natural way to model them: people prefer a simpler representation of the world, where many features are eliminated (see online appendix). The simplification depends on the incentives in the environment, so sparsity

⁵⁸The rich work on sparsity (Candès and Tao (2006), Donoho (2006)), which has established many near-optimality properties of the use of the “ ℓ_1 ” norm. Some of those results could be used to quantify the optimality properties of the sparse max.

⁵⁹I omit here “process” models (Tversky (1972), Gabaix, Laibson, Moloche, and Weinberg (2006)) and automata models (Rubinstein 1998). They are instructive but very complex to use. See also Fudenberg and Levine (2006), Brocas and Carillo (2008) and Cunningham (2013) for very different dual-self models. A vast literature quantifies the empirical importance of inattention, e.g. DellaVigna and Pollet (2007), Cohen and Frazzini (2008), Hirshleifer, Lim and Teoh (2009). Relatedly, a literature studies BR at the level of organizations (Geanakoplos and Milgrom 1991, Radner and Van Zandt 2001).

can model “dynamic attention” to features of the environment, whereas behavioral models with fixed weights cannot. It is useful to have a model where the strength of biases depend on the environment – even to assess how strong that dependence is (an interesting question left for future research). That application is sketched in the online appendix.

Inattention and information acquisition. This paper is related to the literature on modeling inattention (DellaVigna 2009, Veldkamp 2011). One strand uses fixed costs, paid over time (Grossman and Laroque 1990, Gabaix and Laibson 2002, Mankiw and Reis 2002, Abel, Eberly and Panageas 2013, Schwartzstein 2014). Those models are instructive, but quickly become hard to work with as the number of variables increases. Also, those papers require a time dimension, so they don’t naturally apply to problems where the action is taken in one period, such as the basic consumption problem (1).

This paper builds on Chetty, Looney and Kroft (1999)’s insights. They study a consumption problem with two goods, where the agent may not think about the tax. Attention is modeled as paying a fixed cost. The present paper proposes a general sparse max with nonlinear constraints. Also, it derives the whole of basic consumer and equilibrium theory with several goods.

Sim’s “rational” inattention. An influential proposal made by Sims (2003) is to use an entropy-based penalty for the cost of information; this literature is progressing impressively (e.g., Maćkowiak and Wiederholt 2009, Woodford 2012, Caplin and Dean 2013). It has the advantage of a nice mathematical foundation. The main differences are that the sparse max (i) allows for source-dependent inattention, and (ii) is more tractable.

In its pure form, the Sims formulation doesn’t yield source-dependent inattention. For instance, take the basic quadratic problem. With an entropy penalty à la Sims (2003), the solution to the quadratic problem is: $\mathbb{E}[a^{\text{Sims}} | x] = \lambda \sum_i \mu_i x_i$ for some $\lambda \in [0, 1]$. Hence, all dimensions are dampened equally. In contrast, in the sparse model, less important dimensions are dampened more (eq. 9). As a result, the pure Sims approach doesn’t generate any nominal illusion (see earlier). This is why other researchers (Maćkowiak and Wiederholt 2009, Woodford 2012) deviate from the Sims approach, and also have basis-dependent models.

In addition, the entropy-based model is much less tractable. It leads to non-deterministic models (agents take stochastic decisions), and the modeling is very complex when it goes beyond the linear-Gaussian case: the solutions require a computer to be solved, and there’s no closed form (Matějka and Sims 2010). Even when the model is solved with quadratic approximations, the budget constraint remains a hard problem, so researchers use savings as a buffer. As a result, no one has (yet) been able to work out the basic consumption problem (1), much less derive its implications for basic consumption and equilibrium theory.

Limited understanding of strategic interactions. In several models, the bounded rationality comes from the interactions between the decision maker and other players: see Camerer, Ho, and Chong (2004), Crawford and Irriberri (2007), Eyster and Rabin (2005), Jéhiel (2005). These models are useful for capturing naiveté about strategic interactions. However, in a single-person context,

they typically model the agent as fully rational.

Opacity, shrouding, confusion, and frames. A growing literature researches the impact of “confusion” of consumers on market equilibrium (e.g. Gabaix and Laibson 2006). The present model is less general than the “comparison frames” of Piccione and Spiegler (2012) but more specific.

It may be interesting to note the Sims framework is based on Shannon’s information theory of the 1940s. The Hansen and Sargent (2007) framework (which is concerned with robustness rather than simplicity) is influenced by the engineering literature of the 1970s. The present framework is inspired by the sparsity-based literature of the 1990s-2000s.

7 Conclusion

This paper proposes an enrichment of the traditional “max” operator, with some boundedly rational features: the “sparse max” operator. This formulation is quite tractable. At the same time, it arguably has some psychological realism.

The simplicity of the core model allows for the formulation of a sparse, limited attention version of important building blocks of economics: the basic theory of consumer behavior, and competitive equilibrium. We can bring a behavioral enrichment to venerable and often-used concepts such as Marshallian and Hicksian demand, elasticity of substitution, and competitive equilibrium sets. Some surprises emerge. The model allows us to better understand what is robust and non-robust in basic microeconomics. One day, it might even help an experimental investigation of core microeconomics, helped by the existence of a behavioral alternative.

We argued that other models with a more traditional form (e.g. extraction of noisy signals) might lead to fairly similar features, but would be quite intractable. The sparse max allows us to explore features of economic life that hopefully apply to other models. It does it with relatively little effort.

Though many predictions have yet to be tested, the extant evidence is encouraging: the model seems qualitatively correct in predicting (rather than positing) inattention to minor parts of the pricing schemes, fiscal illusion, nominal illusion, and a variety of plausible comparative statics.

No doubt, the model could and should be greatly enriched. It is currently silent about some difficult operations such as memory management (Mullainathan 2002), and mental accounts (Thaler 1985).

As a work in progress, I extend the model to include multi-agent models and dynamic programming (Gabaix 2013a,b) to handle applications in macroeconomics and finance (see also Croce, Lettau and Ludvigson 2012). The sparsity theme proves particularly useful. Agents in a sparsity macro model are often simpler to model than in the traditional model, as their actions depend on a small number of variables and are easier to analyze and interpret. They may be more realistic too. Hence the sparse max might be a useful versatile tool for thinking about the impact of bounded rationality in economics.

Appendix A: Notation

f_x, f_{xy} : derivatives of a function f ; $f_x := \frac{\partial f}{\partial x}$, $f_{xy} := \frac{\partial^2 f}{\partial x \partial y}$

I : identity matrix of the appropriate dimension

Basic sparse model

m : the attention vector. m^* is the attention chosen by the agent.

x : an n -vector of disturbances the agent may pay attention to

x^s : the perception of vector x after simplification by the agent. $x_i^s = m_i x_i$

a^r, a^s, a^d : action under the rational, the sparse model, and default action (typically $\arg \max_a u(a, 0)$)

a_{x_i} : derivative of the action w.r.t. variable x_i , for the rational agent, evaluated at $x = 0$

κ : cost of attention parameter. $\kappa = 0$ is the rational model. $\bar{\kappa}$ is its unitless version

\mathcal{A} : attention function, often parametrized as \mathcal{A}_α

$g(m_i)$: cost of attention function

Λ : cost of inattention matrix

L : Lagrangian

σ_i, σ_{ij} : standard deviation of x_i ; covariance between x_i and x_j

Consumer theory

\mathbf{p} : price vector. \mathbf{p}^d : default price vector

w : wealth

λ : Lagrange multiplier

$\mathbf{c}(\mathbf{p}, w)$: Marshallian demand

S : Slutsky matrix

Competitive equilibrium theory

$\boldsymbol{\omega}$: endowment vector

OC : offer curve

\mathcal{P}^* : set of equilibrium prices

\mathcal{C}^a : set of equilibrium allocations for a consumer a

$\mathbf{Z}(\mathbf{p})$: excess demand vector at price \mathbf{p}

Appendix B: Motivation and Microfoundations

7.1 Motivation for the Basic Sparse Max (Without Constraints)

An attention vector m generates a representation of the world $x^s(m)$; it leads the agent to take an action $a(x^s(m)) = \arg \max_a u(a, x^s(m))$, and get utility $v(m) := u(a(x^s(m)), x)$. The agent

wishes to pick the best model, i.e. the attention vector m that maximizes utility $v(m)$, minus a cognition cost $\mathcal{C}(m)$ which we'll discuss soon:

$$\max_m \mathbb{E}[v(m)] - \mathcal{C}(m). \quad (18)$$

This is potentially a very complex problem: the agent seems to have to calculate the best action for all vectors m , and calculate the utility consequences – a very difficult task. Hence, we need to simplify v . To explore how to do so, we perform a Taylor expansion of $v(m)$ in the limit of small x . Here $\iota := (1, \dots, 1)'$, so that $v(\iota)$ is the utility when the agent is fully attentive.

Lemma 2 *The utility losses from imperfect inattention are: $\mathbb{E}[v(m) - v(\iota)] = -\frac{1}{2}(m - \iota)' \Lambda (m - \iota) + o(\|x\|^2)$ for a cost-of-inattention matrix with $\Lambda_{ij} := -\sigma_{ij} a_{x_i} u_{aa} a_{x_j}$. The $o(\|x\|^2)$ term is 0 when utility is linear-quadratic.*

Hence, in the sparse max we will posit that the agent solves the simpler problem:

$$\max_m -\frac{1}{2}(m - \iota)' \Lambda (m - \iota) - \mathcal{C}(m).$$

This is one way to circumvent Simon's "infinite regress problem" – that optimizing the allocation of thinking cost can be even more complex than the original problem. I avoid this problem by assuming a simpler representation of it, namely a quadratic loss. Using $\mathcal{C}(m) = \sum_i g(m_i)$ leads to the formulation in the paper. *If the utility function is linear-quadratic, Definition 1 is the optimal solution to the cognitive optimization problem (18).*

7.2 Two Models that Generate a Noisy Sparse Max

Under some conditions, some models offer a noisy microfoundation for the sparse max, i.e. their representative agent version is an agent using the sparse max. We emphasize the basic, quadratic case.

Heterogeneous fixed costs Consider next an agent with "all or nothing" attention: each m_i is equal to 0 (no attention) or 1 (full attention). This can be represented as a fixed cost for the penalty ($\alpha = 0$) in (6). The following Proposition indicates that the sparse max can be viewed as the "representative agent equivalent" of many agents with heterogeneous fixed costs.

Proposition 14 (Fixed costs models as a microfoundation of sparse max in the basic case). *Suppose that agents use fixed costs \tilde{k} , and the distribution of \tilde{k} 's is $P(\tilde{k} \leq q) = \mathcal{A}(2q/\kappa)$ for some function \mathcal{A} . Then, in quadratic problems with a one-dimensional action, the average behavior of these "fixed cost" agents is described by the sparse max (with just one agent) with attention function \mathcal{A} .*

Noisy signals Here is a version of this idea with the noisy signals model (it also holds for multi-dimensional actions, see online appendix).

Proposition 15 (Signal-extraction models as a noisy microfoundation of sparse max in the basic case). *Consider a model with quadratic loss (3). The agent receives noisy signals $S_i = x_i + \varepsilon_i$, with (x_i, ε_i) uncorrelated Gaussian random variables, and noise ε_i has the relative precision $T_i = \frac{\text{var}(x_i)}{\text{var}(\varepsilon_i)}$. The agent: (i) decides on signal precision T_i , (ii) receives signals $S = (S_i)_{i=1\dots n}$, and then (iii) takes action $a(S)$, to maximize expected utility. Hence, the agent does: $\max_{T_i \geq 0} \max_{a(S)} \mathbb{E}[u(a(S), x) | S] - \frac{\kappa}{2} \sum_i G_\alpha(T_i)$, where G_α satisfies $G'_\alpha(T) = g'_\alpha\left(\frac{T}{1+T}\right) \frac{1}{1+T}$. Then, for a given x , averaging over the signals, the optimal action $a(S)$ is the sparse max $a^s(x)$ with cost $\kappa g_\alpha(m)$: $\mathbb{E}[a(S) | x] = a^s(x) = \sum_i \mathcal{A}_\alpha(\sigma_i^2/\kappa) x_i$.*

Hence, an economist who favors the paradigm of Bayesian updating with costly signals (which is simply an allegory for some messier, biological reality) may interpret the sparse max as follows: *The sparse max is the representative agent version of a model with noisy signals.* Of course, this holds only under particular assumptions.

Appendix C: Welfare Analysis in Consumer Theory

We complete our behavioral version of textbook consumer theory by studying welfare. We define the attention matrix as $M := \text{diag}(m_1, \dots, m_n)$ (the diagonal matrix with elements m_i), so that (15) becomes $S^s = S^r M$.

Negative semi-definiteness, WARP Recall that a small price change $\delta \mathbf{p}$ leads to a compensated change in consumption $\delta \mathbf{c} = S \delta \mathbf{p}$. In the traditional model, the following holds: $\delta \mathbf{p} \cdot \delta \mathbf{c}^r \leq 0$, “on average, when prices go up, (compensated) demand goes down.” This can be rewritten $\delta \mathbf{p}' S^r \delta \mathbf{p} \leq 0$: the Slutsky matrix is negative semi-definite. Here is the version with a sparse agent.

Proposition 16 (The Slutsky matrix is not negative semi-definite, Violation of WARP). *Suppose that $S M \mathbf{p}^d \neq 0$, and consider a deviation from the default price $\delta \mathbf{p} = \mathbf{p} - \mathbf{p}^d$. The agent’s decisions violate the weak axiom of revealed preferences (WARP): there is a small price change $\delta \mathbf{p}$, such that the corresponding change in consumption $\delta \mathbf{c}^s = S^s \delta \mathbf{p}$ satisfies $\delta \mathbf{p} \cdot \delta \mathbf{c}^s > 0$. In other terms, the Slutsky matrix S^s fails to be negative semi-definitive. However, for all price changes $\delta \mathbf{p}$, we have: $\delta \mathbf{p}^s \cdot \delta \mathbf{c}^s \leq 0$.*

WARP fails, but something like it holds: at \mathbf{p}^d , we have $\delta \mathbf{p}^s \cdot \delta \mathbf{c}^s \leq 0$, i.e. $\sum m_i \delta p_i \delta c_i^s \leq 0$. Hence we do preserve a salience-weighted “law of demand”: “when prices go up, (compensated) demand goes down, but in a salience-weighted sense”.

Here is the intuition. Suppose that the agent pays attention to the car price, but not gas. Suppose that the car price goes down, but gas price goes up by a lot. A rational agent will see that

the total price of transportation (gas+price) has gone up, so he consumes less of it: $\delta \mathbf{c}^r \cdot \delta \mathbf{p} < 0$, with $\mathbf{c}^r = (c_{\text{car}}, c_{\text{gas}}, c_{\text{food}})$. However, a sparse agent just sees that the car price went down, so he consumes more transportation: $\delta \mathbf{c}^s \cdot \delta \mathbf{p} > 0$. This is a violation of WARP.⁶⁰

Hicksian demand, welfare and related notions We call indirect utility function $v(\mathbf{p}, w) = u(\mathbf{c}(\mathbf{p}, w))$ ⁶¹, the expenditure function $e(\mathbf{p}, v) := \min_{\mathbf{c}} \mathbf{p} \cdot \mathbf{c}$ s.t. $u(\mathbf{c}) \geq v$, and the Hicksian demand $\mathbf{h}(\mathbf{p}, v) := \arg \min_{\mathbf{c}} \mathbf{p} \cdot \mathbf{c}$ s.t. $u(\mathbf{c}) \geq v$.

Proposition 17 *The sparse Hicksian demand is: $\mathbf{h}^s(\mathbf{p}, u) = \mathbf{h}^r(\mathbf{p}^s, u)$. The sparse expenditure function is $e^s(\mathbf{p}, u) = \mathbf{p} \cdot \mathbf{h}^s(\mathbf{p}, u) = \mathbf{p} \cdot \mathbf{h}^r(\mathbf{p}^s, u)$.*

Proposition 18 (Link between Slutsky matrix and expenditure function). *At the default price, the expenditure function satisfies: $e^s = e^r$, $e_{\mathbf{p}}^s = e_{\mathbf{p}}^r$, and $e_{\mathbf{p}\mathbf{p}}^s = e_{\mathbf{p}\mathbf{p}}^r - (I - M)' e_{\mathbf{p}\mathbf{p}}^r (I - M)$, i.e. $e_{p_i p_j}^s = e_{p_i p_j}^r \times (m_i + m_j - m_i m_j)$. Hence, $e_{p_i p_j}^s = S_{ij}^s + S_{ji}^s - \frac{S_{ij}^s S_{ji}^s}{S_{ij}^r}$ rather than the traditional $e_{p_i p_j}^r = S_{ij}^r$.*

Let us now discuss the welfare losses from price misperception.

Proposition 19 (Indirect utility function and welfare losses). *At price $p = \mathbf{p}^d$, the quantities $v, v_p, v_w, v_{pw}, v_{ww}$ are the same under the sparse and the traditional model, but v_{pp} differs:*

$$v_{pp}^s - v_{pp}^r = -v_w^r (e_{pp}^s - e_{pp}^r) = v_w^r (I - M)' S^r (I - M). \quad (19)$$

The intuition is simple: the utility loss $(v_{pp}^s - v_{pp}^r)$ is equal to the extra expenditure $e_{pp}^s - e_{pp}^r$ due to suboptimal behavior, times the utility value of money, v_w . This sub-optimal behavior is itself due to a lack of substitution effects, $(I - M)' S^r (I - M)$. Welfare losses are second order (e.g. Krusell-Smith 1996). We next turn to evaluating welfare from choice data.

Proposition 20 (Shephard's lemma, Roy's identity). *Evaluated at the default price, we have Shephard's lemma: $e_{p_i}^s = h_i^s$ at $\mathbf{p} = \mathbf{p}^d$, and Roy's identity: $c_i^s(\mathbf{p}, w) + \frac{v_{p_i}^s(\mathbf{p}, w)}{v_w^s(\mathbf{p}, w)} = 0$. However, when $\mathbf{p} \neq \mathbf{p}^d$, we have the modified Shephard's lemma:*

$$e_{p_i}^s(\mathbf{p}, u) = h_i^s(\mathbf{p}, u) + (\mathbf{p} - \mathbf{p}^s) \cdot \mathbf{h}_{p_i}^r(\mathbf{p}^s, u) m_i, \quad (20)$$

⁶⁰Condition $SMp^d \neq 0$ is quite weak – with two goods it essentially means that $m_1 \neq m_2$. Here is a simple example: $\mathbf{p}^d = (1, 1)$ and $S^r = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ (from $u = \ln c_1 + \ln c_2$), and $m = (1, 0)$. Consider $\delta \mathbf{p} = (1, 2)$ (which the reader may wish to multiply by some small $\varepsilon > 0$ so we deal with small price changes). As the price of good 2 increases more, the rational agent consumes less of good 2, and more of good 1: $\delta \mathbf{c}^s = S^r \delta \mathbf{p} = (1, -1)$. However, the sparse consumer perceives $\delta \mathbf{p}^s = M \delta \mathbf{p} = (1, 0)$, so he perceives only that good 1's price increases. So he consumes less of good 1, and more of good 2: $\delta \mathbf{c}^s = S^r \delta \mathbf{p}^s = (-1, 1) = -\delta \mathbf{c}^r$. Hence, $\delta \mathbf{p} \cdot \delta \mathbf{c}^r = -1 < 0 < \delta \mathbf{p} \cdot \delta \mathbf{c}^s = 1$, a violation of WARP.

⁶¹Here we include only “consumption utility”. Incorporating attention cost would lead to a more complex analysis, e.g. drawing on Bernheim and Rangel (2009).

and the modified Roy's identity:

$$c_i^s(\mathbf{p}, w) + \frac{v_{p_i}^s(\mathbf{p}, w)}{v_w^s(\mathbf{p}, w)} = [(\mathbf{p} \cdot c_{w'}^r(\mathbf{p}^s, w')) \mathbf{p}^s - \mathbf{p}] \cdot c_{p_i}^r(\mathbf{p}^s, w') m_i, \quad (21)$$

Hence, Shephard's lemma and Roy's identity hold for a perfectly naive ($m_i = 0$) or rational ($m_i = 1$) consumer, but fail for intermediate levels of rationality. In addition, (20) gives the following result, which illustrates anew the peril of mis-measuring demand elasticities by assuming perfect rationality.⁶²

Proposition 21 *When the price of good i increases by Δp_i , a naive application of Shephard's lemma (i.e., one assuming perfect rationality) will overestimate the expenditure required to compensate the consumer by $-\frac{1}{2} S_{ii}^r (\Delta p_i)^2 m_i (1 - m_i)$, a bias which is maximal at intermediate sophistication.*

Appendix D: Proofs of Basic Results

Derivation of Example 2 We denote by $D(x)$ and $S(x)$ the supply and demand for labor when the wage change is x (so that before the tax change, employment is $D(0) = S(0)$). The equilibrium wage change Δw satisfies: $D(\Delta w + t^f) = S(\Delta w - t^e)$, i.e. $D(\Delta n + t) = S(\Delta n)$. Using functional forms $D(\Delta w) = D(0) - D_1 \Delta w$ and $S(\Delta w) = S(0) + S_1 \Delta w$ (or reasoning with small changes), we have $-D_1(\Delta n + t) = S_1 \Delta n$. Hence, with $\beta := \frac{D_1}{D_1 + S_1} \in (0, 1)$: $\Delta n = -\beta t$ and $\Delta w = \Delta n + t^e = -\beta t + t^e$.

Endogenous allocation of attention. We have $\frac{t^e}{t} = s$, $\frac{\Delta w}{t} = \frac{-\beta t + t^e}{t} = s - \beta$. Hence, $\frac{\Delta n^s}{t} = \frac{-m_1 t^e + m_2 \Delta w}{t} = -\tau(s) + \tau(-\beta + s) := f(s)$, where $\tau(x) := xA(x)$ is a "truncation" function, with $A(x) := \mathcal{A}_\alpha(x^2 \sigma_t^2 / \kappa)$. Here $\sigma_t = \mathbb{E}[t^2]^{1/2}$ is the size of the tax. We next state a Lemma.

Lemma 3 *For x, y positive, $\tau(x + y) \leq \tau(x) + \tau(y)$, with a strict inequality when κ is small enough and $\alpha > 0$.*

Lemma proof: As $A(x)$ is weakly increasing in x :

$$\tau(x) + \tau(y) = A(x)x + A(y)y \leq A(x + y)x + A(x + y)y = A(x + y)(x + y) = \tau(x + y).$$

The result for κ small enough comes from the fact that $A'(x) > 0$ whenever attention is non-zero and $\alpha > 0$. \square

⁶²This result echoes related analyses of Chetty, Looney and Kroft (2009). The advance is that now those errors via Shephard and Roy are established in full generality, rather than with two goods. The online appendix gives more intuition for those results, and further discussion.

The perceived wage change Δn is (when $t = 1$): $f(s) = -\tau(s) + \tau(s - \beta)$. Consider two shares s, s' with $s \leq \beta < s'$. We want to prove that the perceived wage is higher when the employee pays the lower share, i.e. $f(s) \geq f(s')$. Indeed:

$$\begin{aligned} f(s) - f(s') &= -\tau(s) + \tau(s - \beta) + \tau(s') - \tau(s' - \beta) \\ &= \tau(s') - [\tau(s) + \tau(\beta - s) + \tau(s' - \beta)] \text{ using } \tau(-x) = -\tau(x) \\ &\geq \tau(s') - \tau(s + \beta - s + s' - \beta) = 0 \text{ by applying Lemma 3 twice.} \end{aligned}$$

Proof of Proposition 2 We follow Definition 2. The Lagrangian is $L(\mathbf{c}, \mathbf{x}, \lambda) = u(\mathbf{c}) + \lambda(w - \mathbf{p}^s \cdot \mathbf{c})$, with $\mathbf{p} = \mathbf{p}^d + \mathbf{x}$ and $p_i^s = p_i^d + m_i x_i$. We just need step 2 here. The function $\mathbf{c}(\lambda) = \arg \max_{\mathbf{c}} L(\mathbf{c}, \mathbf{x}^s, \lambda)$ satisfies $u'(\mathbf{c}(\lambda)) = \lambda \mathbf{p}^s$ i.e. $\mathbf{c}(\lambda) = u'^{-1}(\lambda \mathbf{p}^s)$. Hence using the solution λ^* satisfies the budget constraint $\mathbf{p} \cdot \mathbf{c}(\lambda^*) = w$ (Lemma 5 shows formally that it binds), and chosen consumption is $\mathbf{c}^s = \mathbf{c}(\lambda^*)$. Calling $w' := \mathbf{p}^s \cdot \mathbf{c}^s$, we have $\mathbf{c}^s = \mathbf{c}^r(\mathbf{p}^s, w')$, as it satisfies $\mathbf{p}^s \cdot \mathbf{c}^s = w$ and $u'(\mathbf{c}^s) = \lambda \mathbf{p}^s$.

Derivation of Example 4 We have $w = \mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, w') = \mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, 1) w'$, so $w' = \frac{w}{\mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, 1)}$, and $\mathbf{c}^s(\mathbf{p}, w) = \mathbf{c}^r(\mathbf{p}^s, w') = \mathbf{c}^r(\mathbf{p}^s, 1) w' = \frac{\mathbf{c}^r(\mathbf{p}^s, 1)}{\mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, 1)} w$.

Proof of Proposition 3 We have $L = u(\mathbf{c}) + \lambda^d (w - (\mathbf{p}^d + \mathbf{x}) \cdot \mathbf{c})$. So, $L_{\mathbf{c}} = u'(\mathbf{c}) - \lambda \mathbf{x}$, using the notation $\lambda := \lambda^d$ and $\mathbf{c}_{\mathbf{x}} = -L_{cc}^{-1} L_{c\mathbf{x}} = u''^{-1} \lambda$. In Step 1, $\mathbf{c}_{\mathbf{x}} L_{cc} \mathbf{c}_{\mathbf{x}} = \lambda^2 u''^{-1}$, so the problem is:

$$\min_m \frac{1}{2} \sum_i (m_i - 1)^2 \sigma_{p_i}^2 (-u''^{-1})_{ii} \lambda^2 + \kappa \sum_i |m_i|^\alpha,$$

so $m_i = \mathcal{A}(v_i/\kappa)$ with

$$\begin{aligned} v_i &= \sigma_{p_i}^2 (-u''^{-1})_{ii} \lambda^2 = \frac{\sigma_{p_i}^2}{(p_i^d)^2} (p_i^d)^2 \frac{c_i^d \psi_i}{u_{c_i}} \lambda^2 \text{ defining } \psi_i := u_{c_i} (-u''^{-1})_{ii} / c_i^d, \\ &= \left(\frac{\sigma_{p_i}}{p_i^d} \right)^2 \frac{\lambda^2 (p_i^d)^2 c_i^d \psi_i}{\lambda p_i^d} = \frac{\sigma_{p_i}^2}{p_i^d} \lambda p_i^d c_i^d \psi_i \text{ using } u_{c_i} = \lambda p_i^d. \end{aligned}$$

The term ψ_i is a price elasticity. For instance, when $u(\mathbf{c}) = \sum_i A_i c_i^{1-1/\eta_i}$, then $\psi_i = \eta_i$.

Proof of Proposition 4 We have: $\mathbf{c}^s(\mathbf{p}, w) = \mathbf{c}^r((p_i^d + m_i(p_i - p_i^d))_{i=1\dots n}, w'(\mathbf{p}))$. We differentiate w.r.t. p_j :

$$\frac{\partial \mathbf{c}^s}{\partial p_j} = \frac{\partial \mathbf{c}^r}{\partial p_j} m_j + \frac{\partial \mathbf{c}^r}{\partial w} \frac{\partial w'}{\partial p_j}(\mathbf{p}). \quad (22)$$

Proposition 2 implies $\mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}^s, w'(\mathbf{p})) = w$, and differentiating w.r.t. p_j : $0 = c_j^r + \mathbf{p} \cdot \frac{\partial \mathbf{c}^r}{\partial p_j} m_j + \mathbf{p} \cdot \frac{\partial \mathbf{c}^r}{\partial w} \frac{\partial w'}{\partial p_j}$. As $\mathbf{p} \cdot \mathbf{c}^r(\mathbf{p}, w) = w$, we have (differentiating w.r.t. p_j and w respectively): $c_j^r + \mathbf{p} \cdot \frac{\partial \mathbf{c}^r}{\partial p_j} = 0$

and $\mathbf{p} \cdot \frac{\partial \mathbf{c}^r}{\partial w} = 1$, so $0 = c_j^r - c_j^r m_j + \frac{\partial w'}{\partial p_j}$, i.e. $\frac{\partial w'}{\partial p_j} = (m_j - 1) c_j^r$. Finally (22) gives: $\frac{\partial \mathbf{c}^s(\mathbf{p}, w)}{\partial p_j} = \frac{\partial \mathbf{c}^r}{\partial p_j} m_j + \frac{\partial \mathbf{c}^r}{\partial w} (m_j - 1) c_j^r$.

Proof of Proposition 6

$$\begin{aligned} S_{ij}^s &:= \frac{\partial c_i^s}{\partial p_j} + \frac{\partial c_i^s}{\partial w} c_j^s = \frac{\partial c_i^s}{\partial p_j} + \frac{\partial c_i^r}{\partial w} c_j^r, \text{ as for all } w, c_j^s(\mathbf{p}^d, w) = c_j^r(\mathbf{p}^d, w) \\ &= \frac{\partial c_i^r}{\partial p_j} m_j - \frac{\partial c_i^r}{\partial w} c_j^r (1 - m_j) + \frac{\partial c_i^r}{\partial w} c_j^r, \text{ by (13)} \\ &= \left(\frac{\partial c_i^r}{\partial p_j} + \frac{\partial c_i^r}{\partial w} c_j^r \right) m_j = S_{ij}^r m_j, \text{ by (14)}. \end{aligned}$$

Proof of Proposition 8 *Efficiency implies common price misperception.* An agent a 's consumption features $u'(\mathbf{c}_a) = \lambda_a^* \mathbf{p}(m_a)$, where $\mathbf{p}(m_a)$ is the price he perceives. Suppose two consumers a, b don't perceive the same relative prices. So there are two goods – that we can call 1 and 2, such that $\frac{p_1(m_a)}{p_2(m_a)} < \frac{p_1(m_b)}{p_2(m_b)}$. That implies $\frac{u_{c_1}^a}{u_{c_2}^a} < \frac{u_{c_1}^b}{u_{c_2}^b}$. Hence, as is well-known we can design a Pareto improvement by having a and b trade some of good 1 for good 2.

Common price misperception implies efficiency. Consider an allocation $(\tilde{\mathbf{c}}^a)_{a \in A}$ that would Pareto-dominate $(\mathbf{c}^a)_{a \in A}$. Consumer a chose \mathbf{c}^a over $\tilde{\mathbf{c}}^a$, so, calling λ_a^* the Lagrange multiplier used by a in Definition 2: $u(\mathbf{c}^a) - \lambda_a^* \mathbf{p}^s \cdot \mathbf{c}^a \geq u(\tilde{\mathbf{c}}^a) - \lambda_a^* \mathbf{p}^s \cdot \tilde{\mathbf{c}}^a$, so $\lambda_a^* \mathbf{p}^s \cdot (\tilde{\mathbf{c}}^a - \mathbf{c}^a) \geq u(\tilde{\mathbf{c}}^a) - u(\mathbf{c}^a) \geq 0$, and $\mathbf{p}^s \cdot (\tilde{\mathbf{c}}^a - \mathbf{c}^a) \geq 0$, with at least one strict inequality. Summing over the a 's, and using $\sum_a \tilde{\mathbf{c}}^a = \sum_a \mathbf{c}^a = \boldsymbol{\omega}$, we obtain $\mathbf{p}^s \cdot (\boldsymbol{\omega} - \boldsymbol{\omega}) > 0$, a contradiction.

Proof of Proposition 9 In an endowment economy, $c(t) = \omega(t)$. We have $\frac{u_i(\mathbf{c}(t))}{u_1(\mathbf{c}(t))} = \frac{p_i^s(t)}{p_1^s(t)}$ 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 sparse economy (where they're not). Using $p_1^s(t) = p_1^r(t) = p_1(0)$, that implies that *the perceived price need to be the same in the sparse 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. $dp_i^{[s]} = \frac{1}{m_i} dp_i^{[r]}$; hence, with $\sigma_i^k := \frac{\text{var}(dp_i^k)^{1/2}}{p_i^k}$, $\sigma_i^s = \frac{1}{m_i} \sigma_i^r$.

Proof of Proposition 10 Proposition 3 gives: $m_i = \mathcal{A}_1 \left((\sigma_i^s)^2 \psi_i c_i p_i / \kappa \right) = \mathcal{A}_1 \left(\frac{(\sigma_i^r)^2}{m_i^2} \psi_i \frac{c_i p_i}{\kappa} \right) = 1 - \frac{m_i^2}{J_i}$, $J_i := (\sigma_i^r)^2 \psi_i \frac{c_i p_i}{\kappa}$. So $m_i = \frac{-J_i + \sqrt{J_i^2 + 4J_i}}{2}$, which increases in J_i .

References

Abaluck, Jason, and Jonathan Gruber, “Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program,” *American Economic Review* 101 (2011), 1180–1210.

- Abel, Andrew, Janice Eberly, and Stavros Panageas, “Optimal Inattention to the Stock Market with Information Costs and Transactions Costs,” *Econometrica*, 81 (2013), 1455–1481.
- Aguiar, Victor and Roberto Serrano, “Slutsky Matrix Norms and the Size of Bounded Rationality,” Working Paper, 2014.
- Akerlof, George, and Janet Yellen, “Can Small Deviations from Rationality Make Significant Differences in Economic Equilibria?” *American Economic Review*, 75 (1985), 708–720.
- Allcott, Hunt, and Nathan Wozny, “Gasoline Prices, Fuel Economy, and the Energy Paradox,” *Review of Economics and Statistics*, (forthcoming).
- Anagol, Santosh, and Hugh Hoikwang Kim, “The Impact of Shrouded Fees: Evidence from a Natural Experiment in the Indian Mutual Funds Market,” *American Economic Review*, 102 (2012), 576–93.
- Belloni, Alexandre, and Victor Chernozhukov, “ ℓ_1 -Penalized Quantile Regression in High-Dimensional Sparse Models,” *Annals of Statistics*, 39 (2011), 82–130.
- Bernheim, B. Douglas, and Antonio Rangel, “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *Quarterly Journal of Economics*, 124 (2009), 51–104.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, “Salience Theory of Choice under Risk,” *Quarterly Journal of Economics*, 127 (2012), 1243–85.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, “Salience and Consumer Choice”, *Journal of Political Economy*, 121 (2013), 803–843.
- Bolton, Patrick, and Antoine Faure-Grimaud, “Thinking Ahead: The Decision Problem,” *Review of Economic Studies*, 76 (2009), 1205–1238.
- Brocas, Isabelle, and Juan Carillo, “The Brain as a Hierarchical Organization,” *American Economic Review*, 98 (2008), 1312–1346.
- Brown, Jennifer, Tanjim Hossain, and John Morgan, “Shrouded Attributes and Information Suppression: Evidence from the Field,” *Quarterly Journal of Economics*, 125 (2010), 859–876.
- Browning, Martin, and Pierre-André Chiappori, “Efficient Intra-Household Allocations - A General Characterization and Empirical Tests,” *Econometrica*, 66 (1998), 1241–1278.
- Buchanan, James M., *Public Finance in Democratic Process*, (The University of North Carolina Press, 1967)
- Bulow, Jeremy, John Geanakoplos, and Paul Klemperer, “Multimarket Oligopoly: Strategic Substitutes and Complements,” *Journal of Political Economy*, 93 (1985), 488–511.
- Busse, Meghan, Christopher Knittel, and Florian Zettelmeyer, “Are Consumers Myopic? Evidence from New and Used Car Purchases,” *American Economic Review*, 103 (2012), 220–256.
- Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong, “A Cognitive Hierarchy Model of Games,” *Quarterly Journal of Economics*, 119 (2004), 861–898.
- Camerer, C., and Lovallo, D., “Overconfidence and Excess Entry: An Experimental Approach”, *American Economic Review*, 89 (1999), 306–318.

- Candès, Emmanuel, and Terence Tao, “Near-optimal Signal Recovery from Random Projections: Universal Encoding Strategies?” *IEEE Transactions on Information Theory*, 52 (2006), 5406–5425.
- Caplin, Andrew, and Mark Dean, “Rational Inattention and State Dependent Stochastic Choice,” Working paper, 2013.
- Carroll, Gabriel D., James Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick, “Optimal Defaults and Active Decisions,” *Quarterly Journal of Economics*, 124 (2009), 1639–1674.
- Chetty, Raj, “Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, 80 (2012), 969–1018.
- Chetty, Raj, Adam Looney, and Kory Kroft, “Salience and Taxation: Theory and Evidence,” NBER Working Paper no. 13330, 2007.
- Chetty, Raj, Adam Looney, and Kory Kroft, “Salience and Taxation: Theory and Evidence,” *American Economic Review*, 99 (2009), 1145–1177.
- Cohen, Lauren, and Andrea Frazzini, “Economic Links and Predictable Returns,” *Journal of Finance*, 63 (2008), 1977–2011.
- Crawford, Vincent, and Nagore Iriberri, “Level- k Auctions: Can Boundedly Rational Strategic Thinking Explain the Winner’s Curse and Overbidding in Private-Value Auctions?” *Econometrica*, 75 (2007), 1721–1770.
- Croce, Massimiliano, Martin Lettau, and Sydney Ludvigson, “Investor Information, Long-Run Risk, and the Duration of Risky Cash-Flows,” Working Paper, 2012.
- Cunningham, Tom, “Biases and Implicit Knowledge,” Working Paper, 2013.
- Dal Bó, Ernesto, Pedro Dal Bó and Erik Eyster, “The Demand for Bad Policy When Voters Underappreciate Equilibrium Effects,” Working Paper, 2013.
- Debreu, Gérard, “Economies with a Finite Set of Equilibria,” *Econometrica*, 38 (1970), 387–392
- DellaVigna, Stefano, “Psychology and Economics: Evidence from The Field,” *Journal of Economic Literature*, 47 (2009), 315–372
- DellaVigna, Stefano, and Joshua Pollet, “Demographics and Industry Returns,” *American Economic Review* 97 (2007), 1667–1702.
- Donoho, David, “Compressed Sensing.” *IEEE Transactions on Information Theory*, 52 (2006), 1289–1306.
- Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedel, and Joel Sobel, “Other-Regarding Preferences in General Equilibrium,” *Review of Economic Studies*, 78 (2011), 613–639.
- Ellison, Glenn and Sara Fisher Ellison, “Search, Obfuscation, and Price Elasticities on the Internet,” *Econometrica*, 77 (2009), 427–452.
- Echenique, Federico, Daniel Golovin and Adam Wierman, “A Revealed Preference Approach to Computational Complexity in Economics.” Working Paper (2013).
- Eyster, Erik, and Matthew Rabin, “Cursed Equilibrium,” *Econometrica*, 73 (2005), 1623–1672.
- Fehr, Ernst, and Jean-Robert Tyran, “Does Money Illusion Matter?” *American Economic Review*, 91 (2001), 1239–1262.

Fudenberg, Drew, and David Levine, *The Theory of Learning in Games*, (Cambridge, MA: MIT Press, 1998).

Fudenberg, Drew, and David Levine, “A Dual-Self Model of Impulse Control,” *American Economic Review*, 96 (2006), 1449–1476.

Fuster, Andreas, David Laibson, and Brock Mendel, “Natural Expectations and Macroeconomic Fluctuations,” *Journal of Economic Perspectives*, 24 (2010), 67–84.

Gabaix, Xavier, “Sparse Dynamic Programming and Aggregate Fluctuations,” Working Paper, 2013a.

Gabaix, Xavier, “Some Game Theory with Sparsity-Based Bounded Rationality,” Working Paper, 2013b.

Gabaix, Xavier, and David Laibson, “The 6D bias and the Equity Premium Puzzle,” *NBER Macroeconomics Annual* 16 (2002), 257–312.

Gabaix, Xavier, and David Laibson, “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets,” *Quarterly Journal of Economics*, 121 (2006), 505–540.

Gabaix, Xavier, David Laibson, Guillermo Moloche and Stephen Weinberg, “Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model,” *American Economic Review*, 96 (2006), 1043–1068.

Galí, Jordi, “The Return of the Wage Phillips Curve,” *Journal of the European Economic Association*, 9 (2011), 436–461.

Geanakoplos, John, and Andreu Mas-Colell, “Real Indeterminacy with Financial Assets,” *Journal of Economic Theory*, 47 (1989), 22–38.

Geanakoplos, John, and Paul Milgrom, “A Theory of Hierarchies Based on Limited Managerial Attention,” *Journal of the Japanese and International Economies*, 5 (1991), 205–225.

Gennaioli, Nicola, and Andrei Shleifer, “What Comes to Mind,” *Quarterly Journal of Economics*, 125 (2010), 1399–1433.

Greenwood, Robin, and Samuel G. Hanson, “Waves in Ship Prices and Investment,” NBER Working Paper no. 19246, 2013.

Grossman, Sanford, and Guy Laroque, “Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods,” *Econometrica*, 58 (1990), 25–51.

Gul, Faruk, Wolfgang Pesendorfer and Tomasz Strzalecki, “Coarse Competitive Equilibrium and Extreme Prices,” Working Paper, 2014.

Hansen, Lars, and Thomas Sargent, *Robustness*, (Princeton, NJ: Princeton University Press, 2007).

Hassan, Tarek, and Thomas M. Mertens, “Market Sentiment: A Tragedy of the Commons,” *American Economic Review, Papers and Proceedings*, 101 (2011), 402–05.

Hirshleifer, David, Sonya Lim and Siew Teoh, “Driven to Distraction: Extraneous Events and Underreaction to Earnings News,” *Journal of Finance*, 64 (2009), 2289–2325

Jéhiel, Philippe, “Analogy-Based Expectation Equilibrium,” *Journal of Economic Theory*, 123

(2005), 81–10.

Kahneman, Daniel, “Maps of Bounded Rationality: Psychology for Behavioral Economics,” *American Economic Review*, 93 (2003), 1449–1475.

Kahneman, Daniel, *Thinking, Fast and Slow*, (Farrar, Straus and Giroux, 2011).

Kerschbamer Rudolf and Georg Kirchsteiger, “Theoretically Robust but Empirically Invalid? An Experimental Investigation into Tax Equivalence,” *Economic Theory*, 16 (2000), 719–734.

Kőszegi, Botond, and Adam Szeidl, “A Model of Focusing in Economic Choice,” *Quarterly Journal of Economics*, 128 (2013), 53–104.

Kreps, David M., *Microeconomic Foundations I*, (Princeton, NJ: Princeton University Press, 2012).

Krusell, Per, and Anthony Smith, “Rules of Thumb in Macroeconomic Equilibrium: A Quantitative Analysis,” *Journal of Economic Dynamics and Control*, 20 (1996), 527–558.

Loewenstein, George, Ted O’Donoghue and Matthew Rabin, “Projection Bias in Predicting Future Utility,” *Quarterly Journal of Economics*, 118 (2003), 1209–1248.

Lucas, Robert, “Expectations and the Neutrality of Money,” *Journal of Economic Theory*, 4 (1972), 103–124.

Maćkowiak, Bartosz and Mirko Wiederholt, “Optimal Sticky Prices under Rational Inattention,” *American Economic Review*, 99 (2009), 769–803.

Mas-Colell, Andreu, Michael Whinston, and Jerry Green, *Microeconomic Theory*, (Oxford: Oxford University Press, 1995).

Matějka, Filip and Christopher Sims, “Discrete Actions in Information-constrained Tracking Problems,” Working Paper, 2010.

Mankiw, N. Gregory, and Ricardo Reis, “Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *Quarterly Journal of Economics*, 117 (2002), 1295–1328.

Miller, George, “The Magical Number Seven Plus or Minus Two: Some Limits on our Capacity for Processing information,” *Psychological Review*, 63 (1956), 81–97.

Mullainathan, Sendhil, “A Memory-Based Model of Bounded Rationality,” *Quarterly Journal of Economics*, 117 (2002), 735–774.

Piccione, Michele and Ran Spiegler, “Price Competition under Limited Comparability,” *Quarterly Journal of Economics*, 127 (2012), 97–135.

Radner, Roy, and Timothy Van Zandt, “Real-Time Decentralized Information Processing and Returns to Scale,” *Economic Theory*, 17 (2001), 545–575.

Rubinstein, Ariel, *Modeling Bounded Rationality*, (Cambridge, MA: MIT Press, 1998).

Samuelson, Paul, *Foundations of Economic Analysis*, (Cambridge, MA: Harvard University Press, 1947).

Sargent, Thomas, *Bounded Rationality in Macroeconomics*, (Oxford: Oxford University Press, 1993).

Sausgruber, Rupert, and Jean-Robert Tyran, “Testing the Mill Hypothesis of Fiscal Illusion,” *Public Choice*, 122 (2005), 39–68.

Schwartzstein, Joshua, “Selective Attention and Learning,” *Journal of the European Economic Association* (forthcoming).

Shafir, Eldar, Peter Diamond and Amos Tversky, “Money Illusion,” *Quarterly Journal of Economics*, 112 (1997), 341–74.

Sims, Christopher, “Implications of Rational Inattention,” *Journal of Monetary Economics*, 50 (2003) 665–690.

Thaler, Richard H., “Mental Accounting and Consumer Choice,” *Marketing Science*, 4 (1985), 199–214.

Tibshirani, Robert, “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society B*, 58 (1996), 267–288.

Tversky, Amos, “Elimination by Aspects: A Theory of Choice,” *Psychological Review*, 79 (1972), 281–299.

Tversky, Amos, and Daniel Kahneman, “Judgment under Uncertainty: Heuristics and Biases,” *Science*, 185 (1974), 1124–1130.

Varian, Hal, *Microeconomic Analysis*, 3rd edition. (Norton, 1992).

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

Woodford, Michael, “Inattentive Valuation and Reference-Dependent Choice,” Working Paper, 2012.