Online Appendix for "A Sparsity-Based Model of Bounded Rationality"

Xavier Gabaix, July 2014

This appendix presents some thoughts on using sparse max in practice; applications of the sparse max (e.g. classic behavioral biases, pricing by a monopolist and induced sticky prices, attention to shrouded attributes); an analysis of (sub) optimization under constraints; omitted proofs; and some reasonable variants.[69]

VIII. Notes on the Practical Application of the Sparse Max

Here are some reflections on using sparse max, mostly drawn from the consumption theory application in this paper, and from Gabaix (2013a).

How to specify the default, i.e. $x = 0$**?** Typically, the default consists of setting all variables to their expected value until a detailed analysis is possible – e.g. their average. E.g., in the consumption example, $\boldsymbol{p}^d = \mathbb{E}[\boldsymbol{p}]$.[70] This "expected value" could be something replaced by something psychologically richer, like the "value using rounded numbers" (if the average price is $\mathbb{E}[\boldsymbol{p}] = \$4.1$, consumers might round it to a default price $p^d = \$4$). For most economic applications, the expectation seems a sensible default.

What does the agent know for free? In the static consumption case, the agent must exhaust his budget w, so he will know it (even if that's in step 2). In other cases, knowledge of cash-on-hand is particularly plausible, while financial wealth in a retirement account might be an x variable that might be only sparsely accessed before retirement.

What about covariances? Dealing with covariances is a bit delicate with pencil and paper, though it is easy to do it numerically (as we have a convex problem).[71] One worked-out example where covariances matter is the "nominal wage / real wage" problem discussed in the paper; see Lemma 3 below.

In addition, in many cases, people may ignore covariances. So, one may imagine that they "sparsify" the covariance matrix and just keep variance terms. For instance, a statistician might

[69] I thank Deepal Basak for very good research assistance.

[70] It's both sensible and optimal (in linear-quadratic cases) to replace a variable by its certainty equivalent (e.g. its mean), even if the mean is not literally observed.

[71] The quadratic case $\alpha = 2$ is easier, though it does not generate sparsity: the solution is then $\arg\min_m \frac{1}{2}(m-\iota)' \Lambda (m-\iota) + \kappa m'm$, which gives $\Lambda(m-\iota) + 2\kappa m = 0$, hence $m = (2\kappa I + \Lambda)^{-1} \Lambda \iota$. That solution, however, does not necessarily satisfy $m_i \in [0,1]$.

know that, if GDP growth is high, then the price of salmon (a luxury good) is likely to be high. But a consumer, seeing that his wage is high, may not infer that the price of salmon might be high. That could be modeled via a "sparse use" of statistical knowledge, but that would take us too far for now.

Anchoring and adjustment One spectacular demonstration of anchoring and adjustment is in Tversky and Kahneman (1974): a transparently random roulette wheel is stopped, returns a number; then, subjects are asked about the number of African countries in the UN. Surprisingly, if the number returned by the roulette wheel is higher, people guess a higher number of African countries.

A sparse agent may (qualitatively at least) behave much like a Kahneman-Tversky subject. The number spun by the wheel becomes a default value of x^d. Then, there is an adjustment via a more effortful process. The answer is: $a = x^d + m\left(x - x^d\right)$, where x is the truth. This gives partial anchoring on the default.

Here the sparsity model is just one simple way to think about it. It just builds on the classic "Gaussian signal extraction model, with default at x^d, and partial adjustment to the truth", which gives the same answer under some conditions (Proposition 16). I do not wish to claim that sparsity "explains" anchoring – just that its behavior is much like that of anchoring (perhaps for similar reasons, i.e. contextual information is usually helpful and should be relied upon). That "explanation" is a contentious debate. It would be interesting to systematically and quantitatively investigate these "anchoring" effects, comparing different calibrated models. The sparse max is at least one simple model to think about this.

Calibration We can venture a word about calibration. As a rough baseline, we can imagine that people will search for information that accounts for at least $\xi = 10\%$ of the variance of the decision, i.e., if $|\mu_i|^2 \sigma_i^2 < \xi \sigma_a^2$. That means that we must have $\kappa / |u_{aa}| = \xi \sigma_a^2$, i.e. $\kappa = \xi |u_{aa}| \sigma_a^2$. The "cost" κ should scale like $|u_{aa}| \sigma_a^2$, in particular proportionally to marginal utility. That leads (using the scale-free κ) to $\overline{\kappa} = \xi \simeq 0.3$. In general, in applications, it will be useful to use such scaling (see Gabaix 2013a,b).

The default: complements The present model has a very simple model of the default: if $\widetilde{x}$ is draw from a distribution, the default x is $x^d = \mathbb{E}[\widetilde{x}]$. In the paper, we normalized it to $x^d = 0$. More generally, we could write (as in the original definition of the sparse max):

$$\operatorname*{smax}_{a|\kappa,\sigma,x^d} u(a,x) \text{ subject to } b(a,x) \geq 0,$$

where the definition is the one in the paper, replacing x by $x - x^d$.

A truly satisfactory model of the default would entail deep, difficult issues, like: the "evoked set" given a problem at hand, and how a "network of associations of ideas" creates a default (perhaps like in the roulette wheel and African nations of Kahneman-Tversky). I defer this to future research. The present paper, though, makes progress on modelling the choice to deviate from the default, allocating attention parsimoniously, and its economic consequences.

Even though I will not provide a deep elaboration on the default, I submit some remarks on a potential elaboration of the default. One could imagine taking a default action $a^d = \arg\max_a \mathbb{E}\left[u\left(a, \widetilde{x}\right)\right]$, averaging over the stochastic realization of $\widetilde{x}$. For the quadratic problem in the paper, that leads to the same outcome, but for other function that might lead to a different behavior: e.g. the agent will have a "precautionary saving" type of behavior.

What goes into the expectation operator $x^d = \mathbb{E}\left[\widetilde{x}\right]$? One benchmark is the rational expectations: then the $x_i - x_i^d$ is the classic "unexpected inflation". However, a richer way would be to say that the agent may (at some cost) choose to change his default, or to keep it fixed. For instance, take prices, that would be the "expected price". If the agent has gotten "used" to a 10% annual inflation (either "he has paid the cost of learning it" or "he learned it for free from experience"), it will be used in the formation of the expectation. However, if it's a surprise inflation, in the sense that it's above the level anticipated in the default model of the agent, then the extra inflation will be one more variable x_i and will typically only partially be taken into account, generating nominal illusion. This mechanism is worked out in detail in section IX.D.

To very cleanly think about those issues, one might need a language of lazy (or sparse) updating of information. Doing that systematically would require a full paper. I hope that these remarks clarify enough the practical choices in deciding what should be in the default x.

How much endogenization is useful? Sometimes, the structure of attention m matters a good deal (e.g., Example 1, Example 2, Propositions 3 and 10, section IX.B, section IX.C, Gabaix 2013a). Sometimes however, just the existence of $m < 1$ is enough to get an interesting effect. Then, we can recommend using the attention function $\mathcal{A}_1$ (we saw why in section II.A), and an exogenous κ. If κ needs to be endogenized, then I recommend version (19), which is easy to use and sensible.

I now turn to concrete applications of the model.

IX. Some Other Applications of Sparse Max

IX.A. Classic Behavioral Biases

Here I revisit some classic behavioral biases, and see how they can be expressed in the sparsity language – and sometimes we get intriguing predictions.

A full investigation of sparsity's consequence for the "classic" Kahneman-Tversky (henceforth, KT) biases would be an interesting endeavor for future research. Those biases have been modeled now by quite a few researchers (notably Bordalo, Gennaioli and Shleifer (2012, 2014), Kőszegi and Szeidl (2013), Rabin and coauthors), whereas a systematic behavioral of textbook microeconomics is the more distinctive application of the sparsity approach.

Base-rate neglect The agent sees a signal S. There are n disjoint hypotheses (i.e. states) with probability $P(H_i)$. The conditional probability of the signal for each hypothesis is $P(S \mid H_i)$. The agent is asked the probability of hypothesis H_1 given the signal S, $P(H_1 \mid S)$. Tversky and Kahneman (1974) show that people exhibit base-rate neglect, i.e. partially neglect $P(H_i)$. We shall see that a sparse agent will do something similar.

We call $\overline{p}_B = \frac{1}{n}$ the average of base rates, and $\overline{p}_C = \frac{1}{n}\sum_i P(S \mid H_i)$ the average of the conditional probabilities. The agent chooses how much attention to pay to the specific probabilities; if no attention is paid, he replaces them by their average value. Hence, calling m_B the attention to the base rate and m_C the attention to the conditional probability, the agent's perception of the problem is:

$$\begin{aligned} P^s(c; m) &= \overline{p}_B(1 - m_B) + m_B P(c), \\ P^s(S \mid H_i; m) &= \overline{p}_C(1 - m_C) + m_C P(S \mid H_i). \end{aligned}$$

The agent is asked $P(H_1 \mid S)$. He will form $P^s(H_1 \mid S; m)$ as in Bayes' rule, but with attention vector $m = (m_B, m_C)$:

$$P^s(H_1 \mid S; m) := \frac{P^s(S \mid H_1; m)\, P^s(H_1; m)}{\sum_{i=1}^{n} P^s(S \mid H_i; m)\, P^s(H_i; m)}.$$

Applying the sparse max (formally, to $\frac{-1}{2}\left(a - P(H_1 \mid S)\right)^2$) gives:[72]

$$m_B = \mathcal{A}_\alpha\left(\sigma_{p_B}^2/\kappa\right), m_C = \mathcal{A}_\alpha\left(\left(\frac{\overline{p}_B}{\overline{p}_C}\right)^2 \sigma_{p_C}^2/\kappa\right),$$

where $\sigma_{p_B}^2$ (resp. $\sigma_{p_C}^2$) is the variance of base rates (resp. conditional probabilities).

If $\frac{\sigma_{p_B}}{\overline{p}_B} < \frac{\sigma_{p_C}}{\overline{p}_C}$, then $m_B < m_C$, and we obtain "base rate neglect". In this sparse max, the attention weights are endogenous, so they respond to incentives (Gennaioli and Shleifer (2010) and Bordalo, Gennaioli and Shleifer (2014) offer interesting, and very different models, with a more discontinuous response to incentives).

[72] The derivation is simple, and uses: $\partial_{m_B} P^s(H_1|S;m)_{|m=0} = P(H_1) - \overline{p}_B$ and $\partial_{m_C} P^s(H_1|S;m)_{|m=0} = \frac{\overline{p}_B}{\overline{p}_C}\left(P(S|H_1) - \overline{p}_C\right)$.

This is the key difference between this approach and previous procedures with fixed, exogenous weights (e.g. Grether (1980), Bodoh-Creed, Benjamin, Rabin (2013)). For instance, if the base rates become more extreme, the agent is predicted to pay more attention to them. Lynch and Ofir (1989) find that base rates are rated as less relevant only when the conditional probabilities are extreme.[73] Koehler (1996) reviews the literature (which is not unanimous in its findings) and concludes "when base rates are made more extreme or when individuating information is made less diagnostic, the impact of base rates on judgments increases".

This perspective shows that base rate neglect can be expected on average: this is the case if $\frac{\sigma_{p_B}}{\overline{p}_B} < \frac{\sigma_{p_C}}{\overline{p}_C}$. This is the case for many natural setups with "strongly informative signals" (e.g. a cat looks like a cat, and other animals don't). However, in other environments with "weak signals", then people just don't update. For instance, when there is no signal, people just don't update – again, unlike the simple mechanical models.

In sum, a procedure with endogenous m can give additional sensible comparative statics on how biases respond to the environment. A full treatment of these issues (including continuous distributions etc.) is delegated to future research.

Projection bias The agent has to guess a variable y_{t+1} (for instance, his future hunger), given y_t (current hunger). The true relation is $y_{t+1} = (1-\phi)\, y_t + \varepsilon_{t+1}$. A sparse agent may not see the mean-reversion, i.e. he may instead use a relation such as:

$$y_{t+1} = (1 - m\phi)\, y_t + \varepsilon_{t+1},$$

where m indicates the attention to mean-reversion. Loewenstein, O'Donoghue and Rabin (2003) have a related formulation, with an exogenous m (in their formulation, the full utility is extrapolated, not the variables "inside the utility" – which seems useful here). Here, with sparsity we can endogenize the m, which is exogenous in previous work.

This model makes simple, plausible predictions. Suppose the agent is asked either (i) "decide how much you should buy now for today and tomorrow" (which depends on today's and tomorrow's hunger), or (ii) "decide how much you should buy for tomorrow". There will be more projection bias in case (i) than (ii): this because the mean-reversion matters relatively more in question (ii) than question (i), so the agent will pay more attention to it (this is if we use the scale-free κ).[74]

Two other comparative statics emerge (which are generic to this cost-benefit model, but absent

[73] In the lawyer / engineer problem, a rich description tends to attenuate the base rate (TK 1974). It could be that because this rich description looks quite diagnostic of lawyers for some subjects, of engineers for others (though the diagnoses cancel out on average): this strong diagnosticity leads to more attention to the description, less to the base rate (under the "scale-free" κ version).

[74] To see this, take the objective function $u = v(c_1 - n_1 - h) + v(c_2 - n_2 - h(1-\phi)) - p(c_1 + c_2)$, where $a = (c_1, c_2)$ is the decision variable, n_t is a "need" that raises expenditure at t, h is hunger at time 1, and p is the price of consumption. We assume $x = (n_1, n_2, \phi)$. Calling $c_* := v'^{-1}(p)$, and $m := (m_1, m_2, m_\phi)$, we have $c_1^s = c_* + m_1 n_1 + h$,

from a model with exogenous m). First, the projection bias will be bigger for relatively unimportant things. Second, a very distracted agent (to the extent that this increases κ) will suffer more from projection bias.

Insensitivity to predictability / Misconceptions of regression to the mean / Illusion of validity Given performance y_t of an airline pilot (say), an agent predicts the performance next period (TK 1974): they should give $\overline{y}_{t+1} := \mathbb{E}[y_{t+1} \mid y_t]$. Suppose the true model is $y_t = a + \varepsilon_t$, where a is core ability, ε_t is noise. Then, we should have $\overline{y}_{t+1} = \mathbb{E}[a \mid y_t]$.

When people form their judgment, they often "forget about reversion to the mean", i.e. say $\overline{y}_{t+1} = y_t$, rather than $\overline{y}_{t+1} = \lambda y_t$ with a dampening factor $\lambda < 1$ (normatively, $\lambda = \frac{1}{1+\frac{\sigma_\varepsilon^2}{\sigma_a^2}}$).

We can interpret that they have a model where they might "forget about the noise", i.e. in their perceived model, $var^s(\varepsilon) = m\sigma_\varepsilon^2$. This is, they choose whether or not to think about the noise. If they don't think about the existence of the noise, then they will just answer $\overline{y}^s_{t+1} = y_t$. It's as if they used a model with $\sigma_\varepsilon = 0$. In general, the prediction is

$$\overline{y}^s_{t+1}(m) = \frac{1}{1+\frac{\sigma_\varepsilon^2}{\sigma_a^2}m} y_t \text{ with } m = \mathcal{A}_\alpha\left(\frac{u_{aa}^2}{\kappa}\frac{\sigma_\varepsilon^4}{\sigma_a^4}\right).$$

Hence, when the noise isn't very large, people just answer (as $m = 0$): $\overline{y}^s_{t+1} = y_t$. A slightly richer setup will give the conclusion of Tversky and Kahneman (1974) that people will overestimate the consequences of punishment.

Inattention to sample size: conservatism, strength vs weight. People seem insensitive to sample size when making statistical judgments (TK 1974). In the sparsity language, one might say that instead of taking into account the sample as N, the agent perceives the sample to be $N^s(m) = N^m$. This is, the innovation is $x = \ln N$. Suppose that the answer depends on N, i.e. is $a(N, x)$, then attention is $m = \mathcal{A}_\alpha\left(\frac{u_{aa}^2 a_N^2}{\kappa}\mathbb{E}\left[(\ln N)^2\right]\right)$. The dependence on $\ln N$ could be a factor that leads people to be quite generally insensitive to sample size.[75]

This leads to Edwards (1968)'s conservatism, and Griffin and Tversky (1992)'s lack of attention to the "weight" of the evidence. In the sparsity model, this would be because the "weight" (as measured by $\mathbb{E}\left[(\ln N)^2\right]$) generally varies less than the "strength" (as measured by the signal,

$c_2^s = c_* + m_2 n_2 + h(1 - m_\phi \phi)$. Applying the sparse max gives $m_\phi = \mathcal{A}_\alpha\left(\frac{v''(c_*)h^2\phi^2}{\kappa}\right)$, $m_t = \mathcal{A}_\alpha\left(\frac{v''(c_*)\sigma_{n_t}^2}{\kappa}\right)$. Using the scaled-free κ, in question (i), we obtain $m_\phi^{(i)} = \mathcal{A}_\alpha\left(\frac{1}{\kappa}\frac{h^2\phi^2}{\sigma_{n_1}^2+\sigma_{n_2}^2+h^2\phi^2}\right)$, while in question (ii), $m_\phi^{(ii)} = \mathcal{A}_\alpha\left(\frac{1}{\kappa}\frac{h^2\phi^2}{\sigma_{n_2}^2+h^2\phi^2}\right)$, so people are less distracted, and pay more attention to the mean-reversion in hunger in problem (ii) than problem (i).

[75] Of course, some people just don't know the underlying statistical law that the variance of a mean decrease like the inverse of the sample size. That might be model as a very high κ_i for this feature.

which is usually absent, and suddenly strong). A proper calibration would entail looking at the "ecologically relevant" typical weight and signal strength.

Conclusion on KT-style biases Many KT biases seem closely linked to simplification and inattention. Sparsity seems useful to model how inattention varies with the (perceived) benefits of attention (as seems warranted by some experimental evidence). It would take another paper to assess the performance of this modelling approach, including a systematic calibration. We simply note that this is potentially promising, and leave that full study to future research.

IX.B. Optimal Monopoly Pricing and Sparsity-Induced Price Stickiness and Sales

I study the behavior of a monopolist facing a boundedly rational consumer who inattentive to small price changes. This will lead to endogenous price stickiness (even though the monopolist face no menu costs), and "sales".

The consumers has the utility function $u(Q,y) = y + Q^{1-1/\psi}/(1-1/\psi)$ when he consumes a quantity Q of the good and has a residual budget y. So, if the price is p, the demand is $D(p) = p^{-\psi}$ where $\psi > 1$ is the demand elasticity.[76] The consumer uses the sparse max; his demand is:

$$D^s(p) = D\left(p^d + \tau\left(p - p^d, \kappa\right)\right), \tag{26}$$

using the thresholding function defined[77] in section XV.E, where $\kappa = \sqrt{\frac{\kappa^m}{p^d c^d \psi \sigma^2_{\ln p}}}$) (calling κ^m the basic cost parameter κ in Definition 1). Hence, the consumer is insensitive to price changes when $p \in (p^d - \kappa, p^d + \kappa)$.[78] The default price p^d will be endogenized later to be the average price.

The monopolist picks p to maximize profits: $\max_p (p-c) D^s(p)$ where c is the marginal cost (in this section, to conform to the notations of the optimal pricing literature, c denotes a marginal cost rather than consumption). The following proposition describes the optimal pricing policy.

[76] Previous work on rational firms and inattentive consumers includes Heidhues and Kőszegi (2010) with loss-averse consumers, L'Huillier (2010) with differently-informed consumers, and Matejka (2010) with a Sims (2003)-type entropy penalty. Their models are quite different from the one presented here in specific assumptions and results. Still, there is a common spirit that behavioral consumers can lead to interesting behavior by rational firms. At a minimum, the present paper offers a particularly transparent and tractable version of this theme. Chevalier and Kashyap (2011) offer a theory of price stickiness and sales based on agents with heterogeneous search costs.

[77] I use this because it leads to very simple expressions. Other truncation functions would work similarly.

[78] This is a testable implication: the price elasticity of demand is the smaller the closer the price is to its default.

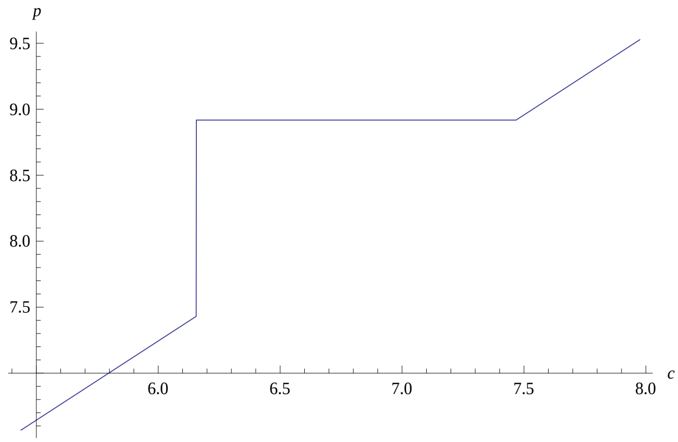

FIGURE V: Optimal price p set by the monopolist facing boundedly rational consumers, as a function of the marginal cost c.

PROPOSITION 23 *With a sparse consumer, the monopolist's optimal price is:*

$$p(c) = \begin{cases} \frac{\psi c+\kappa}{\psi-1} & \text{if } c \leq c_1, \\ p^d + \kappa & \text{if } c_1 < c \leq c_2, \\ \frac{\psi c-\kappa}{\psi-1} & \text{if } c > c_2, \end{cases} \tag{27}$$

where $c_1 = c^d - 2\sqrt{c^d \kappa/\psi} + O(\kappa)$ *solves equation (31), with* $c_2 = c^d + \kappa$ *where* $c^d := (1 - 1/\psi)p^d$ *is the marginal cost that would correspond to the price* p^d *in the model without cognitive frictions. The pricing function is discontinuous at* c_1 *and continuous elsewhere.*

Let us interpret Proposition 23. When $p \in (p^d - \kappa, p^d + \kappa)$, the demand $D^s(p)$ is insensitive to price changes. Therefore, the monopolist will not charge a price $p \in (p^d - \kappa, p^d + \kappa)$: he will rather charge a price $p = p^d + \kappa$. This yields a whole interval of prices that are not used in equilibrium, and significant bunching at $p = p^d + \kappa$. There, the price is locally independent of the marginal cost. This is a real "stickiness."[79] This effect is illustrated in Figure V.[80]

For a low enough marginal cost c, the price falls discretely, like a "sale". There is a discrete jump below the modal price, but not above it. The asymmetry is due to the fact that in the inattention region $(p^d - \kappa, p^d + \kappa]$ the firm wishes to set a high price $p^d + \kappa$ rather than a low price. Hence, when we leave the inattention region, the price rises a bit above $p^d + \kappa$, or otherwise jumps discretely below $p^d - \kappa$.

The cutoff c_1 is much further below c^d than c_2 is above it. It deviates from the baseline c^d

[79] If the consumer's default is in nominal terms and mentally adjusting for inflation is costly, this model can easily yield nominal stickiness.

[80] The assumed values are $\psi = 6$, $p^d = 8.7$, and $\kappa = 0.025p^d$. They imply $\kappa = 0.22$, $c^d = 7.25$, $c_1 = 6.16$, $c_2 = 7.46$, $p(c_1) = 7.43$, and $p(c_2) = 8.92$.

proportionally to $\sqrt{\kappa}$ whereas $c_2 = c^d + \kappa$.[81]

This simple model seems to account for a few key stylized facts. Prices are "sticky," with a wide range being insensitive to marginal cost. This paper predicts "sales": a temporary large fall in the price after which the price reverts to exactly where it was (if c goes back to (c_1, c_2)). This type of behavior is documented empirically in Klenow and Malin (2011). In addition, the model says that the typical size of sales will be $p(c_2) - p(c_1)$, i.e., to the leading order

$$p(c_2) - p(c_1) = 2\sqrt{\frac{\kappa p^d}{\psi - 1}}, \tag{28}$$

where $\kappa = \sqrt{\frac{\kappa^m}{p^d c^d \psi \sigma^2_{\ln p}}}$. Hence, the model makes the testable prediction that *the gap in the distribution of price changes, and the size of sales, is higher for goods with low consumption volatility, for goods that are less price elastic, and for goods with a low expenditure share.*

To close the model, one needs a theory of the default price. In a stationary environment, the most natural way is to specify p^d to be the average empirical price

$$p^d = \mathbb{E}\left[p\left(\widetilde{c}, p^d\right)\right], \tag{29}$$

given the distribution over the marginal costs $\widetilde{c}$. By the implicit function theorem, for sufficiently small κ and a smooth non-degenerate distribution of costs,[82] there is a fixed point p^d. In the small κ limit, one can show that $p^d = \frac{\psi}{\psi-1}\overline{c} + \frac{2\overline{c}f(\overline{c})+2F(\overline{c})-1}{\psi-1}\kappa + o(\kappa)$ with $\overline{c} := \mathbb{E}[c]$ (the derivations are below). Hence, the default price is higher than it would be in the absence of bounded rationality.

The model is robust to some form of consumer heterogeneity. The key is that the aggregate demand function $D(p)$ has kinks. Hence, if there are, for example, two types of agents—two $p_i^d + \kappa_i$ with $i \in \{1, 2\}$—then we might also expect two reference prices.

This example illustrates that it is useful to have a tractable model, such as the sparse max, to think about the consequences of bounded rationality in market settings.[83] Also, the sparse max model is designed to generate inattention in the first place, not price stickiness and sales. Rather, it generates a potential new approach to price stickiness as an unexpected by-product.

[81] There is also a more minor effect. For very low marginal cost, consumers do not see that the price is actually too low: they replace p by $p + \kappa$. Hence, they react less to prices than usually (demand is less elastic), which leads the monopolist to raise prices. For high marginal cost, consumers replace the price by $p - \kappa$, so their demand is more elastic, and the price is less than the monopoly price.

[82] For instance, it is enough to have a density $f(c)$ that is strictly positive and continuous around $\overline{c}$.

[83] For instance, much of the analysis will carry over to a closely related setup where consumers are inattentive to the decimal digits of the price, i.e., $D^s(n + x) = D^s(n)$ for n a positive integer and $x \in [0, 1)$. There will be bunching at a price like \$2.99. Likewise, one can solve the model with a fixed cost of thinking. It still yields price rigidity but loses the "sales" effect: there are two discontinuities in the optimal price function, rather than one.

Proof of Proposition 23. The monopolist solves

$$\max_p \pi(p), \qquad \pi(p) = (p-c)\left(p^d + \tau\left(p - p^d, \kappa\right)\right)^{-\psi}.$$

Consider first the interior solutions with $p \notin \left(p^d - \kappa, p^d + \kappa\right)$. Call $\varepsilon = sign\,(p - p_d)$. Then, $p^d + \tau\left(p - p^d, \kappa\right) = p - \varepsilon\kappa$. Therefore, $\partial_p \tau\left(p - p^d, \kappa\right) = 1$, and the f.o.c. is $p - \varepsilon\kappa - \psi(p - c) = 0$, i.e.,

$$p = p^{int} := \frac{\psi c - \varepsilon\kappa}{\psi - 1}. \tag{30}$$

The profit is

$$\pi\left(p^{int}\right) = \left(\frac{\psi c - \varepsilon\kappa}{\psi - 1} - c\right)\left(\frac{\psi c - \varepsilon\kappa}{\psi - 1} - \varepsilon\kappa\right)^{-\psi} = \psi^{-\psi}\left(\frac{c - \varepsilon\kappa}{\psi - 1}\right)^{1-\psi}.$$

Next, it is not optimal for the monopolist to have $p \in \left(p^d - \kappa, p^d + \kappa\right)$ as $p = p^d + \kappa$ yields the same demand and strictly higher profits. The profit is

$$\pi\left(p^d + \kappa\right) = \left(p^d + \kappa - c\right)\left(p^d\right)^{-\psi}.$$

It is optimal to choose p^{int} rather than $p^d + \kappa$ iff $R \geq 1$ where

$$\begin{aligned} R\left(c, c^d, \kappa\right) &= \frac{\pi\left(p^{int}\right)}{\pi\left(p^d + \kappa\right)} = \frac{\psi^{-\psi}\left(\frac{c - \varepsilon\kappa}{\psi - 1}\right)^{1-\psi}}{\left(\frac{\psi}{\psi - 1}c^d + \kappa - c\right)\left(\frac{\psi}{\psi - 1}c^d\right)^{-\psi}} \\ &= \frac{(c - \varepsilon\kappa)^{1-\psi}}{\left[\psi c^d + (\psi - 1)(\kappa - c)\right]\left(c^d\right)^{-\psi}}. \end{aligned}$$

The cutoffs c_1 and $c_2 > c_1$ are the solution to $R\left(c_i, c^d, \kappa\right) = 1$. The c_2 bound is easy to find because it is clear (as the profit function is increasing for $p < p^{int}$) that c_2 must be such that $p^{int}(c_2) = p^d + \kappa$, i.e., $\frac{\psi c_2 - \kappa}{\psi - 1} = \frac{\psi c^d}{\psi - 1} + \kappa$, so $c_2 = c^d + \kappa$. The more involved case is the one where $c < c^d$ as then there can be two local maxima (this is possible as the demand function is not log-concave). Hence, the cutoff c_1 satisfies, with $\varepsilon = -1$,

$$R\left(c_1, c^d, \kappa\right) = 1, \tag{31}$$

and $c_1 < c_d$. To obtain an approximate value of c_1, note that $R(c, c, 0) = 1$: when $\kappa = 0$, the cutoff corresponds to $c = c^d$. Also, calculations show $R_1(c, c, 0) = 0$ and $R_{11}(c, c, 0) \neq 0$. Hence, a small κ implies a change δc_1 such that, to the leading order, $\frac{1}{2}R_{11} \cdot (\delta c)^2 + R_3 \cdot \kappa = 0$, i.e., $c_1 = c^d - \sqrt{\frac{-2R_3\kappa}{R_{11}}} + +O(\kappa)$. Calculations yield $c_1 = c^d - 2\sqrt{c^d\kappa/\psi} + O(\kappa)$. ■

IX.C. Allocation of Attention with a Base vs Add-On Good

Extant analyses of opaque or shrouded attributes typically assume that more attention is paid to the base good (e.g. the printer) rather than the add-on (e.g. the cartridge, see Gabaix and Laibson 2006, building on Ellison 2005). We shall see how this can be derived, rather than merely assumed – so that we can see more precisely when attention to the add-on good is likely to be non-zero. Note that here, these attributes are overlooked simply because they do not matter very much, not because they are hidden. The latter dimension could be incorporated as a larger $\kappa_i g(m_i)$, with κ_i high for "hidden" things, but we won't use that degree of freedom here.

We take a utility:

$$u(c_1, c_2, c_3) = \frac{g^{1-1/\psi}}{1-1/\psi} + c_3, \qquad g := \left(c_1^{1-1/\phi} + c_2^{1-1/\phi}\right)^{\phi/(\phi-1)},$$

where c_1 is the consumption of the base good, c_2 is the add-on consumption, g the aggregate good made of c_1 and c_2, and c_3 represents linear utility for residual wealth. We consider the case where $\phi < \psi$, i.e. the add-on and base goods are fairly complementary (their elasticity of substitution is ϕ), but the good as a whole has a high elasticity of demand, ψ.

The traditional, rational case is solved as usual. We normalize $p_3 = 1$. The price of composite good g is $p_g = \left(\sum_{i=1}^2 p_i^{1-\phi}\right)^{1/(1-\phi)}$, and its demand is $g = p_g^{-\psi}$. The demand for good $i < 3$ is $c_i = (p_i/p_g)^{-\phi} g = p_i^{-\phi} p_g^{\phi-\psi}$. The share of the total expenditure on the base good, spent on good i, $f_i = \frac{p_i c_i}{\sum_{j=1}^2 p_j c_j}$, is $f_i = \frac{p_i^{1-\phi}}{\sum_{j=1}^2 p_j^{1-\phi}}$.

Consider now the sensitivity to a small price change $d\ln p_i = d\pi_i$. We have: $d\ln p_g = \sum f_j d\pi_j$, and using $c_i = p_i^{-\phi} p_g^{\phi-\psi}$, we obtain $d\ln c_i = -\phi d\pi_i - (\psi - \phi)\, d\ln p_g$, and

$$d\ln c_i = -\sum_{j=1}^{e} \psi_{ij} d\pi_j, \tag{32}$$

where $\psi_{ij} := (\psi - \phi) f_j$ for $j \neq i$, and

$$\psi_{ii} \equiv \psi_i \equiv f_i \psi + (1 - f_i)\, \phi, \tag{33}$$

is own-price elasticity of demand for good i. This is the same ψ_i defined in Proposition 3 (which can also be calculated directly). If $f_i \simeq 1$, so that the base good is really the whole good, then its elasticity should be ψ. However, if $f_i \simeq 0$, so that the good is a small add-on, its own-price elasticity should be the elasticity "inside the composite good," ϕ.

From now on, we consider the case where $d\pi_1, d\pi_2$ are uncorrelated with mean 0 and standard deviations σ_1, σ_2 respectively, and m_i the attention to price i. Applying Proposition 3, there is some attention to the price of the base good, and no attention to the add-on, iff:

$$p_1^d c_1^d \psi_1 \sigma_{\pi_1}^2 > \kappa' \geq p_2^d c_2^d \psi_2 \sigma_{\pi_2}^2,$$

where $\kappa' = \frac{\kappa}{2}$ in the fixed cost case ($\alpha = 0$), and $\kappa' = \kappa$ in the linear case ($\alpha = 1$).

Other psychological sources could be profitably added to this example. For instance, if the add-on is purchased later, this may be an added source of inattention. We defer this extension to future research.

IX.D. Thinking About Nominal vs Real Quantities

Let us develop the "nominal vs real" frame in the paper.[84] Say the real wage increase is R, inflation is π (we normalize expected inflation to 0, and assume that R and π are uncorrelated), and the data are $x_1 = R + \pi$, the nominal wage increase, and $x_2 = \pi$, inflation.

The utility function is $u(a, x) = -\frac{1}{2}(a - R)^2 = -\frac{1}{2}(a - x_1 + x_2)^2$. The consumer is asked to predict his real wage, so $a^r = R = x_1 - x_2$.

He'll pick action:

$$\begin{aligned} a(x^s) &= x_1^s - x_2^s = m_1 x_1 - m_2 x_2 = m_1 (R + \pi) - m_2 \pi \\ &= m_1 R + (m_1 - m_2) \pi. \end{aligned}$$

There is nominal illusion if a^s depends on π, i.e. if $m_1 \neq m_2$.

What m will a sparse consumer pick? Expected utility is:

$$v(m) = \mathbb{E}[u(a(x^s), x)] = -\frac{1}{2}(a(x^s) - R)^2 = -\frac{1}{2}(1 - m_1)^2 \sigma_R^2 - \frac{1}{2}(m_1 - m_2)^2 \sigma_\pi^2,$$

since $cov(R, \pi) = 0$, so the sparse max gives:

$$m^* = \min \frac{1}{2}(1 - m_1)^2 \sigma_R^2 + \frac{1}{2}(m_1 - m_2)^2 \sigma_\pi^2 + \kappa(|m_1| + |m_2|) = f(m_1, m_2). \quad (34)$$

This shows that the consumer thinks more about the nominal wage than inflation (if $m_2 < m_1$, there is money illusion): $m_2 \leq m_1$. Otherwise, if $m_2 > m_1$, one can decrease m_2 a bit and both $(m_1 - m_2)^2 \sigma_\pi^2$ and $\kappa |m_2|$ will fall. Also, if $m_2 > 0$, $f_{m_2|m_2=m_1} = \kappa > 0$, so $m_2 < m_1$. We conclude that *if m_2 is positive, then there is necessarily money illusion ($m_2 < m_1$).*

The explicit solution is the following (using the case $\alpha = 1$).

[84] See also Deaton (1977) for an early analysis of misperception of inflation.

Lemma 3 *In the nominal / real example above, the allocation of attention is (defining* $\kappa_* := \frac{\sigma_R^2\sigma_\pi^2}{\sigma_R^2+2\sigma_\pi^2}$*):*

For $\kappa \in [0, \kappa_*]$*:* $m_1 = 1 - \frac{2\kappa}{\sigma_R^2}, m_2 = 1 - \kappa\left(\frac{2}{\sigma_R^2} + \frac{1}{\sigma_\pi^2}\right)$.

For $\kappa \in [\kappa_*, \sigma_R^2]$*:* $m_1 = \frac{\sigma_R^2 - \kappa}{\sigma_R^2 + \sigma_\pi^2}, m_2 = 0$.

For $\kappa \in [\sigma_R^2, \infty)$*:* $m_1 = m_2 = 0$.

Whenever $m_2 > 0$*,* $m_1 > m_2$*, i.e. there is nominal illusion.*

Proof. The first-order conditions are:

$$\text{FOC}_{m_1} : -(1 - m_1)\,\sigma_R^2 + (m_1 - m_2)\,\sigma_\pi^2 + \kappa s_1 = 0, \tag{35}$$

$$\text{FOC}_{m_2} : -(m_1 - m_2)\,\sigma_\pi^2 + \kappa s_2 = 0, \tag{36}$$

where $s_i = sgn\,(m_i) = 1$ if $m_i > 0$, and is some quantity in $[-1, 1]$ if $m_i = 0$. Adding the two

$$-(1 - m_1)\,\sigma_R^2 + \kappa\,(s_1 + s_2) = 0, \tag{37}$$

First, consider the case m_1, m_2 both positive: $s_i = 1$. Then, FOC_{m_2} gives:

$$m_2 = m_1 - \frac{\kappa}{\sigma_\pi^2}, \tag{38}$$

and (37) gives:

$$m_1 = 1 - \frac{2\kappa}{\sigma_R^2}, \qquad m_2 = 1 - \kappa\left(\frac{2}{\sigma_R^2} + \frac{1}{\sigma_\pi^2}\right). \tag{39}$$

We check that $m_i > 0$, i.e. $\kappa < \kappa_* := \frac{\sigma_R^2\sigma_\pi^2}{\sigma_R^2+2\sigma_\pi^2}\sigma_\pi^2$.

Next, the case where $m_1 > 0 = m_2$. Then, FOC_{m_1} gives:

$$m_1 = \frac{\sigma_R^2 - \kappa}{\sigma_R^2 + \sigma_\pi^2} > 0,$$

which implies $\sigma_R^2 > \kappa$, and

$$\kappa s_2 = m_1\sigma_\pi^2 = \frac{\sigma_R^2 - \kappa}{\sigma_R^2 + \sigma_\pi^2}\sigma_\pi^2 \leq \kappa,$$

i.e.

$$\kappa \geq \frac{\sigma_R^2\sigma_\pi^2}{\sigma_R^2 + 2\sigma_\pi^2}.$$

Finally, check $m_1 = m_2 = 0$. The FOC give: $\kappa s_1 = \sigma_R^2 \leq \kappa$, $\kappa s_2 = 0$. □

X. (Sub)optimization Under Constraints: Marginal Analysis

This section develops machinery that is used in some derivations. It deals with a basic issue: given optimization under constraints, and a shock, how do the various variables (action, Lagrange multiplier, utility) change? If consumers are not fully rational, what are utility losses?

Consider the problem:

$$\max_a u(a,x) \text{ s.t. } b(a,x) \geq 0,$$

where a is the action and x is a "shift" parameter (which is general and could represent a shift in income, price, etc.). We will derive the change in action a_x when there is an infinitesimal parameter shift x, and the losses from a suboptimal action.

Say that $b = \left(b^1, ..., b^K\right)$ stacks together K constraints, and define the Lagrangian:

$$L(a,x,\lambda) = u(a,x) + \lambda b(a,x),$$

with $\lambda \in \mathbb{R}_+^K$, and the value function

$$\begin{aligned} v(x) &= \max_a u(a,x) \text{ s.t. } b(a,x) \geq 0 \\ &= L(a(x), x, \lambda(x)), \end{aligned}$$

where $\lambda(x) \in \mathbb{R}^K$ is the correct Lagrange multiplier corresponding to the problem. We suppose that all constraints bind at the default, $\lambda_k^d > 0$. We assume that the objective function u and the "budget constraints" b^k are concave in a and twice continuously differentiable, and that $L(a,x,\lambda)$ is strictly concave in a.

We state our first proposition.

Proposition 24 (Change in action, Lagrange multiplier and utility after a shift) *The derivative of the Lagrange multiplier λ and action a with respect to shift x are given by:*

$$\lambda_x = \left(b_a' L_{aa}^{-1} b_a\right)^{-1} \left(b_x' - b_a' L_{aa}^{-1} L_{ax}\right), \tag{40}$$

$$\begin{aligned} a_x &= -L_{aa}^{-1}\left(L_{ax} + b_a \lambda_x\right) && (41) \\ &= -\left(L_{aa}^{-1} b_a\right)\left(b_a' L_{aa}^{-1} b_a\right)^{-1} b_x' - \left[I - L_{aa}^{-1} b_a \left(b_a' L_{aa}^{-1} b_a\right)^{-1} b_a'\right] L_{aa}^{-1} L_{ax}. && (42) \end{aligned}$$

In addition, the value function v_x satisfies: $v_x = L_x$ and

$$v_{xx} = L_{xx} - a_x L_{aa} a_x + 2b_x \lambda_x. \tag{43}$$

The impact of a change x on the action can be re-expressed more intuitively as follows. Define

$$a(x, y) = \arg\max_a u(a, x) \text{ s.t. } b(a, x) + y \geq 0, \tag{44}$$

where $y \in \mathbb{R}^K$ can be interpreted at some multidimensional "extra income," and the infinitesimal "income effect" $a_y := \partial_y a(x, y)_{|y=0}$ so that $a_y \delta y$ is the change in the action when the budget constraint is slacker by δy. Define also the compensated demand (around $a^d = a(x, 0)$:

$$\overline{a}(x) := a\left(x, -b\left(a^d, x\right)\right). \tag{45}$$

This is the generalization to arbitrary problems of the compensated demand in the consumer theory.

Proposition 25 (Change in action and Lagrange multiplier after a shift, expressed in terms of income and substitution effects) *The marginal change of the optimal action is:*

$$a_x = \overline{a}_x + a_y b_x, \tag{46}$$

where income effects (a_y) and substitution effects ($\overline{a}_x$) are:

$$a_y = -\left(L_{aa}^{-1} b_a\right)\left(b'_a L_{aa}^{-1} b_a\right)^{-1}, \tag{47}$$

$$\overline{a}_x = -\left(I + a_y b_a\right) L_{aa}^{-1} L_{ax}. \tag{48}$$

The term $\overline{a}_x$ is a "Slutsky matrix" for general problems under constraints. It satisfies $b_a \overline{a}_x = 0$, i.e. induces "budget neutral" changes. In addition, $b_a a_y = -I$.

The interpretation of (41) is as follows. In the right-hand side of (41), the first term, $-L_{aa}^{-1} L_{ax} \delta x$, is the "myopic" change in action, using the same shadow prices (Lagrange multiplier) as before the shift and forgetting about the budget constraint. The second term is the change in action to satisfy the budget constraint. This interpretation motivates Step 2.ii in the sparse max with constraints (Definition 2).

On the right-hand side of (42), the first term (in b_x) is the "income effect" on a_x: it is non-zero iff the budget constraint ceases to bind after the shift x. The second term (in L_{ax}) is a "substitution effect": it measures how the shift changes marginal utility.

In equation (42), the first term is a direct change of income, and the second is the change in the price, L_a, that is orthogonal to the price vector b_a.

In equation (43), the sign of the $-a_x L_{aa} a_x$ term may come as surprise at first, as a mechanical application of the chain rule might suggest a positive sign. The interpretation is the following: take the case where $b_x = 0$, so that the budget constraint is still satisfied after the shift. If a_x didn't change, the utility change would be L_{xx}. However, a_x can change (by a substitution effect), so that utility is increased. It's increased by $-a_x L_{aa} a_x \geq 0$. This intuition is helped by the following Proposition, which states the utility loss from a suboptimal policy.

PROPOSITION 26 *Suppose that the agent chooses a (potentially suboptimal) policy $a^B(x)$, that still satisfies budget balance $b\left(a^B(x), x\right) = 0$, and is twice continuously differentiable around $x = x^d$, with derivative $a_x^B\left(x^d\right)$. Let $v^B(x) = u\left(a^B(x), x\right)$ be the resulting value function. Then, we have, at $x = x^d$:*

$$v_x^B = v_x = L_x, \tag{49}$$

and

$$v_{xx}^B = v_{xx} + \left(a_x^B - a_x\right) L_{aa} \left(a_x^B - a_x\right), \tag{50}$$

where a_x and v_{xx} are given in Proposition 24.

Utility losses from using the sparse max We can now compute the losses from the sparse max. The agent will choose action $a^s = a(x^s, y^s)$, where $y^s(x, x^s)$ makes sure that the budget constraint is exactly satisfied: it is defined by $b(a(x^s, y^s), x) = 0$. The rational answer is $a^r = a(x, 0)$. We want to calculate the utility loss,

$$v^r - v^s = u(a^r, x) - u(a^s, x).$$

Its value is given by the following Proposition.

PROPOSITION 27 *In the sparse max under constraints, where the agent faces true vector x but perceives x^s, utility losses are:*

$$v^r - v^s = -\frac{1}{2}(x^s - x)' \overline{a}_x' L_{aa} \overline{a}_x (x^s - x) + o\left(\|x^s - x\|^2\right). \tag{51}$$

Income vs substitution effects in sparse max

PROPOSITION 28 (Income vs substitution effects with sparse max) *For the traditional action, a change x induces a change (see equation 46):*

$$a^r - a^d = \overline{a}_x x + a_y b_x x + o(x), \tag{52}$$

Equation (86) gives

$$a^s - a^r = \overline{a}_x (x^s - x) + o(x), \tag{53}$$

hence:

$$a^s - a^d = \overline{a}_x x^s + a_y b_x x + o(x). \tag{54}$$

Hence, in sparse max, the "income effect" is fully taken into account ($a_y b_x x$), but the "substitution effect" is only partially taken into account ($\overline{a}_x x^s$).

This is because sparse max respects the budget constraint, hence the income effect works fully. The difference between the two actions a^s and a^r is the substitution effect.

Application to consumer theory

Take the consumption problem $\max_c u(c)$ s.t. $b(c, w, p) = w - p \cdot c \geq 0$. The action a is the consumption vector c, and the shift vector $x = (w, p)$. Hence the Lagrangian is:

$$L(c, x, \lambda) = u(c) + \lambda b(c, x), \qquad b(c, x) = w - p \cdot c.$$

Define $q := L_{cc}^{-1} b_c$, i.e.

$$q = -\lambda^{-1} u''^{-1} p.$$

We have $b_c = -p$ and $b_c' L_{cc}^{-1} b_c = -p'q$, a scalar. Consider first changes in w. Then, $b_w = 1$, $L_{cw} = 0$, and (42) give: $c_w = -q(-p'q)^{-1} - 0$, hence

$$\text{Marshallian demand: } c_w(p, w) = \frac{q}{p'q}. \tag{55}$$

Hence vector q is (up to a factor $1/(p'q)$) the marginal reaction of consumption to a change in wealth.

Other calculations give, for the Slutsky matrix $S = c_p + c_w c' = \overline{c}_p$,

$$\text{Slutsky matrix: } S = \lambda u''^{-1} + \lambda \frac{qq'}{p'q} = \lambda Q' u''^{-1} Q, \tag{56}$$

where

$$Q = I - pc_w' = I - \frac{pq'}{p'q}, \tag{57}$$

is a projection ($Q^2 = Q$) that has the property $Qp = 0$ and $c_w' Q = 0$. Note that

$$\frac{1}{\lambda} S = u''^{-1} Q = Q' u''^{-1} = Q' u''^{-1} Q,$$

which implies:

$$S u'' S = \lambda S. \tag{58}$$

Finally $c_p(p,w) = S - c_w c'$, i.e.

$$\text{Marshallian demand: } c_p(p,w) = \lambda u''^{-1} + \frac{q(\lambda q' - c')}{p'q}.$$

X.A. *Proof of Proposition 14*

We'll use the following Definition.

Definition 3 *Consider the two problems: (i)* $\operatorname{smax}_a u(a,x)$ *subject to* $b(a,x) \geq 0$ *and (ii)* $\operatorname{smax}_a \widetilde{u}(a,x)$ *subject to* $\widetilde{b}(a,x) \geq 0$. *Suppose that they are related by:* $\widetilde{u}(a,x) = f(u(a,x))$ *and* $\widetilde{b}(a,x) = h(b(a,x))$, *where* f *and* h *are increasing,* C^2 *functions,* $h(0) = 0$ *and* $h'(0)$ *is a non-singular matrix.*[85] *We say that the sparse max is "ordinal" or "reparametrization invariant" (RI) if it returns the same attention* m^* *and action* a^s, *for any* f, h.

Note that if the sparse max is ordinal, then u and b can simply be quasi-concave, rather than concave.

We prove here considerations stated in Proposition 14.

We start with some simple results.

Proposition 29 *When maximization is without constraints, the model is invariant to transformations of the utility* $u \mapsto f(u)$.

Proof. With $\widetilde{u} = f(u)$, $\widetilde{u}_a = f'(u)\, u_a$ and $\widetilde{u}_{aa} = f''(u)\, u_a^2 + f'(u)\, u_{aa}$. But as $u_a = 0$ at the default, we have $\widetilde{u}_{aa} = f'(u)\, u_{aa}$. Likewise, $\widetilde{u}_{ax} = f'(u)\, u_{ax}$. So, $\overline{a}_x = \widetilde{u}_{aa}^{-1} \widetilde{u}_{ax} = a_x$. Likewise, in the scaled-κ, we have $\widetilde{\kappa} = f'(u)\,\kappa$. Then, in the objective function the m minimizes (eq. 6), everything is the same, just multiplied by $f'(u)$. Hence m is the same, and so is a^s. ■

Next, we consider linear rescaling.

Proposition 30 *When maximization is with constraints, the model is invariant to linear transformations* $f(x) = Fx$, $F \in \mathbb{R}_{++}$, $h(x) = Hx$, H *an invertible* $K \times K$ *matrix which restricts to a bijection on* $\mathbb{R}_+^K \to \mathbb{R}_+^K$.

Proof. The multiplier for the tilde problem, $\widetilde{\lambda}$, satisfies: $\widetilde{u}_a + \widetilde{\lambda}\widetilde{b}_a = 0$, i.e. $Fu_a + \widetilde{\lambda}' H b_a = 0$. As we had $u_a + \lambda' b_a = 0$, we have $\widetilde{\lambda}' H = F\lambda'$. Hence, $\widetilde{L} = Fu + \widetilde{\lambda}' H b = Fu + F\lambda' b = FL$, so we get $\widetilde{L}_{aa} = FL_{aa}$, $\widetilde{L}_{ax} = FL_{ax}$, and $\overline{a}_x = a_x$, $\widetilde{\kappa} = F\kappa$. So, step 1 gives the same m^*, and step 2 gives the same a^s. ■

The model with constraints is not invariant to non-linear transformations of u or b, when there are constraints. The reason is that, in the Lagrangian $\widetilde{L} = f(u) + \lambda h(b)$, calculation of $\widetilde{L}_{aa}$ yields

[85] When there are K constraints, $h'(0)$ is a $K \times K$ matrix.

extra f'' and h'' terms. However, invariance is restored by adopting the "*compensated sparse max*". The reason, as we shall now see, is that it uses "intrinsic" expressions.

The procedure determining m is:

$$m^* = \arg\min_m -\mathbb{E}\left[\frac{1}{2}(x^s - x)' \overline{a}_x' L_{aa} \overline{a}_x (x^s - x)\right] + \kappa \sum_i g(m_i) \text{ s.t. } x_i^s = m_i x_i. \tag{59}$$

Denote by $v^s(m, x) = u(a^s \mid m, x)$ the utility obtained by the sparse agent with attention m, and call the optimal utility $v^r(x) = v^s(\iota, x)$.

The key reason to use the compensated demand $\overline{a}_x$ is Proposition 27, which gives:

$$v^s(m, x) = v^r(x) + A(m, x) + o\left(\|x\|^2\right),$$
$$\text{where } A(m, x) := \frac{1}{2}(x^s - x)' \overline{a}_x' L_{aa} \overline{a}_x (x^s - x).$$

Hence, (59) can be re-expressed:

$$\min_m \mathbb{E}\left[v^r(x) - v^s(m, x)\right] + \kappa \sum_i g(m_i) + o\left(\|x\|^2\right),$$

(to be more precise, and eliminate the $o\left(\|x\|^2\right)$ terms, we might like to say "m maximizes the quadratic *germ* of the expression," but involving that machinery of germs might be overkill for our purposes).

Now, we have $f(v^s(m)) = f\left(v^r + A(m, x) + o\left(\|x\|^2\right)\right)$, i.e.

$$f(v^s(m)) = f(v^r) + f'(u) A(m, x) + o\left(\|x\|^2\right),$$

while for the transformed "tilde" problem, we have $\widetilde{v}^s(m) = \widetilde{v}^r + \widetilde{A}(m, x) + o\left(\|x\|^2\right)$, with $\widetilde{A}(m, x) := \frac{1}{2}(x^s - x)' \widetilde{\overline{a}}_x' \widetilde{L}_{aa} \widetilde{\overline{a}}_x (x^s - x)$ is the analogue of $A(m, x)$ for the tilde problem. This shows that

$$\widetilde{A}(m, x) = f'(u) A(m, x),$$

i.e., in the sparse max under the transformed problem, the left-hand side is just the old one, times $f'(u)$: indeed, this implies that $\widetilde{\kappa} = f'(u)\kappa$. As the scale-free version of κ gives $\widetilde{\kappa} = f'(u)\kappa$, the whole expression (59) is simply multiplied by $f'(u)$ when we transform the problem. Hence, nothing changes in m^s, nor in a^s.

In short, the problem is invariant because it's expressed in intrinsic terms.

A more computational proof is possible. The key steps are: First, $\overline{a}_x$ is independent of f, h, i.e. is totally intrinsic. Second, $\overline{a}_x L_{aa} \overline{a}_x$ is multiplied by $f'(u)$ in the transformed problem, and so is κ. Hence, nothing changes in Step 1. Likewise, nothing changes in Step 2.

XI. Proofs That Were Omitted In the Paper

Proof of Lemma 1 Call $f(m,v) := \frac{1}{2}(m-1)^2 v + |m|^\alpha$, and $\tilde{\mathcal{A}}_\alpha(v) := \arg\min_{m\in[0,1]} f(m,v)$. So, $\mathcal{A}_\alpha(v) = \sup\left[\tilde{\mathcal{A}}_\alpha(v)\right]$, i.e. we choose the highest minimizer.

Sparsity-Inducing. When $\alpha \le 1$, then

$$f_m = (m-1)v + \alpha m^{\alpha-1} \ge -v + \alpha,$$

so $f_m > 0$ for all $m \in (0,1]$ whenever $v \le \alpha$. Hence, $\mathcal{A}_\alpha(v) = 0$ for $v \le \alpha$: $\mathcal{A}_\alpha$ is sparsity-inducing.

When $\alpha > 1$ and $v > 0$: $f_m(m=0) = -v < 0$, so $m = 0$ can't be a minimizer. Hence, $\mathcal{A}_\alpha(v) > 0$ for all $v > 0$: $\mathcal{A}_\alpha$ is not sparsity-inducing.

Continuity: By Berge's theorem (e.g., Kreps 2013, Proposition A4.7), the correspondence $\widetilde{\mathcal{A}}_\alpha : v \mapsto \arg\min_{m\in[0,1]} f(m,v)$ is upper-hemicontinuous because f is continuous. When $\alpha \ge 1$, $f_{mm} > 0$, so f is strictly convex in m, so that $\widetilde{\mathcal{A}}_\alpha$ is single-valued, i.e. $\widetilde{\mathcal{A}}_\alpha(v) = \{\mathcal{A}_\alpha(v)\}$. Immediately from the definition of upper-hemicontinuity applied to a single-valued correspondence, $\mathcal{A}_\alpha$ is a continuous function.

When $\alpha < 1$, we shall show that $\mathcal{A}_\alpha(v)$ is discontinuous at the point $v_* := \dfrac{2-2\alpha}{{m_*}^{2-\alpha}}$, where $m_* := \dfrac{2-2\alpha}{2-\alpha} \in (0,1)$.

Define $g(m,v) := f(m,v) - f(0,v) = \frac{1}{2}v(m^2 - 2m) + m^\alpha$. First, a simple calculation shows that

$$g(m_*, v_*) = 0, \qquad g_m(m_*, v_*) = 0, \qquad g_{mm}(m_*, v_*) > 0.$$

Indeed m_*, v_* were chosen to satisfy the first two conditions, which signify that $m = m_*$ is a local extremum of f with the same value for f as $m = 0$.

We are now going to show that $\lim_{v\searrow v_*} \mathcal{A}_\alpha(v) = m_*$, but $\mathcal{A}_\alpha(v) = 0$ for $v < v_*$, so that $\mathcal{A}_\alpha$ is discontinuous at v_*.

Note that $g_{mm} > 0$. So $g(\cdot, v_*)$ can have no minimizers other than m_* and 0. First, $g(\cdot, v_*)$ is strictly increasing on $[m_*, 1]$. Second, if $m \in (0, m_*)$ satisfied the second-order condition $g_{mm}(m, v_*) \ge 0$, then $g(\cdot, v_*)$ would be strictly convex on $[m, m_*]$ and would thus have m_* (as a point satisfying the first-order condition) as its unique minimizer. Hence, for $m \notin \{0, m_*\}$, $g(m, v_*) > 0$. This implies that $\tilde{\mathcal{A}}_\alpha(v_*) = \{0, m_*\}$ and $\mathcal{A}_\alpha(v_*) = m_*$.

As $f_{mv} < 0$, $-f$ is supermodular, so $\mathcal{A}_\alpha(v)$ is weakly increasing in v. Therefore, for any $v > v_*$, $\mathcal{A}_\alpha(v) \ge \mathcal{A}_\alpha(v_*) = m_*$. Also $\tilde{\mathcal{A}}_\alpha(v)$ is upper hemi-continuous. Therefore, $\lim_{v\searrow v_*} \mathcal{A}_\alpha(v) = m_*$.

Next take $v = v_* - \varepsilon$, for $\varepsilon > 0$. Then for any $m \in (0,1]$,

$$g(m, v_* - \varepsilon) - g(m, v_*) = -\frac{\varepsilon}{2}\left(m^2 - 2m\right) > 0,$$

Therefore,

$$g(m, v_* - \varepsilon) > g(m, v_*) \geq g(m_*, v_*) = 0$$

So, for $\varepsilon > 0$, $f(m, v_* - \varepsilon) > f(0, v_* - \varepsilon)$ for all $m \in (0, 1]$. Hence, $\tilde{\mathcal{A}}_\alpha(v_* - \varepsilon) = \{0\}$ and $\mathcal{A}_\alpha(v_* - \varepsilon) = 0$ i.e. $\mathcal{A}_\alpha(v) = 0$ for all $v < v_*$.

Proof of Lemma 2

We define $a(x) = \arg\max_x u(a, x)$. We fix an m, and define x^s by $x_i^s := m_i x_i$, and $\widehat{a}(x) = a(x^s(x)) - a(x)$: $\widehat{a}(x)$ is difference between the sparse action and the rational action. Finally, define: $V(x) = u(a^s(x), x) - u(a(x), x)$, the utility differential between the sparse and the rational action. We have:

$$\begin{aligned}
V(x) &= u(a(x) + \widehat{a}(x), x) - u(a(x), x), \\
V_x(x) &= u_a(a(x) + \widehat{a}(x), x)(a_x + \widehat{a}_x) + u_x(a(x) + \widehat{a}(x), x) \\
&\quad - [u_a(a(x), x) a_x + u_x(a(x), x)],
\end{aligned}$$

and taking now the derivative at $x = 0$, and evaluating expressions at $(a, x) = \left(a^d, 0\right)$:

$$\begin{aligned}
V_{xx}(0) &= (a_x + \widehat{a}_x)' u_{aa}(a_x + \widehat{a}_x) + u_a(a_{xx} + \widehat{a}_{xx}) + 2u_{xa}(a_x + \widehat{a}_x) \\
&\quad - [a_x' u_{aa} a_x + u_a a_{xx} + 2u_{xa} a_x] \\
&= \widehat{a}_x' u_{aa} \widehat{a}_x + 2a_x' u_{aa} \widehat{a}_x + 2u_{xa} \widehat{a}_x \text{ using } u_a = 0 \\
&= \widehat{a}_x' u_{aa} \widehat{a}_x + 2(a_x' u_{aa} + u_{xa}) \widehat{a}_x \\
&= \widehat{a}_x' u_{aa} \widehat{a}_x \text{ using } a_x = -u_{aa}^{-1} u_{ax}, \text{ so that } a_x' u_{aa} = -u_{xa}.
\end{aligned}$$

Calling $M = diag(m_1, ..., m_n) = \frac{\partial x^s}{\partial x}$ the diagonal matrix with diagonal m, I the identity matrix of dimension n, so that $\widehat{a}_x = a_x \frac{\partial x^s}{\partial x} - a_x = a_x(M - I)$, we obtain:

$$V_{xx}(0) = (M - I)' a_x' u_{aa} a_x (M - I).$$

Note also that $V(0) = 0$ and $V_x(0) = 0$. So, by Taylor expansion:

$$V(x) = \frac{1}{2} x' V_{xx}(0) x + o\left(\|x\|^2\right).$$

Finally,

$$\begin{aligned}
\mathbb{E}\left[v\left(m\right)-v\left(\iota\right)\right] &= \mathbb{E}\left[u\left(a^{s}\left(x\right),x\right)-u\left(a\left(x\right),x\right)\right]=\mathbb{E}\left[V\left(x\right)\right] \\
&= \mathbb{E}\left[\frac{1}{2}x'V_{xx}\left(0\right)x\right]+o\left(\left\|x\right\|^{2}\right), \\
\mathbb{E}\left[\frac{1}{2}x'V_{xx}\left(0\right)x\right] &= \frac{1}{2}\mathbb{E}\left[x'\left(M-I\right)'a_{x}'u_{aa}a_{x}\left(M-I\right)x\right] \\
&= \frac{1}{2}\mathbb{E}\left[\sum_{i,j}x_{i}\left(m_{i}-1\right)a_{x_{i}}'u_{aa}a_{x_{j}}\left(m_{j}-1\right)x_{j}\right] \\
&= -\frac{1}{2}\sum_{i,j}\mathbb{E}\left[x_{i}x_{j}\right]\left(m_{i}-1\right)a_{x_{i}}'u_{aa}a_{x_{j}}\left(m_{j}-1\right) \\
&= -\frac{1}{2}\sum_{i,j}\left(m_{i}-1\right)\Lambda_{ij}\left(m_{j}-1\right)\ \text{ using }\Lambda_{ij}:=-\mathbb{E}\left[x_{i}x_{j}\right]a_{x_{i}}u_{aa}a_{x_{j}},
\end{aligned}$$

so $\mathbb{E}\left[v\left(m\right)-v\left(\iota\right)\right]=-\frac{1}{2}\sum_{i,j}\left(m_{i}-1\right)\Lambda_{ij}\left(m_{i}-1\right)+o\left(\left\|x\right\|^{2}\right)$, as announced.

If the utility function is quadratic, then the expression in Lemma 2 is exact: the $o\left(\left\|x\right\|^{2}\right)$ term has to be exactly 0, as the right-hand-side must be a linear function of the variance of x, the σ_{ij}.

Proof of Proposition 3 (reparametrization-invariant version) *Let us study the problem in the "reparametrization-invariant" version of sparse max (section V.D).* The step 1 problem is: (with $\lambda=\lambda^{d}$), with compensated demand sensitivity $\overline{\boldsymbol{c}}_{p_i}=\frac{\partial \boldsymbol{c}}{\partial p_i}+\boldsymbol{c}_w\boldsymbol{c}_i$(the vector $\left(S_{ji}\right)_{j=1\dots n}$ in the Slutsky matrix)

$$\min_{m}\frac{1}{2}\sum_{i}\left(m_{i}-1\right)^{2}\sigma_{p_i}^{2}\left(-\overline{\boldsymbol{c}}_{p_i}u''\overline{\boldsymbol{c}}_{p_i}\right)+\kappa\sum_{i}\left|m_{i}\right|^{\alpha}.$$

Equation 58 in the appendix shows that $Su''S=\lambda S$ (with S the Slutsky matrix), i.e. $\overline{\boldsymbol{c}}_{p_i}u''\overline{\boldsymbol{c}}_{p_j}=\lambda S_{ij}$, so that the $\left(m_{i}-1\right)^{2}$ term on the left-hand side is:

$$\sigma_{p_i}^{2}\left(\overline{\boldsymbol{c}}_{p_i}u''\overline{\boldsymbol{c}}_{p_i}\right)=\lambda\sigma_{p_i}^{2}S_{ii}=\lambda\frac{\sigma_{p_i}^{2}}{p_i^{2}}\frac{p_iS_{ii}}{c_i}c_ip_i=-\lambda\frac{\sigma_{p_i}^{2}}{p_i^{2}}\psi_ic_ip_i,$$

where $\psi_i=\frac{-p_i^{d}S_{ii}}{c_i^{d}}$ is the compensated price-elasticity for good i (actually, its absolute value: it's positive). Hence, $m_i=\mathcal{A}\left(v_i\right)$ with $v_i=\lambda\cdot c_ip_i\cdot\frac{\sigma_{p_i}^{2}}{p_i^{2}}\cdot\psi_i/\kappa$.

Proof of Proposition 5 We want to show that there are real number χ such that $\boldsymbol{c}^{s}\left(\chi\boldsymbol{p},\chi w\right)\neq\boldsymbol{c}^{s}\left(\boldsymbol{p},w\right)$. We calculate (borrowing from a later result (15), but there is no logical inconsistency), at $\chi=1$:

$$\frac{dc_i}{d\chi} := \frac{d\boldsymbol{c}_i^s(\chi\boldsymbol{p},\chi w)}{d\chi} = \sum_j \frac{\partial c_i^s}{\partial p_j} p_j + \frac{\partial c_i^s}{\partial w} w = \sum_j \left(S_{ij}^s - \frac{\partial c_i^s}{\partial w} c_j \right) p_j + \frac{\partial c_i^s}{\partial w} w$$
$$= \sum_j S_{ij}^s p_j + \frac{\partial c_i^s}{\partial w} \left(-\sum_j p_j c_j + w \right),$$
$$\frac{dc_i}{d\chi} = \sum_j S_{ij}^r m_j p_j. \tag{60}$$

If all attention parameters were the same ($m_j = \overline{m}$ for some $\overline{m}$), then $\frac{dc_i}{d\chi} = \overline{m} \sum_j S_{ij}^r p_j = 0$ (as $S^r \boldsymbol{p} = 0$). If not all m_j are the same, then generically (something we will not formally prove, but could be proved by standard techniques), $\sum_j S_{ij}^r m_j p_j \neq 0$. To gain intuition for this, take the case of 2 goods, and normalize $p^d = (1, 1)$. Then, we can write $S^r = A \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ for some $A > 0$ (as S^r is symmetric and $S^r \boldsymbol{p} = 0$). So,

$$\frac{dc_1}{d\chi} = \sum_j S_{1j}^r m_j p_j = A(-m_1 + m_2) \neq 0.$$

Proof of Proposition 12 We shall prove the proposition, and also that it holds for any general $m_1 \neq m_2$, when the consumer has the perception in logs, (89).

Clearly the set of equilibrium allocations for consumer a is in a's offer curve, OC^a. Let us now show the converse, i.e. $OC^a \subset \mathcal{C}^a$. Take a point in $\boldsymbol{c} \in OC^a$, a price $\boldsymbol{p}$ that supports it ($\boldsymbol{D}^a(\boldsymbol{p}) = \boldsymbol{c}$) and $\boldsymbol{c}^s = \boldsymbol{\omega} - \boldsymbol{c}$ the corresponding allocation for consumer b. Because $\boldsymbol{p} \cdot \boldsymbol{c} = \boldsymbol{p} \cdot \boldsymbol{\omega}^a$, and $\boldsymbol{\omega} = \boldsymbol{\omega}^a + \boldsymbol{\omega}^b$, we have $\boldsymbol{p} \cdot \boldsymbol{c}^b = \boldsymbol{p} \cdot \boldsymbol{\omega}^b$, so $\boldsymbol{c}^b$ is on consumer b's budget set for any price $\chi\boldsymbol{p}$ with $\chi > 0$. We have to see if there is a $\chi > 0$ such that price $\chi\boldsymbol{p}$ leads consumer b to demand consumption $\boldsymbol{c}^s$. This is the case if there is a Lagrange multiplier λ such that

$$u'(\boldsymbol{c}) = \lambda (\chi\boldsymbol{p})^s,$$

where $(\chi\boldsymbol{p})^s$ is the perceived price corresponding to price $\chi\boldsymbol{p}$. We consider two cases.

Case 1: Using the assumptions of the Proposition 12 (with $m_1 = 1$, $m_2 = 0$), we have $(\chi\boldsymbol{p})_1^s = \chi p_1$, and $(\chi\boldsymbol{p})_2^s = p_2^d$. Then

$$\lambda (\chi\boldsymbol{p})^s = \begin{pmatrix} \lambda\chi p_1 \\ \lambda p_2^d \end{pmatrix}. \tag{61}$$

Case 2: Suppose consumer b has the loglinear perception (89), and general $m_1 \neq m_2$. Then

$$\lambda (\chi\boldsymbol{p})^s = \begin{pmatrix} \lambda\chi^{m_1} p_1^{m_1} \left(p_1^d\right)^{1-m_1} \\ \lambda\chi^{m_2} p_1^{m_2} \left(p_2^d\right)^{1-m_2} \end{pmatrix}. \tag{62}$$

In both cases, the right-hand side spans $\mathbb{R}^2_{++}$ as λ, χ vary in $\mathbb{R}^2_{++}$ Hence, there are λ, χ such that $u'(\boldsymbol{c}) = \lambda(\chi \boldsymbol{p})^s$, i.e. $\boldsymbol{c}^s$ is indeed the consumption demanded by consumer b at price $\chi \boldsymbol{p}$. That means that $\boldsymbol{c} \in \mathcal{C}^a$, and finally $OC^a \subset \mathcal{C}^a$.

In the case with additive perceptions (11), a similar proposition would hold, but $\mathcal{C}^a$ would be a one-dimensional strict subset of OC^a, as with general linear perceptions $\lambda(\chi \boldsymbol{p})^s$ does not span fully $\mathbb{R}^2_{++}$ as λ, χ vary in $\mathbb{R}^2_{++}$.

A lemma on binding constraints

LEMMA 4 *Consider the case with one constraint. In the setup of Definition 2, Step 2, utility $u(a(\lambda), x^s)$ is weakly decreasing in λ. Moreover, given $\lambda^* > 0$, we have $b(a(\lambda^*), x) = 0$.*

Proof. We call

$$L(a, x, \lambda) := u(a, x) + \lambda \cdot b(a, x).$$

So, $L_a(a(\lambda), x^s, \lambda) = 0$, so $L_{aa} a'(\lambda) + L_{\lambda a} = 0$, i.e. (as $L_{\lambda a} = b_a$), $a'(\lambda) = -L_{aa}^{-1} b_a$. Hence, utility $v(\lambda, x) = u(a(\lambda), x)$ satisfies (observing that $u_a + \lambda b_a = 0$), $v_\lambda(\lambda, x) = u_a a'(\lambda) = \lambda b'_a L_{aa}^{-1} b_a \leq 0$ as L_{aa}^{-1} is negative definite, and λ is nonnegative. ■

Proof of Proposition 13 $L(a, x, \lambda) := u(a, x) + \lambda \cdot b(a, x)$ is concave in a, and strictly concave when $\lambda > 0$. If we assume $\lambda^* > 0$ and L is strictly concave throughout, the general result follows (given our assumption of no interior extremum) from continuity. In particular, this implies that $a(\lambda)$ is uniquely defined and inherits twice-differentiability from u, v.

The Lagrangians associated with the two problems are (strategically adding a constant, and denoting by Λ the Lagrange multiplier for the second problem) $L(a, x, \lambda) = u(a, x) - \widehat{u} + \lambda(\widehat{w} - w(a, x))$ and $M(a, x, \Lambda) = \widehat{w} - w(a, x) + \Lambda(u(a, x) - \widehat{u})$. Hence

$$L(a, m, \lambda) = \lambda M(a, m, 1/\lambda). \tag{63}$$

Let us follow Definition 2. At the default, we obtain $\Lambda^d = 1/\lambda^d$.

Next, see step 1. Consider m^*. It comes from $\operatorname{smax}_a L(a, x, \lambda^d)$, which yields the same m^* as $\operatorname{smax}_a \lambda^d M(a, x, \Lambda^d)$, as the two functions are the same. Hence, step 1 applied to L or M, the two m^* selected are the same.[86]

Move on to step 2. We define $a^L(\lambda) := \arg\max_a L(a, x^s, \lambda)$ and $a^M(\Lambda) := \arg\max_a M(a, x^s, \Lambda)$. Then (63) implies: $a^M(\lambda) = a^L(1/\lambda)$.

Finally, as $\lambda^* > 0$, Lemma 4 shows that the constraint binds, i.e. $w(a^L(\lambda^*), x) = \widehat{w}$. It yields a utility $u(a^L(\lambda^*), x) = \widehat{u}$. But this is the same solution as for problem M, which seeks a solution

[86] With the scale-free version of κ, the same holds when maximizing $\operatorname{smax}_a M(a, x, \Lambda^d)$.

$a^M(\Lambda^*)$, that should solve $u\left(a^M(\Lambda^*),\mu\right)=\widehat{u}$ and also yields a cost $w\left(a^M(\Lambda^*),\mu\right)=\widehat{w}$. As the two problems are the same, we get $a^L(\lambda^*)=a^M(\Lambda^*)$ and $\Lambda^*=1/\lambda^*$.

Proof of Proposition 15 An agent with cost k will exert attention iff $\sigma^2\geq 2k$: his attention is $1_{\sigma^2\geq 2k}=\mathcal{A}_0(\sigma^2/k)$. Hence, the average attention is

$$\mathbb{E}\left[\mathcal{A}_0\left(\sigma^2/\widetilde{k}\right)\right]=\mathbb{E}\left[1_{\sigma^2\geq 2\widetilde{k}}\right]=\mathbb{P}\left(\frac{\sigma^2}{2}\geq\widetilde{k}\right)=\mathcal{A}\left(\sigma^2/\kappa\right).$$

So, the average attention (in this heterogeneous population) is the attention in the basic sparse max (with just one agent): $\mathbb{E}\left[1_{\sigma^2\geq 2\widetilde{k}}\right]=\mathcal{A}(\sigma^2/\kappa)$.

In the quadratic problem, $a^r(x)=\sum_i x_i$. The action of a fixed cost agent with cost k is: $a^{s,\text{Fixed Cost}}(x)=\sum_i 1_{\sigma_i^2\geq 2k}x_i$. Hence, average over the $\widetilde{k}$

$$\mathbb{E}\left[a^{s,\text{Fixed Cost}}(x)\right]=\sum_i\mathbb{E}\left[1_{\sigma_i^2\geq 2\widetilde{k}}\right]x_i=\sum_i\mathcal{A}\left(\sigma_i^2/\kappa\right)x_i=a^s(x).$$

Proof of Proposition 16 Given the precision T_i, it is well-known that $\mathbb{E}[x_i\mid S_i]=\lambda_i S_i$ with $\lambda_i=\frac{cov(x_i,S_i)}{var(S_i)}=\frac{\sigma_i^2}{\sigma_i^2+\sigma_{\varepsilon_i}^2}=\frac{T_i}{1+T_i}$. Hence, *the "$m_i$" in a deterministic attention problem is the $\frac{T_i}{1+T_i}$ in a signal-plus-noise model.* The optimal action is $a(S)=\mathbb{E}\left[\sum_i\mu_i x_i\mid S\right]=\sum_i\lambda_i\mu_i S_i$, utility is

$$\mathbb{E}\left[u\left(a(S),x\right)\right]=-\frac{1}{2}\mathbb{E}\left[\left(\sum_i\mu_i^2\left(x_i-\lambda_i S_i\right)\right)^2\right]=-\frac{1}{2}\sum_i\frac{\mu_i^2\sigma_i^2}{1+T_i}$$

after a bit of algebra. Hence, the problem is:

$$\min_{T_i\geq 0}\sum_i\frac{1}{2}\frac{\mu_i^2\sigma_i^2}{1+T_i}+G_\alpha(T_i)\tag{64}$$

Given $m_i=\frac{T_i}{1+T_i}$, we can write $T_i=l(m_i):=\frac{m_i}{1-m_i}$. Hence, problem (64) is equivalent to

$$\min_{m_i\in[0,1]}F(m)\text{ where }F(m):=\sum_i\frac{1}{2}(1-m_i)\sigma_i^2+G_\alpha(l(m_i)).\tag{65}$$

This problem (65) can be compared to the problem:

$$\min_{m_i\in[0,1]}f(m)\text{ where }f(m):=\sum_i\frac{1}{2}(1-m_i)^2\sigma_i^2+g_\alpha(m_i)\tag{66}$$

As we posited $G'_\alpha(T)=g'_\alpha\left(\frac{T}{1+T}\right)\frac{1}{1+T}$, we have $G'_\alpha(l(m))=g'_\alpha(m)(1-m)$. Therefore, $F_{m_i}(m)=\frac{1}{1-m_i}f_{m_i}(m)$. So not only the FOC matches but the sign of first derivatives are also the same for the two problems. Hence, the solutions of problems (64), (65) and (66) coincide.

Some examples. When $g_\alpha(m)/\kappa = m^\alpha$, we have $G'_\alpha(T)/\kappa = \frac{\alpha T^{\alpha-1}}{(1+T)^\alpha}$. The explicit solutions are:

$$G_0(T)/\kappa = 1_{T>0}, G_1(T)/\kappa = \ln(1+T), G_2(T)/\kappa = 2\ln(1+T) - \frac{2T}{1+T} \tag{67}$$

Proposition 16 extends to a multidimensional action.

PROPOSITION 31 (Extension of Proposition 16 to a multidimensional action) *Suppose the action a is a J-dimensional vector, and utility is*

$$u(a,x) = v(a - Bx), \qquad v(a) = -\frac{1}{2}a'\Gamma a, \tag{68}$$

where B is a $J \times n$ matrix, and Γ is a $J \times J$ positive definite matrix (so that the rational action is $a^r = Bx$). Then Proposition 16 holds.

Proof: We follow closely the proof of Proposition 16. We still have $\mathbb{E}[x_i \mid S_i] = \lambda_i S_i$ with $\lambda_i = \frac{T_i}{1+T_i}$, and the optimal action is $a(S) = \mathbb{E}[Bx \mid S] = BMS$, where $M := diag(\lambda_i)$. Hence

$$\mathbb{E}[a(S) \mid x] = Bx^s \text{ with } x_i^s = \lambda_i x_i.$$

Defining $\eta_i := x_i - \lambda_i S_i$, we have (by the standard arguments on Gaussian spaces) the orthogonal decomposition $x_i = \mathbb{E}[x_i \mid S_i] + \eta_i = \lambda_i S_i + \eta_i$, with $cov(\eta_i, S_i) = 0$. Hence,

$$\begin{aligned}
\mathbb{E}[u(a,x) \mid S] &= \mathbb{E}[v(a(S) - B(MS + \eta)) \mid S] \\
&= \mathbb{E}[v(-B\eta) \mid S] = \mathbb{E}[v(-B\eta)] = -\frac{1}{2}\mathbb{E}[\eta' B'\Gamma B\eta] \\
&= -\frac{1}{2}\sum_i V_{ii} var(\eta_i), \\
V &:= B'\Gamma B.
\end{aligned}$$

Hence, the problem is as above

$$\min_{T_i \geq 0} \sum_i \frac{1}{2}\frac{\sigma_i'^2}{1+T_i} + G_\alpha(T),$$

with $\sigma_i'^2 := \sigma_i^2 V_{ii}$. The equivalence follows, by the same reasoning as above: $\lambda_i = \mathcal{A}_\alpha(\sigma_i'^2/\kappa)$, so that:

$$\mathbb{E}[a(S) \mid x] = a^s(x) = Bx^s \text{ with } x_i^s = \mathcal{A}_\alpha\left(\sigma_i'^2/\kappa\right)x_i.$$

Proof of Proposition 17 Call $f(\boldsymbol{x}) = \boldsymbol{x}' S^s \boldsymbol{x} = \boldsymbol{x}' S^r M \boldsymbol{x}$. Then, $f'(\boldsymbol{x}) = S^r M \boldsymbol{x} + \boldsymbol{x}' S^r M$, and as $\boldsymbol{p}^{d\prime} S^r = 0$, $f'(\boldsymbol{p}^d) = S^r M p^d =: q$, while $f(\boldsymbol{p}^d) = 0$. As $q \neq 0$, $f'(\boldsymbol{p}^d) q = q'q > 0$. Hence, a vector $\boldsymbol{x} = \boldsymbol{p}^d + \varepsilon q$ satisfies $f(\boldsymbol{x}) > 0$ for $\varepsilon > 0$ small enough (as $f(\boldsymbol{x}(\varepsilon)) = q'q\varepsilon + O(\varepsilon^2)$).

Finally, $\delta \boldsymbol{c}^s = S^r \delta \boldsymbol{p}^s$ implies $\delta \boldsymbol{p} \cdot \delta \boldsymbol{c}^s = \delta \boldsymbol{p} \cdot S^r \delta \boldsymbol{p}^s \leq 0$. □

Here is more analysis of the car plus gas example. Goods 1, 2 and 3 are car, gas and food. The default price is $\boldsymbol{p}^d = (1, 1, 2)$, and expenditure shares are $(1/4, 1/4, 1/2)$. The Slutsky matrix is: $S^r = \begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$, times a constant that we normalize to 1. This can be rationalized by a utility function $\ln(\min(c_1, c_2)) + \ln c_3$: car and gas are both needed for transportation. Hence, for "transportation", the price is $p_1 + p_2$, the price of car plus gas.

Let's say that attention is $m = (1, 0, 1)$, i.e. people pay attention to the car price and food price, but not to the price of gas. This is meant to capture lower attention to energy consumption, when the agent buys the car. Suppose now that there is a decrease in the car price, and an increase in the price of gas, say $\delta \boldsymbol{p} = (-1, 2, 0)$. The rational agent sees that the total price of transportation has increased by $-1 + 2 = 1$, so he should consume less transportation – less car and gas. However, a sparse agent perceives a price $\delta \boldsymbol{p}^s = (m_i \delta p_i)_{=1\ldots 3} = (-1, 0, 0)$: he rejoices as the price of a car has decreased, but he does not see the increase of the gas price. He thinks that the price of transportation has decreased, so he consumes *more* of car plus gas. The price of transportation has truly increased, but he consumes more of it. Mathematically, the $\delta \boldsymbol{c}^r = S^r \delta \boldsymbol{p} = (-1, -1, 1)$, while $\delta \boldsymbol{c}^s = S^r \delta \boldsymbol{p}^s = (1, 1, -1)$: so $\delta \boldsymbol{p} \cdot \delta \boldsymbol{c}^r = -1$, while $\delta \boldsymbol{p} \cdot \delta \boldsymbol{c}^s = 1$.

Proof of Proposition 19 We have: $e^s(\boldsymbol{p}, u) = \boldsymbol{p} \cdot \boldsymbol{h}^s(\boldsymbol{p}, u) = \boldsymbol{p} \cdot \boldsymbol{h}^r(\boldsymbol{p}^s, u)$. As the default, $\boldsymbol{p} = \boldsymbol{p}^d = \boldsymbol{p}^s$, hence $e^s(\boldsymbol{p}^d) = e^r(\boldsymbol{p}^d)$ and $e^s_{\boldsymbol{p}}(\boldsymbol{p}^d, u) = \boldsymbol{p}^d \cdot \boldsymbol{h}^r(\boldsymbol{p}^d, u) + \boldsymbol{h}^r_{\boldsymbol{p}}(\boldsymbol{p}^d, u) = e^r_{\boldsymbol{p}}(\boldsymbol{p}^d, u)$.

We have $e^s_{\boldsymbol{p}} = \boldsymbol{h}^s + \boldsymbol{p} \cdot \boldsymbol{h}^s_{\boldsymbol{p}}(\boldsymbol{p})$ and, taking the derivative at $\boldsymbol{p}^d$,

$$\begin{aligned} e^s_{pp} &= \boldsymbol{h}^s_{\boldsymbol{p}} + \boldsymbol{h}^{s\prime}_{\boldsymbol{p}} + \boldsymbol{p} \cdot \boldsymbol{h}^s_{pp} = \boldsymbol{h}^{R\prime}_{\boldsymbol{p}} M + M h^s_{\boldsymbol{p}} + \boldsymbol{p} \cdot \boldsymbol{h}^s_{pp} \\ &= \boldsymbol{h}^{R\prime}_{\boldsymbol{p}} M + M h^r_{\boldsymbol{p}} + \boldsymbol{p} \cdot \boldsymbol{h}^s_{pp}. \end{aligned}$$

(Here, the expression $\boldsymbol{p} \cdot \boldsymbol{h}^s_{pp}$ is understood as the matrix with terms $\boldsymbol{p} \cdot \boldsymbol{h}^s_{p_i p_j}$). As

$$\begin{aligned} &\boldsymbol{h}^s = \boldsymbol{h}^r\left((I - M)\boldsymbol{p}^d + Mp\right), \qquad \boldsymbol{h}^s_{\boldsymbol{p}} = \boldsymbol{h}^r_{\boldsymbol{p}}\left((I - M)\boldsymbol{p}^d + Mp\right) M, \\ &\boldsymbol{h}^s_{pp} = M h^r_{pp} M, \end{aligned}$$

and as $\boldsymbol{h}^r(\lambda \boldsymbol{p}) = \boldsymbol{h}^r(\boldsymbol{p})$ for all $\lambda > 0$, we have (differentiating with respect to $\boldsymbol{p}$), $\lambda h^r_{\boldsymbol{p}}(\lambda \boldsymbol{p}) = \boldsymbol{h}^r_{\boldsymbol{p}}(\boldsymbol{p})$, and (differentiating with respect to λ at $\lambda = 1$), $\boldsymbol{h}^r_{\boldsymbol{p}} + \boldsymbol{h}^r_{pp} \cdot \boldsymbol{p} = 0$, so

$$\boldsymbol{p} \cdot \boldsymbol{h}^s_{pp} = M\left(\boldsymbol{p} \cdot \boldsymbol{h}^r_{pp}\right) M = -M h^r_{\boldsymbol{p}} M = -M e^r_{pp} M \text{ where } \boldsymbol{h}^r_{\boldsymbol{p}} = e^r_{pp} \text{ is standard,}$$

$$e^s_{pp} = e^r_{pp} M + M e^r_{pp} - M e^r_{pp} M.$$

This expression can be written coordinate-wise: $e^s_{p_i p_j} = e^r_{p_i p_j}\left(m_i + m_j - m_i m_j\right)$, and $e^s_{pp} = e^r_{pp} - (I - M)\, e^r_{pp}\, (I - M)$. The last expression can be understood as: the sparse expenditure function (in its second derivative) is the rational one plus an extra cost due to the lack of perfect optimization. That cost is $(I - M)\, e^r_{pp}\, (I - M)$.

This and Proposition 6 (which can be stated $S^s_{ij} = e^r_{ij} m_j$) imply: $e^s_{p_i p_j} = S^s_{ij} + S^s_{ji} - \frac{S^s_{ij} S^s_{ji}}{S^r_{ij}}$.

Proof of Proposition 20 The fact that $\boldsymbol{c}^s\left(\boldsymbol{p}^d, w\right) = \boldsymbol{c}^r\left(\boldsymbol{p}^d, w\right)$ implies that $v^s\left(\boldsymbol{p}^d, w\right) = v^r\left(\boldsymbol{p}^d, w\right)$, hence v, v_w, v_{ww} is identical for both models. To go further, we use Proposition 26. The fact that $v_{\boldsymbol{p}}$ is the same for both models comes from (49). Then, applying (50) to the shift $x = (w, \boldsymbol{p})$ gives:

$$v^s_{wp} - v_{wp} = \left(\boldsymbol{c}^s_w - \boldsymbol{c}_w\right)' L_{cc}\left(\boldsymbol{c}^s_w - \boldsymbol{c}_w\right) = 0,$$

where we use $\boldsymbol{c}^s_w - \boldsymbol{c}_w = 0$, which comes from $\boldsymbol{c}^s\left(\boldsymbol{p}^d, w\right) = \boldsymbol{c}^r\left(\boldsymbol{p}^d, w\right)$ for all w.

For v^s_{pp} we rely on duality: we have, for all $\boldsymbol{p}$ and w, and for both the behavioral and traditional model: $v\left(\boldsymbol{p}, e\left(\boldsymbol{p}, w\right)\right) = w$, so

$$v_{\boldsymbol{p}}\left(\boldsymbol{p}, e\left(\boldsymbol{p}, w\right)\right) + v_w\left(\boldsymbol{p}, e\left(\boldsymbol{p}, w\right)\right) \cdot e_{\boldsymbol{p}}\left(\boldsymbol{p}, w\right) = 0,$$

and taking the derivative at $\boldsymbol{p} = \boldsymbol{p}^d$,

$$v_{pp} + 2V_{pw} e_{\boldsymbol{p}} + e_{\boldsymbol{p}} v_{ww} e_{\boldsymbol{p}} + v_w e_{pp} = 0.$$

This is true for both the sparse (s) and rational (r) models: Also, note that $e_{\boldsymbol{p}}, v_w, v_{ww}$ and v_{wp} are the same for both models. Hence

$$v^s_{pp} + 2V_{pw} e_{\boldsymbol{p}} + e_{\boldsymbol{p}} v_{ww} e_{\boldsymbol{p}} + v_w e^s_{pp} = 0,$$
$$v^r_{pp} + 2V_{pw} e_{\boldsymbol{p}} + e_{\boldsymbol{p}} v_{ww} e_{\boldsymbol{p}} + v_w e^r_{pp} = 0.$$

The common values are left without a superscript, e.g. $e_{\boldsymbol{p}} = e^r_{\boldsymbol{p}} = e^s_{\boldsymbol{p}}$. Subtracting those two equations yields: $v^s_{pp} - v^r_{pp} = -v_w\left(e^s_{pp} - e^r_{pp}\right)$.

Finally, the expression in S comes from: $e^s_{pp} = e^r_{pp} - (I - M)'\, e^r_{pp}\, (I - M)$.

Proof of Proposition 21 *Shephard's lemma.* We have: $e^s(\boldsymbol{p}, u) = \boldsymbol{p} \cdot \boldsymbol{h}^r(\boldsymbol{p}^s, u)$, so

$$e^s_{p_j} = h^r_j + \boldsymbol{p} \cdot \boldsymbol{h}^r_{p_j}(\boldsymbol{p}^s, u) m_j = h^r_j + (\boldsymbol{p} - \boldsymbol{p}^s) \cdot \boldsymbol{h}^r_{p_j}(\boldsymbol{p}^s, u) m_j,$$

as $\boldsymbol{h}^r(\boldsymbol{p}', u)$ is homogenous of degree 0 in $\boldsymbol{p}'$, so that $\boldsymbol{p}' \cdot \boldsymbol{h}^r_{p'_j}(\boldsymbol{p}', u) = 0$, and $\boldsymbol{p}^s \cdot \boldsymbol{h}^r_{p_j}(\boldsymbol{p}^s, u) = 0$.

Roy's identity. Let us first calculate $\frac{\partial w'}{\partial w}$ and $\frac{\partial w'}{\partial p_j}$. Proposition 2 gives $\boldsymbol{p} \cdot \boldsymbol{c}^r(\boldsymbol{p}^s, w') = w$, so $\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'} \frac{\partial w'}{\partial w} = 1$, and $\frac{\partial w'}{\partial w} = \frac{1}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}$. Also, taking the derivative with respect to p_j, $c^r_j + \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j} m_j + \boldsymbol{p} \cdot \boldsymbol{c}^r_{w'} \frac{\partial w'}{\partial p_j} = 0$, so:

$$\frac{\partial w'}{\partial p_j} = \frac{-c^r_j - \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j}(\boldsymbol{p}^s, w') m_j}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}.$$

Now, note that $\boldsymbol{c}^s_w = \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w') \frac{\partial w'}{\partial w}$ (from eq. 12) implies:

$$\boldsymbol{c}^s_w = \frac{\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}.$$

Next, given $v^s(\boldsymbol{p}, w) = u(\boldsymbol{c}^s(\boldsymbol{p}, w))$ we have: $v^s_w = u'(\boldsymbol{c}^s) \cdot \boldsymbol{c}^s_w$, hence $v^s_w = \lambda \boldsymbol{p}^s \cdot \frac{\boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}$, and as $q \cdot \boldsymbol{c}^r_{w'}(q, w') = 1$ (which comes from $q^r \cdot \boldsymbol{c}(q, w') = w'$),

$$v^s_w = \frac{\lambda}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}. \tag{69}$$

Finally, $\boldsymbol{c}^s(\boldsymbol{p}, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w')$ gives:

$$\begin{aligned} \boldsymbol{c}^s_{p_j} &= \boldsymbol{c}^r_{p_j}(\boldsymbol{p}, w') m_j + \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w') \frac{\partial w'}{\partial p_j}, \\ \boldsymbol{c}^s_{p_j} &= \boldsymbol{c}^r_{p_j}(\boldsymbol{p}, w') m_j - \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w') \frac{c^r_j + \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j}(\boldsymbol{p}^s, w') m_j}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')}, \end{aligned} \tag{70}$$

so

$$\begin{aligned} v^s_{p_j} &= u'(\boldsymbol{c}) \cdot \boldsymbol{c}^s_{p_j}(\boldsymbol{p}, w) \\ &= \lambda \boldsymbol{p}^s \cdot \left[\boldsymbol{c}^r_{p_j}(\boldsymbol{p}, w') m_j - \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w') \frac{c^r_j + \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j}(\boldsymbol{p}^s, w') m_j}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')} \right] \\ &= \lambda \boldsymbol{p}^s \cdot \boldsymbol{c}^r_{p_j}(\boldsymbol{p}, w') m_j - \lambda \frac{\boldsymbol{c}^r_j + \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j}(\boldsymbol{p}^s, w') m_j}{\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}(\boldsymbol{p}^s, w')} \\ &= \lambda \boldsymbol{p}^s \cdot \boldsymbol{c}^r_{p_j}(\boldsymbol{p}, w') m_j - \left(c^r_j + \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j}(\boldsymbol{p}^s, w') m_j \right) v^s_w. \end{aligned}$$

Hence

$$\begin{aligned}
\frac{v^s_{p_j}}{v^s_w} &= \boldsymbol{p}^s \cdot \boldsymbol{c}^r_{p_j}\left(\boldsymbol{p}, w'\right) m_j \times \boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}\left(\boldsymbol{p}^s, w'\right) - \boldsymbol{c}^r_j - \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j}\left(\boldsymbol{p}^s, w'\right) m_j \\
&= -c^r_j + \left[\boldsymbol{p}^s \cdot \boldsymbol{c}^r_{p_j}\left(\boldsymbol{p}, w'\right) \times \boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}\left(\boldsymbol{p}^s, w'\right) - \boldsymbol{p} \cdot \boldsymbol{c}^r_{p_j}\left(\boldsymbol{p}^s, w'\right)\right] m_j \\
&= -c^r_j + \left[\left(\boldsymbol{p} \cdot \boldsymbol{c}^r_{w'}\left(\boldsymbol{p}^s, w'\right)\right) \boldsymbol{p}^s - \boldsymbol{p}\right] \cdot \boldsymbol{c}^r_{p_j}\left(\boldsymbol{p}^s, w'\right) m_j.
\end{aligned}$$

Note that in (23) at $\boldsymbol{p} = \boldsymbol{p}^d$, we recover the traditional Roy's identity because then $\boldsymbol{p}^s = \boldsymbol{p}$ and $\boldsymbol{p} \cdot c^r_{w'}\left(\boldsymbol{p}, w'\right) = 1$.

Proof of the Proposition 22 Suppose over $t \in [0,1]$, $p_i(t) = p^d_i + x_i t$: the price increases continuously, and the other prices remain constant. Following the usual Shephard's lemma, the econometrician will measure the welfare loss (in a money metric) as $\Delta e^{\text{naive}} = \int_{t=0}^1 h^s_i\left(\boldsymbol{p}(t), u\right) dp_1$. However, the true loss is: (with $S(t)$ the rational Slutsky matrix at $h\left(p^s(t), u\right)$)

$$\begin{aligned}
\Delta e^{\text{true}} &= e\left(p_i(1), u\right) - e\left(p_i(0), u\right) = \int_{t=0}^1 e_{p_i}\left(p(t), u\right) dp_i(t) \\
&= \int_{t=0}^1 \boldsymbol{h}^s_i\left(\boldsymbol{p}(t), u\right) dp_i(t) + \int_0^1 (1-m_i) x_i t S_{ii}(t)\, m_i dp_i(t) \text{ by (22) and } \boldsymbol{h}^r_p = S(t), \\
\Delta e^{\text{true}} &= \Delta e^{\text{naive}} + m_i(1-m_i) x_i^2 \int_0^1 t S_{ii}(t)\, dt. \qquad (71)
\end{aligned}$$

Hence, as $S_{ii} \leq 0$, $\Delta e^{\text{true}} \leq \Delta e^{\text{naive}}$, whatever the sign of x_i. For a small x_i, $S_{ii}(t)$ is approximately constant and:

$$\Delta e^{\text{true}} = \Delta e^{\text{naive}} + \frac{1}{2} m_i(1-m_i) x_i^2 S_{ii} + o\left(x_i^2\right). \qquad (72)$$

An illustrative example To understand the situation, a quasi-linear example is useful. Take:

$$u(\boldsymbol{c}) = c_n + \sum_{i=1}^{n-1} \frac{c_i^{1-1/\psi_i}}{1 - 1/\psi_i}, \qquad (73)$$

with $p_n = 1$ and $\psi_i > 0$. The demand is $c_i = \left(p^s_i\right)^{-\psi_i}$, indirect utility:

$$\begin{aligned}
v(\boldsymbol{p}, w) &= w - \sum_{i<n} p_i c_i + \sum_{i<n} \frac{c_i^{1-1/\psi_i}}{1 - 1/\psi_i} \\
&= w - \sum_{i<n} \left[\frac{\psi_i}{1-\psi_i} \left(p^s_i\right)^{1-\psi_i} + \left(p^s_i\right)^{1-\psi_i}\right], \qquad (74)
\end{aligned}$$

and the expenditure function

$$e(p,v) = v + \sum_{i<n} \frac{\psi_i}{1-\psi_i} \left(p_i^s\right)^{1-\psi_i} + \left(p_i^s\right)^{1-\psi_i}, \quad (75)$$

so $e_{p_i}^r = c_i = p_i^{-\psi_i}$ and $S_{ii}^r = e_{p_i p_i}^r = -\psi_i \frac{c_i}{p_i}$.

Suppose now that we increase the price of good i from p_i^d to $p_i^d + \Delta p_i$, $\Delta p_i = x_i$. In the rational model, total expenditure increases by (neglecting here and below terms of the third order and higher):

$$\Delta e^r = e(p_i + x_i) - e(p_i) = e_{p_i}^r x_i + \frac{1}{2} e_{p_i p_i}^r x_i^2,$$

$$\Delta e^r = c_i x_i + \frac{S_{ii} x_i^2}{2}, \quad (76)$$

$$\Delta e^r = c_i \Delta p_i + \frac{\Delta c_i^r \Delta p_i}{2}, \quad (77)$$

as a change $\Delta p_i = x_i$ induces a consumption change (for good i) $\Delta c_i^r = S_{ii} \Delta p_i$ for the rational agent; recall also that it induces a change $\Delta c_i^s = S_{ii} \Delta p_i m_i$ for the sparse agent.

For the sparse agent, Proposition 19 gives:

$$\Delta e^s = c_i x_i + \frac{S_{ii} x_i^2}{2} \left(2m_i - m_i^2\right). \quad (78)$$

The naive Δe^{naive} will use the changes at time t:

$$h_i(t) - h_i(0) = h_{p_i}^i x_i m_i t = S_{ii} x_i m_i t. \quad (79)$$

The quantity change at time t is muted by the attention m_i. Hence,

$$\Delta e^{\text{naive}} = \int_{t=0}^{1} \boldsymbol{h}_1^s(\boldsymbol{p}(t), u)\, dp_1(t) = \int_0^1 \left(c_1^d + S_{ii} x_i m_i t\right) x_i dt,$$

$$\Delta e^{\text{naive}} = c_1^d x_1 + \frac{S_{ii} x_i^2}{2} m_i,$$

$$\Delta e^{\text{naive}} = c_i \Delta p_i + \frac{\Delta c_i^s \Delta p_i}{2}. \quad (80)$$

This is, Δe^{naive} naively uses the procedure (77), applied to the actual change Δc_i^s, but forgetting about inattention.

Hence, the difference between the two gives: $\Delta e^s - \Delta e^{\text{naive}} = \frac{S_{ii} x_i^2}{2} \left[\left(2m_i - m_i^2\right) - m_i\right]$, i.e.

$$\Delta e^s - \Delta e^{\text{naive}} = \frac{S_{ii} x_i^2}{2} m_i (1 - m_i). \quad (81)$$

This is the result previously obtained via Shephard's lemma analysis (72) (with $\Delta e^s = \Delta e^{\text{true}}$). It can also be expressed:

$$\Delta e^s - \Delta e^{\text{naive}} = \Delta p_i \Delta c_i^s (1 - m_i). \tag{82}$$

This explains the "U-shape": the underestimation is always $\Delta p_i \Delta c_i^s (1 - m_i)$, so it is proportional to $1 - m_i$ for a given consumption change; but the consumption is proportional to m_i: so the total effect is proportional to $m_i (1 - m_i)$.

Finally, in the quasi-linear case with $c_w^r(\boldsymbol{p}, w) = 1$, the modified Roy's identity formula gives, where there is just a change δp_i

$$\begin{aligned} c_i^s(\boldsymbol{p}, w) + v_{p_i}^s(\boldsymbol{p}, w) &= \left[\left(\boldsymbol{p} \cdot c_{w'}^r(\boldsymbol{p}^s, w')\right) \boldsymbol{p}^s - \boldsymbol{p}\right] \cdot c_{p_i}^r(\boldsymbol{p}^s, w') m_i \\ &= \left[\boldsymbol{p}^s - \boldsymbol{p}\right] \cdot c_{p_i}^r(\boldsymbol{p}^s, w') m_i \\ &= (p_i^s - p_i) h_{p_i}^r(\boldsymbol{p}^s, u) m_i, \end{aligned}$$

which is the right-hand side of the (22). Hence, a naive use of Roy's identity will lead to the same bias, with the same loss as in Shephard (72).

Proofs of the Statements on (Sub)optimization

Proof of Proposition 24 Differentiating $L_a(a, x, \lambda) = 0$ with respect to x,

$$0 = L_{aa} a_x + L_{ax} + L_{a\lambda} \lambda_x, \tag{83}$$

so as $L_{a\lambda} = b_a$,

$$a_x = -L_{aa}^{-1}(L_{ax} + b_a \lambda_x). \tag{84}$$

The budget constraint $b(a, x) = 0$ gives:

$$0 = b_a' a_x + b_x' = -b_a' L_{aa}^{-1}(L_{ax} + b_a \lambda_x) + b_x',$$

This allows us to find λ_x :

$$\lambda_x = \left(b_a' L_{aa}^{-1} b_a\right)^{-1} \left(b_x' - b_a' L_{aa}^{-1} L_{ax}\right).$$

Note that if there is just one budget constraint, then $b_a' L_{aa}^{-1} b_a$ is simply a scalar. When there are K budget constraints, then $b_a' L_{aa}^{-1} b_a$ is an invertible $K \times K$ matrix (as L is strictly concave and the constraints bind), while, for a change δx, both $b_x' \delta x$ and $b_a' L_{aa}^{-1} L_{ax} \delta x$ are K-dimensional vectors. Hence $\lambda_x \delta x$ is a K-dimensional vector.

Finally, we have:

$$
\begin{aligned}
a_x &= -L_{aa}^{-1}\left(L_{ax}+b_a\lambda_x\right) \\
&= -L_{aa}^{-1}\left[L_{ax}+b_a\left(b_a'L_{aa}^{-1}b_a\right)^{-1}\left(b_x'-b_a'L_{aa}^{-1}L_{ax}\right)\right] \\
&= -L_{aa}^{-1}b_a\left(b_a'L_{aa}^{-1}b_a\right)^{-1}b_x'-\left[I-L_{aa}^{-1}b_a\left(b_a'L_{aa}^{-1}b_a\right)^{-1}b_a'\right]L_{aa}^{-1}L_{ax}.
\end{aligned}
$$

Part concerning the value function.

Differentiating with respect to x:

$$
v_x = L_x + L_a a_x + L_\lambda \lambda_x, \tag{85}
$$

so at $x=0$, $v_x = L_x$. The other part comes from differentiating (85), and observing that $L_a a_{xx} = 0$ and $L_\lambda \lambda_{xx} = 0$ because $L_a = L_\lambda = 0$ at the default:

$$
\begin{aligned}
v_{xx} &= v_{xx|x=x^d} \\
&= L_{xx} + a_x L_{aa} a_x + \lambda L_{\lambda\lambda}\lambda_x + 2\left(a_x L_{ax} + L_{\lambda x}\lambda_x + a_x L_{a\lambda}\lambda_x\right) \\
&= L_{xx} + a_x L_{aa} a_x + 0 + 2\left(-a_x\left(L_{aa}a_x + L_{a\lambda}\lambda_x\right) + L_{\lambda x}\lambda_x + a_x L_{\lambda a}\lambda_x\right) \text{ using } L_{ax} + L_{aa}a_x + L_{a\lambda}\lambda_x = 0 \\
&= L_{xx} + a_x L_{aa} a_x + 0 + 2\left(-a_x L_{aa} a_x + b_x \lambda_x\right) \text{ using } L_{\lambda x} = b_x \\
&= L_{xx} - a_x L_{aa} a_x + 2 b_x \lambda_x.
\end{aligned}
$$

□

Proof of Proposition 25. Note that $\overline{a}(x) := a\left(x, -b\left(a^d, x\right)\right)$ implies directly $\overline{a}_x = a_x - a_y b_x$. To derive a_y, apply equation (42) to $b(a,y) := b(a,0) + y$ (and replacing "x" by "y"). As $b_y' = 1$ and $L_{ay} = 0$, this gives $a_y = -\left(L_{aa}^{-1}b_a\right)\left(b_a'L_{aa}^{-1}b_a\right)^{-1}$.

Next, equation (42) applied to the original problem with $b(a,x) \geq 0$ gives:

$$
\begin{aligned}
a_x &= -\left(L_{aa}^{-1}b_a\right)\left(b_a'L_{aa}^{-1}b_a\right)^{-1}b_x' - \left[I - L_{aa}^{-1}b_a\left(b_a'L_{aa}^{-1}b_a\right)^{-1}b_a'\right]L_{aa}^{-1}L_{ax}, \\
&= a_y b_x' - \left(I + a_y b_a'\right)L_{aa}^{-1}L_{ax},
\end{aligned}
$$

which gives

$$
\overline{a}_x = a_x - a_y b_x = -\left(I + a_y b_a'\right)L_{aa}^{-1}L_{ax}.
$$

From $b\left(\overline{a}(x), x\right) = b\left(a^d, x\right)$, we get $b_a \overline{a}_x = 0$.

From $b\left(a(x,y), x\right) + y = 0$, we get $b_a a_y + I = 0$.

Proof of Proposition 26. Call $h(x) = a^x(x) - a(x)$, $g(a,x) = L(a,\lambda(x),x)$

$$\begin{aligned}
f(x) &:= v^B(x) - v(x) = L(a^x(x),\lambda(x),x) - L(a(x),\lambda(x),x) \\
&= g(a(x)+h(x),x) - g(a(x),x), \\
f_x(x) &= [g_a(a_x+h_x)+g_x]_{|(a(x)+h(x),\lambda(x)x)} - [g_a(a_x+h_x)+g_x]_{|(a(x),\lambda(x)x)}, \\
f_{xx}\left(x^d\right) &= (a_x+h_x)\,g_{aa}(a_x+h_x) + 2g_{ax}(a_x+h_x) + g_{xx} - [a_x g_{aa} a_x + 2g_{ax}a_x + g_{xx}] \\
&= h_x g_{aa} h_x + 2h_x(g_{aa}a_x + g_{ax}) \\
&= h_x g_{aa} h_x,
\end{aligned}$$

as $g_a(a(x),x) = 0$ implies $g_{aa}a_x + g_{ax} = 0$. That proves:

$$v_{xx}^B - v_{xx} = \left(a_x^B - a_x\right) L_{aa}\left(a_x^B - a_x\right).$$

Also, as $g_a = 0$, $f_x\left(x^d\right) = g_a h_x = 0$, which proves $v_x^B = L_x$.

Proof of Proposition 27

We first calculate $a^s - a^r$. We do a Taylor expansion (using infinitesimal x, x^s, y)[87]:

$$\begin{aligned}
a^s - a^r &= a(x^s, y^s) - a(x,0) \\
&= a_x(x^s - x) + a_y y^s.
\end{aligned}$$

To find the budget compensation y^s, do a Taylor expansion of: $b(a(x^s,y^s),x) = b(a(x^s,0),x) = 0$:

$$b_a[a_x(x^s - x) + a_y y^s] = 0.$$

As for all y, $b(a(x,y)) + y = 0$, which gives (taking the derivative with respect to y), $b_a a_y = -I$,

$$y^s = b_a a_x(x^s - x).$$

Finally,

$$\begin{aligned}
a^s - a^r &= a_x(x^s - x) + a_y y^s \\
&= a_x(x^s - x) + a_y b_a a_x(x^s - x) \\
&= (1 + a_y b_a)\,a_x(x^s - x), \\
a^s - a^r &= \overline{a}_x(x^s - x), \qquad (86)
\end{aligned}$$

[87] This is a bit unorthodox, but it simplifies the notation nicely. It's very easy, of course, to add $+o(x)+o(y)+o(x^s)$ everywhere, to use conventional rather than infinitesimal calculus.

where we used the identity:

$$(1 + a_y b_a) a_x = \overline{a}_x, \tag{87}$$

proven now (by Proposition 25)

$$\begin{aligned}
(1 + a_y b_a) a_x &= (1 + a_y b_a)(\overline{a}_x + a_y b_a) \\
&= \overline{a}_x + a_y b_a \overline{a}_x + a_y b_a + a_y b_a a_y b_a \\
&= \overline{a}_x + 0 + a_y b_a - a_y b_a \text{ as } b_a \overline{a}_x = 0 \text{ and } b_a' a_y = -1 \\
&= \overline{a}_x.
\end{aligned}$$

Finally,

$$\begin{aligned}
v^s - v^r &= \left(v_x^B - v_x\right) x + \frac{1}{2} x' \left(v_{xx}^B - v_{xx}\right) x + o\left(\|x^s - x\|^2\right) \\
&= 0 + (a^s - a^r)' L_{aa} (a^s - a^r) + o\left(\|x^s - x\|^2\right) \text{ by Proposition 26} \\
&= \frac{1}{2} (x^s - x)' \overline{a}_x' L_{aa} \overline{a}_x (x^s - x) + o\left(\|x^s - x\|^2\right) \text{ by (86).}
\end{aligned}$$

XII. Complements to Consumer and Equilibrium Theory

XII.A. Complements to Consumer Theory

In the paper, the price perceived by a sparse agent varies additively with the true price:

$$p_i^s(m) = m_i p_i + (1 - m_i) p_i^d. \tag{88}$$

More general functions $p_i^s(m)$ could be devised. E.g., perceptions can be in percentage terms, i.e. be loglinear,

$$\ln p_i^s(m) = m_i \ln p_i + (1 - m_i) \ln p_i^d. \tag{89}$$

Quantitatively, this makes very little difference, as in both cases, $\frac{\partial p_i^s}{\partial p_t}_{|p=p^d} = m_i$. The log version is a bit more convenient in the most theoretical analyses, because when p_i ranges from 0 to ∞, so does p_i^s in the loglinear version (89), but not in the additive version (88). All properties in section III apply also with this log-linear formulation, because the key property employed in the proof is $\frac{\partial p_i^s}{\partial p_t}_{|p=p^d} = m_i$. Indeed, even the optimal m_i^* is the same in both versions.[88]

On the offer curve The next Proposition formalizes the notion that, if m does not have all equal components, then the offer curve has "one extra dimension" compared to the traditional

[88] To verify this intuition, here is the proof of Proposition 3 in the loglinear case. We have $L = u(\boldsymbol{c}) + \lambda(w - \boldsymbol{p}^s(\boldsymbol{x}) \cdot \boldsymbol{c})$ with $x_i := \ln \frac{p_i}{p_i^d}$ so that $p_i(\boldsymbol{x}) = p_i^d e^{x_i}$. So, $L_{\boldsymbol{c}} = u'(\boldsymbol{c}) - \lambda p(\boldsymbol{x})$, $L_{\boldsymbol{cc}} = u''(c)$, $L_{\boldsymbol{cx}} = -\lambda D$,

model.

Proposition 32 (Extra-dimensional offer curve). *Take a price $\boldsymbol{p}$ such that $\boldsymbol{p}^s \cdot \left(\frac{\partial \boldsymbol{p}^s}{\partial \boldsymbol{p}}\right)^{-1} (\boldsymbol{c} - \boldsymbol{\omega}) \neq 0$. Then around $\boldsymbol{D}(\boldsymbol{p})$, the offer curve of the sparse agent has one extra dimension compared to the traditional model, i.e. it has dimension n.*

Proof It is enough to show that for small $\delta \boldsymbol{c}$, we can find a small $\delta \boldsymbol{p}$ such that $\boldsymbol{D}(\boldsymbol{p} + \delta \boldsymbol{p}) = \boldsymbol{c} + \delta \boldsymbol{c}$, i.e. there is a $\delta\lambda$ such that $\Phi(\boldsymbol{p} + \delta \boldsymbol{p}, \lambda + \delta\lambda, \boldsymbol{c} + \delta \boldsymbol{c}) = 0$, with

$$\Phi(\boldsymbol{p}, \lambda, \boldsymbol{c}) = \begin{pmatrix} u_{\boldsymbol{c}}(\boldsymbol{c}) - \lambda \boldsymbol{p}^s(\boldsymbol{p}) \\ \boldsymbol{p} \cdot (\boldsymbol{c} - \boldsymbol{\omega}) \end{pmatrix}.$$

It is enough that $\partial\Phi$ (the derivative with respect to $(\boldsymbol{p}, \lambda)$) has rank $n+1$. We calculate $\partial\Phi = \begin{pmatrix} \frac{\partial \boldsymbol{p}^s}{\partial \boldsymbol{p}} & -\boldsymbol{p}^s \\ \boldsymbol{c} - \boldsymbol{\omega} & 0 \end{pmatrix}$. To see if $\partial\Phi$ has rank $n+1$, we show that, given arbitrary $\delta u, \delta b$, we can find $\delta \boldsymbol{p}, \delta\lambda$ such that $\partial\Phi \binom{\delta \boldsymbol{p}}{\delta\lambda} = \binom{\delta u}{\delta b}$, i.e. $\begin{pmatrix} \frac{\partial \boldsymbol{p}^s}{\partial \boldsymbol{p}} & -\boldsymbol{p}^s \\ \boldsymbol{c} - \boldsymbol{\omega} & 0 \end{pmatrix} \binom{\delta \boldsymbol{p}}{\delta\lambda} = \binom{\delta u}{\delta b}$. The first equation gives $\delta \boldsymbol{p} = \left(\frac{\partial \boldsymbol{p}^s}{\partial \boldsymbol{p}}\right)^{-1} (\delta u + \boldsymbol{p}^s \delta\lambda)$, and the second equation gives $(\boldsymbol{c} - \boldsymbol{\omega}) \left(\frac{\partial \boldsymbol{p}^s}{\partial \boldsymbol{p}}\right)^{-1} (\delta u + \boldsymbol{p}^s \delta\lambda) = \delta b$. This has a solution in $\delta\lambda$ if $(\boldsymbol{c} - \boldsymbol{\omega}) \left(\frac{\partial \boldsymbol{p}^s}{\partial \boldsymbol{p}}\right)^{-1} \boldsymbol{p}^s \neq 0$. $\square$

The restriction implies that $\boldsymbol{c} \neq \boldsymbol{\omega}$: we do not start at the endowment (this can be seen in the "pinch" at $\boldsymbol{\omega}$ in Figure III, right panel). It also implies that in the log linear model, m does not have all identical components – i.e., the consumer pays more attention to some goods than others.[89]

We show that when inattention is unlimited, the offer curve is very wide indeed. This "wide" OC effect relies on potentially extreme prices and misperceptions. In slight variants, the OC does not cover the whole space, e.g. if there is "limited misperception."[90]

Example 5 (Wide offer curves with unbounded inattention). *Suppose that there two goods, with different attention ($m_1 \neq m_2$) in the loglinear specification (89). Then, any consumption that does not dominate the endowment nor is dominated by it, is in the consumer's offer curve.*

---

with $D := Diag\left(p_i^d\right)$, and $\boldsymbol{c_x} = -L_{cc}^{-1} L_{c\boldsymbol{x}} = u''^{-1}\lambda D$. In Step 1, $\boldsymbol{c_x} L_{cc} \boldsymbol{c_x} = \lambda^2 D u''^{-1} D$, so the problem is:

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

where $\sigma_{p_i}^2 := \left(p_i^d\right)^2 var(x_i) = \left(p_i^d\right)^2 var(\ln p_i)$. The rest is as in the proof of Proposition 3.

[89] Indeed, if $m = \overline{m}\iota$, then $\boldsymbol{p}^s \cdot \left(\frac{\partial \boldsymbol{p}^s}{\partial \boldsymbol{p}}\right)^{-1} = \overline{m}^{-1} \boldsymbol{p}$, and the condition of Proposition 32 is $\boldsymbol{p} \cdot (\boldsymbol{c} - \boldsymbol{\omega}) \neq 0$, which is not satisfied.

[90] Such a limited misperception can come from "attention allocated ex post": $p_i(m) = p_i^d \cdot \exp \tau(\ln \frac{p_i}{p_i^d}, \alpha)$, where $\alpha > 0$ and τ is as in (107). Then, $\left|\ln \frac{p_i(m)}{p_i}\right| \leq \alpha$. With that model, we have a less extreme OC, but it retains its extra-dimensional shape. It is very similar to Figure III.

Proof Let $\boldsymbol{c} = (c_1, c_2)$ be the candidate consumption. Suppose that $(c_1, c_2) \neq (\omega_1, \omega_2)$. We're looking to see if there are prices $\boldsymbol{p}$ such that $\boldsymbol{D}(\boldsymbol{p}) = c$, i.e. if there is a $\lambda > 0$ and a $\boldsymbol{p}$ such that: $u_{c_i}(\boldsymbol{c}) = \lambda p_i^{m_i}$ and $\boldsymbol{p} \cdot (\boldsymbol{\omega} - \boldsymbol{c}) = 0$. Given λ, we just set $\boldsymbol{p}_i = (u_{c_i}(\boldsymbol{c})/\lambda)^{1/m_i}$, so the equation to solve is $f(\lambda) = 0$ where $f(\lambda) = \boldsymbol{p}(\lambda) \cdot (\boldsymbol{\omega} - \boldsymbol{c})$, i.e. $f(\lambda) = \sum_i \lambda^{-1/m_i} u_{c_i}(\boldsymbol{c})^{1/m_i} (\omega_i - c_i)$.

Suppose, without loss of generality, that $m_1 > m_2$. Then as $\lambda \to 0$, $f(\lambda) \sim \lambda^{-1/m_1} u_{c_1}(\boldsymbol{c})^{1/m_1} (\omega_1 - c_1)$, and $sign(f(\lambda)) = sign(\omega_1 - c_1)$. Likewise, as $\lambda \to \infty$, $sign(f(\lambda)) = sign(\omega_2 - c_2)$. By lack of domination, $\omega_1 - c_1$ and $\omega_2 - c_2$ have opposite signs. Hence, $f(\lambda)$ has opposite signs in near 0 and near infinity. By the intermediate value theorem, there is a λ^* such that $f(\lambda^*) = 0$.

The proof suggests a way this property might fail with more than two goods. Order them such that $m_1 > ... > m_n$. Then, if $\omega_1 - c_1$ and $\omega_n - c_n$ have the same sign, the λ might fail to exist. On the other hand, if $\omega_1 - c_1$ and $\omega_n - c_n$ have different signs, then c is on the offer curve. $\square$

XII.B. Complements to Competitive Equilibrium

Second welfare theorem: complements to Proposition 9 and its proof The proof in the main paper established that the allocation can be implemented in a decentralized equilibrium iff there are A numbers $\delta^a$and n numbers q_i such that

$$\delta^a \pi_i = p_i^d + m_i^a q_i \text{ for all } a \leq A, i \leq n, \text{ with } \delta^a > 0, q_i > -p_i^d. \tag{90}$$

Here, we show that this is generically impossible to satisfy when $n > 2$ or $A > 2$ – when we allow $(\pi_i)_{i \leq n}$ and $(m_i^a)_{i \leq n, a \leq A}$ to take generic values in $\mathbb{R}_{++}^n \times [0,1]^{n \times A}$. We proceed by contradiction.

First, we observe that generically all q_i must be different from 0. Indeed, suppose that this is not the case, and that for instance $q_1 = 0$. Then (90) implies $\delta^a = \frac{p_1^d}{\pi_1}$, and we have, for $i > 1$, $\frac{p_1^d}{\pi_1} \pi_i = p_i^d + m_i^a q_i$. Given that generically $\frac{p_1^d}{\pi_1} \pi_i \neq p_i^d$, we have generically $q_i \neq 0$ (for $i > 1$), so that we can write: $m_i^a = \frac{\frac{p_1^d}{\pi_1} \pi_i - p_i^d}{q_i}$. This must hold for all agents a, which implies $m_i^a = m_i^b$ for all agents a, b, which is generically not true. We obtain a contradiction.

Hence, we proceed in the generic case where all q_i are different from 0. Then, (90) becomes:

$$m_i^a = \frac{p_i^d}{q_i} + \frac{\pi_i \delta^a}{q_i}. \tag{91}$$

Define the function Ψ by

$$\Psi : E := \mathbb{R}_{++}^A \times \prod_{i=1}^n (-p_i^d, \infty) \to F := \mathbb{R}^{n \times A}$$

$$\Psi(\delta, q) = \left(\frac{p_i^d}{q_i} + \frac{\pi_i \delta^a}{q_i} \right)_{i \leq n, a \leq A}$$

Equation (91) is equivalent to solving in (δ, q) :

$$\Psi(\delta, q) = (m_i^a)_{i \le n, a \le A} \tag{92}$$

Now, Ψ has an image of dimension at most $\dim E = A + n$. Hence, (92) is generically impossible to satisfy if $\dim E < \dim F$. We have

$$\begin{aligned} \dim E < \dim F \Leftrightarrow A + n < nA \Leftrightarrow (n-1)(A-1) > 1 \\ \Leftrightarrow A > 2 \text{ or } n > 2 \end{aligned}$$

(recall that $A \ge 2$ and $n \ge 2$). Hence, with strictly more than 2 consumers or 2 goods, the second welfare theorem generically fails.

We note that (90) does have a solution (so that the allocation is implementable in competitive equilibrium) in some interesting (though non-generic) cases. For instance, *when agents have the same misperceptions ($m_i^a = m_i^b$ for all agents a, b, and goods i), the second welfare theorem holds.*

The case with 2 goods and 2 agents. This case is the remaining one not handled by genericity arguments. We can analyze it by directly solving for (90), and find (we normalize $\boldsymbol{p}^d = (1,1)$ for simplicity, and without loss of generality):

$$\begin{aligned} \delta_a &= \frac{m_1^a \left(m_2^a - m_2^b\right) \pi_1 - m_2^a \left(m_1^a - m_1^b\right) \pi_2}{\pi_1 \pi_2 \Delta} \\ q_1 &= \frac{\left(m_2^a - m_2^b\right)(\pi_1 - \pi_2)}{\pi_2 \Delta} \\ \Delta &:= m_1^b m_2^a - m_2^b m_1^a \end{aligned}$$

and symmetric expressions for δ_b, q_2. As an illustration, assume that $\Delta \neq 0$ and $m_i^a > m_i^b$ for $i = 1, 2$. Then, the above expression implies that there are price parameters π_1, π_2 (hence target allocations) that force $\delta_a < 0$; however, we needed to have $\delta_a > 0$.[91] Hence, these allocations (parameterized by this $\boldsymbol{\pi}$) are not implementable in competitive equilibrium: the second welfare theorem does not hold for these allocations. In sum, the second welfare theorem fails for a non-trivial (positive measure) set of allocations (parameterized by $\boldsymbol{\pi}$) and attention vectors $\boldsymbol{m}^a$.

On the other hand, there are allocations that allow for implementation, as can be shown for instance numerically.[92]

In sum, with 2 agents and 2 goods, the second welfare theorem fails for a non-trivial (positive measure) set of allocations and attention vectors. It also holds for a complementary non-trivial set

[91] Recall that $\boldsymbol{\pi}$ is an index of the target allocation, in the sense that $u^{a\prime}(\boldsymbol{c}^a) = \xi^a \boldsymbol{\pi}$ for all agents a's.

[92] For instance, take $\boldsymbol{m}^a = (1,1), \boldsymbol{m}^b = (1,.9), \boldsymbol{\pi} = (1.1, 1)$. Then, the solutions are $\delta^a = \delta^b = 1$ and $\boldsymbol{q} = (.1, 0)$. By perturbation, there is a positive measure of parameters around these that ensure a solution to (90).

of allocations and attention vectors.

We can summarize the situation as follows.[93] *The second welfare theorem fails with probability 1 when there are strictly more than 2 goods or 2 consumers, and fails with probability strictly between 0 and 1 when there are 2 goods and 2 consumers.*

Other aspects of competitive equilibrium We next discuss other aspects of competitive equilibrium with sparse agents.

When the price level is high, the relative price of the non-salient good is high. Consider a sparse economy in the representative agent case (all agents have the same utility, endowment, and perceptions) with endowment $\boldsymbol{\omega}$, and define $\boldsymbol{p}^d = \overline{\boldsymbol{p}} := u'(\boldsymbol{\omega})$. Then, the set of equilibrium prices is:[94]

$$\mathcal{P}^* = \left\{ \left(\overline{p}_i \left(1 + \frac{\chi}{m_i} \right) \right)_{i=1\ldots n} : \chi \in \left(-\min_j m_j, \infty \right) \right\}, \tag{93}$$

while in the traditional model, the set of equilibrium prices is: $\mathcal{P}^* = \{ (\overline{p}_i (1+\chi))_{i=1\ldots n} : \chi \in (-1, \infty) \}$.[95]

This means that, "when the price level is high (χ is high), the relative price of the obscure good is high." When χ is high, the consumer perceives the high price of the obscure good less, hence demands more of it. That increases the price of the obscure good.

A corollary is the following.

When the price level is high, the agent with the relatively higher endowment of the obscure good becomes relatively better off. To illustrate this effect, we consider a polar case.

Example 6 *Assume that both agents have identical perceptions ($m^a = m^b$) and homothetic preferences ($c^r(p,w) = c^r(p,1)\,w$), with the log specification (89). The set of equilibrium prices is $\mathcal{P}^* = \{ (\overline{p}_i \chi^{1/m_i})_{i=1\ldots n} : \chi \in \mathbb{R}_{++} \}$, for some $\overline{p}$. The set of allocations becomes an interval: $\mathcal{C}^a = [s^a_{\min}, s^a_{\max}]\,\boldsymbol{\omega}$, where $s^a_{\min} = \min_i \frac{\boldsymbol{\omega}^a_i}{\boldsymbol{\omega}_i}$, $s^a_{\max} = \max_i \frac{\omega^a_i}{\omega_i}$. When the price level is high, the agent with the relatively higher endowment of the non-salient good (the good i with lowest m_i) is relatively better off.*

Derivation: Given p, the share of total income that goes to consumer a is $s^a = \frac{p\cdot\boldsymbol{\omega}^a}{p\cdot\omega}$. We clearly have $s^a \in [s^a_{\min},\ s^a_{\max}]$. With two goods, say that $\frac{\omega^a_1}{\omega_1} \geq \frac{\omega^a_2}{\omega_2}$.

[93] We are assuming a probability measure on the $\boldsymbol{\pi}$ and $(\boldsymbol{m}^a)_{a\leq A}$ that has a non-zero density in the interior of $\mathbb{R}^n_{++} \times [0,1]^{n\times A}$.

[94] The derivation of (93) is as follows: A price $\boldsymbol{p}$ is in $\mathcal{P}^*$ iff the associated $\boldsymbol{p}^s$ satisfies $\boldsymbol{p}^s = \lambda u'(\boldsymbol{\omega})$, for some number $\lambda > 0$, i.e. iff $\boldsymbol{p}^s = \lambda \overline{\boldsymbol{p}}$, i.e. iff for all i, $(1-m_i)\,p^d_i + m_i p_i = \lambda \overline{p}_i$, e.g. (using the assumption $\boldsymbol{p}^d = \overline{\boldsymbol{p}}$), $p_i = \left(1 + \frac{\lambda-1}{m_i}\right)\overline{p}_i$. Prices are positive iff $\lambda - 1 > -\min_j m_j$. Equation (93) follows, with $\chi := \lambda - 1$.

[95] In the case of the "perception in logs", eq. (89), we have $\mathcal{P}^* = \left\{ (\overline{p}_i \chi^{1/m_i})_{i=1\ldots n} : \chi \in \mathbb{R}_{++} \right\}$. The economic message is the same as with the additive specification, with a somewhat cleaner specification.

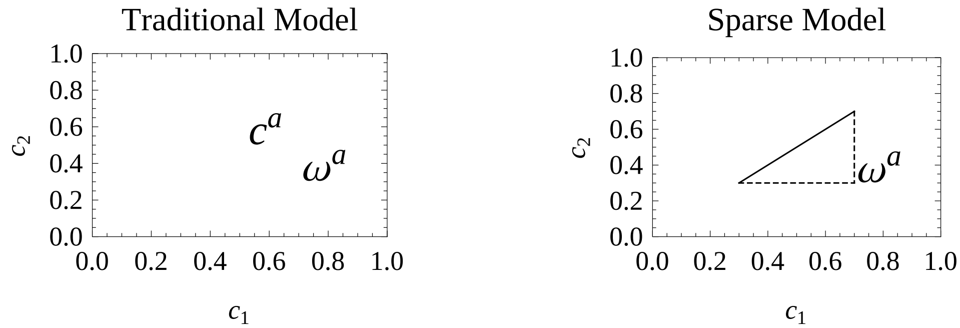

FIGURE VI: These Edgeworth boxes show competitive equilibria when both agents have Leontief preferences. The left panel illustrates the traditional model with rational agents: there is just one equilibrium, $\boldsymbol{c}^a = (1/2, 1/2)$. The right panel illustrates the situation when type a is rational, and type b is boundedly rational: there is a one-dimensional continuum of competitive equilibria. Agent a's share of the total endowment ($\boldsymbol{\omega}^a$) is the same in both cases.

Now, let p vary according to $p(\chi) = \overline{p}\chi^{1/m_i}$ for $\chi > 0$. So, as $p_1/p_2 \to 0$, $s^a \to s^a_{\min} = \frac{\omega^a_2}{\omega_2}$, while as $p_1/p_2 \to \infty$, $s^a \to \frac{\omega^a_1}{\omega_1} = s^a_{\max}$. So, as p varies, s^a covers the whole range $[s^a_{\min}, s^a_{\max}]$. We now prove the uniqueness part. Given consumers have homothetic preferences and identical perceptions, $u'(\boldsymbol{\omega}) = kp^s$ for some χ. So $\mathcal{P}^* = \{(\overline{p}_i\chi)_{i=1\ldots n} : \chi \in \mathbb{R}_{++}\}$ for $\overline{p}$ s.t. $\overline{p}^m = u'(\boldsymbol{\omega})$ (i.e. $\overline{p}_i^{m_i}\left(p^d_i\right)^{1-m_i} = u'(\boldsymbol{\omega})$). $\square$

As an illustration of Example 6, Figure VI shows competitive equilibria as in Figure IV, but for Leontief consumers, i.e. with $u(c_1, c_2) = \min_i c_i$.[96] In the Cobb-Douglas case, across equilibria, a high consumption of good 1 corresponds to a low consumption of good 2. In the Leontief case, a high consumption of good 1 corresponds to a high consumption of good 2. This is because of the relative endowment effect mentioned above.

For the existence of the equilibrium, the key reference is Debreu (1970). See also Shafer and Sonnenschein (1975) for equilibrium existence in non-standard economies.

Finally, let us examine two sparse consumers, with heterogeneous attention. Differential attention leads to different allocations, even though agents have the same preferences and endowments. Some agents may pay no attention to some prices, as long as other agents do pay attention to that price.

EXAMPLE 7 *Consider a case with identical preferences $u(c_1, c_2) = v(c_1) + v(c_2)$, for some concave function v, and identical endowments, $\boldsymbol{\omega}^a = \boldsymbol{\omega}^b = (1/2, 1/2)$, but asymmetric perceptions (using the log specification ()): Type a has attention $m^a = (\alpha, \beta)$, and type b has attention $m^b = (\beta, \alpha)$, with $\alpha \neq \beta$. Assume $p^d = (1, 1)$ and $\lim_{c\to 0} v'(c) = \infty$. Then, the equilibrium set contains*

[96] For the Cobb-Douglas, case, the equilibrium set is curve is the set of c such that $\prod_{i=1}^{2}\left(\frac{2c_i}{\omega^a_i} - 1\right) = 1$ and $0 \preceq c \preceq \omega$. For the Leontief, it's the set of c such that $\{c_1 = c_2\}$ and $c_1 \in [\min_i \omega^a_i, \max_i \omega^a_i]$.

$\{(c, 1-c), c \in (0,1)\}$ *and the set of equilibrium prices contains* $\{(p,p), p > 0\}$.

Derivation: Let (c_1, c_2) be the allocation of consumer a, so that consumer b's allocation is $(1-c_1, 1-c_2)$. We look for equilibria with prices (p,p) and allocations $c^a = (c, 1-c)$. The first-order condition for consumer a is: $\frac{v'(c_1)}{v'(c_2)} = \frac{p_1^\alpha}{p_2^\beta} = p^{\alpha-\beta}$, and for consumer b: $\frac{v'(1-c_1)}{v'(1-c_2)} = \frac{p_1^\beta}{p_2^\alpha} = p^{\beta-\alpha}$. This is possible if $\frac{v'(c)}{v'(1-c)} = p^{\alpha-\beta}$.

Equilibrium selection What pins down the price? One approach is via a quantity theory of money M, e.g. chosen by the central bank: then the price $\boldsymbol{p} \in \mathcal{P}^*$ is the (often unique) one that ensures that nominal GDP is M ($\boldsymbol{p} \cdot \boldsymbol{\omega} = M$, normalizing velocity to 1, and equating "output" with $\boldsymbol{\omega}$). Other approaches would rely on expectations in a dynamic model.

Endogenous default price Endogenizing the default price would be interesting. In the context of a static model, one might hypothesize that a good default price would be such that it is also an equilibrium price of the related sparse economy, i.e. such that $\boldsymbol{Z}\left(\boldsymbol{p}^d; \boldsymbol{p}^d\right) = 0$ (rewrite excess demand as $\boldsymbol{Z}\left(\boldsymbol{p}; \boldsymbol{p}^d\right)$). Then $\boldsymbol{p}^d$ must be an equilibrium price of the underlying rational model. Alternatively, $\boldsymbol{p}^d$ might be better thought of as some expectation of the price, given the past prices and recent shocks. The proper locus of the endogenization of $\boldsymbol{p}^d$ may be in an explicitly dynamic model, something I tackle in companion work.

Generality of the effects Many effects simply stem from the fact that the Marshallian demand $\boldsymbol{c}^s(\boldsymbol{p}, w)$ is not homogenous of degree 0 in $(\boldsymbol{p}, w)$, which is a form of "nominal illusion" – or really, a price illusion, as there is no "money", simply a default price (Fehr and Tyran 2001, Shafir, Diamond and Tversky 1997 analyze money illusion). In many ways, many effects here do not depend on the specifics of money illusion. However, it is still useful to have a specific model, as the economic intuition for many effects does depend on the spirit of the sparsity model (e.g. salient prices), and some conditions need to be specified (e.g. Proposition 32).

We conclude with a simple example, suggested by Oliver Hart.

Example 8 *(Exchange economy with misperceptions). There are two agents types* a, b, *with the same utility* $u(c_1, c_2) = \ln c_1 + c_2$, *and the same endowment* $(1,1)$. *Hence, in the traditional model, there is just one equilibrium allocation:* $(1,1)$ *to each agent. In the behavioral version, assume that agents correctly perceive the price of good 1, but have attention* m^a, m^b *to the price of good* 2*: type* $\theta \in \{a, b\}$ *perceives it to be:* $p_2^\theta = 1 + m^\theta (p_2 - 1)$ *(the default price is* $(1,1)$*). The demand of agent* θ *is* $c_1^\theta = \frac{p_2^\theta}{p_1}$, *and the market-clearing condition for good 1 is:* $c_1^a + c_1^b = 2$ *(the total endowment of good 1 is 2). Hence,* $\frac{p_2^a}{p_1} + \frac{p_2^b}{p_1} = 2$, *i.e.* $p_1 = \frac{p_2^a + p_2^b}{2} = 1 + \frac{m^a + m^b}{2}(p_2 - 1)$. *Hence, the set of equilibrium prices*

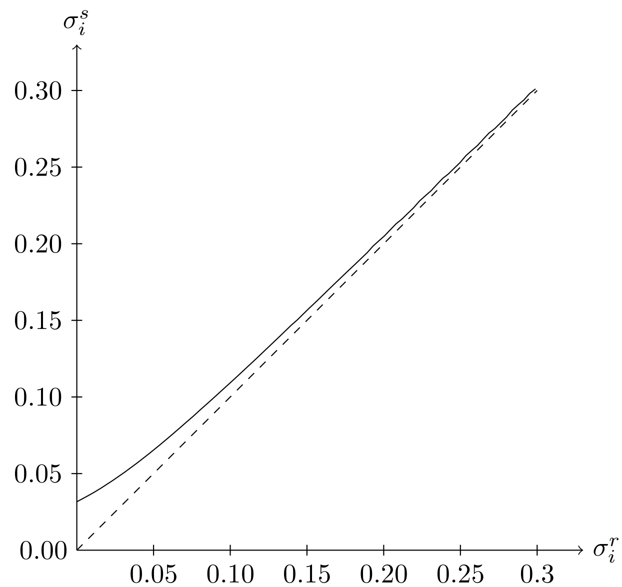

FIGURE VII: This Figure illustrates the link between fundamental volatility σ_i^r in an economy with traditional agents, and volatility σ_i^s in an economy with sparse agents (solid line). The dashed line is the 45 degree line. Parameters: $\alpha = 1$, $\frac{\psi_i c_i^d p_i^d}{\kappa} = 10^3$.

is $\mathcal{P}^* = \left\{ (p_1, p_2) \text{ s.t. } p_1 - 1 = \frac{m^a + m^b}{2} (p_2 - 1), \, p_2 > 0 \right\}$. *Equilibrium consumption for consumer b is:* $c_1^b = \frac{p_2^b}{p_1}$*:*

$$c_1^b = \frac{1 + m^b (p_2 - 1)}{1 + \frac{m^a + m^b}{2} (p_2 - 1)}. \tag{94}$$

When consumers a and b have different misperceptions, there is a different equilibrium allocation for each p_2*. Suppose* $m^b < m^a$*. Economically, if the price level increases (say* p_2 *increases, so that* p_1 *increases also), agent b sees clearly the increase in the price of good 1, but less the price of good 2, so that he consumes less of good 1: To each price level corresponds a different real equilibrium, as in a Phillips curve.*

Further analysis of Proposition 11 Figure VII illustrates Proposition 11, and shows the link between fundamental volatility σ_i^r in an economy with traditional agents, and volatility σ_i^s in an economy with sparse agents. The two are very close for high fundamental volatility (as agents anyway pay attention to the good). However, when $\sigma_i^r \to 0$, $\sigma_i^s \to \sqrt{\frac{\kappa}{\psi_i c_i^d p_i^d}}$, a sort of minimum size of volatility. This is because of sparsity: volatility needs to be high enough so that (at least some) agents think about it.[97] (Propositions 3 and Proposition 10).

[97] Things would quantitatively 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).

XIII. Further Analysis of Sparse Max

XIII.A. Relation to Other Models of Inattention

"No Improving Action Switches" Criterion and Sparse Max Caplin and Martin (2014) have proposed a criterion for the rationality of inattention, the "No Improving Action Switches" (NIAS); See also Caplin and Dean (2013). Let a be the action chosen by the agent and $\widehat{a}$ a candidate action. It states that for all actions a

$$\text{NIAS: } \forall \widehat{a},\ \widehat{a} \in \arg\max_{a'} \mathbb{E}\left[u(a', x) \mid a = \widehat{a}\right],$$

i.e., conditional on states where an agent chose $\widehat{a}$, there is no action a' that leads to an average improvement: in other terms, given an action, that action is optimal over the average circumstances in which the agent take that action.

First, in the plainest version of the sparsity model, NIAS is violated. When $a = mx$, with $m = 1/2$, the agent would be better of doing $a' = x = 2a$.

This may sound disturbing at first, but it should not be upon further reflection. The reason is that in general, *representative-agent models do not satisfy the NIAS criterion*.

To see this, take the "noisy signals" version of section VII.B. Say that the signal is x, and each agent j receives a signal $s_j = x + \varepsilon_j$. Say there is a continuum of agents $j \in [0, 1]$, with i.i.d. noises ε_j. Then, the rational action is $a_j = m(x + \varepsilon_j)$, with $m = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_{\varepsilon_i}^2}$, and the average action is $a = \int_0^1 a_j dj = mx$. Hence, given state x, the optimal action of the representative agent is mx. Hence, *even under rational Bayesian updating, the representative agent doesn't satisfy the NIAS criterion*. This is the case even though each agent satisfies NIAS (for his action $a_j = m(x + \varepsilon_j)$). Hence, it makes good sense that the sparse max, interpreted as the representative-agent version of a model with noisy perception, does not satisfy NIAS. Plainly, individual agents have extra noise in their perception, and that contributes to the dampening in their actions ($m < 1$).

Relation to Ellis (2013). Ellis (2013)'s model has no "cost" of attention: in his framework, attention is chosen optimally, from a fixed set of attention policies. Hence, it doesn't have the "trade-off" as in this paper. Hence, the sparsity model is not a special case of Ellis. By the way, neither is Sims (2003), as Ellis indeed discusses. In addition, in Ellis there's no "partial attention" to prices, say.[98]

[98] This discusses Ellis (2013). This paper is evolving with a recent new version (2014, LSE) which has a trade-off, but the paper (and its link with sparse max) was not stabilized yet when this appendix was written.

XIII.B. *Alternative Formulation of the Budget Adjustment*

Proposition 2 gives a simple, concrete way to understand the budget adjustment. We restate it here: Given the true price vector $\boldsymbol{p}$ and the perceived price vector $\boldsymbol{p}^s$, the Marshallian demand of a sparse agent is

$$\boldsymbol{c}^s(\boldsymbol{p}, w) = \boldsymbol{c}^r(\boldsymbol{p}^s, w'), \tag{95}$$

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

We generalize this to arbitrary problems. We define:

$$a(x, y) := \arg\max_a u(a, x) \text{ s.t. } b(a, x) + y \geq 0.$$

This is the optimal action when the budget constraint is relaxed by a $y \in \mathbb{R}^K$ (in a consumption application, $y = w' - w$). This allows us to give the more general Proposition behind Proposition 2.

PROPOSITION 33 *The following variant of step 2 is equivalent to the action in Definition 2: define* $y^* \in \mathbb{R}^K$ *by:*

$$y^* := \arg\max_y u(a(x^s, y), x) \text{ s.t. } b(a(x^s, y), x) \geq 0. \tag{96}$$

Then, choose the action $a^s = a(x^s, y^*)$.

In simpler terms, the procedure is: Find the budget adjustment y^* such that $b(a(x^s, y^*), x) = 0$ (take the utility-maximizing one if there are several such y^*). The action is $a^s = a(x^s, y^*)$. In other terms, find the budget adjustment y^* such that, when the agent chooses the action $a(x^s, y^*)$ with that budget adjustment, he hits the actual constraint. This is the generalization of Proposition 2.

Proof of the Proposition 33: Given x, the agent perceives x as $x^s(x)$. For $y \in \mathbb{R}^K, \lambda \in \mathbb{R}_+^K$, let

$$L^y(a, x^s, \lambda) := u(a, x^s) + \lambda \cdot [b(a, x^s) + y].$$

Then, letting $L := L^0$, we have $L^y(\cdot, \lambda) = L(\cdot, \lambda) + \lambda \cdot y$.

Let us define $a(\lambda, x)$ as the action that maximizes $L(\cdot, \lambda)$ for any λ. Let $\lambda(x)$ be the value of λ that maximizes agents perceived utility given the budget constraint. Let $a^s(x)$ be the action corresponding to $\lambda = \lambda(x)$. Formally,

$$
\begin{aligned}
a(\lambda, x) &:= \arg\max_a L(a, x^s(x), \lambda), \\
\lambda(x) &:= \arg\max_{\lambda \in \mathbb{R}^K_+} u(a(\lambda, x), x^s(x)) \text{ s.t. } b(a(\lambda, x), x) \geq 0, \\
a^s(x) &:= a(\lambda(x), x).
\end{aligned}
$$

Let us define $\hat{a}(x^s, y)$ as the action that maximizes agents perceived utility subject to perceived budget y. Let $y(x)$ be the budget adjustment such that $\hat{a}(x^s, y)$ is actual budget feasible. Let $\hat{a}^s(x)$ be the optimal action corresponding to perceived budget $y(x)$. Formally,

$$
\begin{aligned}
\hat{a}(x^s, y) &:= \arg\max_a u(a, x^s) \text{ s.t. } b(a, x^s) + y \geq 0, \\
y^*(x^s, x) &:= \arg\max_{y \in \mathbb{R}^K} u(\hat{a}(x^s, y), x^s) \text{ s.t. } b(\hat{a}(x^s, y), x) \geq 0, \\
y(x) &:= y^*(x^s(x), x), \\
\hat{a}^s(x) &:= a(x^s(x), y(x)).
\end{aligned}
$$

Note that $L^y(\cdot, x^s, \cdot)$ is the Lagrangian associated with $\hat{a}(x^s, y)$: The Lagrangian for the constrained optimization problem $\max_a u(a, x^s)$ s.t. $b(a, x^s) + y \geq 0$ (with Lagrange multiplier λ) is $L^y(a, x^s, \lambda)$.

Let $x^{s*} = x^s(x)$ be as in Definition 2. Let

$$
A := \{a(\lambda, x)\}_{\lambda \geq 0}, \quad \hat{A} := \{\hat{a}(x^{s*}, y)\}_y.
$$

Claim $A = \hat{A}$.

Proof. For all $\lambda \in \mathbb{R}^K_+$, let

$$
\begin{aligned}
A_\lambda &:= \{a : \ L_a(a, x^{s*}, \lambda) = 0\} \\
&= \{a : \ L^y_a(a, x^{s*}, \lambda) = 0\} \text{ for } y \in \mathbb{R}^K, \\
\hat{A}_\lambda &:= \{a \in A_\lambda : \ \exists y \in \mathbb{R}^K \text{ s.t. } c := b(a, x^{s*}) + y \geq 0 \text{ and } \lambda_i c_i = 0 \ \forall i\}.
\end{aligned}
$$

Obviously, $\hat{A}_\lambda \subseteq A_\lambda$. For $a \in A_\lambda$, $\ y = -b(a, x^{s*})$ witnesses $a \in \hat{A}_\lambda$. Thus $A_\lambda = \hat{A}_\lambda$.

Hence

$$
\begin{aligned}
\hat{A} &= \bigcup_{y \in \mathbb{R}^K, \lambda \in \mathbb{R}^K_+} \{a : \ L^y_a(a, x^{s*}, \lambda) = 0, \ \ c := b(a, x^{s*}) + y \geq 0 \text{ and } \lambda_i c_i = 0 \ \forall i\} \\
&= \bigcup_{\lambda \in \mathbb{R}^K_+} \hat{A}_\lambda = \bigcup_{\lambda \in \mathbb{R}^K_+} A_\lambda = \{a : \ \exists \lambda \in \mathbb{R}^K_+ \text{ s.t. } L_a(a, x^{s*}, \lambda) = 0\} = A.
\end{aligned}
$$

■

Now we go back to the proof of the Proposition:

$$\begin{aligned}
u(a^s(x), x^{s*}) &= \max_{\lambda\in\mathbb{R}_+^K} u(a(\lambda, x), x^{s*}) \text{ s.t. } b(a(\lambda, x)) \geq 0 \\
&= \max_{a\in A} u(a, x^{s*}) \text{ s.t. } b(a, x) \geq 0 \\
&= \max_{a\in \hat{A}} u(a, x^{s*}) \text{ s.t. } b(a, x) \geq 0 \\
&= \max_{y\in\mathbb{R}^K} u(\hat{a}(x^{s*}, y), x^{s*}) \text{ s.t. } b(\hat{a}(x^{s*}, y), x) \geq 0 \\
&= u(\hat{a}^s(x), x^{s*}).
\end{aligned}$$

Either both 'true' budget constraints bind or neither does. Therefore, by our strict concavity conditions, $a^s = \hat{a}^s$.

□

XIII.C. Envelope Theorem with Sparse Max

With the sparse max, the envelope theorem holds at the default model, but needs to be modified away from the default model. An implication is that Shephard's lemma and Roy's identity hold at the default price, but need to be modified away from the default price.

Proposition 34 *(Envelope theorem for sparse max – without constraint). Consider the problems* $v^r(x) = \max_a u(a,x)$ *and* $v^s(x) = \operatorname{smax}_a u(a,x)$.

Then, we have the envelope theorem, and the modified envelope theorem, respectively:

$$v_x^r = u_x(a^r, x), \tag{97}$$

$$v_x^s = u_x(a^s, x) + u_a(a^s, x)\, a_x(x^s)\, M, \tag{98}$$

with $M = \frac{\partial x_s}{\partial x}$.

Proof. It follows directly comes from the chain rule: with $a(x) = \arg\max u(a,x)$,

$$v^s(x) = u(a(x^s), x),$$

so

$$v^s(x) = u_x + u_a a_x(x^s) \frac{\partial x_s}{\partial x}. \tag{99}$$

■

Note that $u_a(a^s, x^s) = 0$, but typically $u_a(a^s, x) \neq 0$.

Likewise, the envelope theorem will have extra terms in the sparse max with constraints.

XIV. Reparametrization Properties of Sparse Max

XIV.A. Pros and Cons of the Compensated Sparse Max

Here we complement the discussion in section V.D.

The regular vs compensated sparse max just differ in attention m that they generate. Given that attention m, they give the same action.

The compensated sparse max has a quirk when dealing with maximization with respect to a one-dimensional action a, with one constraint.

Example 9 *Consider the problem $u(a,x) = -\frac{1}{2}(a - \sum_{i=1}^n \mu_i x_i)^2$ subject to $a \le B$, for some real number B. The traditional, non-sparse action is: $a^r = \min\left(\sum_i \mu_i x_i, B\right)$. The sparse action is $a^s = \min\left(\sum_i \mathcal{A}_\alpha\left(\mu_i^2\sigma_i^2/\kappa\right)\mu_i x_i, B\right)$.*

Derivation The Lagrangian is $L(a,x) = -\frac{1}{2}(a - \sum_{i=1}^n \mu_i x_i)^2 + \lambda^d (B-a)$, so $a_{x_i} = -L_{aa}^{-1} L_{ax_i} = \mu_i$. Hence, the first step gives $x_i^s = \mathcal{A}_\alpha\left(\mu_i^2\sigma_i^2/\kappa\right)x_i$. In the second step, call $a^{s\prime} = \sum_i \mathcal{A}_\alpha\left(\mu_i^2\sigma_i^2/\kappa\right)\mu_i x_i$. We have $a(\lambda) = \arg\max_a -\frac{1}{2}\left(a - a^{s\prime}\right)^2 + \lambda(B-a)$, i.e. $a(\lambda) = a^{s\prime} - \lambda$. We next need to solve $\lambda^* = \arg\max_\lambda -\frac{1}{2}\left(a(\lambda) - a^{s\prime}\right)^2$ s.t. $B \ge a(\lambda)$, i.e. $\max_\lambda -\frac{1}{2}\lambda^2$ s.t. $B \ge a^{s\prime} - \lambda$. That gives: $a^s = \min\left(a^{s\prime}, B\right)$. □

However, with the "compensated sparse max", $a^s = \min\left(\sum_i \mathcal{A}_\alpha\left(\mu_i^2\sigma_i^2/\kappa\right)\mu_i x_i, B\right)$ when $B \ge 0$, but $a^s = 0$ when $B < 0$.

Let us now consider the compensated sparse max. When $B > 0$, the constraint doesn't bind at the default ($x = 0$), and $\overline{a}_{x_i} = a_{x_i} = \mu_i$. The compensated and plain sparse max gives the same answer.

However, when $B < 0$, with the compensated sparse max, $a^s = B$, independently of x_i – which is a bit odd. The mechanical reason is as follows. At the default the constraint binds. This does not matter for the plain sparse max. However, with the compensated smax: $\overline{a}_{x_i} = 0$, and $m_i = 0$, and $a^s = B$, independently of x_i. As the constraint binds, a is "stuck" at $a = B$, and paying attention to x_i won't make any difference. This is because we only consider local changes.[99]

A conclusion is that: The regular sparse max is easier to use. The compensated sparse max has some extra nice properties (reparametrization invariance), but is a bit more delicate to use.

[99] One could imagine some solutions. One would be to examine the losses not just at a^d, but at another point. E.g., in the example above, that would be the value of $\overline{a}_{x_i}$ at a point with $B > 0$.

Consumer problem with compensated smax If we use the compensated smax, we have $\overline{c}^i_{p_j} = S_{ij}$, the Slutsky term. Hence, Proposition 3 is verified, with ψ_i the compensated own-price elasticity $\psi_i = -\frac{p_i^d}{c_i^d} S_{ii}$. In slight contrast, the ψ_i with the basic sparse max is the one in the proof of Proposition 3.

XV. Some Variants of the Sparse Max

Here are some variants that can be useful in some contexts, but that I did not choose for the core model.

XV.A. Different Defaults

Implicitly, the default model is $m = 0$. However, one could imagine a different default, say m^d. Hence, the penalty becomes $\left|m_i - m_i^d\right|^\alpha$ rather than $|m_i|^\alpha$.

XV.B. Different Costs

We replace the penalty $\mathcal{C}(m) = \kappa \sum_i m_i^\alpha$ by:

$$\mathcal{C}(m) = \sum_i \kappa_i m_i^\alpha. \tag{100}$$

Indeed, it is clear that some tasks (e.g., computing the 100th decimal of $\sqrt{2}$) are much harder than others. In some economic situations, this is an important force, which could be formulated with a higher κ_i. For instance, a smaller font size describing dimension i might be associated with a higher κ_i.

XV.C. Meta-Meta Cognition

Meta-meta-cognition: when the agent refines the determinants of his attention In the sparse max, step 1 is a form of meta-cognition (which could be partially unconscious): it decides about what to think. We can go one (and more) steps further, and do meta-meta-cognition. If step 1 is "given estimated size (variance) of x_i, how much should I think about x_i", the meta-meta-step would be "should I update my estimate of the estimated size of x_i". Mathematically, this is updating the estimate of $v_i = \sigma_i^2 = \mathbb{E}\left[x_i^2\right]$.

To do so, one can simply use the sparse max, twice: first, to choose how much to refine on the determinants on attention, then choosing attention itself. Take the case with just one variable (the case with n variables is essentially identical). The value of m_x (the attention to x) comes from doing:

$$\max_{m_x} V(m_x, \sigma_x) := -\frac{1}{2} b v_x (1-m_x)^2 - \kappa g(m_x), \qquad b := \left| u_{aa} a_x^2 \right|.$$

Suppose now that the agent starts from a default estimate of v^d, while the true value is v. Should the agent pay attention to this new estimate? The agent decides how what attention m_v he should pay to v so that his perceived v is:

$$v^s(m_v) := v^d + m_v \left(v - v^d\right). \tag{101}$$

How to pick m_v? Well, we just apply the sparse max to the "meta" function $V(v_x, m_x)$, where m_x is the action, and m_v is the attention to v:

$$\operatorname*{smax}_{a=m_x; v_x - v_x^d} V(m_x, \sigma_x).$$

We calculate: $V_{m_x m_x} = -b v_x - \kappa g''(m_x)$, $V_{m_x v_x} = b(1-m_x)$, so the optimal attention to m_v is:[100]

$$m_v^* = \mathcal{A}\left(V_{m_x v_x}^2 V_{m_x m_x}^{-1} \mathbb{E}\left[\left(v_x - v_x^d\right)^2\right] / \kappa\right). \tag{102}$$

Hence, where the agent thinks he might have a wrong estimate of the variance ($\mathbb{E}\left[\left(v_x - v_x^d\right)^2\right]$ high), and the stakes are high (high b), he thinks more about updating that variance, i.e. of updating the determinants of his attention. Then, the attention to x itself is:

$$m_x = \mathcal{A}\left(b^2 v^s(m_v^*) / \kappa\right). \tag{103}$$

XV.D. *Paying Attention to Different Things for Consumption vs Investment*

Suppose that the agent must take two contemporaneous actions, say consume c and find the ratio θ of stocks in his investments (investing decision). He might pay attention to different things in his consumption decision, and in his investing decision.

How should we represent this? Let $u(c, \theta, x)$ be the full utility (including continuation utility), where x is the vector of disturbances, and let $a = (c, \theta)$ be the vector of actions. Call m^1 the attention vector in the consumption decision, and m^2 the attention vector in the investment decision. The optimal action $a^* = (c^*, \theta^*)$ satisfies:

$$u_c\left(c^*, \theta^*, x\left(m^1\right)\right) = 0, \qquad u_\theta\left(c^*, \theta^*, x\left(m^2\right)\right) = 0.$$

[100] The κ there could be a different "meta" κ^M.

This means that when the agent is choosing c, he uses the attention vector m^1, but when he is choosing θ, he uses attention vector m^2.

More generally, if m^i is the attention vector for action i, the optimality condition is

$$\forall i, u_{a_i}\left(a^*, m^i\right) = 0. \tag{104}$$

In non-derivative form: a^* satisfies:

$$\forall i, a_i^* \in \arg\max_{a_i} u\left(a_i, a_{-i}^*, m^i\right). \tag{105}$$

The allocation of attention m^i would be the attention in the reduced problem $\text{smax}_{a_i} u\left(a_i, a_{-i}^d, x\right)$, i.e. the attention vector given by the sparse max for that reduced problem of just choosing action a_i.

In that sense this models a mildly (but perhaps realistically) schizophrenic agent. For instance, when the agent feels that equities are overvalued (i.e. we take the case where x is one-dimensional, and is the deviation of the equity premium from its average value), he will lower the allocation to equities, θ (as he pays attention to the expected return to equities, $m^2 > 0$). But the same agent, if rational-consistent, should also change his consumption (indeed, he should consume more, as as rate of return of his investment is lower; this assumes some non-extreme values of the intertemporal elasticity of substitution that are classic and we'll not delve into here). However, a very sparse agent will not change his consumption, because when choosing his consumption, he doesn't pay attention to the rate of return of his investments ($m^1 = 0$; under reasonable calibrations, this would be true for intermediary values of κ, which ensure $0 = m^1 < m^2$).

XV.E. Other Cost Functions

Piecewise linear attention function, as a function of variance

Other cost functions Can we get a cost function g that ensures that $\mathcal{A}\left(\sigma^2\right) = 0$ for $\sigma^2 > 0$ low enough, and $\mathcal{A}\left(\sigma^2\right) = 1$ for σ^2 high enough? Yes indeed, though g cannot be convex anymore. To get sparsity, you need $g'(0) > 0$. Let us state a result coming directly from (66):

$$\frac{g'_\alpha(m)}{1-m} = \frac{\sigma^2}{\kappa} = \mathcal{A}^{-1}(m).$$

Suppose we want: $\mathcal{A}\left(\sigma^2\right) = \left(\sigma^2 - B\right)/C$, in a range $\sigma^2 \in [B, B+C]$. Then $\mathcal{A}^{-1}(m) = B + Cm$, hence we want:

$$g'(m) = (1-m)(B+Cm)$$
$$= -Cm^2 + (C-B)m + B,$$

$$g(m) = \frac{-C}{3}m^3 + \frac{C-B}{2}m^2 + Bm,$$

with the domain $m \in [0,1]$.

That cost function g gives

$$\mathcal{A}\left(\sigma^2\right) = \max\left(0, \min\left(1, \left(\sigma^2 - B\right)/C\right)\right),$$

i.e. attention is 0 iff $\sigma^2 \leq B$, and is 1 iff $\sigma^2 \geq B + C$, with a linear ramp in between.

We can easily check that the solution of $\mathcal{A}(\sigma^2) = \arg\min_{m\in[0,1]} \frac{1}{2}(m-1)^2 \sigma^2 + g(m)$ is unique.

This function has the advantage of allowing for both 0 and full attention.

One disadvantage is that it's not convex. Hence, it is possible that $\min_m \frac{1}{2}(\iota - m)' \Lambda (\iota - m) + \sum g(m_i)$ will have multiple local equilibria.

Piecewise linear truncation function One very useful convex cost function is: $g(m) = -\kappa \ln(1-m)$. The corresponding attention function is $\mathcal{A}(\sigma^2) = \arg\min_m \frac{1}{2}(m-1)^2 \sigma^2 + g(m)$, i.e. (assuming at first an interior solution) $(m-1)\sigma^2 - \frac{1}{1-m} = 0$, i.e. $m = 1 - \frac{1}{\sigma}$, and finally

$$\mathcal{A}\left(\sigma^2\right) = \max\left(1 - \frac{1}{\sigma}, 0\right). \tag{106}$$

This is the function initially used in the NBER working paper version of this paper. (The $\tau\left(\mu, \frac{\kappa}{\sigma}\right)$ function for $\mu = 1$ and $\kappa = 1$).

Hence, g induces sparsity (this is because $g'(0) > 0$ – a function differentiable at 0^+ induces sparsity iff $g'(0) > 0$), like the function $g_1(m) = m$. However it has a somewhat simpler thresholding function, as we shall now see.

In the "ex post attention", with $\sigma_i = |x_i|$, we have $x^s = \mathcal{A}(x^2/\kappa)\, x = x\left(1 - \frac{\sqrt{\kappa}}{|x|}\right)_+ = sign(x)\left(|x| - \sqrt{\kappa}\right)_+$, i.e.

$$x^s = \tau\left(x, \sqrt{\kappa}\right),$$

where we define:

$$\tau(\mu, \kappa) = \max(|\mu| - |\kappa|, 0)\, sign(\mu), \tag{107}$$

i.e., for $\kappa > 0$, $\tau(\mu, \kappa) = 0$ for $|\mu| \leq |\kappa|$, $\mu - \kappa$ for $\mu > \kappa$ and $\mu + \kappa$ for $\mu < -\kappa$.

The piecewise linear thresholding function is particularly useful in some applications, because it

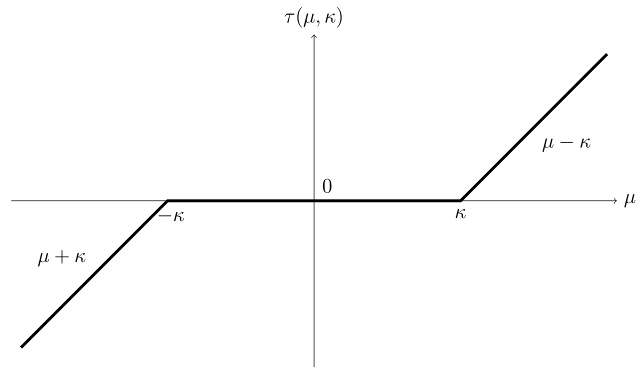

FIGURE VIII: The truncation function τ

is piecewise affine in μ.

XV.F. Discrete Actions

The model is formulated with a Euclidean action space, which is the substrate in many economic problems and confers a nice structure (e.g., a metric) on them. It extends to a discrete action space, as I illustrate here. See Gabaix (2013b) for more examples.

Action $a \in \{1, ..., A\}$ generates utility $u(a, x)$, but the agent may use an imperfect $u(a, x^s(m))$. To formulate the model, some notations are useful: for a function $f(a)$, define $\|\Delta_a f(a)\| := \left(\frac{1}{A}\sum_{a=1}^{A} \mathbb{E}\left[\left(f(a) - f\left(a^d\right)\right)^2\right]\right)^{1/2}$ to be the dispersion of f across actions. Then, define $\sigma_i^m = \|\Delta_a u_{x_i}(a, 0)\|$: it is the typical size of the marginal enrichment m_i. A natural analogue of Step 1 is:

$$\text{Step } 1' : \max_m \sum_i \frac{1}{2}\left(\sigma_i^m\right)^2 (m_i - 1)^2 + \kappa \sum_i |m_i|^\alpha .$$

It yields:

$$m_i^* = \mathcal{A}_\alpha\left(\left(\sigma_i^m\right)^2 / \kappa\right). \tag{108}$$

To illustrate this formalism, consider the choice between A goods: good $a \in \{1...A\}$ has a value:

$$u(a, x) = \sum_{i=1}^{n} \mu_i x_{ia},$$

with the x_{ia}'s i.i.d. across goods a, normalized to have mean 0 and standard deviations σ_i. The dimensions $i \in \{1, ..., n\}$ are (normalized) hedonic dimensions, e.g., price, weight, usefulness, aesthetic appeal of each good. Applying the above Step 1', we obtain $\left(\sigma_i^m\right)^2 = \mu_i^2 \sigma_i^2$ and finally:

PROPOSITION 35 *Suppose that the agent chooses among A goods where good $a \in \{1...A\}$ has value*

$u(a,x) = \sum_{i=1}^{n} x_{ia}$. *Then, the boundedly rational perception of a good a is*

$$u^s(a,x) = \sum_{i=1}^{n} \mathcal{A}_\alpha\left(\mu_i^2\sigma_i^2/\kappa\right)\mu_i x_{ia}. \tag{109}$$

Hence, we obtain a dimension-by-dimension dampening, with small dimensions (small σ_i) dampened more or fully, very much in the spirit of the initial example we started from, but for discrete actions. Compared to process models of discrete choice with partial attention (e.g., Tversky 1972, Payne, Bettman and Johnson 1993, Gabaix, Laibson, Moloche and Weinberg 2006), this model eschews sequential search (which typically does not lead to a closed form for the perceived value) and is thus much more tractable. Indeed, an equation such as (109) could be fairly directly estimated.

XVI. References

Bodoh-Creed, Aaron, Dan Benjamin, and Matthew Rabin, "The Dynamics of Base-Rate Neglect," Working Paper, 2013.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," Working Paper, Harvard University, 2013.

Caplin, Andrew, and Daniel Martin, "A Testable Theory of Imperfect Perception," *Economic Journal* (forthcoming).

Deaton, Angus, "Involuntary Saving Through Unanticipated Inflation," *American Economic Review*, 67 (1977), 899–910.

Edwards, Ward, "Conservatism in Human Information Processing,"in *Formal Representation of Human Judgement*, ed. B. Kleinmuntz, (New York: Wiley, 1968).

Ellis, Andrew, "Foundations for Optimal Inattention," Working Paper, Boston University, 2013.

Ellison, Glenn, "A Model of Add-on Pricing," *Quarterly Journal of Economics*, 120 (2005), 585–637.

Grether, David, "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of Economics*, 95 (1980), 537–557.

Griffin, Dale, and Amos Tversky, "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology*, 24 (1992) 411-435.

Klenow, Peter and Benjamin Malin, "Microeconomic Evidence on Price-Setting," in *Handbook of Monetary Economics* 3A, B. Friedman and M. Woodford, (Amsterdam: Elsevier, 2011, 231-284).

Koehler, Jonathan J, "The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges," *Behavioral and Brain Sciences*, 19 (1996), 1–17.

L'Huillier, Jean-Paul, "Consumers' Imperfect Information and Price Rigidities," Working Paper, 2013.

Lynch, John G., and Chezy Ofir, "Effects of Cue Consistency and Value on Base-rate Utilization" *Journal of Personality and Social Psychology*, 56 (1989): 170–181.

Matějka, Filip, "Rigid Pricing and Rationally Inattentive Consumer," Working Paper, 2013.

Ofir, Chezy, "Pseudodiagnosticity in Judgement under Uncertainty," *Organizational Behavior and Human Decision Processes*, 42 (1988), 343–363.

Shafer, Wayne and Hugo Sonnenschein, "Equilibrium in Abstract Economies without Ordered Preferences," *Journal of Mathematical Economics*, 2 (1975), 345–348.