Behavioral Inattention

Xavier Gabaix

Motivation

It is clear that our attention is limited.

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

Modifying this assumption is empirically relevant, theoretically doable, and has great consequences in making economics more psychologically realistic, understanding markets, and designing better policies.

A Simple Framework for Modeling Attention

Simple unifying framework for behavioral inattention in economic modeling

Useful in unifying several themes of behavioral economics, at least in a formal sense.

An introduction: Anchoring and adjustment via Gaussian signal extraction

There is a true value x, drawn from a Gaussian distribution $\mathcal{N}\left(x^d, \sigma_x^2\right)$

> x^d is the default value (here, the prior mean)
>
> σ_x^2 is the variance.

Agent does not know this true value, but receives the signal

$$s = x + \varepsilon$$

where ε is drawn from an independent distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Agent takes the action a.

Objective function:

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

If she's rational, solves:

$$\max_a \mathbb{E}\left[-\frac{1}{2}(a-x)^2 \mid s\right]$$

That is, the agent wants to guess the value of x given the noisy signal s.

First-order condition:

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

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

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

with the dampening factor

$$m = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \in [0,1]$$

Agent should anchor at the prior mean x^d, and partially adjust (with a shrinkage factor m) toward the signal s.

Average action $\bar{a}(x) := \mathbb{E}[a(s) \mid x]$ is then:

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

This is "anchoring and adjustment". As Tversky and Kahneman (1974, p. 1129):

"People make estimates by starting from an initial value that is adjusted to yield the final answer [...]. Adjustments are typically insufficient".

Here, agents start from the default value x^d and on expectation adjusts it toward the truth x.

Adjustments are insufficient, as $m \in [0,1]$, because signals are generally imprecise.

Most models are variants or vast generalizations of the model above, with different weights m (endogenous or not) on the true value.

A first class of models eliminates the noise, as not central, at least for the prediction of the average behavior.

A second keeps the noise as central - which often leads to more complicated models.

Turn now to simple formal frame- work for modeling inattention.

Models with deterministic attention and action

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

$$\max_a \; u(a, x)$$

Again, a is an action (possibly multidimensional), and x is a vector of "attributes", e.g. price innovations, characteristics of goods, additional taxes, deviations from the steady state and so on.

Rational agent will choose $a^r(x) = \text{argmax}_a \; u(a, x)$.

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

$$\max_a u(a, x, m)$$

where m is a vector that will characterize the **degree of attention**.

She takes the action:

$$a(x, m) = \underset{a}{\text{argmax}} \; u(a, x, m).$$

In inattention models, we will often take:

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

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

$$x_i^s := m_i x_i + (1 - m_i) x_i^d$$

with an attention parameter $m_i \in [0,1]$, and where x_i^d is the "default value" of variable i.

When $m_i = 0$, the agent "does not think about x_i "

Replaces x_i by $x_i^s = 0$

When $m_i = 1$, she perceives the true value ($x_i^s = x_i$).

$m = (m_i)_{i=1\ldots n}$ is the attention vector.

The default x_i^d is typically the prior mean of x_i.

However, it can be psychologically more sophisticated.

If the mean price of good i is $\mathbb{E}[x_i] = \$10.85$, then the normatively simplest default is $x_i^d = \mathbb{E}[x_i] = \$10.85$.

But the default might be a truncated price, e.g. $x_i^d = \$10$.

Quadratic example:

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

Traditional optimal action:

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

where the r superscript is as in the traditional rational actor model.

For instance, to choose a, the decision maker should consider not only changes x_1 in her wealth, but also:

> Deviation of GDP from its trend, x_2
>
> Impact of interest rate, x_{10}
>
> Demographic trends in China, x_{100}
>
> Recent discoveries in the supply of copper, x_{200}, etc.

There are, say, $n > 10{,}000$ factors that should in principle be taken into account.

A sensible agent will "not think" about most of these factors, especially the less important ones.

We will formalize this notion.

After attention m is chosen, the behavioral agent optimizes under her simpler representation of the world:

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

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

Unifying behavioral biases: Much of behavioral economics reflects a form of inattention

Many behavioral biases share a common structure: people anchor on a simple perception of the world, and partially adjusts toward it.

Conceptually, there is a "true model", and there is a "default, simple model" that spontaneously comes to mind.

Attention m parameterizes the particular convex combination of the default and true models that corresponds to the agent's perception.

Inattention to true prices and shrouding of add-on costs

Default price p^d

New price: p

Price perceived by the agent:

$$p^s(p,m) = mp + (1-m)p^d$$

Take the case without income effect, where the rational demand is $c^r(p)$.

Demand of a behavioral agent is $c^s(p) = c^r(p^s(p,m))$

Sensitivity of demand to price is $c^s(p)' = mc^r(p^s)'$.

Demand sensitivity is muted by a factor m.

Logarithmic space: perceived price is

$$p^s = (p)^m (p^d)^{1-m}$$

Psychology of numbers shows that the latter formulation (in log space) is psychologically more accurate.

Similar reasoning applies to the case of goods sold with separate add-ons.

Price of a base good is p, and the price of an add-on is $\hat{p}$.

Consumer might only partially see the add-on, such that she perceives the add-on cost to be $\hat{p}^s = m\hat{p}$.

Myopic consumer perceives total price to be only $p + m\hat{p}$, while the full price is $p + \hat{p}$.

Such myopic behavior allows firms to shroud information on add-on costs from consumers in equilibrium (Gabaix and Laibson 2006).

Inattention to taxes

Price of a good is p, and the tax on that good is τ.

Full price is $q = p + \tau$

But consumer may pay only partial attention to the tax

Perceived tax is $\tau^s = m\tau$, and the perceived price is $q^s = p + m\tau$.

Neglected risks

Probability of a bad state of the world happening is p.

Perceived probability is $p^s = mp$, if the default probability is $p^d = 0$.

This generates underreaction to neglected risks.

Hyperbolic discounting: inattention to the future

In an intertemporal choice setting, suppose that true utility is:

$$U_0 = \sum_{t=0}^{\infty} \delta^t u_t$$

Let $U_1 = \sum_{t=1}^{\infty} \delta^{t-1} u_t$ be the continuation utility, so that

$$U_0 = u_0 + \delta U_1$$

A present-biased agent will instead see a perceived utility

$$U_0^s = u_0 + m\delta U_1.$$

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

But normative interpretation is different.

If the $m = \beta$ is about misperception, then the favored normative criterion is to maximize over the preferences of the rational agents, i.e maximize $u_0 + \delta U_1$.

With hyperbolic discounting or a planner-doer model (Thaler and Shefrin 1981; Fudenberg and Levine 2012) the welfare criterion is not so clear as one needs to trade off the utility of several "selves".

Prospect theory: Inattention to the true probability

Literature in psychology that finds that probabilities are mentally represented in "log odds space".

Perceptual bias is "ubiquitous" and gives a unified account of many phenomena.

If $p \in (0,1)$ is the probability of an event, the log odds are $q := \ln \frac{p}{1-p} \in (-\infty, \infty)$.

Then, people may misperceive numbersi.e. their median perception is

$$q^s = mq + (1-m)q^d$$

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

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

which is the median perception of a behavioral agent.

We have derived a probability weighting function $\pi(p)$.

This yields overweighting of small probabilities (and symmetrically underweighting of probabilities close to 1).

Intuition: a probability of 10^{-6} is just too strange and unusual, so the brain "rectifies it" by dilating it toward a more standard probability such as $p^d \simeq 0.36$

Overweight!

This is exactly as in the simple Gaussian updating model, done in the log odds space.

This gives a probability weighing function much like in prospect.

Distortions of payoff + distortions of probability = prospect theory.

How to obtain loss aversion?

Assume a "pessimistic prior": typical gamble in life has negative expected value.

For instance the default probability for loss events is higher than the default probability in gains events.

This will create loss aversion.

This thinking is all somewhat ex-post.

But still, makes sense that nature made people prospect-theoretic.

Maybe we should make robots prospect-theoretic if their perceptions were noisy

They would misperceive payoffs and probabilities (because of the inherent noisiness of the intuitive treatment of numbers in the mind)

They would do so in an environment where gambles have in general negative expected values.

Optimal correction of such features creates respectively: diminishing sensitivity, distortion of probabilities, and loss aversion.

Essentially the same as prospect theory.

Tension:

Tendency to neglect lots of small probability events in the "editing phase" of Kahneman and Tversky (1979)

Agents decides which states of the world to take into account at all

And tendency to overestimate them in the decision phase.

This is the kind of tension that irks non-behavioral economists, and embarrasses behavioral economists.

Endogenous attention and sparsity solve this.

Overconfidence: Inattention to my true ability

If x is my true driving ability, with overoptimism my prior x^d may be a high ability value

Perhaps the ability of the top 10% of drivers.

Rosy perceptions come from this high default ability (for myself), coupled with behavioral neglect to make the adjustment.

Related bias: "overprecision"

I think that my beliefs are more accurate than they are

Then x is the true precision of my signals, and x^d is a high precision.

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

Cursedness: Inattention to the conditional probability

Players underestimate the correlation between their strategies and those of their opponents.

The structure is formally similar, with cursedness χ being $1-m$:

Agent forms a belief that is an average of m times to the true probability, and $1-m$ times a simplified, naïve probability distribution.

Projection bias: Inattention to future circumstances by anchoring too much on present circumstances

Need to forecast x_t, a variable at time t.

I might use its time-zero value as an anchor, i.e. $x_t^d = x_0$.

Perception at time zero of the future variable is

$$x_t^s = mx_t + (1-m)x_0$$

hence the agent exhibits projection bias.

Base-rate neglect: Inattention to the base rate

True base probability P is replaced by

$$P^s(y) = mP(y) + (1-m)P^d(y)$$

$P^d(y)$ is a uniform distribution on the values of y.

Correlation neglect

Simplify a situation assuming random variables are uncorrelated

True probability of variables $y = (y_1, \dots, y_n)$ is a joint probability $P(y_1, \dots, y_n)$

> Marginal distribution of y_i is $P_i(y_i)$.

Simpler default probability: joint density assuming no correlation:

$$P^d(y) = P_1(y_1) \dots P_n(y_n)$$

Correlation neglect is captured by a subjective probability:

$$P^s(y) = mP(y) + (1-m)P^d(y).$$

Insensitivity to sample size

True sample size N is replaced by a perceived sample size:

$$N^s = \left(N^d\right)^{1-m} N^m$$

Agents update based on that perceived sample size.

Insensitivity to predictability / Misconceptions of regression to the mean / Illusion of validity: Inattention to the stochasticity of the world

When people see a fighter pilot's performance, they fail to appreciate reversion to the mean.

> If the pilot does less well the next time, they attribute this to lack of motivation, for instance, rather than reversion to the mean.

Call x the pilot's core ability, and $s_t = x + \varepsilon_t$ the performance on day t, where ε_t is an i.i.d. Gaussian noise term and x is drawn from a $N(0, \sigma_x^2)$ distribution.

Given the performance s_t of, say, an airline pilot, an agent predicts next period's performance (Tversky and Kahneman 1974).

Rationally, she predicts $\bar{x}_{t+1} := \mathbb{E}[x_{t+1} \mid x_t] = \lambda x_t$ with

$$\lambda = \frac{1}{1+\sigma_\varepsilon^2/\sigma_x^2}.$$

However, a behavioral agent may "forget about the noise":

In her perceived model, $\mathrm{Var}^s\,(\varepsilon) = m\sigma_\varepsilon^2$.

If $m = 0$, they don't think about the existence of the noise, and answer $\bar{y}_{t+1}^s = y_t$.

Such agent will predict:

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

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

When will see overreaction vs. underreaction?

Variable y_{it} follows a process $y_{i,t+1} = \rho_i y_{it} + \varepsilon_{it}$, and ε_{it} an i.i.d. innovation with mean zero.

The decision-maker deals with many such processes, with various autocorrelations, that are ρ^d on average.

Hence, for a given process, she may not fully perceive the autocorrelation.

Instead use the subjectively perceived autocorrelation ρ_i^s, as in

$$\rho_i^s = m\rho_i + (1-m)\rho^d$$

Instead of seeing precisely the fine nuances of many AR(1) processes…

the agent anchors on a common autocorrelation ρ^d…

…and then adjusts partially toward the true autocorrelation of variable y_{it}, which is ρ_i.

Agent's prediction is $\mathbb{E}_t^s[y_{i,t+k}] = (\rho_i^s)^k y_{i,t}$

So:

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

$\mathbb{E}_t^s$ is the subjective expectation

$\mathbb{E}_t$ is the rational expectation.

Hence, the agent exhibits:

overreaction for processes that are less autocorrelated than ρ^d, as $\frac{\rho_i^s}{\rho_i} > 1$

underreaction for processes that are more autocorrelated than ρ^d, as $\frac{\rho_i^s}{\rho_i} < 1$.[11]

If:

the growth rate of a stock price is almost not autocorrelated

and the growth rate of earnings has a very small positive autocorrelation

Then:

people will overreact to past returns by extrapolating too much.

On the other hand, processes that are quite persistent (say, inflation) will be perceived as less autocorrelated than they truly are

Agents will underreact by extrapolating too little (as found by Mankiw, Reis, and Wolfers 2003).

Left-digit bias: Inattention to non-leading digits

A number, in decimal representation, is $x = a + \frac{b}{10}$, with $a \geq 1$ and $b \in [0,1)$.

An agent's perception of the number might be

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

where a low value of $m \in [0,1]$ indicates left-digit bias.

Lacetera, Pope, and Sydnor (2012) find compelling evidence of left-digit bias in the perception of the mileage of used cars sold at auction.

Exponential growth bias

Difficult to compound interest rates

$x = (1 + r)^t$ is the future value of an asset

But simpler perceived as is $x^d = 1 + rt$

Perceived growth is just $x^s = mx + (1 - m)x^d$.

Taking stocks of all these examples

Illustration that the very simple framework above allows one to think in a relatively unified way about a wide range of behavioral biases.

Four directions in which such baseline examples can be extended.

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

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

3. In the "basic psychology" direction, researchers think more deeply about the "default perception of the world", i.e. what an agent perceives spontaneously. Psychology helps determine this default. [13]

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

Psychological underpinnings

See Pashler (1998) and Styles (2006) for book-length surveys on the psychology of attention.

Conscious versus unconscious attention

Systems 1 and 2 (Kahneman):

System 1: intuitive, fast, largely unconscious

System 2: analytical, slow, conscious system.

System 2, working memory, and conscious attention

We do not handle thousands of variables when dealing with a specific problem.

But in our long term memory, we know about many variables, x.

Hence, we can handle **consciously** relatively few m_i that are different from 0.

System 1 / Unconscious attention monitoring.

At the same time, the mind contemplates unconsciously thousands of variables x_i, and decides which handful it will bring up for conscious examination (that is, whether they should satisfy $m_i > 0$).

For instance, my system is currently monitoring if I'm too hot, thirsty, low in blood sugar, but also in the presence of a venomous snake, and so forth.

This is not done consciously.

If a variable becomes very alarming (e.g. a snake just appeared), it will be "brought to consciousness"

That is, to the attention of system 2 .

Those are the variables with an $m_i > 0$.

In general: dual decision systems.

Based on similarity to past situations

Also holds for memory retrival

Possible to endogenize it?

Reliance on defaults

What guess does one make when there is no time to think?

This is represented by the case $m = 0$

Variables x are are replaced by their default valu

Default model ($m = 0$), and the default action a^d (which is the optimal action under the default model) corresponds to "system 1 under extreme time pressure".

The importance of default actions has been shown in a growing literature.

Here, the default model is very simple (basically, it is "do not think about anything"), but it could be enriched, following other models (e.g. Gennaioli and Shleifer 2010).

Other themes

If choice of attention is unconscious, then there is choice of "attentional blindness".

Canonical experiment for this is the gorilla experiment of Simons and Chabris (1999).

When asked to perform a time-consuming task, subject often didn't see a gorilla in the midst of the experiment.

Just done again

Another theme - not well integrated by the economics literature: "extreme seriality of thought" (see Huang and Pashler 2007)

In the context of visual attention, it means that people can process things only one color at the time.

In other contexts, like the textbook rabbit / duck visual experiment, it means that one can see a rabbit or a duck in a figure, but not both at the same time.

From an economic point of view, serial models that represent the agent's action step by step tend to be complicated but instructive

More "outcome based models", that directly give the action rather than the intermediary steps, can be useful.

Measuring Attention: Methods and Findings

Measuring attention: Methods

There are essentially five ways to measure attention:

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

Methods 3-5 can show that attention is not full

Help reject the naïve rational model

1 and 2 truly measure attention (i.e., measure the parameter m)

Measuring inattention via deviation from an optimal action

Suppose the optimal action function is $a^{BR}(x) = a^r(mx)$

So the derivative with respect to x is:

$$a_x^{BR}(x) = m a_x^r(mx)$$

Therefore attention can be measured as

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

The attention parameter m is identified by the ratio of the sensitivities to the signal x of the boundedly-rational action function a^{BR} and of the rational action function a^r.

This requires knowing the normatively correct slope, a_x^r.

How does one do that?

1. Achievable in "clear and understood" context, e.g. where all prices are very clear, with just a simple task (hence $m = 1$).

   Allows us to measure a_x^r

2. Maybe the "normatively correct answer" is the attention of experts.

   Should one buy generic drugs (e.g. aspirins) or more expensive "branded drugs" - with the same basic molecule?

   Bronnenberg, Dubé, Gentzkow, and Shapiro (2015): health care professionals are less likely to pay extra for premium brands.

Deviations from Slutsky symmetry

Deviations from Slutsky symmetry allow one in principle to measure inattention.

Abaluck and Adams (2017): Slutsky symmetry should also hold in random demand models.

Utility for good i is $v_i = u_i - \beta p_i$

Consumer chooses $a = \text{argmax}_i\ (u_i - \beta p_i + \varepsilon_i)$

ε_i are arbitrary noise terms, maybe correlated.

The probability of choosing i:

$$c_i(p) = \mathbb{P}\left(u_i - \beta p_i + \varepsilon_i = \max_j u_j - \beta p_j + \varepsilon_j\right)$$

Slutsky term $S_{it} = \frac{\partial c_i}{\partial p_j}$

Result: $S_{ij} = S_{ji}$ again, under the rational model.

With inattention to prices, and $c^s(p) = c^r\left(Mp + (1 - M)p^d\right)$, where $M = \text{diag}\ (m_1, \dots, m_n)$ is the diagonal matrix of attention:

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

exactly like in the basic model.

Abaluck and Adams (2017) use this to study inattention to complex health care plans.

A priori obscure idea (the deviation from Slutsky symmetry in limited attention models, as in Gabaix 2014) can lead to concrete real-world measurement of the inattention to health-care plans characteristics.

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

Typical methods:

Process-tracing experiment: Mouselab

Subjects need to click on boxes to see which information they contain

Eye tracking methods

Researchers can follow which part of the screen subjects look at.

Many other physiological methods of measurement:

Pupil dilation (Kahneman 1973).

Schulte-Mecklenbeck et al. (2017): recent review.

Measures are useful, but not ideal:

They measure attentional inputs, not attention itself.

To see this, call T the time spend on dimension i

Time here is a stand in for other measures, e.g. time gazing at the dimension, fMRI intensity, pupil dilatation, and so forth.

Model "attention processing function" as a function of time:

$$m = f(T)$$

Time spent is an input in the attention production function, but it is not attention per se.

Also, attention is limited to $[0,1]$ and time T is unbounded, so function f cannot be linear.

Moreover, the function must be modulated by some "mental effort" M:

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

One may look at a whole lecture/seminar without effort (low M), so total amount learned (indexed by m) is very low.

It would be great to measure the production function of attention, $f(T, M)$.

Arieli, Ben-Ami, and Rubinstein (2011):

Eye-tracking experiment to trace the decision process of experiment participants in the context of choice over lotteries

Individuals rely on separate evaluations of prizes and probabilities in making their decisions.

Krajbich and Rangel (2011):

Drift-diffusion model is a good predictor of choice and reaction times when subjects are faced with choices over two or three alternatives.

Lahey and Oxley (2016):

Eye tracking techniques

Examine recruiters, and see what information they look at in resumes, in particular from white vs African-American applicants

Bartoš, Bauer, Chytilová, and Matějka (2016):

Statistical discrimination guides information acquisition.

Surveys

Difficulty:

Take an economist.

When surveyed, she knows the level of interest rate.

But doesn't mean that she takes the interest rate into account when buying a sweater - so as to satisfy her rational Euler equation for sweaters.

Ignorance in a survey: evidence that they are inattentive.

Even if people show knowledge, does not mean that they take it into account in their decision.

Information, as measured in surveys, is an input into attention

Not the actual attention metric.

While people know their average tax rate, they often don't know their marginal one, and often use the average tax rate as a default proxy for the marginal tax rate (De Bartolomé 1995; Liebman and Zeckhauser 2004).

Impact of reminders, advice

If people don't pay attention, perhaps a reminder will help.

Reminder is a "free signal"

Or an increase in the default attention m_i^d to a dimension.

A reminder could come, for instance, from the newspaper.

Huberman and Regev (2001) show how a New York Times article creates a big impact for one company's stock price.

Reminders have an impact on savings (Karlan, McConnell, Mullainathan, and Zinman 2016) and medical adherence (Pop-Eleches et al. 2011).

Hanna, Mullainathan, and Schwartzstein (2014):

Provide summary information to seaweed farmers.

Allows them to improve their practice, and achieve higher productivity.

This is consistent with a model in which farmers were not optimally using all the information available to them.

Model: if an agent is pessimistic about the fact that some piece of information is useful, she won't pay attention to it, so that she won't be able to realize that it is useful.

Knowledge about the informativeness of the piece of information) leads to paying more attention, and better learning.

Measuring attention: Findings

Inattention to taxes

People don't fully pay attention to taxes

First experimental measure of attention to taxes : Chetty, Looney, and Kroft (2009)

Field experiment.

Mean attention of between 0.06 (by computing the ratio of the semi-elasticities for sales taxes, which are not included in the sticker price, vs. excise taxes, which are included in the sticker price) and 0.35 (computing the ratio of the semi-elasticities for sales taxes vs. more salient sticker prices).

Taubinsky and Rees-Jones (2017):

Online experiment and elicit the maximum tag price that agents would be willing to pay when there are no taxes or when there are standard taxes corresponding to their city of residence.

The ratio of these two prices is $1 + m\tau$, where τ is the tax.

This allows the estimation of tax salience m.

They find $\mathbb{E}[m] = 0.25$ and $\mathrm{Var}\ (m) = 0.13$.

Mean attention is quite small, but the variance is high.

The variance of attention is important, because when attention variance is high, optimal taxes are generally lower (Farhi and Gabaix 2017)

Because heterogeneity in attention creates heterogeneity in response, and additional misallocations, which increase the dead-weight cost of the tax.

Shrouded attributes

People won't pay attention to "shrouded attributes", such as "surprise" bank fees, minibar fees, shipping charges, and the like.

Gabaix and Laibson (2006):

Work out the market equilibrium implication of such attributes with naïve consumers.

That is, consumers who are not paying attention to their existence when buying the "base good" product.

If there are enough naïves there is an inefficient equilibrium where shrouded attributes are priced much above marginal costs.

In this equilibrium, naïve consumers are "exploited": they pay higher prices and subsidize the non-naïves.

Brown, Hossain, and Morgan (2010): consumers are inattentive to shrouded shipping costs in eBay online auctions.

Grubb (2009) and Grubb and Osborne (2015): consumers don't pay attention to sharp marginal charges in three-part tariff pricing schemes and predict their future demand with excessive ex-ante precision

For example, individuals frequently exhaust their cellular plans' usage allowance, and incur high overage costs.

Jin, Luca, and Martin (2017):

Laboratory experiments

Consumers form overly optimistic expectations of product quality when sellers choose not to disclose this information.

People are behavioral rather than Bayesian

Bayesian agent should be suspicious of any non-disclosed item, rather than just ignore it like a behavioral agent.

Lliterature on firms' incentives to hide these attributes (Heidhues and Kőszegi 2010, 2017), and competition with boundedly rational agents (Spiegler 2011; Tirole 2009; Piccione and Spiegler 2012; De Clippel, Eliaz, and Rozen 2014).

Check Behavioral Industrial Organization, by Paul Heidhues and Botond Kőszegi.

Inattention in health plan choices

Confusion and inattention in the choice of health care plans.

McFadden (2006): misinformation in health plan choices.

Abaluck and Gruber (2011):

> People choose Medicare plans more often if premiums are increased by \$100 than if expected out of pocket cost is increased by \$100.

Handel and Kolstad (2015):

> Choice of health care plans at a large firm.
>
> Poor information about plan characteristics has a large impact on employees' willingness to pay for the different plans available to them
>
> On average, they overvalue plans with more generous coverage and lower deductibles.
>
> Abaluck and Adams (2017):
>
> Consumers' inertia in health plan choices is largely attributable to inattention.

Inattention to health consequences

We do not always attend to the health consequences of our choices.

How big is this effect?

Hyperbolic discounting with $m \simeq 0.7$ (Gruber and Kőszegi 2001).

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

Lacetera, Pope, and Sydnor (2012):

> Estimate inattention via buyers' "left-digit bias" in evaluating the mileage of used cars sold at auction.
>
> x is the true mileage of a car
>
> x^d the mileage rounded to the leading digit
>
> $r = x - x^d$: "mileage remainder"
>
> Perceived mileage is $x^s = x^d + m(x - x^d)$.
>
> Lacetera, Pope, and Sydnor (2012) find a mean attention parameter of $m = 0.69$.

Busse, Lacetera, Pope, Silva-Risso, and Sydnor (2013b) find that attention is lower for older and cheaper cars, and lower for lower-income retail buyers.

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

When you buy a car, you should pay attention to both the sticker price of the car, and the present value of future gasoline payments.

But people will pay less than full attention to the future value of gas payments

The full price of the car $p_{\text{car}} + p_{\text{gas}}$ will be perceived as $m_{\text{car}}\, p_{\text{car}} + m_{\text{gas}}\, p_{\text{gas}}$.

Somewhat inconsistent empirical findings.

Allcott and Wozny (2014): partial inattention to gas prices

Estimate is $\frac{m_{\text{gas}}}{m_{\text{price}}} = 0.76$.

Busse, Knittel, and Zettelmeyer (2013a):

Cannot reject the null hypothesis of equal attention, $\frac{m_{\text{gas}}}{m_{\text{price}}} = 1$.

One can conjecture that people likewise do not fully pay attention to the cost of car parts - this remains to be seen.

Inattention in finance

Large amount of evidence of partial inattention in finance.

Hirshleifer, Lim, and Teoh (2009):

When investors are more distracted (as there are more events that day), inefficiencies are stronger

DellaVigna and Pollet (2007):

Investors have a limited ability to incorporate some subtle forces (predictable change in demand because of demographic forces) into their forecasts, especially at long horizons.

DellaVigna and Pollet (2009):

Investors are less attentive on Fridays

When companies report their earnings on Fridays, the immediate impact on the price (as a fraction of the total medium run impact) is lower.

Hirshleifer, Lim, and Teoh (2009):

Investors are less attentive to a given stock when there are lots of other news in the market.

Cohen and Frazzini (2008):

Investors are quick at pricing the "direct" impacts on an announcement, but slower at pricing the "indirect" impact

E.g. a new plane by Boeing gets reflected in Boeing's stock price, but less quickly in that of Boeing's supplier network.

Baker, Pan, and Wurgler (2012):

When thinking about a merger or acquisition price, investors put a lot of attention on recent (trailing 52 weeks) prices.

This has real effects: merger waves occur when high returns on the market and likely targets make it easier for bidders to offer a peak price.

Malmendier and Nagel (2011):

Generations who experienced low stock market returns invest less in the stock market. People seem to put too much weight on their own experience when forming their beliefs about the stock market.

Evidence of reaction to macro news with a lag

Delayed reaction in macro data.

Friedman (1961):

"long and variable lags" in the impacts of monetary stimulus.

Also what motivated models of delayed adjustment, e.g. Taylor (1980).

Empirical macro research in the past decades has frequently found that a variable (e.g. price) reacts to shocks in other variables (e.g. nominal interest rate) only after a significant delay.

Delayed reaction is confirmed by the more modern approaches of Romer and Romer (1989) and Romer and Romer (2004)

Monetary policy shocks using the narrative account of Federal Open Market Committee (FOMC) Meetings

Price level would only start falling 25 months after a contractionary monetary policy shock.

This is confirmed also by more formal econometric evidence with identified VARs.

Sims (2003) notes that in nearly all Vector Autoregression (VAR) studies, a variable reacts smoothly and with delay when responding to shocks in other variables, but contemporaneously and significantly different from zero when responding to its own shocks.

Such finding is robust in VAR specifications of various sizes, variable sets, and identification method (Leeper, Sims, and Zha 1996; Christiano, Eichenbaum, and Evans 2005).

Micro survey data suggest that macro sluggishness is not just the result of delayed action, but rather the result of infrequent observation as well.

Alvarez, Guiso, and Lippi (2012) and Alvarez, Lippi, and Paciello (2017):

Infrequent reviewing of portfolio choice and price setting

Median investor reviews her portfolio 12 times and makes changes only twice annually

Median firm in many countries reviews price only 2-4 times a year.

Attention across stakes and studies

Attention over many studies Table 1 and Figure 1 contain a synthesis of ten studies of attention (i.e., gave an estimate of the parameter m).

Tribute to the hard work of many behavioral economists.

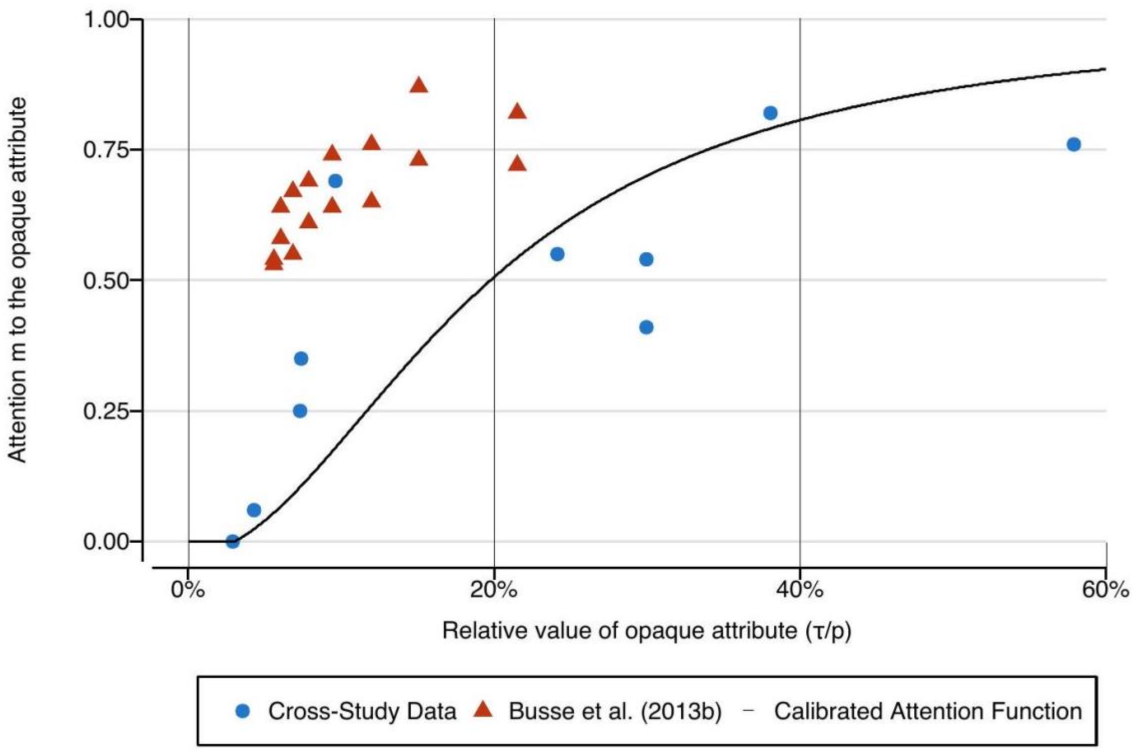
m is measured as the degree to which individuals underperceive the value of an opaque add-on attribute τ to a quantity or price p

Subjectively perceived total value of the quantity is $p^s(m) = p + m\tau$

Study	Good or Quantity	Opaque Attribute	Attentio n Estimat e (m)	Attribute Importanc e (τ/p)

Allcott and Wozny (2014)	Expense associated with car purchase	Present value of future gasoline costs	0.76	0.58
Hossain and Morgan (2006)	Price of CDs sold at auction on eBay	Shipping costs	0.82	0.38
DellaVigna and Pollet (2009)	Public company equity value	Value innovation due to earnings announcements	0.54	0.30
DellaVigna and Pollet (2009)	Public company equity value	Value innovation due to earnings announcements that occur on Fridays	0.41	0.30
Hossain and Morgan (2006)	Price of CDs sold at auction on eBay	Shipping costs	0.55	0.24

Lacetera, Pope, and Sydnor (2012)	Mileage of used cars sold at auction	Mileage left-digit remainder	0.69	0.10
Chetty, Looney, and Kroft (2009)	Price of grocery store items	Sales tax	0.35	0.07
Taubinsky and Rees-Jones (2017)	Price of products purchased in laboratory experiment	Sales tax	0.25	0.07
Chetty, Looney, and Kroft (2009)	Price of retail beer cases	Sales tax	0.06	0.04
Brown, Hossain, and Morgan (2010)	Price of iPods sold at auction on eBay	Shipping costs	0.00	0.03
Mean	-	-	0.44	0.21
Standard Deviation	-	-	0.28	0.18



Models of Endogenous Attention: Deterministic Action

Sparsity model: emphasizes the absolute importance of effects.

Salience model: mostly interested in relative importance.

Paying more attention to more important variables: The sparsity model

Gabaix (2014):

High degree of applicability

Generalization of the max operator used in economics, allowing agents to be less than fully attentive.

This helps write a behavioral version of…

Basic textbook microeconomics

Basic theory of taxation (Farhi and Gabaix (2017))

Basic dynamic macroeconomics (Gabaix (2016a))

Macroeconomic fiscal and monetary policy (Gabaix (2016b)).

Agent faces a maximization problem

Traditional version: $\max_a u(a, x)$ subject to $b(a, x) \geq 0$

u is a utility function

b is a constraint.

Define the "sparse max" operator (Gabaix 2014):

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

less than fully attentive version of the "max" operator.

Variables a, x and function b have arbitrary dimensions.

Default parameter: $x = 0$.

Default action:

Optimal action under the default parameter

$a^d := \arg \ \max_a u(a, 0)$ subject to $b(a, 0) \geq 0$

u and b are concave in a (and at least one of them strictly concave) and twice continuously differentiable around $(a^d, 0)$.

We will typically evaluate the derivatives at the default action and parameter, $(a, x) = (a^d, 0)$.

The sparse max: First, without constraints

Define first the sparse max without constraints

i.e. study $\text{smax}_a \ u(a, x)$

Optimal action:

$$a(x, m) := \arg \ \max_a u(a, x, m)$$

Indirect utility:

$$v(x, m) = u(a(x, m), x)$$

Assume attention creates a psychic cost:

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

with $\alpha \geq 0$.

If $\alpha = 0$, there is a fixed cost κ paid each time m_i is non-zero.

Parameter $\kappa \geq 0$ is a penalty for lack of sparsity.

If $\kappa = 0$, the agent is the traditional, rational agent model.

Allocate attention m as:

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

This is complicated!

Key step: agent will solve a version of this problem.

Definition 4.1 (Sparse max - abstract definition).

Agents does two things:

Step 1: solves the optimal problem above, but in a simplified version:

(i) she replaces her utility by a linear-quadratic approximation
(ii) imagines that the vector x is drawn from a mean 0 distribution, with no correlations, but the accurate variances.

Step 2: picks the best action, seen above: $a(x,m) := \arg \max_a u(a,x,m)$

Introduce some notation:

Expected size of x_i is $\sigma_i = \mathbb{E}\left[x_i^2\right]^{1/2}$, in the "ex ante" version of attention.

In the "ex post allocation of attention" version, we set $\sigma_i := |x_i|$.

Define $a_{x_i} := \frac{\partial a}{\partial x_i} := -u_{aa}^{-1} u_{ax_i}$,

Indicates by how much a change x_i should change the action, for the traditional agent.

Derivatives are evaluated at the default action and parameter:

i.e. at $(a,x) = \left(a^d, 0\right)$

$V(m) = \mathbb{E}[u(a(x,m),x)]$ is the expected consumption utility.

Taylor expansion shows that, for small x

(call $\iota = (1, \ldots, 1)$ the vector corresponding to full attention, like the traditional agent):

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

defining $\Lambda_{ij} := -\sigma_{ij} a_{x_i} u_{aa} a_{x_j}, \sigma_{ij} := \mathbb{E}[x_i x_j]$ and $\sigma^2 = \left\|(\sigma_i^2)_{i=1\ldots}\right\|$.

Agent drops the non-diagonal terms (this is an optional, but useful, feature of the sparse max).

The agent solving simplified problem picks:

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

Attention function

Start with the case with just one variable, $x_1 = x$.

Problem becomes:

$$\min_m \frac{1}{2}(1 - m)^2 \sigma^2 + \kappa |m|^\alpha$$

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

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

Figure 2 plots how attention varies with the variance σ^2 for fixed, linear and quadratic cost:

$$\mathcal{A}_0(\sigma^2) = 1_{\sigma^2 \geq 2}, \mathcal{A}_1(\sigma^2) = \max\left(1 - \frac{1}{\sigma^2}, 0\right), \mathcal{A}_2(\sigma^2) = \frac{\sigma^2}{2 + \sigma^2}$$

What if a^s indeed induces no attention to many variables?

Lemma 4.1 (Special status of linear costs).

When $\alpha \leq 1$ (and only then)…

attention function $\mathcal{A}_\alpha(\sigma^2)$ induces sparsity:

when the variable is not very important, then the attention weight is 0 ($m = 0$).

When $\alpha \geq 1$ (and only then)…

the attention function is continuous.

Hence, only for $\alpha = 1$ do we obtain both sparsity and continuity.

For this reason $\alpha = 1$ is recommended for most applications.

Below I state most results in their general form, making clear when $\alpha = 1$ is required. [31]

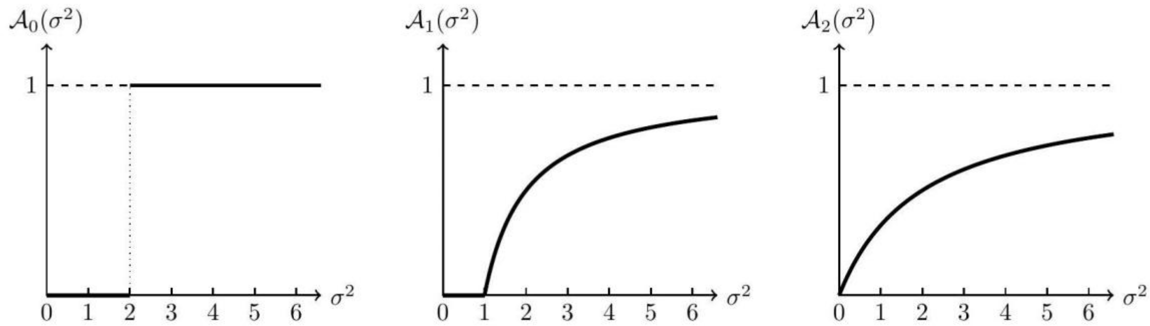


Figure 2: Three attention functions $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$, corresponding to fixed cost, linear cost and quadratic cost respectively. We see that $\mathcal{A}_0$ and $\mathcal{A}_1$ induce sparsity - i.e. a range where attention is exactly 0. $\mathcal{A}_1$ and $\mathcal{A}_2$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

The sparse max: Values of attention

Proposition 4.1 The sparse max is done in two steps.

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

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

$\mathcal{A}: \mathbb{R} \to [0,1]$ is the attention function

σ_i^2 is the perceived variance of x_i^2

$a_{x_i} = -u_{aa}^{-1} u_{ai}$ is the traditional marginal impact of a small change in x_i, evaluated at $x = 0$,

κ is the cost of cognition.

Step 2: Choose the action

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

Hence more attention is paid to variable x_i if…

it is more variable (high σ_i^2)

if it should matter more for the action (high $\left|a_{x_i}\right|$)

if an imperfect action leads to great losses (high $|u_{aa}|$)

and if the cost parameter κ is low.

The sparse max procedure entails (for $\alpha \leq 1$):

"Eliminate each feature of the world that would change the action by only a small amount"

(i.e., when $\alpha = 1$, eliminate the x_i such that $\left|\sigma_i \cdot \frac{\partial a}{\partial x_i}\right| \leq \sqrt{\frac{\kappa}{|u_{aa}|}}$).

This is how a sparse agent sails through life:

> for a given problem, out of the thousands of variables that might be relevant, he takes into account only a few that are important enough to significantly change his decision.
>
> He also devotes "some" attention to those important variables, not necessarily paying full attention to them.

(**Maybe brain representation of world follows similar pattern?**)

Revisit the initial example.

Quadratic loss problem: Traditional and the sparse actions are:

$a^r = \sum_{i=1}^{n} b_i x_i$, and

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

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

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

Sparse max: Full version, allowing for constraints

Extend the sparse max so that it can handle maximization under $K (= \dim\ b)$ constraints.

For example:

$$\max_{c_1,\ldots,c_n} u(c_1, \ldots, c_n)$$

$$\text{subject to } p_1 c_1 + \cdots + p_n c_n \leq w$$

Start from a default price $\boldsymbol{p}^d$.

New price: $p_i = p_i^d + x_i$

Price perceived by the agent: $p_i^s(m) = p_i^d + m_i x_i$

That is:

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

How to satisfy the budget constraint?

Agent who underperceives prices will tend to spend too much

But he's not allowed to do so!

Traditional model:

ratio of marginal utilities optimally equals the ratio of prices: $\frac{\partial u/\partial c_1}{\partial u/\partial c_2} = \frac{p_1}{p_2}$

Preserve that idea, **but in the space of perceived prices**.

Hence, the ratio of marginal utilities equals the ratio of perceived prices:

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

i.e. $u'(\boldsymbol{c}) = \lambda \boldsymbol{p}^s$, for some scalar λ

The agent will tune λ so that the constraint binds,

the value of $\boldsymbol{c}(\lambda) = u'^{-1}(\lambda \boldsymbol{p}^s)$ satisfies $\boldsymbol{p} \cdot \boldsymbol{c}(\lambda) = w \cdot$

In step 2, the agent "hears clearly" whether the budget constraint binds.

This agent is boundedly rational, but smart enough to exhaust his budget.

Generalize this idea to arbitrary problems (heavier notation):

Define the Lagrangian:

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

with $\lambda^d \in \mathbb{R}_+^K$ the Lagrange multiplier when $x = 0$ (the optimal action in the default model is $a^d = \arg\max_a L(a, 0)$).

Marginal action is: $a_x = -L_{aa}^{-1} L_{ax}$.

To turn a problem with constraints into an unconstrained problem, we add the "price" of the constraints to the utility.

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

Step 1: Choose the attention m^* as before, using $\Lambda_{ij} := -\sigma_{ij} a_{x_i} L_{aa} a_{x_j}$, with $a_{x_i} = -L_{aa}^{-1} L_{a x_i}$.

Define $x_i^s = m_i^* x_i$ the associated sparse representation of x.

Step 2: Choose the action.

Form a function $a(\lambda) := \arg\max_a u(a, x^s) + \lambda b(a, x^s)$.

Then, maximize utility under the true constraint: $\lambda^* = \arg\max_{\lambda \in \mathbb{R}_+^K} u(a(\lambda), x^s)$ subject to $b(a(\lambda), x) \geq 0$.

(With just one binding constraint this is equivalent to choosing λ^* such that $b(a(\lambda^*), x) = 0$; in case of ties, we take the lowest λ^*.)

The resulting sparse action is $a^s := a(\lambda^*)$.

Utility is $u^s := u(a^s, x)$.

Step 2 of Definition 4.2 allows quite generally for…

translation of a boundedly rational maximum without constraints…

…into a boundedly maximum with constraints.

For intuition, turn to consumer theory.

Consequences for consumption

We will develop consumer demand from the above procedure.

For instance, Marshallian demand of a behavioral agent is

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

where the as-if budget w' solves $\boldsymbol{p} \cdot \boldsymbol{c}^r(\boldsymbol{p}^s, w') = w$

i.e. ensures that the budget constraint is hit under the true price.

Determination of the attention to prices, m^*

Recall that λ^d is the Lagrange multiplier at the default price.

Proposition 4.2 (Attention to prices).

The sparse agent's attention to price i is:

$$m_i^* = \mathcal{A}_\alpha\left(\left(\frac{\sigma_{p_i}}{p_i^d}\right)^2 \psi_i \lambda^d p_i^d c_i^d / \kappa\right)$$

where ψ_i is the price elasticity of demand for good i.

Hence attention to prices is greater for goods…

(i) with more volatile prices $\left(\frac{\sigma_{p_i}}{p_i^d}\right)$

(ii) with higher price elasticity ψ_i (i.e. for goods whose price is more important in the purchase decision)

(iii) with higher expenditure share $\left(p_i^d c_i^d\right)$.

These predictions seem sensible, though not extremely surprising.

What is important is that we have some procedure to pick the m, so that the model is closed.