BEHAVIORAL ECONOMICS · LECTURE 8

Bounded rationality III: salience and sparse attention

Niels Boissonnet – Bielefeld University

Master Economics – Winter 2026/27

Before we start: which wine?

Two bottles of the same two wines. Wine H is clearly **better** than wine L.

In a supermarket

- Wine L: €10
- Wine H: €20

In a restaurant

- Wine L: €50
- Wine H: €60

Your choice

1. Which bottle do you buy in the supermarket?
2. Which bottle do you order in the restaurant?

In both places, H costs exactly €10 more than L.

Learning objectives

By the end of this lecture you should be able to:

- Explain how Chetty et al. (2009) test whether consumers **attend to taxes** as much as to prices, and read their difference-in-differences estimates.
- Derive the optimal action in the **Gaussian noisy-signal** model and interpret it as anchoring and adjustment.
- Set up and solve the **sparsity model** of attention (Gabaix, 2014) with quadratic losses, and do its comparative statics.
- Compute **salience weights** (Bordalo et al., 2013) and predict choices in simple examples.
- Contrast attention allocated by **stakes** (sparsity) with attention drawn by **contrasts** (salience).

Roadmap

Salience of taxes: evidence

A Bayesian model of inattention

The sparsity model

Applications and limits of sparsity

Salience theory

Wrap-up

Where we are in the course

The assumptions of the Robinson model

0. **Consistent preferences:** Robinson's choices can be described by a utility function u
1. **Costless optimization:** he always finds the option that maximizes u
2. **Correct beliefs & expected utility:** he knows p and maximizes expected utility
3. **Absolute, self-regarding consumption:** only his own (b, y) enters u
4. **Stable tastes:** the same u in both periods

This lecture relaxes Assumption 1 **within** an option: the consumer sees all options, but does not attend equally to all of their **attributes** (price, tax, quality).

Part 2: four ways to make optimization costly

What is costly?

What is costly?	Model	Lecture
Considering *options*	Satisficing	5
	Consideration sets	6
Considering *attributes* of an option	**Sparse attention, salience**	**8 (today)**
Acquiring *information*	Rational inattention	9

Overlaps

These approaches overlap: today's Bayesian and sparsity models are close cousins of rational inattention.

Running question: is attention **chosen** (optimally, given the stakes) or **captured** (by what stands out)?

1

Salience of taxes: evidence

Two studies by Chetty, Looney and Kroft (2009) ask whether consumers react to taxes as much as to prices.

A central assumption in public economics

The standard assumption

- Agents optimize fully with respect to tax policies.
- Ramsey (1927)'s analysis of optimal commodity taxation assumes that agents respond to a **tax change** exactly as to the equivalent **price change**.
- Canonical results on tax incidence, efficiency costs and optimal income taxation rely on the same assumption (Harberger, 1964; Mirrlees (1971); Atkinson and Stiglitz (1976)).

A central assumption in public economics

The standard assumption

- Agents optimize fully with respect to tax policies.
- Ramsey (1927)'s analysis of optimal commodity taxation assumes that agents respond to a **tax change** exactly as to the equivalent **price change**.
- Canonical results on tax incidence, efficiency costs and optimal income taxation rely on the same assumption (Harberger, 1964; Mirrlees (1971); Atkinson and Stiglitz (1976)).

The test

Chetty et al. (2009) test it: are people **less attentive to taxes** than to prices?

- Do tax changes have a smaller effect on demand than price changes?
- If so, some taxes are "**painless**": they raise revenue with little behavioral response.

A simple model of tax attention

Setup

A good has pre-tax price p and an ad valorem tax t: the consumer pays $p(1+t)$.

Rational consumer

Demand depends only on $p(1+t)$, so a 1% rise in $1+t$ has the same effect as a 1% rise in p:

$$\varepsilon_{x,t} \equiv \frac{\partial \log x}{\partial \log(1+t)} = \varepsilon_{x,p} \equiv \frac{\partial \log x}{\partial \log p}.$$

A simple model of tax attention

Setup

A good has pre-tax price p and an ad valorem tax t: the consumer pays $p(1+t)$.

Rational consumer

Demand depends only on $p(1+t)$, so a 1% rise in $1+t$ has the same effect as a 1% rise in p:

$$\varepsilon_{x,t} \equiv \frac{\partial \log x}{\partial \log(1+t)} = \varepsilon_{x,p} \equiv \frac{\partial \log x}{\partial \log p}.$$

Inattentive consumer

She reacts to the *perceived* price $p(1+\theta t)$, where $\theta \in [0,1]$ is her **attention to the tax**. Then, for small t, $\varepsilon_{x,t} \approx \theta\, \varepsilon_{x,p}$.

A simple model of tax attention

Setup

A good has pre-tax price p and an ad valorem tax t: the consumer pays $p(1+t)$.

Rational consumer

Demand depends only on $p(1+t)$, so a 1% rise in $1+t$ has the same effect as a 1% rise in p:

$$\varepsilon_{x,t} \equiv \frac{\partial \log x}{\partial \log(1+t)} = \varepsilon_{x,p} \equiv \frac{\partial \log x}{\partial \log p}.$$

Inattentive consumer

She reacts to the *perceived* price $p(1+\theta t)$, where $\theta \in [0,1]$ is her **attention to the tax**. Then, for small t, $\varepsilon_{x,t} \approx \theta\, \varepsilon_{x,p}$.

Example

Shampoo at $p = \$5.79$, $t = 7.375\%$: true price $\$6.22$. With $\theta = \frac{1}{2}$ the consumer acts as if it cost $\$6.00$.

Two tests of $\theta = 1$: (i) make the tax visible and see whether demand changes; (ii) compare responses to tax and price changes.

Experiment 1: posting tax-inclusive prices

Context

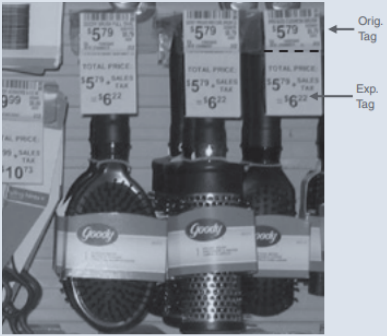
In US stores, shelf prices exclude the **sales tax** (7.375% in the store studied).

Treatment

Add tags showing the **tax-inclusive** price (next to the pre-tax price).

- Three taxable categories: cosmetics, hair-care accessories, deodorants.
- About 750 products.
- The tags stayed for 3 weeks.

The tags



Original tag and experimental tag (Chetty et al., 2009).

Identification: a triple difference

Design

- **Treated:** 13 product categories in the aisle, in the treatment store.
- **Control group 1:** other toiletries in the **same aisle** of the **same store** (95 categories).
- **Control group 2:** the same categories in **two similar stores** without tags.
- Outcome: mean quantity sold per category per week, before vs. during the experiment.

Identification: a triple difference

Design

- **Treated:** 13 product categories in the aisle, in the treatment store.
- **Control group 1:** other toiletries in the **same aisle** of the **same store** (95 categories).
- **Control group 2:** the same categories in **two similar stores** without tags.
- Outcome: mean quantity sold per category per week, before vs. during the experiment.

Triple difference

$$\text{DDD} = \underbrace{(\Delta\text{treated} - \Delta\text{control})_{\text{treatment store}}}_{DD_{TS}} - \underbrace{(\Delta\text{treated} - \Delta\text{control})_{\text{control stores}}}_{DD_{CS}}$$

DD_{CS} removes trends specific to the treated categories (seasonality of deodorants...); the identifying assumption is that such trends are the same in all stores.

Results of Experiment 1

Mean quantity sold per category and week

TABLE 3— EFFECT OF POSTING TAX-INCLUSIVE PRICES: *DDD* ANALYSIS OF MEAN QUANTITY SOLD

Period	Control categories	Treated categories	Difference
Panel A. Treatment store			
Baseline (2005:1–2006:6)	26.48	25.17	−1.31
	(0.22)	(0.37)	(0.43)
	[5,510]	[754]	[6,264]
Experiment (2006:8–2006:10)	27.32	23.87	−3.45
	(0.87)	(1.02)	(0.64)
	[285]	[39]	[324]
Difference over time	0.84	−1.30	$DD_{TS} = -2.14$
	(0.75)	(0.92)	(0.68)
	[5,795]	[793]	[6,588]
Panel B. Control stores			
Baseline (2005:1–2006:6)	30.57	27.94	-2.63
	(0.24)	(0.30)	(0.32)
	[11,020]	[1,508]	[12,528]
Experiment (2006:8–2006:10)	30.76	28.19	−2.57
	(0.72)	(1.06)	(1.09)
	[570]	[78]	[648]
Difference over time	0.19	0.25	$DD_{CS} = 0.06$
	(0.64)	(0.92)	(0.95)
	[11,590]	[1,586]	[13,176]
DDD Estimate			−2.20
			(0.59)
			[19,764]

Notes: Each cell shows mean quantity sold per category per week, for various subsets of the sample. Standard errors (clustered by week) in parentheses, number of observations in square brackets. Experimental period spans week 8 in 2006 to week 10 in 2006. Baseline period spans week 1 in 2005 to week 6 in 2006. Lower panel reflects averages across the two control stores.

Treatment store (units/week)

- treated: $25.17 \rightarrow 23.87 \;\; (-1.30)$
- control: $26.48 \rightarrow 27.32 \;\; (+0.84)$
- $DD_{TS} = -1.30 - 0.84 = -2.14$

Results of Experiment 1

Mean quantity sold per category and week

TABLE 3— EFFECT OF POSTING TAX-INCLUSIVE PRICES: *DDD* ANALYSIS OF MEAN QUANTITY SOLD

Period	Control categories	Treated categories	Difference
Panel A. Treatment store			
Baseline (2005:1–2006:6)	26.48	25.17	−1.31
	(0.22)	(0.37)	(0.43)
	[5,510]	[754]	[6,264]
Experiment (2006:8–2006:10)	27.32	23.87	−3.45
	(0.87)	(1.02)	(0.64)
	[285]	[39]	[324]
Difference over time	0.84	−1.30	$DD_{TS} = -2.14$
	(0.75)	(0.92)	(0.68)
	[5,795]	[793]	[6,588]
Panel B. Control stores			
Baseline (2005:1–2006:6)	30.57	27.94	-2.63
	(0.24)	(0.30)	(0.32)
	[11,020]	[1,508]	[12,528]
Experiment (2006:8–2006:10)	30.76	28.19	−2.57
	(0.72)	(1.06)	(1.09)
	[570]	[78]	[648]
Difference over time	0.19	0.25	$DD_{CS} = 0.06$
	(0.64)	(0.92)	(0.95)
	[11,590]	[1,586]	[13,176]
DDD Estimate			−2.20
			(0.59)
			[19,764]

Notes: Each cell shows mean quantity sold per category per week, for various subsets of the sample. Standard errors (clustered by week) in parentheses, number of observations in square brackets. Experimental period spans week 8 in 2006 to week 10 in 2006. Baseline period spans week 1 in 2005 to week 6 in 2006. Lower panel reflects averages across the two control stores.

Treatment store (units/week)

- treated: $25.17 \rightarrow 23.87 \ \ (-1.30)$
- control: $26.48 \rightarrow 27.32 \ \ (+0.84)$
- $DD_{TS} = -1.30 - 0.84 = -2.14$

Triple difference

Control stores: $DD_{CS} = +0.06.$
DDD $= -2.14 - 0.06 = -2.20$ (s.e. 0.59).

Results of Experiment 1

Mean quantity sold per category and week

TABLE 3— EFFECT OF POSTING TAX-INCLUSIVE PRICES: *DDD* ANALYSIS OF MEAN QUANTITY SOLD

Period	Control categories	Treated categories	Difference
Panel A. Treatment store			
Baseline (2005:1–2006:6)	26.48	25.17	−1.31
	(0.22)	(0.37)	(0.43)
	[5,510]	[754]	[6,264]
Experiment (2006:8–2006:10)	27.32	23.87	−3.45
	(0.87)	(1.02)	(0.64)
	[285]	[39]	[324]
Difference over time	0.84	−1.30	$DD_{TS} = -2.14$
	(0.75)	(0.92)	(0.68)
	[5,795]	[793]	[6,588]
Panel B. Control stores			
Baseline (2005:1–2006:6)	30.57	27.94	-2.63
	(0.24)	(0.30)	(0.32)
	[11,020]	[1,508]	[12,528]
Experiment (2006:8–2006:10)	30.76	28.19	−2.57
	(0.72)	(1.06)	(1.09)
	[570]	[78]	[648]
Difference over time	0.19	0.25	$DD_{CS} = 0.06$
	(0.64)	(0.92)	(0.95)
	[11,590]	[1,586]	[13,176]
DDD Estimate			−2.20
			(0.59)
			[19,764]

Notes: Each cell shows mean quantity sold per category per week, for various subsets of the sample. Standard errors (clustered by week) in parentheses, number of observations in square brackets. Experimental period spans week 8 in 2006 to week 10 in 2006. Baseline period spans week 1 in 2005 to week 6 in 2006. Lower panel reflects averages across the two control stores.

Treatment store (units/week)

- treated: $25.17 \rightarrow 23.87 \;\; (-1.30)$
- control: $26.48 \rightarrow 27.32 \;\; (+0.84)$
- $DD_{TS} = -1.30 - 0.84 = -2.14$

Triple difference

Control stores: $DD_{CS} = +0.06$.
DDD $= -2.14 - 0.06 = -2.20$ (s.e. 0.59).

$\Rightarrow$ posting the tax lowers demand by about **8%**, although nothing changed in the real price.

Experiment 2: taxes on alcohol

Two taxes on beer

- In the US, beer is subject to two taxes:
  - an **excise tax** e, *included* in the posted price;
  - a **sales tax** t, *added at the register*.
- Total price: $p(1+e)(1+t)$. Under full attention, $\log(1+e)$ and $\log(1+t)$ must have the **same** effect on demand.
- Both taxes change regularly and **independently** across US states (1970–2003).

Experiment 2: taxes on alcohol

Two taxes on beer

- In the US, beer is subject to two taxes:
  - an **excise tax** e, *included* in the posted price;
  - a **sales tax** t, *added at the register*.
- Total price: $p(1+e)(1+t)$. Under full attention, $\log(1+e)$ and $\log(1+t)$ must have the **same** effect on demand.
- Both taxes change regularly and **independently** across US states (1970–2003).

Estimation

Estimate, on state–year data,

$$\Delta \log x = \beta_e \, \Delta \log(1+e) + \beta_t \, \Delta \log(1+t) + \text{controls}.$$

Full attention: $\beta_e = \beta_t$.
Inattention to the sales tax: $\theta = \beta_t / \beta_e < 1$.

The data

Tax changes across US states

TABLE 5— SUMMARY STATISTICS FOR STATE BEER CONSUMPTION, TAXES, AND REGULATION

Per capita beer consumption (cans)	243.2 (46.1)
State beer excise tax ($/case)	0.51 (0.50)
State beer excise tax (percent)	6.5 (8.2)
Sales tax (percent)	4.3 (1.9)
Drinking age is 21	0.73 (0.44)
Drunk driving standard	0.65 (0.47)
Any alcohol regulation change	0.19 (0.39)
N (number of state-year pairs)	1,666

Notes: Statistics reported are means with standard deviations in parentheses. Observations are by state for each year from 1970 to 2003. "Drinking age is 21" is an indicator for whether the state-year has a legal drinking age of 21. "Drunk driving standard" indicates state-year has a threshold blood alcohol content level above which one is automatically guilty of drunk driving. "Any alcohol regulation change" is a dummy variable equal to one in any year where a state has raised the drinking age or implemented a stricter drunk driving standard, an administrative license revocation law, or a zero tolerance youth drunk driving law. See Web Appendix A for data sources and sample definition.

Variation

- US states, 1970–2003.
- 153 legislated changes in sales taxes, 131 in excise taxes.
- Correlation between the two kinds of changes: 0.06, so plenty of **independent variation**.

Results: consumption responds to the excise tax, not to the sales tax

Beer consumption and tax changes

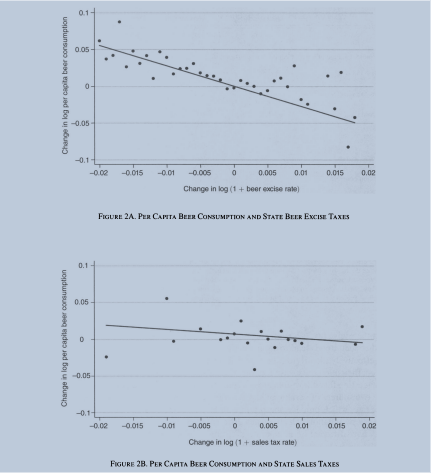


FIGURE 2A. PER CAPITA BEER CONSUMPTION AND STATE BEER EXCISE TAXES

FIGURE 2B. PER CAPITA BEER CONSUMPTION AND STATE SALES TAXES

Reading the figure

- **Top:** change in log beer consumption vs. change in $\log(1 + e)$: clear negative slope.
- **Bottom:** the same against $\log(1 + t)$: almost flat.

Taxes that are not in the posted price barely affect demand.

Results: regression estimates

Regression table

TABLE 6— EFFECT OF EXCISE AND SALES TAXES ON BEER CONSUMPTION

	Baseline (1)	Business cycle (2)	Alcohol regulations (3)	Region trends (4)
Dependent variable: Change in log (per capita beer consumption)				
Δ log (1 + excise tax rate)	−0.88	−0.91	−0.89	−0.71
	(0.17)	(0.17)	(0.17)	(0.18)
Δ log (1 + sales tax rate)	−0.20	−0.01	−0.02	−0.05
	(0.30)	(0.30)	(0.30)	(0.30)
Δ log (population)	0.03	−0.07	−0.07	−0.09
	(0.06)	(0.07)	(0.07)	(0.08)
Δ log (income per capita)		0.22	0.22	0.22
		(0.05)	(0.05)	(0.05)
Δ log (unemployment rate)		−0.01	−0.01	−0.01
		(0.01)	(0.01)	(0.01)
Alcohol regulation controls			x	x
Year fixed effects	x	x	x	x
Region fixed effects				x
F-test for equality of tax elasticities (prob > *F*)	0.05	0.01	0.01	0.06
Sample size	1,607	1,487	1,487	1,487

Notes: Standard errors, clustered by state, in parentheses. All specifications are estimated on full sample for which data are available (state unemployment rate data are unavailable in early years). Column 3 includes three indicators for whether the state implemented per se drunk driving standards, administrative license revocation laws, or zero tolerance youth drunk driving laws, and the change in the minimum drinking age (measured in years). Column 4 includes fixed effects for each of nine census regions. *F*-test tests null hypothesis that coefficients on excise and sales tax rate variables are equal.

Baseline (column 1)

- $\hat{\beta}_e = -0.88$ (s.e. 0.17)
- $\hat{\beta}_t = -0.20$ (s.e. 0.30)
- equality rejected at 5%

Results: regression estimates

Regression table

TABLE 6— EFFECT OF EXCISE AND SALES TAXES ON BEER CONSUMPTION

	Baseline (1)	Business cycle (2)	Alcohol regulations (3)	Region trends (4)
Dependent variable: Change in log (per capita beer consumption)				
Δ log (1 + excise tax rate)	−0.88	−0.91	−0.89	−0.71
	(0.17)	(0.17)	(0.17)	(0.18)
Δ log (1 + sales tax rate)	−0.20	−0.01	−0.02	−0.05
	(0.30)	(0.30)	(0.30)	(0.30)
Δ log (population)	0.03	−0.07	−0.07	−0.09
	(0.06)	(0.07)	(0.07)	(0.08)
Δ log (income per capita)		0.22	0.22	0.22
		(0.05)	(0.05)	(0.05)
Δ log (unemployment rate)		−0.01	−0.01	−0.01
		(0.01)	(0.01)	(0.01)
Alcohol regulation controls			x	x
Year fixed effects	x	x	x	x
Region fixed effects				x
F-test for equality of tax elasticities (prob > *F*)	0.05	0.01	0.01	0.06
Sample size	1,607	1,487	1,487	1,487

Notes: Standard errors, clustered by state, in parentheses. All specifications are estimated on full sample for which data are available (state unemployment rate data are unavailable in early years). Column 3 includes three indicators for whether the state implemented per se drunk driving standards, administrative license revocation laws, or zero tolerance youth drunk driving laws, and the change in the minimum drinking age (measured in years). Column 4 includes fixed effects for each of nine census regions. *F*-test tests null hypothesis that coefficients on excise and sales tax rate variables are equal.

Baseline (column 1)

- $\hat{\beta}_e = -0.88$ (s.e. 0.17)
- $\hat{\beta}_t = -0.20$ (s.e. 0.30)
- equality rejected at 5%

$\hat{\theta} = \hat{\beta}_t / \hat{\beta}_e \approx 0.23$;
with controls, $\hat{\beta}_t$ is close to 0.

Recap: tax salience

Recap

- Standard public economics assumes $\theta = 1$: taxes and prices are equivalent.

Recap: tax salience

Recap

- Standard public economics assumes $\theta = 1$: taxes and prices are equivalent.
- **Store experiment:** making the tax visible lowers demand by about 8%.

Recap: tax salience

Recap

- Standard public economics assumes $\theta = 1$: taxes and prices are equivalent.
- **Store experiment:** making the tax visible lowers demand by about 8%.
- **Alcohol taxes:** demand reacts to the posted excise tax, much less to the sales tax.

Recap: tax salience

Recap

- Standard public economics assumes $\theta = 1$: taxes and prices are equivalent.
- **Store experiment:** making the tax visible lowers demand by about 8%.
- **Alcohol taxes:** demand reacts to the posted excise tax, much less to the sales tax.
- Consequences: “painless” taxes, and tax formulas (incidence, deadweight loss) that must take θ into account.

Recap: tax salience

Recap

- Standard public economics assumes $\theta = 1$: taxes and prices are equivalent.
- **Store experiment:** making the tax visible lowers demand by about 8%.
- **Alcohol taxes:** demand reacts to the posted excise tax, much less to the sales tax.
- Consequences: “painless” taxes, and tax formulas (incidence, deadweight loss) that must take θ into account.

Recap: tax salience

Recap

- Standard public economics assumes $\theta = 1$: taxes and prices are equivalent.
- **Store experiment:** making the tax visible lowers demand by about 8%.
- **Alcohol taxes:** demand reacts to the posted excise tax, much less to the sales tax.
- Consequences: “painless” taxes, and tax formulas (incidence, deadweight loss) that must take θ into account.

Consumers under-react to **non-salient** components of prices. *Why?* Two families of models: attention is **chosen** given the stakes (next sections) or **captured** by contrasts (last section).

2

A Bayesian model of inattention

With a noisy signal, the optimal action shrinks the signal towards a default; the weight on the signal can be read as attention.

A noisy signal

Assumptions

1. An unknown payoff-relevant value $x \sim \mathcal{N}(x^d, \sigma_x^2)$; x^d is the **default** (prior mean).
2. The decision maker (DM) observes a **signal** $s = x + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ independent of x.
3. She chooses an action $a \in \mathbb{R}$ to maximize

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

i.e. she wants her action to match x.

A noisy signal

Assumptions

1. An unknown payoff-relevant value $x \sim \mathcal{N}(x^d, \sigma_x^2)$; x^d is the **default** (prior mean).
2. The decision maker (DM) observes a **signal** $s = x + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ independent of x.
3. She chooses an action $a \in \mathbb{R}$ to maximize

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

i.e. she wants her action to match x.

The DM's problem

Since x is unknown, she solves

$$\max_a \ \mathbb{E}\big[-\tfrac{1}{2}(x - a)^2 \,\big|\, s\big].$$

Optimal action: shrink the signal towards the default

First-order condition

$\mathbb{E}[x - a \mid s] = 0$, so the optimal action is the posterior mean:

$$a(s) = \mathbb{E}[x \mid s] = \alpha\, s + (1 - \alpha)\, x^d,$$

$$\alpha \equiv \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \in [0, 1].$$

Optimal action: shrink the signal towards the default

First-order condition

$\mathbb{E}[x - a \mid s] = 0$, so the optimal action is the posterior mean:

$$a(s) = \mathbb{E}[x \mid s] = \alpha\, s + (1 - \alpha)\, x^d,$$

$$\alpha \equiv \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \in [0, 1].$$

Average action given the true x

Since $\mathbb{E}[s \mid x] = x + \mathbb{E}[\varepsilon \mid x] = x$,

$$\mathbb{E}[a(s) \mid x] = \alpha\, x + (1 - \alpha)\, x^d.$$

Optimal action: shrink the signal towards the default

First-order condition

$\mathbb{E}[x - a \mid s] = 0$, so the optimal action is the posterior mean:

$$a(s) = \mathbb{E}[x \mid s] = \alpha\, s + (1 - \alpha)\, x^d,$$

$$\alpha \equiv \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \in [0, 1].$$

Average action given the true x

Since $\mathbb{E}[s \mid x] = x + \mathbb{E}[\varepsilon \mid x] = x$,

$$\mathbb{E}[a(s) \mid x] = \alpha\, x + (1 - \alpha)\, x^d.$$

Example

$x^d = 10$, $\sigma_x^2 = 4$, $\sigma_\varepsilon^2 = 1$: $\alpha = 0.8$. A signal $s = 20$ gives $a = 0.8 \cdot 20 + 0.2 \cdot 10 = 18$.

Noisy signal (large σ_ε): stay close to the default x^d. Precise signal (small σ_ε): move close to the true x.

Interpretation: anchoring and adjustment

Average action

$$\mathbb{E}[a(s) \mid x] = \alpha\, x + (1 - \alpha)\, x^d,$$

$$\alpha = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2}$$

- Perfectly informed ($\sigma_\varepsilon = 0$): $\alpha = 1$, the DM acts on the true x.
- Uninformed ($\sigma_\varepsilon \to \infty$): $\alpha \to 0$, the DM acts on the default x^d.

Interpretation: anchoring and adjustment

Average action

$$\mathbb{E}[a(s) \mid x] = \alpha\, x + (1-\alpha)\, x^d,$$

$$\alpha = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2}$$

- Perfectly informed ($\sigma_\varepsilon = 0$): $\alpha = 1$, the DM acts on the true x.
- Uninformed ($\sigma_\varepsilon \to \infty$): $\alpha \to 0$, the DM acts on the default x^d.

Anchoring and adjustment

"People make estimates by starting from an initial value that is adjusted to yield the final answer [. . .]. Adjustments are typically insufficient." (Tversky and Kahneman, 1974)

The **anchor** is the prior x^d; the signal produces an **insufficient adjustment** towards x.

Interpretation: anchoring and adjustment

Average action

$$\mathbb{E}[a(s) \mid x] = \alpha\, x + (1-\alpha)\, x^d,$$

$$\alpha = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2}$$

- Perfectly informed ($\sigma_\varepsilon = 0$): $\alpha = 1$, the DM acts on the true x.
- Uninformed ($\sigma_\varepsilon \to \infty$): $\alpha \to 0$, the DM acts on the default x^d.

Anchoring and adjustment

"People make estimates by starting from an initial value that is adjusted to yield the final answer [...]. Adjustments are typically insufficient." (Tversky and Kahneman, 1974)

The **anchor** is the prior x^d; the signal produces an **insufficient adjustment** towards x.

From now on

Normalize $x^d = 0$ (measure x as a deviation from the default), so that $\mathbb{E}[a \mid x] = \alpha x$.

From signal precision to attention

Precision as attention

- Read α as the **attention** paid to x: $\alpha = 1$ full attention, $\alpha = 0$ none.
- A more precise signal is costly: cost $c(\sigma_\varepsilon)$, decreasing in σ_ε.
- $\alpha = \sigma_x^2/(\sigma_x^2 + \sigma_\varepsilon^2)$ is strictly decreasing in σ_ε: choosing $\sigma_\varepsilon \Leftrightarrow$ choosing $\alpha \in [0, 1]$, at cost $\tilde{c}(\alpha) \equiv c(\sigma_\varepsilon)$.

From signal precision to attention

Precision as attention

- Read α as the **attention** paid to x: $\alpha = 1$ full attention, $\alpha = 0$ none.
- A more precise signal is costly: cost $c(\sigma_\varepsilon)$, decreasing in σ_ε.
- $\alpha = \sigma_x^2/(\sigma_x^2 + \sigma_\varepsilon^2)$ is strictly decreasing in σ_ε: choosing $\sigma_\varepsilon \Leftrightarrow$ choosing $\alpha \in [0,1]$, at cost $\tilde{c}(\alpha) \equiv c(\sigma_\varepsilon)$.

Deterministic version (used from now on)

The DM acts as $a = \alpha x$, the *average* Bayesian action, and chooses

$$\max_{\alpha \in [0,1]} \mathbb{E}\big[-\tfrac{1}{2}(x - \alpha x)^2\big] - \tilde{c}(\alpha)$$
$$= \max_{\alpha \in [0,1]} -\tfrac{1}{2}(1-\alpha)^2 \sigma_x^2 - \tilde{c}(\alpha).$$

Attention is a choice: more attention reduces mistakes but costs effort.

3

The sparsity model

With several attributes, attention to each one trades off its stakes against the cost of attention.

Multi-attribute decisions

Assumptions

(Gabaix, 2014, 2019)

1. The option has n attributes $x = (x_1, \ldots, x_n)$, measured as deviations from defaults ($\mathbb{E}x_i = 0$), with $\mathrm{Var}(x_i) = \sigma_i^2$, **independent** across i. *E.g.* $x = (x_1, x_2) = (\text{price}, \text{ tax})$.
2. Quadratic utility:
   $u(a, x) = -\frac{1}{2}\left(a - \sum_i \beta_i x_i\right)^2$.
   The rational action is $a^r(x) = \sum_i \beta_i x_i$.

Multi-attribute decisions

Assumptions

(Gabaix, 2014, 2019)

1. The option has n attributes $x = (x_1, \dots, x_n)$, measured as deviations from defaults ($\mathbb{E}x_i = 0$), with $\mathrm{Var}(x_i) = \sigma_i^2$, **independent** across i. *E.g.* $x = (x_1, x_2) = (\text{price},\ \text{tax})$.
2. Quadratic utility:
   $u(a, x) = -\frac{1}{2}\left(a - \sum_i \beta_i x_i\right)^2$.
   The rational action is $a^r(x) = \sum_i \beta_i x_i$.

Behavioral DM

3. The behavioral DM perceives $x_i^s = \alpha_i x_i$, with **attention** $\alpha = (\alpha_1, \dots, \alpha_n) \in [0, 1]^n$, and acts on her perception:
   $$a(\alpha) = \arg\max_a u(a, \alpha \cdot x) = \sum_i \beta_i\, \alpha_i\, x_i.$$
4. Attention costs $c(\alpha) = \kappa \sum_i \alpha_i^\gamma$, with $\kappa > 0$, $\gamma \geq 0$.

Endogenous attention

The DM proceeds in two steps

1. **Choose attention** $\alpha^\star$ to solve $\max_{\alpha \in [0,1]^n} \mathbb{E}\big[u\big(a(\alpha), x\big)\big] - c(\alpha), \qquad a(\alpha) = \arg\max_a u(a, \alpha \cdot x).$
2. **Act** on her perception: $a(\alpha^\star)$.

Endogenous attention

The DM proceeds in two steps

1. **Choose attention** $\alpha^\star$ to solve $\max_{\alpha\in[0,1]^n} \mathbb{E}\big[u\big(a(\alpha),x\big)\big] - c(\alpha), \qquad a(\alpha) = \arg\max_a u(a, \alpha\cdot x).$
2. **Act** on her perception: $a(\alpha^\star)$.

Quadratic utility, independent attributes

Expand the square; cross terms vanish since $\mathbb{E}[x_i x_j] = 0$:

$$\begin{aligned}\mathbb{E}\big[u(a(\alpha),x)\big] &= -\tfrac{1}{2}\,\mathbb{E}\Big[\Big(\sum_i \beta_i(1-\alpha_i)x_i\Big)^2\Big] \\ &= -\tfrac{1}{2}\sum_i \beta_i^2(1-\alpha_i)^2\sigma_i^2.\end{aligned}$$

Endogenous attention

The DM proceeds in two steps

1. **Choose attention** $\alpha^\star$ to solve $\max_{\alpha\in[0,1]^n} \mathbb{E}\big[u\big(a(\alpha),x\big)\big] - c(\alpha), \qquad a(\alpha) = \arg\max_a u(a, \alpha\cdot x).$
2. **Act** on her perception: $a(\alpha^\star)$.

Quadratic utility, independent attributes

Expand the square; cross terms vanish since $\mathbb{E}[x_i x_j] = 0$:

$$\begin{aligned}\mathbb{E}\big[u(a(\alpha),x)\big] &= -\tfrac{1}{2}\,\mathbb{E}\Big[\Big(\sum_i \beta_i(1-\alpha_i)x_i\Big)^2\Big] \\ &= -\tfrac{1}{2}\sum_i \beta_i^2(1-\alpha_i)^2\sigma_i^2.\end{aligned}$$

So attention to each attribute is chosen **separately**:

$$\alpha_i^\star \in \arg\min_{\alpha_i\in[0,1]} \tfrac{1}{2}\,\beta_i^2\sigma_i^2\,(1-\alpha_i)^2 + \kappa\,\alpha_i^\gamma.$$

Worked example: price and tax with linear cost ($\gamma = 1$)

Linear cost: first-order condition

With $\gamma = 1$, the first-order condition $-\beta_i^2 \sigma_i^2 (1 - \alpha_i) + \kappa = 0$ gives

$$\alpha_i^\star = \max\left\{0,\ 1 - \frac{\kappa}{\beta_i^2 \sigma_i^2}\right\}.$$

Worked example: price and tax with linear cost ($\gamma = 1$)

Linear cost: first-order condition

With $\gamma = 1$, the first-order condition $-\beta_i^2\sigma_i^2(1-\alpha_i) + \kappa = 0$ gives

$$\alpha_i^\star = \max\left\{0,\ 1 - \frac{\kappa}{\beta_i^2\sigma_i^2}\right\}.$$

Example: $\kappa = 1$

- **Price:** large variation across goods, stakes $\beta_1^2\sigma_1^2 = 4 \ \Rightarrow\ \alpha_1^\star = 1 - \frac{1}{4} = \frac{3}{4}$.
- **Tax:** small variation, stakes $\beta_2^2\sigma_2^2 = \frac{1}{2} \ \Rightarrow\ 1 - 2 < 0$, so $\alpha_2^\star = 0$.

Worked example: price and tax with linear cost ($\gamma = 1$)

Linear cost: first-order condition

With $\gamma = 1$, the first-order condition $-\beta_i^2\sigma_i^2(1-\alpha_i) + \kappa = 0$ gives

$$\alpha_i^\star = \max\left\{0,\ 1 - \frac{\kappa}{\beta_i^2\sigma_i^2}\right\}.$$

Example: $\kappa = 1$

- **Price:** large variation across goods, stakes $\beta_1^2\sigma_1^2 = 4 \ \Rightarrow\ \alpha_1^\star = 1 - \frac{1}{4} = \frac{3}{4}$.
- **Tax:** small variation, stakes $\beta_2^2\sigma_2^2 = \frac{1}{2} \ \Rightarrow\ 1 - 2 < 0$, so $\alpha_2^\star = 0$.

The DM **ignores** the low-stakes attribute completely ($\alpha_2^\star = 0$): attention is **sparse**. She under-reacts to, but does not ignore, the high-stakes one.

Beyond quadratic utility: the sparse max

The problem

With general u, the two-step problem is hard: computing the value of attention requires solving the problem one wanted to simplify, an **infinite regress**.

Beyond quadratic utility: the sparse max

The problem

With general u, the two-step problem is hard: computing the value of attention requires solving the problem one wanted to simplify, an **infinite regress**.

The sparse max

Gabaix (2014): choose attention as if the loss were **quadratic** (a second-order approximation),

$$\min_{\alpha \in [0,1]^n} \sum_i \left[\tfrac{1}{2} (1-\alpha_i)^2 \sigma_i^2 \left| u_{aa} \right| a_{x_i}^2 + \kappa \, \alpha_i^\gamma \right],$$

then pick the action that maximizes the **true** u given the perceived $\alpha \cdot x$.
u_{aa} and $a_{x_i} = \partial a^r / \partial x_i$ are evaluated at the default ($x = 0$).

Beyond quadratic utility: the sparse max

The problem

With general u, the two-step problem is hard: computing the value of attention requires solving the problem one wanted to simplify, an **infinite regress**.

The sparse max

Gabaix (2014): choose attention as if the loss were **quadratic** (a second-order approximation),

$$\min_{\alpha\in[0,1]^n} \sum_i \left[\tfrac{1}{2}(1-\alpha_i)^2\,\sigma_i^2\,|u_{aa}|\,a_{x_i}^2 + \kappa\,\alpha_i^\gamma\right],$$

then pick the action that maximizes the **true** u given the perceived $\alpha \cdot x$.
u_{aa} and $a_{x_i} = \partial a^r/\partial x_i$ are evaluated at the default ($x = 0$).

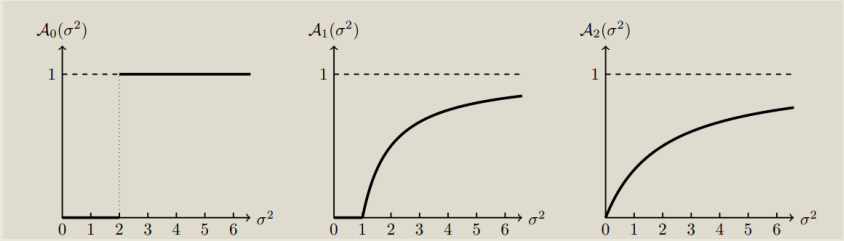
Check

For $u(a,x) = -\frac{1}{2}\left(a - \sum_i \beta_i x_i\right)^2$: $u_{aa} = -1$ and $a^r = \sum_i \beta_i x_i$, so $a_{x_i} = \beta_i$: we recover exactly the problem of the previous slides.

Comparative statics: the stakes

Optimal attention as a function of the stakes ($\kappa = 1$, $\beta = 1$)

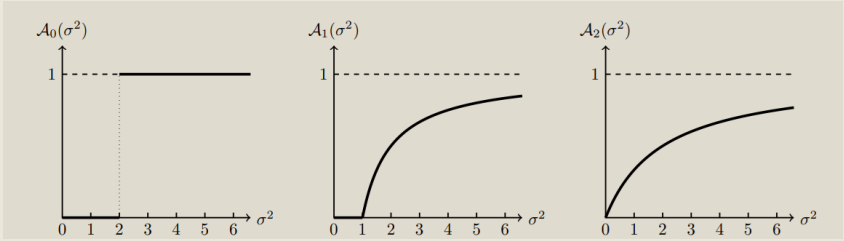
$$\mathcal{A}_\gamma(\sigma^2) \equiv \arg\min_{\alpha \in [0,1]} \left\{ \tfrac{1}{2}(1-\alpha)^2\sigma^2 + \alpha^\gamma \right\}$$



Comparative statics: the stakes

Optimal attention as a function of the stakes ($\kappa = 1$, $\beta = 1$)

$$\mathcal{A}_\gamma(\sigma^2) \equiv \arg\min_{\alpha \in [0,1]} \left\{ \tfrac{1}{2}(1-\alpha)^2\sigma^2 + \alpha^\gamma \right\}$$

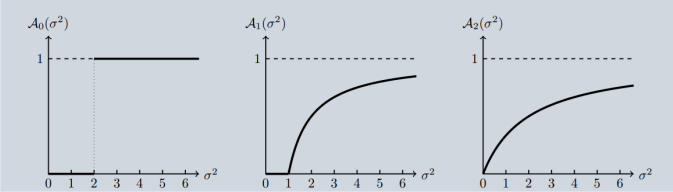


Result

- Low stakes (small σ^2) $\Rightarrow$ little or no attention.
- Attention **increases** with the stakes, and tends to 1 for high stakes.

Comparative statics: the curvature of the cost

Optimal attention $\mathcal{A}_\gamma(\sigma^2)$

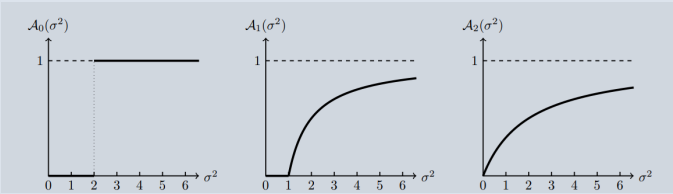


Three cases

- $\gamma < 1$ (e.g. a fixed cost, $\gamma = 0$): attention is **sparse** and **discontinuous** (all or nothing).

Comparative statics: the curvature of the cost

Optimal attention $\mathcal{A}_\gamma(\sigma^2)$

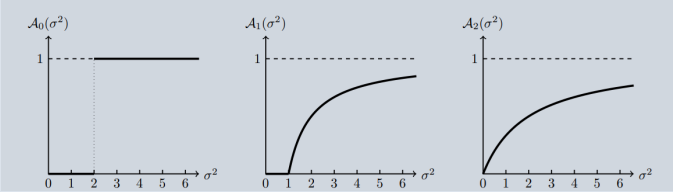


Three cases

- $\gamma < 1$ (e.g. a fixed cost, $\gamma = 0$): attention is **sparse** and **discontinuous** (all or nothing).
- $\gamma = 1$: attention is **sparse** and **continuous**: the only case with both.

Comparative statics: the curvature of the cost

Optimal attention $\mathcal{A}_\gamma(\sigma^2)$

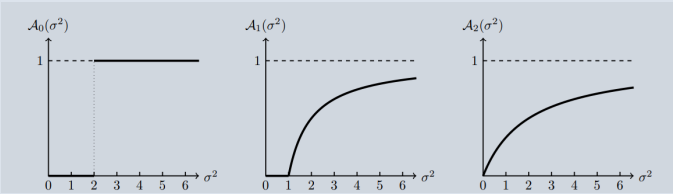


Three cases

- $\gamma < 1$ (e.g. a fixed cost, $\gamma = 0$): attention is **sparse** and **discontinuous** (all or nothing).
- $\gamma = 1$: attention is **sparse** and **continuous**: the only case with both.
- $\gamma > 1$: attention is **continuous** but **not sparse** ($\alpha > 0$ whenever $\sigma^2 > 0$).

Comparative statics: the curvature of the cost

Optimal attention $\mathcal{A}_\gamma(\sigma^2)$



Three cases

- $\gamma < 1$ (e.g. a fixed cost, $\gamma = 0$): attention is **sparse** and **discontinuous** (all or nothing).
- $\gamma = 1$: attention is **sparse** and **continuous**: the only case with both.
- $\gamma > 1$: attention is **continuous** but **not sparse** ($\alpha > 0$ whenever $\sigma^2 > 0$).

$\gamma = 1$ is the most interesting case: small stakes are ignored, large stakes get partial, smoothly rising attention.

Bonus: why is it a second-order approximation?

Setup

One attribute, x small, $a^r(x) = \arg\max_a u(a, x)$. Let $v(\alpha) \equiv \mathbb{E}\big[u\big(a^r(\alpha x), x\big)\big]$ be the value of attention α.

Bonus: why is it a second-order approximation?

Setup

One attribute, x small, $a^r(x) = \arg\max_a u(a, x)$. Let $v(\alpha) \equiv \mathbb{E}\big[u\big(a^r(\alpha x), x\big)\big]$ be the value of attention α.

Step 1: the action error

By a first-order Taylor expansion of a^r around 0,
$h \equiv a^r(\alpha x) - a^r(x) = a_x\,\alpha x - a_x\,x + o(x) = a_x(\alpha - 1)\,x + o(x)$.

Bonus: why is it a second-order approximation?

Setup

One attribute, x small, $a^r(x) = \arg\max_a u(a, x)$. Let $v(\alpha) \equiv \mathbb{E}\big[u\big(a^r(\alpha x), x\big)\big]$ be the value of attention α.

Step 1: the action error

By a first-order Taylor expansion of a^r around 0,
$h \equiv a^r(\alpha x) - a^r(x) = a_x\,\alpha x - a_x\,x + o(x) = a_x(\alpha - 1)\,x + o(x)$.

Step 2: the loss

Expand u in a around $a^r(x)$, using the first-order condition $u_a\big(a^r(x), x\big) = 0$:

$$v(\alpha) - v(1) = \mathbb{E}\Big[u\big(a^r(x) + h, x\big) - u\big(a^r(x), x\big)\Big] = \mathbb{E}\Big[\underbrace{u_a}_{=0}\,h + \tfrac{1}{2}u_{aa}h^2\Big] + o(\sigma^2)$$

$$= -\tfrac{1}{2}\,|u_{aa}|\,a_x^2\,(1-\alpha)^2\,\sigma^2 + o(\sigma^2).$$

The sparse max uses exactly this quadratic loss, with $u_{aa} < 0$ and a_x evaluated at the default.

Recap: Bayesian and sparse attention

Bayesian attention

- **Noisy signal:** the optimal action shrinks the signal towards the default, $\mathbb{E}[a \mid x] = \alpha x + (1 - \alpha)x^d$.

Sparse attention

Recap: Bayesian and sparse attention

Bayesian attention

- **Noisy signal:** the optimal action shrinks the signal towards the default, $\mathbb{E}[a \mid x] = \alpha x + (1 - \alpha)x^d$.
- α = **attention**: choosing a more precise signal = choosing a higher α, at a cost.

Sparse attention

Recap: Bayesian and sparse attention

Bayesian attention

- **Noisy signal:** the optimal action shrinks the signal towards the default, $\mathbb{E}[a \mid x] = \alpha x + (1-\alpha)x^d$.
- α = **attention**: choosing a more precise signal = choosing a higher α, at a cost.

Sparse attention

- **Sparsity model:** attention to each attribute solves a stakes-vs-cost trade-off, $\min \ \frac{1}{2}\beta_i^2\sigma_i^2(1-\alpha_i)^2 + \kappa\alpha_i^\gamma$.

Recap: Bayesian and sparse attention

Bayesian attention

- **Noisy signal:** the optimal action shrinks the signal towards the default, $\mathbb{E}[a \mid x] = \alpha x + (1-\alpha)x^d$.
- α = **attention**: choosing a more precise signal = choosing a higher α, at a cost.

Sparse attention

- **Sparsity model:** attention to each attribute solves a stakes-vs-cost trade-off, $\min \ \frac{1}{2}\beta_i^2\sigma_i^2(1-\alpha_i)^2 + \kappa\alpha_i^\gamma$.
- With $\gamma = 1$: low-stakes attributes are **ignored** ($\alpha_i = 0$), others are under-weighted.

Recap: Bayesian and sparse attention

Bayesian attention

- **Noisy signal:** the optimal action shrinks the signal towards the default, $\mathbb{E}[a \mid x] = \alpha x + (1-\alpha)x^d$.
- α = **attention**: choosing a more precise signal = choosing a higher α, at a cost.

Sparse attention

- **Sparsity model:** attention to each attribute solves a stakes-vs-cost trade-off, $\min \ \frac{1}{2}\beta_i^2\sigma_i^2(1-\alpha_i)^2 + \kappa\alpha_i^\gamma$.
- With $\gamma = 1$: low-stakes attributes are **ignored** ($\alpha_i = 0$), others are under-weighted.
- Sales taxes are small and vary little: the model predicts $\theta < 1$, as Chetty et al. (2009) find.

4

Applications and limits of sparsity

Sparse attention gives attention-based readings of many biases and of macroeconomic puzzles, but its foundations are debated.

Applications in behavioral economics (I)

Money illusion

The real price change is $\Delta q = \Delta p - \pi$ (Δp: nominal price change, π: inflation). A DM with attention α to inflation perceives

$$\Delta q^s = \Delta p - \alpha\,\pi.$$

Example

$\pi = 10\%$, $\Delta p = 8\%$, $\alpha = \frac{1}{2}$: perceived $+3\%$, true -2%.

Applications in behavioral economics (I)

Money illusion

The real price change is $\Delta q = \Delta p - \pi$ (Δp: nominal price change, π: inflation). A DM with attention α to inflation perceives

$$\Delta q^s = \Delta p - \alpha\,\pi.$$

Example

$\pi = 10\%$, $\Delta p = 8\%$, $\alpha = \frac{1}{2}$: perceived $+3\%$, true -2%.

Correlation neglect

Sparse attention to correlations: correlated signals are treated as independent.

Attention to the future

Distant payoffs get less attention, which produces present-biased choices, as in Laibson (1997) and O'Donoghue and Rabin (1999).

Attention-based readings of classic biases (Gabaix, 2019)

Probability weighting

In prospect theory (Kahneman and Tversky, 1979): with partial attention to probabilities, perceived $p^s = \alpha p + (1-\alpha)p^d$ (default p^d, e.g. $\frac{1}{2}$): **small probabilities are over-weighted, large ones under-weighted**.

Base-rate neglect

Base rates are abstract background information and receive little attention (Kahneman & Tversky, 1973).

Insensitivity to sample size

Sample size is a subtle attribute and is not attended to (Tversky and Kahneman, 1971).

Overconfidence

Too little attention to one's own uncertainty (Moore and Healy, 2008).

Application to macroeconomics

Sparse agents in a dynamic economy (Gabaix, *sparse dynamic programming*, JEEA):

Households

- react more to **near** than to distant shocks;
- start saving for retirement **too late**; spending falls in the years before retirement;
- keep a **too small buffer** of savings (partial attention to income risk);
- hardly react to the **interest rate** for small purchases;
- neglect the **hedging** motive in portfolio choice.

Policy

- **Ricardian equivalence** partly fails: a transfer today, taxed back later, raises consumption today.
- The **Lucas critique** loses (some or all of) its bite: small, temporary policy changes are not attended to.

Assessing the sparsity model

Pros

- Elegant and easy to apply: taxation, macro, finance.
- A possible **unified foundation** for many biases.
- Attention weights α_i can be **measured** in data and the model estimated (Gabaix, 2019).

Theoretical limitations

- Specific micro-foundations: Gaussian signals, quadratic utility.
- The cost function is not justified outside the Gaussian setting.
- The sparsity operator (setting small weights to 0) can look **ad hoc**.
- No axiomatic foundation: hard to relate to other models or derive sharp falsifiable predictions.

Is attention really chosen?

Attention can be guided, but not fully controlled

- **Emotional intrusions:** after a breakup, one cannot stop thinking about it during an exam.
- **Curiosity:** attention is drawn automatically to surprising or interesting stimuli.
- **Pain:** usually useful to attend to, but impossible to “turn off”.

Quote

“Why it is harder to pay attention to tax changes than to a funny story remains unmodeled.”

Sparsity models **top-down** attention (chosen given the stakes). Salience theory models **bottom-up** attention: what *stands out* in the choice context captures it.

5

Salience theory

Attention is not chosen but drawn by contrasts: the attributes that stand out in the choice set get more weight.

Salience and consumer choice

Model

Bordalo et al. (2012), Bordalo et al. (2013), Bordalo et al. (2016):

- A good has attributes $x = (x_1, \ldots, x_n)$, e.g. (quality, price).
- **Rational DM:** $U(x) = \sum_i b_i\, x_i$, where b_i is the objective importance of attribute i (for a price, $b_i < 0$).
- **Salient thinker:** $U^s(x) = \sum_i \alpha_i(x)\, b_i\, x_i$, where the weights $\alpha_i(x)$ **distort** the b_i.

Salience and consumer choice

Model

Bordalo et al. (2012), Bordalo et al. (2013), Bordalo et al. (2016):

- A good has attributes $x = (x_1, \ldots, x_n)$, e.g. (quality, price).
- **Rational DM:** $U(x) = \sum_i b_i\, x_i$, where b_i is the objective importance of attribute i (for a price, $b_i < 0$).
- **Salient thinker:** $U^s(x) = \sum_i \alpha_i(x)\, b_i\, x_i$, where the weights $\alpha_i(x)$ **distort** the b_i.

Key difference with sparsity: $\alpha_i(x)$ is not chosen; it depends on how much attribute i of good x **stands out** relative to the other goods in the choice set.

The salience function

Reference point

For attribute i, $\hat{x}_i$ = the average of x_i over the goods in the choice set.

The salience function

Reference point

For attribute i, $\hat{x}_i$ = the average of x_i over the goods in the choice set.

Salience function

$\sigma(x_i, \hat{x}_i)$: how much x_i stands out from $\hat{x}_i$. It is symmetric and satisfies

1. **Ordering:** if $[\min, \max]$ of $\{x, y\}$ is contained in that of $\{x', y'\}$, then $\sigma(x, y) < \sigma(x', y')$: a bigger contrast is more salient.
2. **Diminishing sensitivity:** for $x, y > 0$ and $\epsilon > 0$, $\sigma(x + \epsilon, y + \epsilon) < \sigma(x, y)$: the same contrast is less salient at higher levels.

The salience function

Reference point

For attribute i, $\hat{x}_i$ = the average of x_i over the goods in the choice set.

Salience function

$\sigma(x_i, \hat{x}_i)$: how much x_i stands out from $\hat{x}_i$. It is symmetric and satisfies

1. **Ordering:** if $[\min, \max]$ of $\{x, y\}$ is contained in that of $\{x', y'\}$, then $\sigma(x, y) < \sigma(x', y')$: a bigger contrast is more salient.
2. **Diminishing sensitivity:** for $x, y > 0$ and $\epsilon > 0$, $\sigma(x + \epsilon, y + \epsilon) < \sigma(x, y)$: the same contrast is less salient at higher levels.

Example

For $\theta > 0$,

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

With $\theta = 5$: $\sigma(20, 15) = \frac{5}{40} = \frac{1}{8}$ but $\sigma(60, 55) = \frac{5}{120} = \frac{1}{24}$.

From salience to decision weights

Salience ranking

For each good x, rank its attributes by $\sigma(x_i, \hat{x}_i)$: $r_i(x) = 1$ for the most salient, etc.

From salience to decision weights

Salience ranking

For each good x, rank its attributes by $\sigma(x_i, \hat{x}_i)$: $r_i(x) = 1$ for the most salient, etc.

Decision weights

With $\delta \in (0, 1]$,

$$\alpha_i(x) = \frac{\delta^{r_i(x)}}{\sum_j \delta^{r_j(x)}}.$$

- $\delta = 1$: equal weights, the rational ranking of goods.
- $\delta < 1$: more salient attributes get **more weight**. Ties: equal ranks, equal weights.

From salience to decision weights

Salience ranking

For each good x, rank its attributes by $\sigma(x_i, \hat{x}_i)$: $r_i(x) = 1$ for the most salient, etc.

Decision weights

With $\delta \in (0, 1]$,

$$\alpha_i(x) = \frac{\delta^{r_i(x)}}{\sum_j \delta^{r_j(x)}}.$$

- $\delta = 1$: equal weights, the rational ranking of goods.
- $\delta < 1$: more salient attributes get **more weight**. Ties: equal ranks, equal weights.

Two attributes

The salient one gets $\frac{1}{1+\delta}$, the other $\frac{\delta}{1+\delta}$.
E.g. $\delta = \frac{1}{2}$: weights $\frac{2}{3}$ and $\frac{1}{3}$.

Example: wine in the supermarket and in the restaurant

Two wines

- Two wines: high quality H with $q_H = 30$, low quality L with $q_L = 20$.
- Supermarket prices: $p_H = 20$, $p_L = 10$. Restaurant prices: $p_H = 60$, $p_L = 50$.
- Rational utility: $U = q - p$ (so $b_q = 1$, $b_p = -1$).

Example: wine in the supermarket and in the restaurant

Two wines

- Two wines: high quality H with $q_H = 30$, low quality L with $q_L = 20$.
- Supermarket prices: $p_H = 20$, $p_L = 10$. Restaurant prices: $p_H = 60$, $p_L = 50$.
- Rational utility: $U = q - p$ (so $b_q = 1$, $b_p = -1$).

Rational DM: indifferent in both places

$$U_H = 30 - 20 = 10 = 20 - 10 = U_L,$$
$$U_H = 30 - 60 = -30 = 20 - 50 = U_L.$$

Example: wine in the supermarket and in the restaurant

Two wines

- Two wines: high quality H with $q_H = 30$, low quality L with $q_L = 20$.
- Supermarket prices: $p_H = 20$, $p_L = 10$. Restaurant prices: $p_H = 60$, $p_L = 50$.
- Rational utility: $U = q - p$ (so $b_q = 1$, $b_p = -1$).

Rational DM: indifferent in both places

$$U_H = 30 - 20 = 10 = 20 - 10 = U_L,$$
$$U_H = 30 - 60 = -30 = 20 - 50 = U_L.$$

Reference goods (averages)

Supermarket $(\hat{q}, \hat{p}) = (25, 15)$, restaurant $(\hat{q}, \hat{p}) = (25, 55)$.

Which attribute is salient? ($\theta = 5$)

Salience of quality and price

	Quality $\sigma(q, \hat{q})$	**Price** $\sigma(p, \hat{p})$	**Salient**
Supermarket, $(\hat{q}, \hat{p}) = (25, 15)$			
H	$\frac{5}{30+25+5} = \frac{1}{12}$	$\frac{5}{20+15+5} = \frac{1}{8}$	price
L	$\frac{5}{20+25+5} = \frac{1}{10}$	$\frac{5}{10+15+5} = \frac{1}{6}$	price
Restaurant, $(\hat{q}, \hat{p}) = (25, 55)$			
H	$\frac{1}{12}$	$\frac{5}{60+55+5} = \frac{1}{24}$	quality
L	$\frac{1}{10}$	$\frac{5}{50+55+5} = \frac{1}{22}$	quality

In the supermarket **price** stands out; in the restaurant the same €10 gap is drowned in high prices, so **quality** stands out (diminishing sensitivity).

Choices of the salient thinker

Decision weights

Salient attribute $\frac{1}{1+\delta}$, other $\frac{\delta}{1+\delta}$. Take $\delta = \frac{1}{2}$ (weights $\frac{2}{3}, \frac{1}{3}$).

Supermarket (price salient)

- $U_H^s = \frac{1}{3} \cdot 30 - \frac{2}{3} \cdot 20 = -\frac{10}{3}$
- $U_L^s = \frac{1}{3} \cdot 20 - \frac{2}{3} \cdot 10 = 0$
- $\Rightarrow$ buys the **cheap** wine L.

Choices of the salient thinker

Decision weights

Salient attribute $\frac{1}{1+\delta}$, other $\frac{\delta}{1+\delta}$. Take $\delta = \frac{1}{2}$ (weights $\frac{2}{3}, \frac{1}{3}$).

Supermarket (price salient)

- $U_H^s = \frac{1}{3} \cdot 30 - \frac{2}{3} \cdot 20 = -\frac{10}{3}$
- $U_L^s = \frac{1}{3} \cdot 20 - \frac{2}{3} \cdot 10 = 0$
- $\Rightarrow$ buys the **cheap** wine L.

Restaurant (quality salient)

- $U_H^s = \frac{2}{3} \cdot 30 - \frac{1}{3} \cdot 60 = 0$
- $U_L^s = \frac{2}{3} \cdot 20 - \frac{1}{3} \cdot 50 = -\frac{10}{3}$
- $\Rightarrow$ orders the **good** wine H.

Choices of the salient thinker

Decision weights

Salient attribute $\frac{1}{1+\delta}$, other $\frac{\delta}{1+\delta}$. Take $\delta = \frac{1}{2}$ (weights $\frac{2}{3}, \frac{1}{3}$).

Supermarket (price salient)

- $U_H^s = \frac{1}{3} \cdot 30 - \frac{2}{3} \cdot 20 = -\frac{10}{3}$
- $U_L^s = \frac{1}{3} \cdot 20 - \frac{2}{3} \cdot 10 = 0$
- $\Rightarrow$ buys the **cheap** wine L.

Restaurant (quality salient)

- $U_H^s = \frac{2}{3} \cdot 30 - \frac{1}{3} \cdot 60 = 0$
- $U_L^s = \frac{2}{3} \cdot 20 - \frac{1}{3} \cdot 50 = -\frac{10}{3}$
- $\Rightarrow$ orders the **good** wine H.

For any $\delta < 1$

$U_H^s - U_L^s = \frac{10(\delta-1)}{1+\delta} < 0$ in the supermarket and $\frac{10(1-\delta)}{1+\delta} > 0$ in the restaurant.

Choices of the salient thinker

Decision weights

Salient attribute $\frac{1}{1+\delta}$, other $\frac{\delta}{1+\delta}$. Take $\delta = \frac{1}{2}$ (weights $\frac{2}{3}, \frac{1}{3}$).

Supermarket (price salient)

- $U_H^s = \frac{1}{3} \cdot 30 - \frac{2}{3} \cdot 20 = -\frac{10}{3}$
- $U_L^s = \frac{1}{3} \cdot 20 - \frac{2}{3} \cdot 10 = 0$
- $\Rightarrow$ buys the **cheap** wine L.

Restaurant (quality salient)

- $U_H^s = \frac{2}{3} \cdot 30 - \frac{1}{3} \cdot 60 = 0$
- $U_L^s = \frac{2}{3} \cdot 20 - \frac{1}{3} \cdot 50 = -\frac{10}{3}$
- $\Rightarrow$ orders the **good** wine H.

For any $\delta < 1$

$U_H^s - U_L^s = \frac{10(\delta-1)}{1+\delta} < 0$ in the supermarket and $\frac{10(1-\delta)}{1+\delta} > 0$ in the restaurant.

Same goods, same price difference, opposite choices: the **context** shifts attention. Experimental support: Dertwinkel-Kalt, Köhler, Lange and Wenzel (2017).

6

Wrap-up

Three models of limited attention to attributes, and how they differ on what drives attention.

Recap: three views of limited attention to attributes

Three models compared

	Bayesian signal	Sparsity	Salience
Attention weight	$\alpha = \frac{\sigma_x^2}{\sigma_x^2+\sigma_\varepsilon^2}$	α_i minimizes stakes loss + cost	$\alpha_i(x)$ from salience rank
Attention is…	a by-product of noise	*chosen* (top-down)	*captured* (bottom-up)
Driven by	signal precision	stakes $\beta_i^2\sigma_i^2$, cost κ, γ	contrasts in the choice set
Key prediction	anchoring on x^d	small stakes ignored	context effects
Fits Chetty et al.?	yes ($\theta < 1$)	yes (small tax stakes)	yes (non-posted tax)

All three keep **utility maximization** on *perceived* attributes; they differ on **what drives** the perception.

Review questions

Concepts

1. How does salience differ from **consideration** (Lectures 5–6)?
2. What do salience models and rational inattention have in common? How do they differ?
3. Is attention something people fully control? Why does this matter for modeling?

Applications

4. Explain the idea of a "**painless tax**". Why might governments prefer such taxes?
5. In Experiment 1, what are the treatment and control groups, and why does a triple difference identify the effect of tax salience?
6. Why do low-stakes decisions receive little or no attention in the sparsity model? How does γ matter?

Exercises

1. Easy

Noisy signal with $x^d = 20$, $\sigma_x^2 = 9$, $\sigma_\varepsilon^2 = 3$. Compute α, the action after a signal $s = 28$, and the average action when the true value is $x = 24$.

2. Intermediate

Sparsity with two attributes, stakes $\beta_1^2\sigma_1^2 = 5$ (price) and $\beta_2^2\sigma_2^2 = 0.8$ (tax), $\kappa = 1$. Find $\alpha^\star$ for $\gamma = 1$. For which κ does the DM attend to the tax? Redo for $\gamma = 2$.

3. Intermediate

In the wine example, let both prices carry a mark-up $m \geq 0$: $p_L = 10 + m$, $p_H = 20 + m$ (and $\theta = 5$). For which m is quality salient for both wines? Which wine does a salient thinker with $\delta < 1$ choose, depending on m?

4. Discussion

Can firms or governments **manipulate** salience? Give two examples, and say whether sparsity or salience theory explains them better.

Final take-aways

1. Consumers under-react to **non-salient** price components: posting taxes lowers demand by about 8%, and sales taxes move demand much less than posted excise taxes.
2. In a **Gaussian signal** model the optimal action anchors on the default and adjusts **insufficiently**; the weight on the signal is an attention parameter.
3. In the **sparsity** model attention follows the **stakes**: small stakes are ignored (with $\gamma \leq 1$).
4. In **salience theory** attention is **captured by contrasts**: the same goods are ranked differently in different contexts.

References I

Atkinson, A. B. and Stiglitz, J. E. (1976). The design of tax structure: direct versus indirect taxation. *Journal of Public Economics*, 6(1–2):55–75.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3):1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2016). Competition for attention. *The Review of Economic Studies*, 83(2):481–513.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American economic review*, 99(4):1145–77.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gabaix, X. (2019). Behavioral inattention. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 261–343. Elsevier.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, 115(2):502.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

Ramsey, F. P. (1927). A contribution to the theory of taxation. *The Economic Journal*, 37(145):47–61.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2):105–110.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.