BEHAVIORAL ECONOMICS · LECTURE 9

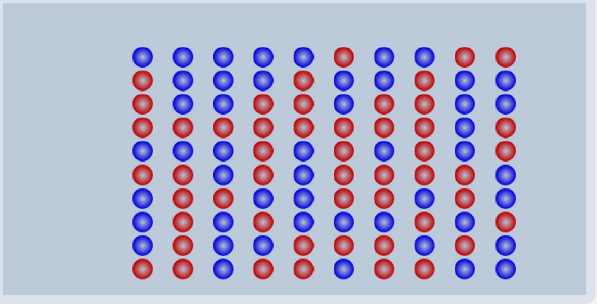
Bounded rationality IV: rational inattention

Niels Boissonnet – Bielefeld University

Master Economics – Winter 2026/27

Before we start: how hard would you look?

The screen



The choice

The screen shows either **47** or **53** red dots. You choose an act:

Act	47 red dots	53 red dots
a	20	0
b	0	10

Question

How much effort would you spend counting the dots? Would your answer change if the payoffs were 2 and 1? Or 2000 and 1000?

Learning objectives

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

- Represent **information structures** and update beliefs using Bayes' rule.
- Calculate the **value of information**.
- Measure information using **entropy**, **mutual information**, and **KL divergence**.
- Set up and interpret the **rational-inattention** optimization problem.
- Apply the model to discrimination and other economic settings.

Roadmap

Motivation and information structures

The value of information

Entropy and information costs

Rational inattention and applications

Exercise workshop

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

Last week: limited attention to the **attributes** of options (sparsity, salience). Today: the **cost of acquiring information** about the state of the world, with any prior, any payoffs, and a general cost of information.

1

Motivation and information structures

A signal moves beliefs from the prior to a posterior; information matters only through these posteriors.

Why rational inattention?

Costly information

- Information improves decisions, but acquiring and processing it is costly.
- Costs may be monetary, cognitive, temporal, or organizational.
- The decision-maker chooses how much and what type of information to acquire.

Why rational inattention?

Costly information

- Information improves decisions, but acquiring and processing it is costly.
- Costs may be monetary, cognitive, temporal, or organizational.
- The decision-maker chooses how much and what type of information to acquire.

Rational inattention is not ignorance: it is optimal inattention when information is costly.

The choice problem

Timing

1. Nature chooses a state $\omega \in \Omega$.
2. The decision-maker has prior μ.
3. She chooses an information structure π.
4. A signal $\gamma \in \Gamma$ is observed.
5. She chooses an action $a \in A$.

Information affects actions through **posterior beliefs**.

States, actions, and payoffs

Setup

- States: $\Omega = \{\omega_1, \ldots, \omega_m\}$.
- Prior: $\mu_\omega = \Pr(\omega)$.
- Actions: A.
- Payoff: a_ω is the payoff from action a in state ω.

Running example

Demand is good, medium, or bad (g, m, b); a firm chooses a high, average, or low price (h, a, l).

Information structures

Definition

An **information structure** is a mapping $\pi : \Omega \to \Delta(\Gamma)$ assigning $\pi_\omega(\gamma) = \Pr(\gamma \mid \omega)$.

Properties

- Rows sum to one: $\sum_\gamma \pi_\omega(\gamma) = 1$ for every ω.
- The signal is generated conditional on the state.

Bayes plausibility / martingale property

Bayes plausibility

$$\mu(\omega) = \sum_{\gamma \in \Gamma} \pi(\gamma) \Pr(\omega \mid \gamma), \qquad \pi(\gamma) \equiv \sum_{\omega} \pi_{\omega}(\gamma)\, \mu(\omega).$$

- $\pi(\gamma)$ is the unconditional probability of signal γ.
- The information structure must be consistent with Bayesian updating.

Average posterior beliefs equal prior beliefs.

Example: Consumer Confidence Index

The CCI as an information structure

	High CCI γ_1	Low CCI γ_2
Good demand g	1	0
Medium demand m	.5	.5
Bad demand b	0	1

Posteriors

- With $\mu_g = \mu_m = \mu_b = 1/3$, each signal has probability $1/2$.
- $\Pr(g \mid \gamma_1) = \dfrac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = 2/3$ and, symmetrically, $\Pr(b \mid \gamma_2) = 2/3$.

Exercise 1: Bayes' rule

Tasks

1. Compute $\Pr(\gamma_1)$ and $\Pr(\gamma_2)$.
2. Compute all posterior probabilities.
3. Verify Bayes plausibility.
4. Is the signal informative? Explain economically.

Exercise 1: Bayes' rule

Tasks

1. Compute $\Pr(\gamma_1)$ and $\Pr(\gamma_2)$.
2. Compute all posterior probabilities.
3. Verify Bayes plausibility.
4. Is the signal informative? Explain economically.

Solution

$\Pr(\gamma_1) = \Pr(\gamma_2) = 1/2$; posteriors over (g, m, b) are $(2/3, 1/3, 0)$ and $(0, 1/3, 2/3)$.

Recap

Information structures

- A signal is a probabilistic mapping from states to observations.
- Bayes' rule maps signals into posterior beliefs.
- Bayes plausibility says posterior beliefs average to the prior.

Information is useful only insofar as it changes optimal actions.

Take-away: information structures

States $\rightarrow$ signals $\rightarrow$ posteriors $\rightarrow$ actions

2

The value of information

Information is worth the gain in expected payoff it brings, which is never negative and positive only if it changes what you do.

Payoff matrix

Payoffs of the firm

	High price h	Average a	Low l
Good g	10	3	1
Medium m	1	2	1
Bad b	-10	-3	-1

Question

What price should be chosen with no information?

No information

With the uniform prior

$$\mathbb{E}[h_\omega] = \tfrac{1}{3}(10+1-10) = \tfrac{1}{3}, \qquad \mathbb{E}[a_\omega] = \tfrac{1}{3}(3+2-3) = \tfrac{2}{3}, \qquad \mathbb{E}[l_\omega] = \tfrac{1}{3}(1+1-1) = \tfrac{1}{3}.$$

Without information, choose the action with the highest prior expected payoff: a, with expected payoff $\frac{2}{3}$.

After observing a signal

After γ_1 (posterior $(\frac{2}{3}, \frac{1}{3}, 0)$)

$$\mathbb{E}[h \mid \gamma_1] = 7$$
$$\mathbb{E}[a \mid \gamma_1] = \frac{8}{3}$$
$$\mathbb{E}[l \mid \gamma_1] = 1$$

Choose h.

After γ_2 (posterior $(0, \frac{1}{3}, \frac{2}{3})$)

$$\mathbb{E}[h \mid \gamma_2] = -\frac{19}{3}$$
$$\mathbb{E}[a \mid \gamma_2] = -\frac{4}{3}$$
$$\mathbb{E}[l \mid \gamma_2] = -\frac{1}{3}$$

Choose l.

Value of information

Definition

The **value of information** is

$$V(\pi) = \mathbb{E}_{\pi}\Big[\max_{a \in A} \mathbb{E}[a_\omega \mid \gamma]\Big] - \max_{a \in A} \mathbb{E}[a_\omega].$$

Value of information

Definition

The **value of information** is

$$V(\pi) = \mathbb{E}_\pi\Big[\max_{a\in A}\mathbb{E}[a_\omega \mid \gamma]\Big] - \max_{a\in A}\mathbb{E}[a_\omega].$$

In the example

$$V(\pi) = \underbrace{\tfrac{1}{2}\cdot 7 + \tfrac{1}{2}\cdot\big(-\tfrac{1}{3}\big)}_{=10/3} - \tfrac{2}{3} = \tfrac{8}{3}.$$

The firm would pay up to $\frac{8}{3}$ for the Consumer Confidence Index.

Why information cannot hurt

Result

$$V(\pi) = \mathbb{E}\Big[\max_a \mathbb{E}[a_\omega \mid \gamma]\Big] - \max_a \mathbb{E}[a_\omega] \geq 0.$$

Why

The decision-maker can always ignore the signal.

When is it zero?

Value is zero if the signal is independent of the state or payoffs are constant.

Exercise 2: buy the information?

Data

- No-information payoff: 2.
- Noisy report: payoff 5, cost 1.
- Perfect report: payoff 8, cost 4.

Tasks

Compute net payoffs; choose the optimal information strategy; repeat if the perfect-report cost falls to 2.

Exercise 2: buy the information?

Data

- No-information payoff: 2.
- Noisy report: payoff 5, cost 1.
- Perfect report: payoff 8, cost 4.

Tasks

Compute net payoffs; choose the optimal information strategy; repeat if the perfect-report cost falls to 2.

Solution

Net payoffs are $2, 4, 4$; both information strategies are optimal initially, while perfect information is optimal when its cost is 2 (net payoff 6).

Recap

Value of information

- Information changes beliefs and potentially actions.
- The value of information is the gain in expected payoff.
- $V(\pi) \geq 0$ because information can be ignored.

Positive value requires decision-relevant information.

Take-away: value of information

Information is valuable when it changes what you do.

3

Entropy and information costs

Entropy measures uncertainty; mutual information measures how much a signal moves beliefs, and prices it.

Why measure information?

Three measures

- **Shannon entropy** measures uncertainty.
- **Mutual information** measures the expected reduction in uncertainty.
- **KL divergence** measures how far posterior beliefs move from the prior.

Information costs need a quantitative measure.

Shannon entropy

Definition

$$H(X) = -\mathbb{E}[\ln p(X)] = -\sum_i p_i \ln p_i.$$

- Entropy is high when probabilities are balanced.
- Entropy is zero when the state is known with certainty.
- With logarithm base 2, units are bits.

Example

A fair coin, $H = \ln 2 \approx 0.69$ nats $= 1$ bit; a coin with $(.9, .1)$: $H \approx 0.33$ nats ≈ 0.47 bits.

Entropy as yes/no questions

Example: $(.5, .3, .2)$

Ask first whether $X = x_1$. If no, distinguish x_2 from x_3.

$$\text{Expected questions} = 1(.5) + 2(.3) + 2(.2) = 1.5.$$

Bound

$H(X)/\ln 2 \leq$ average number of questions (bits) with the best strategy $\leq H(X)/\ln 2 + 1$.

Entropy as yes/no questions

Example: $(.5, .3, .2)$

Ask first whether $X = x_1$. If no, distinguish x_2 from x_3.

$$\text{Expected questions} = 1(.5) + 2(.3) + 2(.2) = 1.5.$$

Bound

$H(X)/\ln 2 \leq$ average number of questions (bits) with the best strategy $\leq H(X)/\ln 2 + 1$.

Check: $H(X)/\ln 2 \approx 1.49$, so $1.49 \leq 1.5 \leq 2.49$. ✓

Exercise 3: entropy

Tasks

1. Compute entropy for $(.5, .5)$.
2. Compute entropy for $(.9, .1)$.
3. Compute entropy for $(1/3, 1/3, 1/3)$.
4. Rank the distributions by uncertainty.

Exercise 3: entropy

Tasks

1. Compute entropy for $(.5, .5)$.
2. Compute entropy for $(.9, .1)$.
3. Compute entropy for $(1/3, 1/3, 1/3)$.
4. Rank the distributions by uncertainty.

Solution

$\ln 2 \approx 0.69$, $-.9 \ln .9 - .1 \ln .1 \approx 0.33$, and $\ln 3 \approx 1.10$ respectively. Ranking: $(1/3, 1/3, 1/3) \succ (.5, .5) \succ (.9, .1)$.

Mutual information

Definition

$$I(X;Y) = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}.$$

$$I(X;Y) = H(X) - \mathbb{E}_Y[H(X \mid Y)].$$

Interpretation

- Expected reduction in uncertainty.
- $I(X;Y) = 0$ iff X and Y are independent.

KL divergence connection

KL divergence

$$D_{\mathrm{KL}}(p_y \| p) = \sum_x p_y(x) \ln \frac{p_y(x)}{p(x)}.$$

$$I(X;Y) = \mathbb{E}_Y[D_{\mathrm{KL}}(p_Y \| p)].$$

p: prior (marginal) distribution of X; p_y: posterior after observing $Y = y$.

Mutual information is the expected distance between posterior and prior beliefs.

Shannon information cost

Shannon cost

For a distribution Q over posteriors ρ (see the next section) and prior μ, the **Shannon cost** is (Sims, 2003)

$$\begin{aligned} K(Q,\mu) &= \kappa\, \mathbb{E}_Q[D_{\mathrm{KL}}(\rho\|\mu)] \\ &= \kappa\Big[-\sum_{\omega}\mu_\omega \ln \mu_\omega + \sum_{\rho} Q(\rho)\sum_{\omega}\rho(\omega)\ln\rho(\omega)\Big], \end{aligned}$$

i.e. $\kappa\times$(entropy of the prior $-$ expected entropy of the posteriors), with $\kappa > 0$ the unit cost of information.

Exercise 4: compare signals

Two signals

	$\Pr(R \mid H)$	$\Pr(R \mid L)$
Signal A	.9	.1
Signal B	.6	.4

Tasks

1. Compute posteriors under a uniform prior.
2. Which signal has greater mutual information?
3. How does κ affect the cost?

Recap

Measuring information

- Entropy measures uncertainty.
- Mutual information measures expected learning.
- KL divergence measures posterior-prior distance.

Shannon costs are proportional to mutual information.

Take-away: information costs

The model prices information by how much beliefs move.

4

Rational inattention and applications

Choosing information optimally yields logit choice, and explains why attention, and hence discrimination, depends on the prior.

Choosing a distribution over posteriors

Bayes plausibility in terms of posteriors

$$\mu(\omega) = \sum_{\rho} Q(\rho)\rho(\omega).$$

Information structures as distributions over posteriors

An information structure can be represented by a distribution over posterior beliefs: $Q(\rho) = \pi(\gamma)$ when $\rho = \Pr(\cdot \mid \gamma)$.

Feasible set

Let $\mathcal{Q}$ be the set of distributions satisfying Bayes plausibility.

The rational-inattention problem

The problem

$$\max_{Q \in \mathcal{Q}} \; G(Q, A) - K_\mu(Q), \qquad G(Q, A) = \mathbb{E}_Q\Big[\max_{a \in A} \mathbb{E}_\rho[a_\omega]\Big].$$

- **Benefit:** expected payoff after information.
- **Cost:** information-processing cost.

The value of information from Section 2, net of the cost of information from Section 3.

Two states and two actions

Merging posteriors

- If two posteriors induce the same action, they are behaviorally equivalent.
- Keeping both is wasteful under strictly concave entropy costs: merging them into their average saves cost without changing behavior.

Hence an optimum needs at most one posterior per action.

Key equality

Optimality condition (Shannon cost)

For any two actions a, b chosen with positive probability, at the optimum:

$$\frac{\rho^a(\omega)}{\rho^b(\omega)} = \exp\left(\frac{a_\omega - b_\omega}{\kappa}\right) \qquad \text{for every } \omega.$$

Implications

- Large utility differences imply stronger posterior separation.
- Higher κ smooths choices.
- Lower κ produces more state-contingent behavior.

Foundation of logit (Matějka and McKay, 2015)

Choice probabilities

$$P(a \mid \omega) = \frac{P(a)\exp(a_\omega/\kappa)}{\sum_{c\in A} P(c)\exp(c_\omega/\kappa)}.$$

- $P(a)$: unconditional probability of choosing a; with equal $P(a)$, this is the multinomial **logit** of Lecture 3.
- As $\kappa \to 0$, choose the best action in each state.
- As $\kappa \to \infty$, choose the best action ex ante.

Foundation of logit (Matějka and McKay, 2015)

Choice probabilities

$$P(a \mid \omega) = \frac{P(a)\exp(a_\omega/\kappa)}{\sum_{c\in A} P(c)\exp(c_\omega/\kappa)}.$$

- $P(a)$: unconditional probability of choosing a; with equal $P(a)$, this is the multinomial **logit** of Lecture 3.
- As $\kappa \to 0$, choose the best action in each state.
- As $\kappa \to \infty$, choose the best action ex ante.

Example

Two equally likely states, $a = (1, 0)$, $b = (0, 1)$, so by symmetry $P(a) = P(b) = \frac{1}{2}$:

$$P(a \mid \omega_1) = \frac{e^{1/\kappa}}{e^{1/\kappa} + 1},$$

i.e. 0.73 for $\kappa = 1$ and 0.88 for $\kappa = \frac{1}{2}$.

Exercise 5: comparative statics

Tasks

1. Explain what happens when κ decreases.
2. What happens when mistakes become more costly?
3. Why does the model predict probabilistic choice?
4. Relate the result to logit choice probabilities.

Application: discrimination (Bartoš et al., 2016)

The recruiter

- A recruiter observes an applicant's group g and chooses how precisely to inspect the CV.
- Quality: $q \sim N(q_g, \sigma^2)$; hiring gives payoff q, not hiring 0.

The signal

- Signal: $y = q + \epsilon$, $\epsilon \sim N(0, \sigma_\epsilon^2)$.
- Lower signal variance means more precise and more costly information.

Posterior quality and hiring

Posterior mean

$$q' = \alpha y + (1-\alpha) q_g, \qquad \alpha = \frac{\sigma^2}{\sigma^2 + \sigma_\epsilon^2}.$$

- Hire iff $q' > 0$.
- More precise signals put more weight on evidence and less on group-based priors.

Posterior quality and hiring

Posterior mean

$$q' = \alpha y + (1-\alpha) q_g, \qquad \alpha = \frac{\sigma^2}{\sigma^2 + \sigma_\epsilon^2}.$$

- Hire iff $q' > 0$.
- More precise signals put more weight on evidence and less on group-based priors.

Example

$q_g = -1$, $\sigma^2 = \sigma_\epsilon^2 = 1$, so $\alpha = \frac{1}{2}$: hire iff $\frac{1}{2}y - \frac{1}{2} > 0$, i.e. iff $y > 1$.

Cherry picking versus lemon dropping

Cherry picking ($q_g < 0$ for both groups)

The average candidate is not hired; only sufficiently good signals lead to hiring (e.g. jobs).

Lemon dropping ($q_g > 0$ for both groups)

The average candidate is hired; only sufficiently bad signals lead to rejection (e.g. rental housing).

Attention is more valuable near the hiring threshold.

In both markets rational inattention hurts the group with the lower prior mean (Bartoš et al., 2016).

Exercise 6: discrimination and attention

Tasks

1. Which group receives more attention in cherry-picking markets?
2. Which group receives more attention in lemon-dropping markets?
3. Why can rational inattention hurt the disadvantaged group in both cases?
4. What changes if signal precision must be equal across groups?

Other applications

Macroeconomics and finance

- Consumption and savings: gradual responses to income news (Sims, 2003).
- Portfolio choice: selective processing of asset information, e.g. home bias (Van Nieuwerburgh and Veldkamp, 2009).

Firms and search

- Pricing: firms acquire information about demand (Matějka and McKay, 2015).
- Search and attention: decisions depend on costly processing.

Recap

Rational inattention

- Rational inattention balances payoff gains and information costs.
- Shannon costs generate smooth, logit-like behavior (Lecture 3).
- Attention focuses on states where information changes actions.

Applications

Applications include discrimination, macroeconomics, and industrial organization.

5

Exercise workshop

Five exercises, from Bayes' rule to optimal attention, to practise the whole lecture in groups.

Workshop instructions

How to work

- Work in groups of 3–4.
- Write down assumptions before calculating.
- Interpret every numerical result economically.

Prepare a 3-minute explanation of one exercise.

Workshop A (Easy): information structures

Setup

Given three states with prior $(1/3, 1/3, 1/3)$ and the CCI matrix from Section 1:

Tasks

1. Derive signal probabilities.
2. Derive posteriors.
3. Verify Bayes plausibility.
4. Identify which actions are optimal after each signal.

Workshop B (Easy): value of information

Setup

Using the price payoff matrix:

Tasks

1. Compute the no-information benchmark.
2. Compute expected payoff with CCI.
3. Calculate the value of information.
4. Find the maximum price the firm should pay for CCI.

Workshop C (Intermediate): entropy and mutual information

Setup

For a binary state with prior $1/2$ and a signal satisfying $\Pr(R \mid H) = .8$, $\Pr(R \mid L) = .3$:

Tasks

1. Compute $H(\Omega)$.
2. Compute posterior entropy after each signal.
3. Compute $I(\Omega; \Gamma)$.
4. Compute the Shannon cost for $\kappa = 2$.

Workshop D (Advanced): optimal attention

Setup

Suppose two actions have state-contingent payoffs $a = (4, 0)$ and $b = (0, 3)$.

Tasks

1. Write the optimal posterior likelihood-ratio condition.
2. Explain how it changes with κ.
3. Discuss the limiting cases $\kappa \to 0$ and $\kappa \to \infty$.
4. Interpret the result as a model of probabilistic choice.

Workshop E (Discussion): application

Choose one application

- discrimination in hiring;
- consumption and savings;
- firm pricing;
- portfolio choice.

Task

Identify states, actions, signals, benefits, costs, and one testable prediction.

Workshop solution strategy

Beliefs

1. Define the state space and prior.
2. Write the signal structure or posterior distribution.
3. Apply Bayes' rule.

Decisions

4. Compute action values.
5. Subtract information costs.
6. Interpret comparative statics.

Closing discussion

The model

- What does the model explain well?
- Which assumptions are restrictive?

Beyond Shannon

- When is Shannon information a poor cost function?
- How could one model biased or motivated attention?

Final take-aways

1. **Information structures** determine posterior beliefs.
2. **Bayes plausibility** links priors and posteriors.
3. Information has non-negative value but positive acquisition cost.
4. **Mutual information** is a natural measure of learning.
5. **Rational inattention** predicts selective, probabilistic, and costly attention.

References I

Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–75.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Van Nieuwerburgh, S. and Veldkamp, L. (2009). Information immobility and the home bias puzzle. *The Journal of Finance*, 64(3):1187–1215.