

Attention and Perception

Lecture 3: Rational Inattention (General Costs)¹

Daniel Martin

danielmartin@ucsb.edu

¹A thanks to Mark Dean, whose slides these ones are based on!

Different approaches to limited attention in economics

1. Limited attention from *mental gaps*

- ▶ Ignore low-salience information (Chetty, Looney & Kroft 2009)
- ▶ Ignore readily available information (DellaVigna & Pollet 2009, Lacetera, Pope & Sydnor 2012)

2. Limited attention from *information frictions*

- ▶ Consideration sets (Caplin, Dean & Martin 2011, Manzini & Mariotti 2014)
- ▶ Rational inattention (Sims 2003, Matějka & McKay 2015)

Modeling Goal and Solution

- ▶ Return to general approach to modeling perception in Caplin & Martin (2015) (CM15 hereafter)
 - ▶ Recall that perception is summarized by information structures
 - ▶ In that paper, they take information structures as fixed or given
- ▶ Ideally, we would like DMs to choose their information structure based on setting
- ▶ Why would DMs not choose the most informative information structure?
 - ▶ Information is costly!

Caplin & Dean (2015), "Revealed preference, rational inattention, and costly information acquisition."

- ▶ *Rational inattention*: Choose information to acquire to maximize utility net of information costs
- ▶ People are free to choose
 - ▶ How much information to acquire
 - ▶ What type of information to acquire
- ▶ Standard approach to modeling rational inattention:
 - ▶ Model the cost of choosing each information structure
 - ▶ Costs usually increasing in "informativeness"
- ▶ Caplin & Dean (2015) (CD15 hereafter) provide testable conditions for a very general form of this model (**will cover this today**)

Set Up: Model Objects

- ▶ Ω : Objective states of the world (finite)
 - ▶ Objective prior probabilities given by $\mu^*(\omega)$
 - ▶ Subjective prior beliefs given by $\mu(\omega)$
- ▶ a : An action
 - ▶ $x(a, \omega) \in X$ is the prize from action a in state ω
 - ▶ Note: CD15 use decision-theoretic $a(\omega)$
- ▶ \mathcal{A} : Grand set of actions
 - ▶ $A \subset \mathcal{A}$: Set of actions (finite)
 - ▶ Note: Unlike CM15, CD15 allow action set to change
- ▶ u : Utility function over prizes
 - ▶ $u(a, \omega) := u(x(a, \omega))$ is the utility of action a in state ω
 - ▶ Note: Unlike CM15, CD15 take u as given for now (relax later)

Set Up: Decision Problems

- ▶ Let D be a collection of decision problems, with arbitrary element $i \in D$
- ▶ Objective states of the world Ω are fixed across decision problems
- ▶ Objective prior can depend on decision problem: $\mu_i^*(\omega)$
 - ▶ Thus, the subjective prior $\mu_i(\omega)$ can too
- ▶ Prizes can depend on decision problem: $x_i(a, \omega)$
 - ▶ Thus, the utility function can too: $u_i(a, \omega) := u(x_i(a, \omega))$
- ▶ Action set can depend on decision problem: A_i
 - ▶ Thus, the decision rule can too: $\sigma_i : \Gamma \rightarrow A_i$
- ▶ In CD15, they just vary the action set across decision problems

The Model

- ▶ For each decision problem $i \in D$
 1. Choose π_i : information structure
 - ▶ $\pi_i(\gamma|\omega)$ is the probability of receiving each posterior γ from each state ω
 - ▶ Note: CD15 calls γ a “signal”
 - ▶ Note: Caplin, Dean & Leahy (2022) use Q instead
 2. Choose σ_i : decision rule
 - ▶ $\sigma_i(\gamma)$ is the probability distribution over actions in A_i given posterior γ
 - ▶ Note: CD15 use C and call “choice function”
 - ▶ Note: Caplin, Dean & Leahy (2022) use q instead
- ▶ To maximize expected value of actions taken minus cost of information K

$$\sum_{\omega \in \Omega} \mu_i(\omega) \sum_{\gamma \in \Gamma} \pi_i(\gamma|\omega) \left(\sum_{a \in A_i} \sigma_i(a|\gamma) u_i(a, \omega) \right) - K(\mu_i, \pi_i)$$

Value and Cost of Information

- ▶ Easy to calculate the *value* of an information structure π in decision problem i

$$G_i(\pi) = \max_{\sigma_i} \sum_{\omega \in \Omega} \mu_i(\omega) \sum_{\gamma \in \Gamma} \pi(\gamma|\omega) \left(\sum_{a \in A_i} \sigma_i(a|\gamma) u_i(a, \omega) \right)$$

- ▶ But what is the correct way to determine the *cost* of an information structure?
- ▶ Several proposals:
 - ▶ Costly sequential search (e.g. McCall 1970)
 - ▶ Cost to reduce variance of normal signal (e.g. Verrecchia 1982)
 - ▶ Shannon mutual information costs (e.g. Sims 2003)

Aim

- ▶ As usual, have two possible approaches
 1. Make further assumptions
 2. Ask if there is *any* cost function that can explain the data
- ▶ Today we take approach 2
- ▶ Later we will follow approach 1

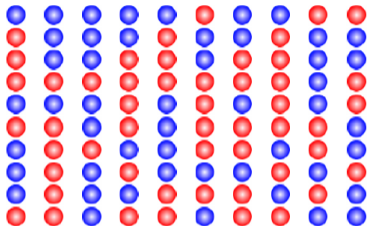
A Caveat

- ▶ We **will** assume throughout that costs are additively separable from utilities
- ▶ Is this assumption restrictive?
- ▶ Yes, see Chambers, Liu & Rehbeck (2020)'s "Costly Information Acquisition"

Observability

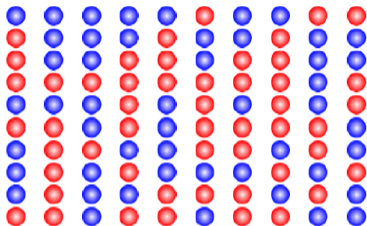
- ▶ What do we observe for all decision problems $i \in D$?
- ▶ States Ω , Actions A_i
- ▶ **State-dependent stochastic choice data** P_i
 - ▶ $P_i(a, \omega)$ probability of choosing action a and being in state ω
- ▶ Also assume we observe:
 - ▶ Subjective prior beliefs μ_i (relax observability later)
 - ▶ Observability a non-issue if assume prior beliefs correct!
 - ▶ And assume fixed across decision problems: $\mu = \mu_i$ for all $i \in D$
 - ▶ Utilities u_i (relax observability later)
 - ▶ And assume fixed across decision problems: $u = u_i$ for all $i \in D$
- ▶ Do **not** observe
 - ▶ Information structures π_i
 - ▶ Posteriors γ
 - ▶ Information costs K

An Experimental Example



- ▶ In Dean & Neligh (2023), subjects were presented with 100 balls
- ▶ State is determined by the number of red balls
- ▶ Subjects are told the prior distribution of red balls

An Experimental Example



Action	Payoff 49 red balls	Payoff 51 red balls
a	10	0
b	0	10

- ▶ No time limit: trade off between effort and financial rewards

An Experimental Example

- ▶ Data: State-dependent stochastic choice
 - ▶ Probability of choosing each action $a, b \in A$ and being in each state of the world

	State = 49 red balls	State = 51 red balls
Action = a	$P_i(a, 49)$	$P_i(a, 51)$
Action = b	$P_i(b, 49)$	$P_i(b, 51)$

- ▶ Observe subject making “same” choice 50 times
- ▶ Can use this to estimate P_i
 - ▶ But we will not be able to observe P_i perfectly
 - ▶ Will only be able to make probabilistic statements
- ▶ Can collect this type of data in the lab
 - ▶ What about outside?

Question

- ▶ What data are consistent with optimal costly information acquisition (with RI)?
 - ▶ i.e., for which there exists a cost function K
- ▶ For each decision problem $i \in D$, there must be an information structure π_i and decision rule σ_i s.t.
 - ▶ σ_i is optimal for each γ
 - ▶ π_i is optimal given K
 - ▶ σ_i and π_i are consistent with P_i

$$P_i(a, \omega) = \mu(\omega) \sum_{\gamma \in \Gamma} \pi_i(\gamma|\omega) \sigma_i(a|\gamma)$$

- ▶ Also, restrict π_i to be “Bayes-plausible” (follow Bayes’ rule)
- ▶ So satisfies the same conditions as the BEU representation (Data Matching, Bayesian Updating, and σ_i Maximization) plus a new one: π_i Maximization

Notes

- ▶ This approach is very flexible
- ▶ Nests *many* models of information acquisition
 - ▶ Fully flexible
 - ▶ Fixed signals
 - ▶ Partitions
 - ▶ Fixed capacity
 - ▶ Sequential search
- ▶ How nest hard constraints?
 - ▶ e.g., a model in which subjects choose the variance of a normal signal, set the cost of all other information structures to ∞

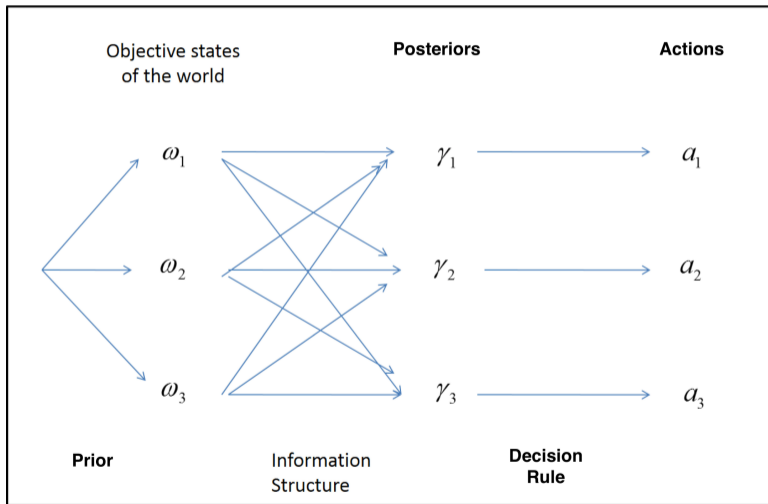
Observing Information Structures

- ▶ Key observation: State-dependent stochastic choice data tells us a lot about the information structure a decision maker has used
- ▶ What if the decision maker is ‘well behaved’?
 - ▶ Chooses each action in response to at most one posterior
 - ▶ No mixed strategies - one action per posterior
- ▶ Information structure can be observed directly from this dataset
 - ▶ For each chosen action a there is an associated “revealed posterior” $\bar{\gamma}^a$
 - ▶ Probability of posterior $\bar{\gamma}^a$ in state ω is the probability of choosing a in ω

$$\bar{\pi}(\bar{\gamma}^a|\omega) = P(a|\omega)$$

- ▶ Call $\bar{\pi}$ the “revealed information structure”

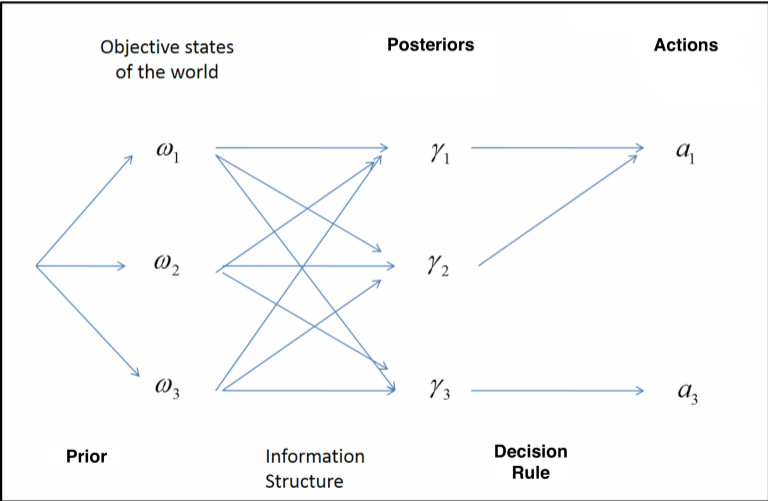
Revealed Information Structure



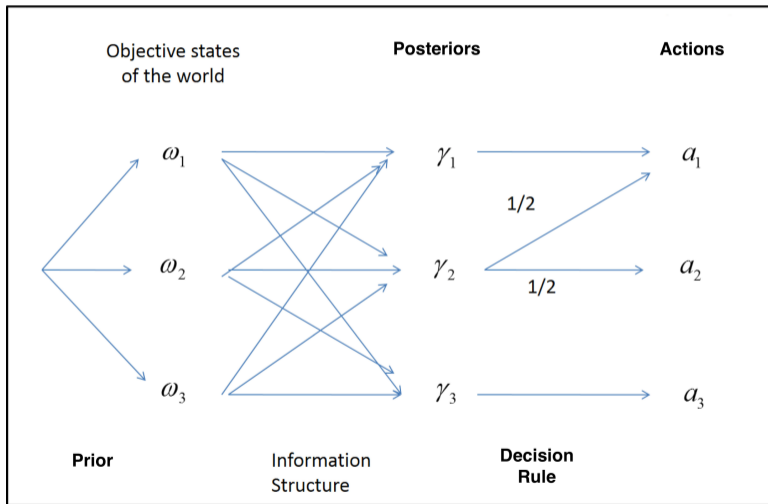
Observing Attentional Strategies

- ▶ What if the decision maker is not well behaved?
 - ▶ Chooses some act in more than one **subjective** state
 - ▶ Mixed strategies - more than one act in an **subjective** state

Same Act in Different States



Mixing



Observing Information Structures

- ▶ Can still recover revealed information structure $\bar{\pi}$
- ▶ Not necessarily the same as true information structure π
- ▶ But will be a **garbling** of the true information structure
 - ▶ i.e., π is statistically sufficient for $\bar{\pi}$
- ▶ There exists a stochastic $|\Gamma(\pi)| \times |\Gamma(\bar{\pi})|$ matrix B such that if we
 - ▶ Apply π
 - ▶ For each γ^i move to $\bar{\gamma}^j$ with probability B^{ij}
 - ▶ We obtain $\bar{\pi}$

$$\sum_j B^{ij} = 1 \quad \forall i$$
$$\bar{\pi}(\bar{\gamma}^j | \omega) = \sum_i B^{ij} \pi(\gamma^i | \omega) \quad \forall j, \omega$$

An Aside: Blackwell's Theorem

- ▶ Recall $G_i(\pi)$ is the *gross value* of using information structure π in decision problem i

$$\begin{aligned} G_i(\pi) &= \max_{\sigma_i} \sum_{\omega \in \Omega} \mu(\omega) \sum_{\gamma \in \Gamma} \pi(\gamma|\omega) \left(\sum_{a \in A_i} \sigma_i(a|\gamma) u(a, \omega) \right) \end{aligned}$$

- ▶ Blackwell's theorem (Blackwell 1953): information structure π' is a garbling of information structure π if and only if

$$G_i(\pi) \geq G_i(\pi') \quad \forall i$$

Observing Information Structures

- ▶ $\bar{\pi}$ may not be the agent's true information structure
 - ▶ But the $\bar{\pi}$ must be a garbling of the true information structure π
 - ▶ So π will be at least as valuable as $\bar{\pi}$ in any decision problem
- ▶ Turns out that this is all we need for now
- ▶ But Caplin, Martin & Marx (2025) show how to recover the full set of possible information structures

Characterizing Rational Inattention

Rational inattention has two parts:

- ▶ Choice of act optimal given information structure
- ▶ Choice of information structure optimal

Characterizing Rational Inattention

Rational inattention has two parts:

- ▶ Choice of act optimal given information structure
- ▶ Choice of information structure optimal

Condition 1

Condition 1 (No Improving Action Switches) For every chosen action $a \in A_i$

$$\sum_{\omega \in \Omega} P_i(a, \omega) (u(a, \omega) - u(b, \omega)) \geq 0$$

for all $b \in A_i$

- ▶ Introduced in CM15
- ▶ Intuition: Cannot improve utility by making a wholesale switch from any chosen action
- ▶ Note: Same as

$$\sum_{\omega \in \Omega} P_i(a, \omega) u(a, \omega) \geq \sum_{\omega \in \Omega} P_i(a, \omega) u(b, \omega)$$

Characterizing Rational Inattention

Rational inattention has two parts:

- ▶ Choice of act optimal given information structure
- ▶ Choice of information structure optimal

- ▶ $G_i(\pi)$ is the gross value of using information structure π in decision problem i
- ▶ Assume two decision problems: $A_1 = \{a_1, b_1\}$ and $A_2 = \{a_2, b_2\}$
- ▶ Cost function must satisfy

$$\begin{aligned}G_1(\pi_1) - K(\pi_1) &\geq G_1(\pi_2) - K(\pi_2) \\G_2(\pi_2) - K(\pi_2) &\geq G_2(\pi_1) - K(\pi_1)\end{aligned}$$

- ▶ Which implies

$$\begin{aligned}G_1(\pi_1) - G_1(\pi_2) &\geq \\K(\pi_1) - K(\pi_2) &\geq \\G_2(\pi_1) - G_2(\pi_2)\end{aligned}$$

Optimal Choice of Information Structures

- ▶ Surplus must be maximized by correct assignments

$$\begin{aligned}G_1(\pi_1) + G_2(\pi_2) \\ \geq G_1(\pi_2) + G_2(\pi_1)\end{aligned}$$

- ▶ Problem: Don't observe π_1 or π_2 !
- ▶ But we know that revealed and true information structure must give same value in DP it was observed

$$G_i(\bar{\pi}_i) = G_i(\pi_i)$$

- ▶ Also, as π weakly Blackwell dominates $\bar{\pi}$

$$G_i(\bar{\pi}_j) \leq G_i(\pi_j)$$

Optimal Choice of Information Structures

- ▶ Surplus must be maximized by correct assignments

$$\begin{aligned} G_1(\bar{\pi}_1) + G_2(\bar{\pi}_2) \\ \geq G_1(\pi_2) + G_2(\pi_1) \end{aligned}$$

- ▶ Problem: Don't observe π_1 or π_2 !
- ▶ But we know that revealed and true information structure must give same value in DP it was observed

$$G_i(\bar{\pi}_i) = G_i(\pi_i)$$

- ▶ Also, as π weakly Blackwell dominates $\bar{\pi}$

$$G_i(\bar{\pi}_j) \leq G_i(\pi_j)$$

Optimal Choice of Information Structures

- ▶ Surplus must be maximized by correct assignments

$$\begin{aligned} G_1(\bar{\pi}_1) + G_2(\bar{\pi}_2) \\ \geq G_1(\bar{\pi}_2) + G_2(\bar{\pi}_1) \end{aligned}$$

- ▶ Problem: Don't observe π_1 or π_2 !
- ▶ But we know that revealed and true information structure must give same value in DP it was observed

$$G_i(\bar{\pi}_i) = G_i(\pi_i)$$

- ▶ Also, as π weakly Blackwell dominates $\bar{\pi}$

$$G_i(\bar{\pi}_j) \leq G_i(\pi_j)$$

Condition 2

- ▶ We can rearrange this as a no-improvement condition

$$\begin{aligned} & G_1(\bar{\pi}_1) - G_1(\bar{\pi}_2) \\ + & G_2(\bar{\pi}_2) - G_2(\bar{\pi}_1) \\ \geq & 0 \end{aligned}$$

- ▶ To guarantee the existence of a cost function requires a stronger condition

Condition 2 (No Improving Attention Cycles) For an observed sequence of decision problems $1, \dots, K$ and associated revealed information structures $\bar{\pi}_1, \dots, \bar{\pi}_K$

$$\begin{aligned} & G_1(\bar{\pi}_1) - G_1(\bar{\pi}_2) \\ + & G_2(\bar{\pi}_2) - G_2(\bar{\pi}_3) \\ + & \dots \\ + & G_K(\bar{\pi}_K) - G_K(\bar{\pi}_1) \\ \geq & 0 \end{aligned}$$

Taking to Data

- ▶ This can easily be rewritten in terms of the data because

$$G_i(\bar{\pi}_j) = \sum_{a \in A_j} \max_{\hat{a} \in A_i} \sum_{\omega \in \Omega} P_j(a, \omega) u(\hat{a}, \omega)$$

Condition 2 (No Improving Attention Cycles) For an observed sequence of decision problems A_1, \dots, A_K

$$\begin{aligned} & \sum_{a \in A_1} \max_{\hat{a} \in A_1} \sum_{\omega \in \Omega} P_1(a, \omega) u(\hat{a}, \omega) - \sum_{a \in A_2} \max_{\hat{a} \in A_1} \sum_{\omega \in \Omega} P_2(a, \omega) u(\hat{a}, \omega) \\ & + \sum_{a \in A_2} \max_{\hat{a} \in A_2} \sum_{\omega \in \Omega} P_2(a, \omega) u(\hat{a}, \omega) - \sum_{a \in A_3} \max_{\hat{a} \in A_2} \sum_{\omega \in \Omega} P_3(a, \omega) u(\hat{a}, \omega) \\ & + \dots \\ & + \sum_{a \in A_K} \max_{\hat{a} \in A_K} \sum_{\omega \in \Omega} P_K(a, \omega) u(\hat{a}, \omega) - \sum_{a \in A_1} \max_{\hat{a} \in A_K} \sum_{\omega \in \Omega} P_1(a, \omega) u(\hat{a}, \omega) \\ & \geq 0 \end{aligned}$$

Testing NIAC

- ▶ We test NIAC by looking at the impact of cycling revealed information structures
- ▶ For example, consider these unconditional joint distributions of states and actions:

$$P_1 = \begin{pmatrix} 49 & 51 \\ \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} \end{pmatrix} \begin{matrix} a_1 \\ b_1 \end{matrix} \quad \& \quad P_2 = \begin{pmatrix} 49 & 51 \\ \frac{2}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{6} \end{pmatrix} \begin{matrix} a_2 \\ b_2 \end{matrix}$$

- ▶ And these utilities:

$$u(a, \omega) = \begin{pmatrix} 49 & 51 \\ 10 & 0 \\ 0 & 10 \end{pmatrix} \begin{matrix} a_1 \\ b_1 \end{matrix} \quad \& \quad u(a, \omega) = \begin{pmatrix} 49 & 51 \\ 20 & 0 \\ 0 & 20 \end{pmatrix} \begin{matrix} a_2 \\ b_2 \end{matrix}$$

The Value of Information

- ▶ The value of the chosen information structure in A_1 in decision problem A_1 is

$$\max_{a \in \{a_1, b_1\}} [P_1(a_1, 49)u(a, 49) + P_1(a_1, 51)u(a, 51)] + \max_{a \in \{a_1, b_1\}} [P_1(b_1, 49)u(a, 49) + P_1(b_1, 51)u(a, 51)]$$
$$\frac{3}{8}10 + \frac{1}{8}0 + \frac{1}{8}0 + \frac{3}{8}10 = \frac{60}{8} = 7\frac{1}{2}$$

- ▶ Switching to the chosen information structure in A_2 gives

$$\max_{a \in \{a_1, b_1\}} [P_2(a_2, 49)u(a, 49) + P_2(a_2, 51)u(a, 51)] + \max_{a \in \{a_1, b_1\}} [P_2(b_2, 49)u(a, 49) + P_2(b_2, 51)u(a, 51)]$$
$$\frac{2}{6}10 + \frac{1}{6}0 + \frac{1}{6}0 + \frac{2}{6}10 = \frac{40}{6} = 6\frac{2}{3}$$

The Value of Information

- ▶ The value of the chosen information structure in A_2 in decision problem A_2 is

$$\max_{a \in \{a_2, b_2\}} [P_2(a_2, 49)u(a, 49) + P_2(a_2, 51)u(a, 51)] + \max_{a \in \{a_2, b_2\}} [P_2(b_2, 49)u(a, 49) + P_2(b_2, 51)u(a, 51)]$$
$$\frac{2}{6}20 + \frac{1}{6}0 + \frac{1}{6}0 + \frac{2}{6}20 = \frac{80}{6} = 13\frac{1}{3}$$

- ▶ Switching to the chosen information structure in A_1 gives

$$\max_{a \in \{a_2, b_2\}} [P_1(a_1, 49)u(a, 49) + P_1(a_1, 51)u(a, 51)] + \max_{a \in \{a_2, b_2\}} [P_1(b_1, 49)u(a, 49) + P_1(b_1, 51)u(a, 51)]$$
$$\frac{3}{8}20 + \frac{1}{8}0 + \frac{1}{8}0 + \frac{3}{8}20 = \frac{120}{8} = 15$$

NIAS and NIAC

- ▶ P_1 and P_2 satisfy NIAS because no wholesale improving switches of actions
- ▶ But do they satisfy NIAC? The change from switching in the first problem is $6\frac{2}{3} - 7\frac{1}{2}$ and the change from switching in the second problem is $15 - 13\frac{1}{3}$, so there is a net improvement
- ▶ This means that this dataset does not satisfy NIAC, so there is no K that rationalizes the data

Theorem 1: CD15

Theorem

For any dataset $\{D, P\}$ the following two statements are equivalent

1. $\{D, P\}$ satisfy NIAS and NIAC
2. There exists a $K : \Pi \rightarrow \mathbb{R}$, $\{\pi_i\}_{i \in D}$ and $\{\sigma_i\}_{i \in D}$ such that π_i and σ_i are optimal and generate P_i for every $i \in D$

Proof.

2 \rightarrow 1 *Trivial*

1 \rightarrow 2 *Rochet (1987) (literature on implementation)*



Costs and Blackwell Ordering

- ▶ So far we have been completely agnostic about the cost function
- ▶ Perhaps we want to impose that more Blackwell-informative information structures are (weakly) more expensive
- ▶ Turns out we get this 'for free'
- ▶ Say we observe π_i in A_i and π_j in A_j such that π_j is a garbling of π_i (π_i is sufficient for π_j)
- ▶ By NIAC must be the case that

$$\begin{aligned}G_j(\pi_j) - K(\pi_j) &\geq G_j(\pi_i) - K(\pi_i) \Rightarrow \\K(\pi_i) - K(\pi_j) &\geq G_j(\pi_i) - G_j(\pi_j)\end{aligned}$$

- ▶ And by Blackwell's theorem (Blackwell 1953)

$$G_j(\pi_i) \geq G_j(\pi_j)$$

Restrictions on the Cost Function

- ▶ Any behavior that can be rationalized can be rationalized with a cost function that
 - ▶ Is weakly monotonic with respect to Blackwell
 - ▶ Allows mixing
 - ▶ Positive with free inattention
- ▶ Reminiscent of Afriat's theorem (Afriat 1967)
- ▶ Can also extend to 'sequential rational inattention'

Recovering Costs

- ▶ Say $\bar{\pi}_i$ is the revealed information structure in decision problem A
- ▶ Assuming weak monotonicity, $\bar{\pi}$ is optimal, so it must be that

$$\begin{aligned}G_i(\pi) - K(\pi) &\leq G_i(\bar{\pi}_i) - K(\bar{\pi}_i) \\K(\bar{\pi}_i) - K(\pi) &\leq G_i(\bar{\pi}_i) - G_i(\pi)\end{aligned}$$

- ▶ If $\bar{\pi}_j$ is used in decision problem B then we can bound relative costs

$$G_j(\bar{\pi}_i) - G_j(\bar{\pi}_j) \leq K(\bar{\pi}_i) - K(\bar{\pi}_j) \leq G_i(\bar{\pi}_i) - G_i(\bar{\pi}_j)$$

- ▶ Caplin, Martin & Marx (2025) provide an efficient method to generate these costs and a representative cost (see later)

What If Utility and Priors Are Unobservable?

- ▶ Can add 'there exists' to the statement of the NIAS and NIAC conditions
- ▶ The dataset has an optimal costly attention representation if **there exists** $\mu \in \Delta(\Omega)$ and $u : X \rightarrow \mathbb{R}$ such that
 - ▶ NIAS is satisfied
 - ▶ NIAC is satisfied
- ▶ If μ is known but u is unknown, conditions are linear and (relatively) easy to check
- ▶ If μ and u are unknown, conditions are harder to check
 - ▶ Still not vacuous
- ▶ Possible to recover all consistent utility functions using geometric approach introduced in CM21

Rational Inattention vs Random Utility

- ▶ Alternative model of random choice: Random Utility
 1. Agent receives some information about the state of the world
 2. Draws a utility function from some set
 3. Chooses in order to maximize utility given information
- ▶ Key differences between Random Utility and Rational Inattention
 1. Random Utility allows for multiple utility functions
 2. Rational Inattention allows attention to vary with choice problem
- ▶ How can we differentiate between the two?

Monotonicity

- ▶ Random Utility implies monotonicity
 - ▶ In fact, fully characterized by Block–Marschak monotonicity (Block & Marschak 1960)

- ▶ For any two decision problems $A_1 = A$ and $A_2 = A \cup b$ for $b \notin A$

$$P_1(a|\omega) \geq P_2(a|\omega)$$

- ▶ Rational Inattention can lead to violations of monotonicity (Matějka & McKay 2015)

Act	Payoff 49 red dots	Payoff 51 red dots
a	23	23
b	20	25
c	40	0

- ▶ Adding act c to $\{a, b\}$ can increase the probability of choosing b in state 51 (there exists a prior for which adding c makes it worth paying attention)

Other Approaches

- ▶ Special class of RI model: Capacity-Constrained Learning (Caplin, Martin, Marx, Morozova & Xu 2026)
- ▶ Variation on RI model: Costly “experiments”
- ▶ Special class of RI model: Uniformly Posterior Separable (UPS) (Caplin, Dean & Leahy 2022, Bloedel & Zhong 2025)

$$K(\mu, \pi) = T(\mu) - E_{\pi} [T(\gamma)]$$

- ▶ Special class of UPS model: “Shannon model” with mutual information costs

$$\begin{aligned} K(\mu, \pi) &= \lambda (H(\mu) - E_{\pi} [H(\gamma)]) \\ &= \lambda \left(\begin{array}{c} \sum_{\gamma \in \Gamma} \pi(\gamma) \sum_{\omega \in \Omega} \gamma(\omega) \ln \gamma(\omega) \\ - \sum_{\omega \in \Omega} \mu(\omega) \ln \mu(\omega) \end{array} \right) \end{aligned}$$

References I

- Afriat, S. N. (1967), 'The construction of utility functions from expenditure data', *International Economic Review* **8**(1), 67–77.
- Blackwell, D. (1953), 'Equivalent comparisons of experiments', *Annals of Mathematical Statistics* **24**(2), 265–272.
- Block, H. D. & Marschak, J. (1960), Random orderings and stochastic theories of response, in I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow & H. B. Mann, eds, 'Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling', Stanford University Press, Stanford, CA, pp. 97–132.
- Bloedel, A. W. & Zhong, W. (2025), The cost of optimally acquired information. arXiv:2511.05466, working paper.
- Caplin, A. & Dean, M. (2015), 'Revealed preference, rational inattention, and costly information acquisition', *American Economic Review* **105**(7), 2183–2203.
- Caplin, A., Dean, M. & Leahy, J. (2022), 'Rationally inattentive behavior: Characterizing and generalizing Shannon entropy', *Journal of Political Economy* **130**(6), 1676–1715.
- Caplin, A., Dean, M. & Martin, D. (2011), 'Search and satisficing', *American Economic Review* **101**(7), 2899–2922.

References II

- Caplin, A. & Martin, D. (2015), 'A testable theory of imperfect perception', *The Economic Journal* **125**(582), 184–202.
- Caplin, A., Martin, D. & Marx, P. (2025), 'Rationalizable learning', *Economic Theory* **80**(1), 171–202.
- Caplin, A., Martin, D., Marx, P., Morozova, A. & Xu, L. (2026), Testing capacity-constrained learning. Working paper; revised March 6, 2026; revise and resubmit at Experimental Economics; arXiv:2502.00195.
- Chambers, C. P., Liu, C. & Rehbeck, J. (2020), 'Costly information acquisition', *Journal of Economic Theory* **186**, 104979.
- Chetty, R., Looney, A. & Kroft, K. (2009), 'Salience and taxation: Theory and evidence', *American Economic Review* **99**(4), 1145–1177.
- Dean, M. & Neligh, N. (2023), 'Experimental tests of rational inattention', *Journal of Political Economy* **131**(12), 3415–3461.
- DellaVigna, S. & Pollet, J. M. (2009), 'Investor inattention and Friday earnings announcements', *Journal of Finance* **64**(2), 709–749.
- Lacetera, N., Pope, D. G. & Sydnor, J. R. (2012), 'Heuristic thinking and limited attention in the car market', *American Economic Review* **102**(5), 2206–2236.

References III

- Manzini, P. & Mariotti, M. (2014), 'Stochastic choice and consideration sets', *Econometrica* **82**(3), 1153–1176.
- Matějka, F. & McKay, A. (2015), 'Rational inattention to discrete choices: A new foundation for the multinomial logit model', *American Economic Review* **105**(1), 272–298.
- McCall, J. J. (1970), 'Economics of information and job search', *The Quarterly Journal of Economics* **84**(1), 113–126.
- Rochet, J.-C. (1987), 'A necessary and sufficient condition for rationalizability in a quasi-linear context', *Journal of Mathematical Economics* **16**(2), 191–200.
- Sims, C. A. (2003), 'Implications of rational inattention', *Journal of Monetary Economics* **50**(3), 665–690.
- Verrecchia, R. E. (1982), 'Information acquisition in a noisy rational expectations economy', *Econometrica* **50**(6), 1415–1430.