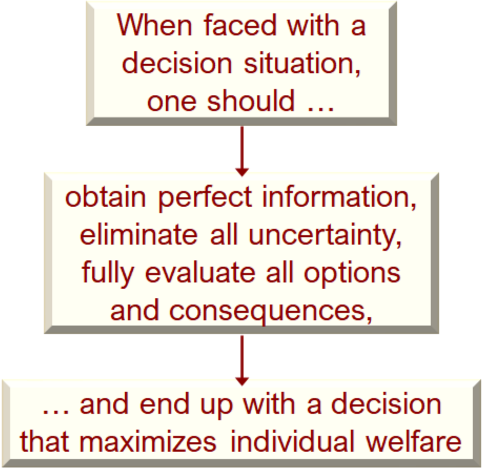
Inattentive Decision-Making In a Changing Environment

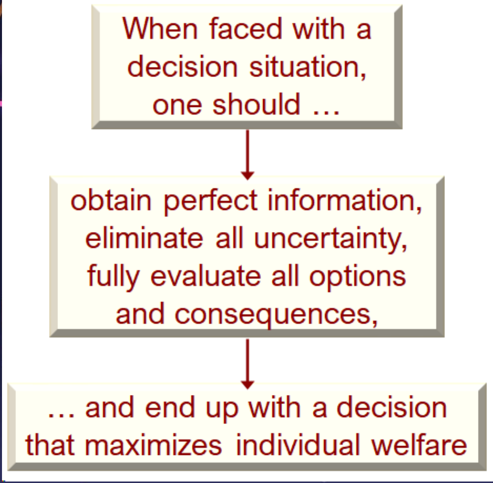
Luminita Stevens
Department of Economics
University of Maryland

Sloan-Nomis Summer School
June 2022

The Classical Model of Decision-Making



The Classical Model of Decision-Making



Some Evidence of Seemingly Non-Normative Behavior in Economic Contexts

At the individual level:

- There seems to be incomplete use of information
    - even when it is both useful & easily available

  e.g., Coibion & Gorodnichenko (2012) inflation surveys

- Choices exhibit heterogeneity, stochasticity & discreteness
    - even in fairly homogeneous environments
    - and for problems with well-defined, unique optima
    - over continuous variables

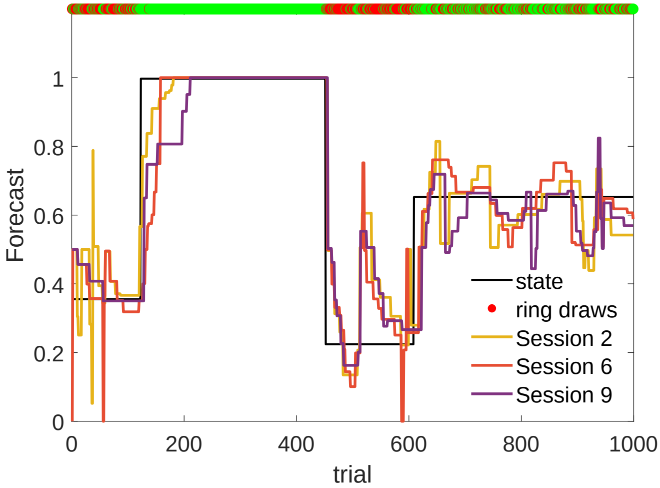
  e.g., Khaw, Stevens & Woodford (2017) in the lab

- Many decision variables are adjusted infrequently
    - even though conditions change constantly

  (prices, wages, hiring, physical capital, retirement portfolios)

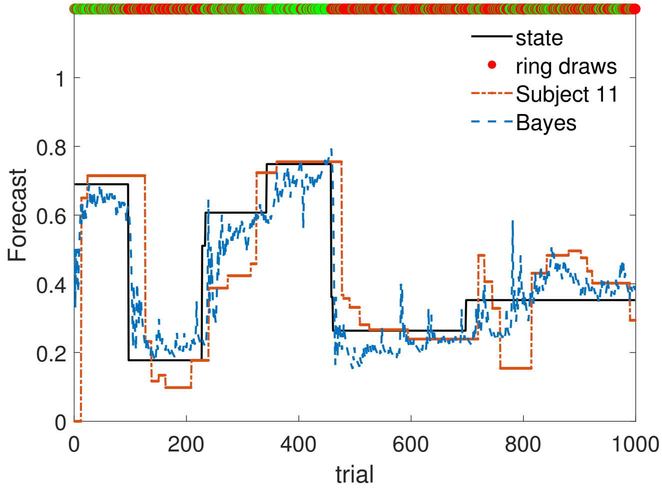
Khaw, Stevens & Woodford (2017)

Stochastic Choices



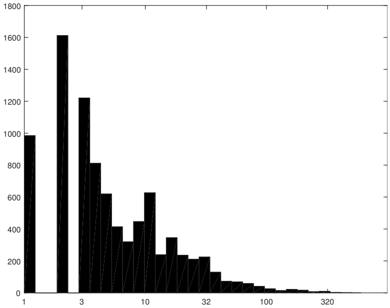
Subject 1, three repetitions of same sequence of ring draws

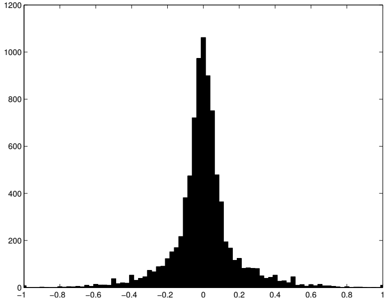
Khaw, Stevens & Woodford (2017)
Discrete Adjustment



Subject 11, session 10

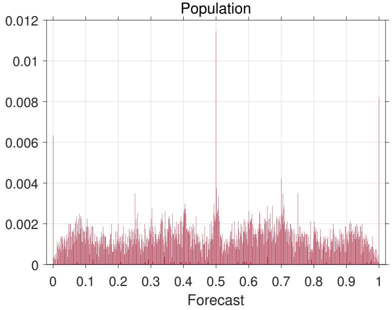
Khaw, Stevens & Woodford (2017)
Discrete Adjustment

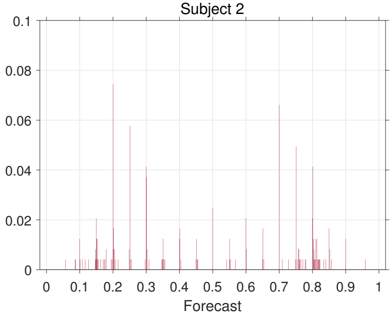




substantial fixed-position spells common
but both **spell length** and **size of adjustment** vary widely

Khaw, Stevens & Woodford (2017)
Discrete Action Space





near-multiples of .05 chosen more often though
neither slider motion nor reward function favor this

Some Economics Explanations of Seemingly Non-Normative Behavior

One could rationalize many of these patterns of behavior within classical, fully optimizing frameworks, for example:

- Stochasticity in response to idiosyncratic shocks to prefs or technology, or due to deliberate (exploratory) randomization

- Inaction due to adjustment costs — though estimated to be small relative to the degree of inaction for many decision variables *e.g.*, Morales-Jimenez & Stevens (2022) for prices

Alternative hypothesis: These patterns reflect inattentive decision-making that is shaped by cognitive limitations

- Stochasticity, discreteness, inaction $\because$ *e.g.*, rational inattention

Rational Inattention: General Principles

Abstract model of constrained-optimal decision-making:

1. Information is abundant but hard to gather and incorporate
2. Its acquisition is a choice that responds to incentives
3. Hence we can apply cost-benefit analysis

→ Decision-makers do not have complete knowledge of their environment (they are inattentive to it), but they **choose** what to focus on, given their objectives and the costs of having a more precise awareness of their environment

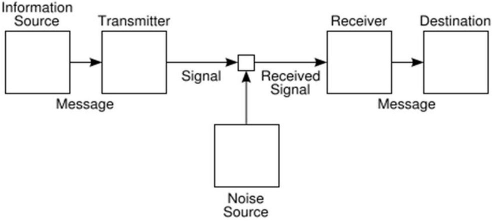
How to formalize and operationalize these principles?

RI: Applying Information Theory to Economic Decision-Making

Sims (1998, 2003) : Think of DMs as having limited capacity to process information and apply Shannon's (1948; 1959) information theory to endogenize their information choices

"Information theory, which formally models physical limits to the rate of transfer of information may provide a way for us to capture the intuitive appeal of the signal-extraction story **while neither introducing implausible constraints on the observability of data nor abandoning entirely the strategy of modeling behavior as reflecting optimization***." (Sims, 1998)*

RI: Applying Information Theory to Economic Decision-Making



Shannon (1948):

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; (...). The significant aspect is that the actual message is **one selected from a set** *of possible messages."*

Rational Inattention: Specific Features

Model of constrained-optimal decision-making and info acquisition:

1. All uncertainty can be formalized with stochastic processes characterized by objectively true probability distributions

2. Information is the reduction of Shannon (1948) uncertainty

3. Decisions are constrained by the capacity DMs devote to reducing uncertainty relevant to particular decision problems

4. DMs can design arbitrary representations dedicated to any source of uncertainty or combination thereof and can choose any (feasible) degree of precision for these representations

5. In fact, DMs design representations that are optimally adapted to each decision problem at hand

Outline

A Static RI Decision Problem
  - stochastic, coarse choices

B Dynamic RI Decision Problems
  - dampened, delayed, infrequent adjustments

RI: Static Decision Problem

- Suppose DM faces decision that leads to an uncertain payoff:
  actions .... $a \in \mathcal{A}$
  payoffs .... $u(a, x)$
  states .... $x \in \mathcal{X}$, $x \sim p(x)$
  prior .... $p(x)$

- Action under prior ...... $\bar{a} = \arg\max_a \sum_x p(x) u(a, x)$ for all x

- Should DM try to differentiate states more ?

Measuring Uncertainty (And Its Reduction) (Shannon (1948))

- Entropy $\equiv$ uncertainty about the state $x \sim p(x)$

$$\mathcal{H}(x) \equiv E_x \log \frac{1}{p(x)} \qquad \leftarrow \text{prior uncertainty}$$

- Conditional entropy $\equiv$ uncertainty after some random var s

$$\mathcal{H}(x|s) \equiv E_{x,s} \log \frac{1}{\tilde{p}(x|s)} \qquad \leftarrow \text{residual uncertainty}$$

- Information $=$ reduction in uncertainty

$$\mathcal{I}(x,s) = \mathcal{H}(x) - \mathcal{H}(x|s) = E_{x,s} \log \frac{\tilde{p}(x|s)}{p(x)}$$

Compression if $\mathcal{I}(x,s) < \mathcal{H}(x)$ $\rightarrow$ decomposition of the state into simpler (lower entropy) representation + data loss

Reducing Uncertainty (Application of Shannon, 1959)

- Suppose DM can reduce uncertainty about x, at a cost, by designing a representation that consists of
  - a set of signals $s \in S$
  - a conditional probability $q(s|x)$ for each x and s

  and an action rule $a^*(s)$ that maps signals into actions

  (note: we can allow for stochastic action rules)

  to max expected payoff net of cost of uncertainty reduction

Reducing Uncertainty (Application of Shannon, 1959)

- Suppose that the cost to the DM of such a representation is linear in the entropy reduction it achieves

$$\begin{aligned}\mathcal{C}(x,s) &= \theta\mathcal{I}(x,s) \\ &= \theta\left[\mathcal{H}(x) - \mathcal{H}(x|s)\right] \\ &= \theta\left[E_{x,s}\log\frac{\tilde{p}(x|s)}{p(x)}\right]\end{aligned}$$

(and there is no cost of implementing the desired action rule)

Reducing Uncertainty (Application of Shannon, 1959)

- Suppose that the cost to the DM of such a representation is linear in the entropy reduction it achieves

$$\begin{aligned} \mathcal{C}(x,s) &= \theta \mathcal{I}(x,s) & &= \theta \mathcal{I}(s,x) \\ &= \theta \left[\mathcal{H}(x) - \mathcal{H}(x|s) \right] & &= \theta \left[\mathcal{H}(s) - \mathcal{H}(s|x) \right] \\ &= \theta \left[E_{x,s} \log \frac{\tilde{p}(x|s)}{p(x)} \right] & &= \theta \left[E_{x,s} \log \frac{q(s|x)}{\bar{q}(s)} \right] \end{aligned}$$

(and there is no cost of implementing the desired action rule)

A Two-Step Problem

$$\max_{q,S,a^*} E_{x,s}\left[u(a(s),x)\right] - \theta\mathcal{I}(x,s)$$

s.t. $\sum_{s\in S} q(s|x) = 1$ for all x and $q(s|x) \geq 0$ for all x, s

where expectations are over the joint distribution $p(x)q(s|x)$

(2) Taking as given posterior beliefs $\tilde{p}(x|s)$, find the action rule

$$a^*(s) = \arg\max_a E_{x|s}\left[u(a,x)\right] \quad \text{for each } s \in S$$

(1) Given the optimal action rule, find the representation to solve

$$\max_{q,S} E_{x,s}\left[u\left(a^*(s),x\right)\right] - \theta\mathcal{I}(x,s)$$

A Simplification

- Solution generates distribution of actions with the same mutual information: $\mathcal{I}(x, a) = \mathcal{I}(x, s)$

- The joint distribution of a and x generated by the solution to our 2-step problem is the same as the solution to the alt. problem $\max_{q^a, A} E_{x,a}[u(a, x)] - \theta \mathcal{I}(x, a)$
  - wasteful to differentiate among s that result in same a
  - or to randomize a upon receipt of any s
  - or to design separate representations for different sources of uncertainty (vector x) relevant to a singleton action
  - (it also means that DMs with different objectives can end up with different posteriors on x)

Woodford (2009); Matějka & McKay (2015); Steiner, Stewart & Matějka (2017)

The Simplified Formulation

- So we can rewrite the problem in terms of a representation that directly indicates the action to be taken

- With some abuse of notation, the problem becomes

$$\max_{q,A} E_{x,a}\left[u(a,x)\right] - \theta \mathcal{I}(x,a)$$

$$\sum_{a \in A} q(a|x) = 1 \text{ for all } x \in \mathcal{X}$$

$$q(a|x) \geq 0 \text{ for all } x \in \mathcal{X}, a \in A$$

where expectations are over the joint distribution $p(x)q(a|x)$ and θ is the unit cost of uncertainty reduction

- We can use Karush-Kuhn-Tucker to find the solution

The Optimal Representation

(1) $$\frac{q(a|x)}{\bar{q}(a)} = \frac{\exp\left\{\frac{u(a,x)}{\theta}\right\}}{\sum_{\tilde{a}\in A} \bar{q}(\tilde{a}) \exp\left\{\frac{u(\tilde{a},x)}{\theta}\right\}} \text{ for all } x \in \mathcal{X},\ a \in A$$

(2) $$\bar{q}(a) = \sum_{x\in\mathcal{X}} p(x) q(a|x) > 0 \text{ for all } a \in A$$

(3) $$Z(a;\bar{q}) \begin{cases} = 1 & \text{for all } a \in A \\ \leq 1 & \text{for all } a \in \mathcal{A}, \quad \text{where} \end{cases}$$

$$Z(a;\bar{q}) = \sum_{x\in\mathcal{X}} p(x) \frac{\exp\left\{\frac{u(a,x)}{\theta}\right\}}{\sum_{\tilde{a}\in A} \bar{q}(\tilde{a}) \exp\left\{\frac{u(\tilde{a},x)}{\theta}\right\}}$$

Illustration

Consider a simple tracking problem

Suppose $u(a,x) = -(a-x)^2, \; x \sim U[76,124]$

Starting with the no-info solution, the optimal representation can be traced for lower and lower values of θ

An efficient algorithm combines

- the Blahut-Arimoto algorithm that iterates (1),(2) to convergence for a guess of the support (Arimoto, 1972; Blahut, 1972; Csiszár, 1974)
- the complementary slackness check on Z and the necessary conditions for the optimal points of support (Rose, 1994)

Rose (1994)

- A sufficient condition for information acquisition is

$$(4)\quad \theta \leq \overline{\theta}, \quad \overline{\theta} \equiv \frac{\sum_x p(x) \left(\frac{\partial}{\partial a} u(a,x) \Big|_{a=\bar{a}} \right)^2}{\sum_x p(x) \left(\frac{\partial^2}{\partial a^2} u(a,x) \Big|_{a=\bar{a}} \right)}$$

- A necessary condition for the points of support is

$$(5)\quad \sum_{x \in \mathcal{X}} \tilde{p}(x|a) \frac{\partial u(a,x)}{\partial a} \Big|_{a \in A} = 0$$

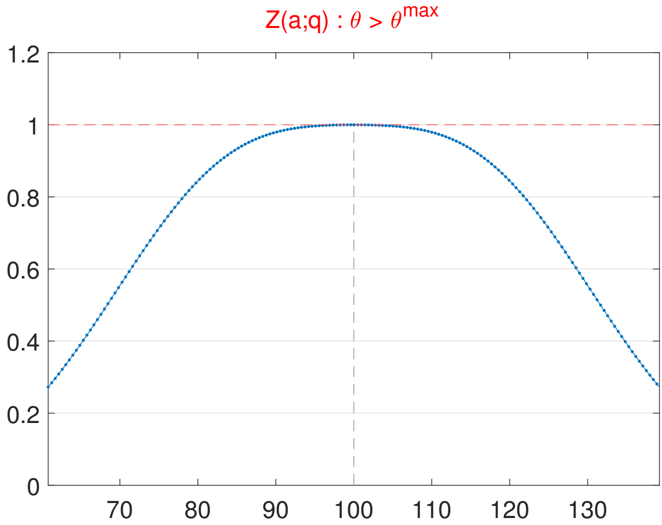
(each action a must maximize the expected payoff under the posterior distribution of x implied by that action)

Growth of Mass Points as Unit Cost Falls

Remember the slack function

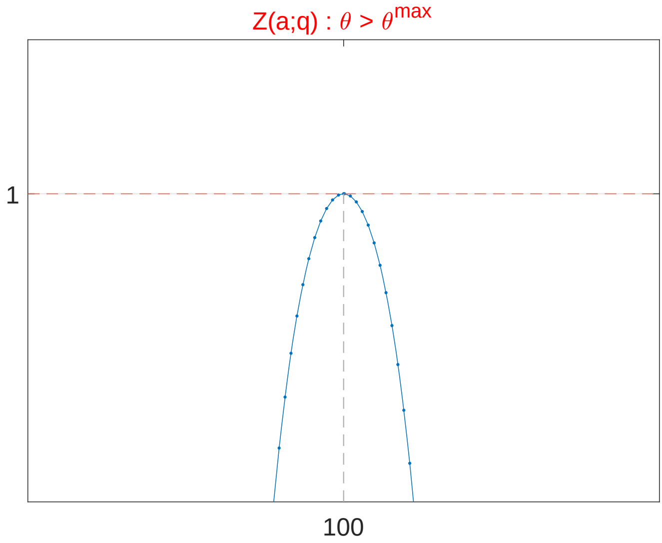
$$Z(a;\bar{q}) = \sum_{x\in\mathcal{X}} p(x) \frac{\exp\left\{\frac{u(a,x)}{\theta}\right\}}{\sum_{\tilde{a}\in A} \bar{q}(\tilde{a}) \exp\left\{\frac{u(\tilde{a},x)}{\theta}\right\}} \overset{?}{\leq} 1$$

Growth of Mass Points as Unit Cost Falls



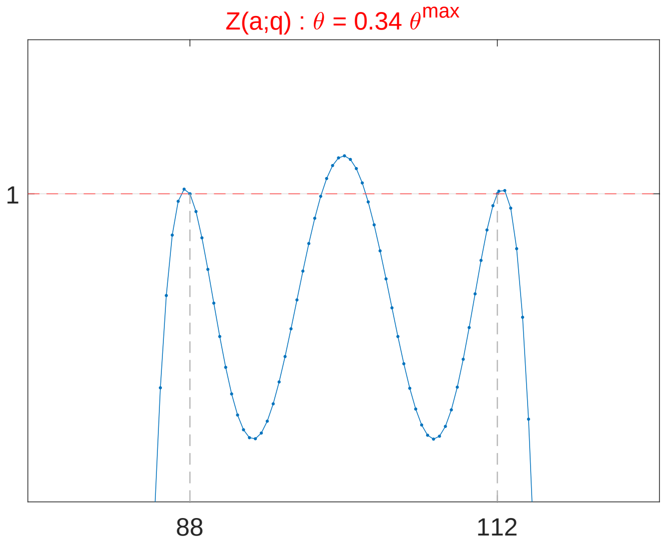
$$u(a, x) = -(a - x)^2, \;\; x \sim U[76, 124]$$

Growth of Mass Points as Unit Cost Falls



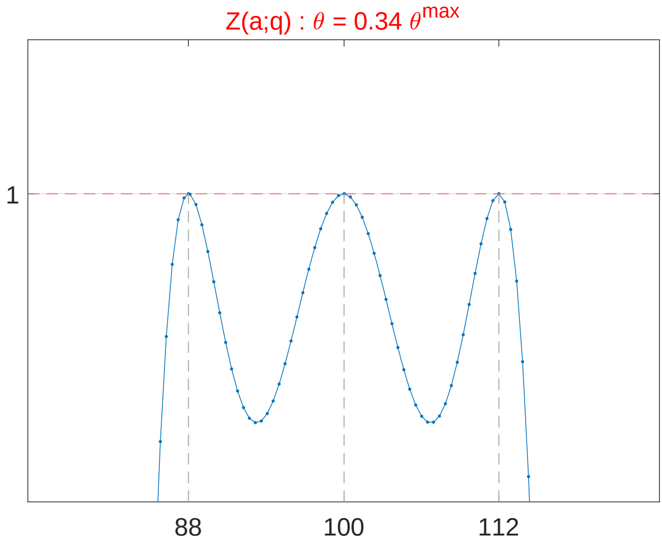
$$u(a, x) = -(a - x)^2, \; x \sim U[76, 124]$$

Growth of Mass Points as Unit Cost Falls



$$u(a,x) = -(a-x)^2, \; x \sim U[76, 124]$$

Growth of Mass Points as Unit Cost Falls



$$u(a,x) = -(a-x)^2, \ x \sim U[76,124]$$

Illustration of Optimal Representation

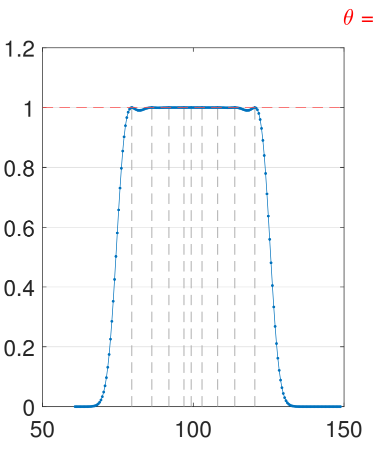
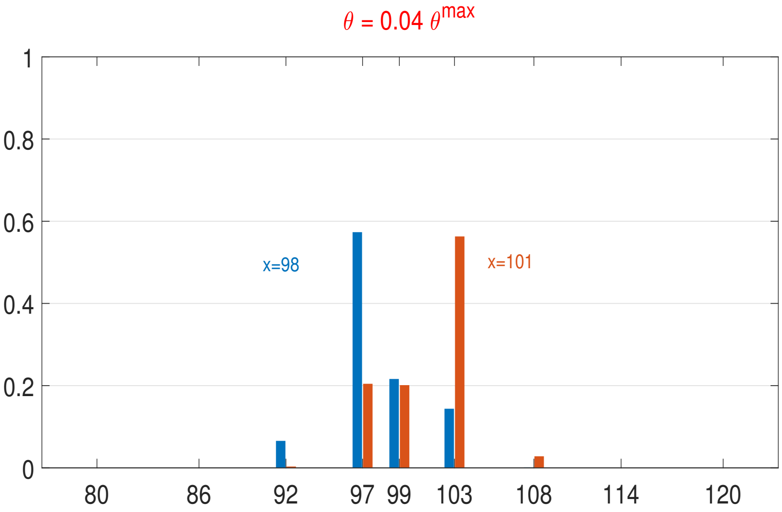


Illustration of Optimal Representation



Brief Extension to Multiple Representations

- The efficiency of info acquisition in RI implies a single representation per decision, even if the state consists of a vector of random variables
- In practice, DMs might be constrained in how they can acquire information on different sources of uncertainty
- DMs also often make multiple decisions, subject to different sources of uncertainty, simultaneously
- Then allocation of attention across sources based on similar principles to allocation of attention across states for a given source, so, similarly, sparsity can arise

  Cover & Thomas (2006); Maćkowiak & Wiederholt (2009) on firms ignoring macro conditions; Van Nieuwerburgh & Veldkamp (2010) on under-diversification of investor portfolios

A Quadratic-Normal Setup

- Suppose $x_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, ..., n$ independent

$$\max_{\{q_{ai}\}} \sum_{i=1}^{n} E\left[-(a_i - x_i)^2\right] - \theta \mathcal{I}(x^n, a^n)$$

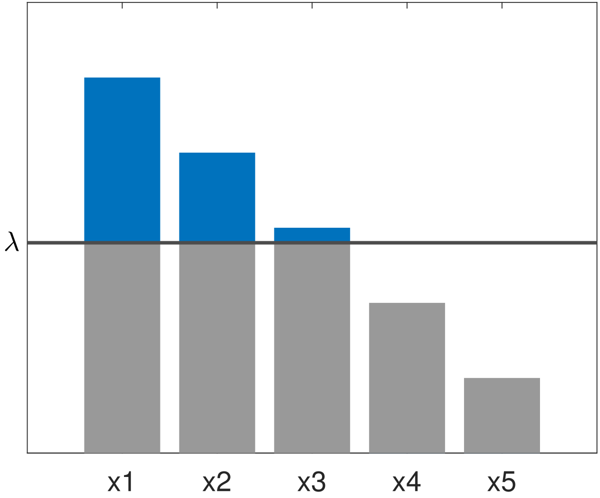
- Solution: Choose orthogonal $a_i \sim \mathcal{N}(0, \sigma_{ai}^2)$ with

$$x_i = a_i + \varepsilon_i$$

$$\sigma_i^2 = \sigma_{ai}^2 + \sigma_{\varepsilon i}^2$$

$$\sigma_{\varepsilon i}^2 = \begin{cases} \lambda_\theta & \texttt{if} \quad \lambda_\theta < \sigma_i^2 \\ \sigma_i^2 & \texttt{otherwise} \end{cases}$$

Reverse Water Filling Illustration



Static RI Summary

For solutions with positive information flow

1. The representation is informative: an action is more likely in a particular state if it leads to a higher payoff in that state

2. An endogenous prior about which action might be best anchors the representation, but loses its pull as θ declines

3. To economize on information costs, the representation is stochastic: $q(a|x) \in (0,1)$ for all $a \in A$ and all $x \in \mathcal{X}$

Matějka & McKay (2015); Caplin & Dean (2013)

Static RI Summary

5. The optimal representation is often coarse, even when compressing a continuous state

4. The action set is endogenously chosen s.t. ex-ante payoffs cannot be increased by reallocating attention to any $a \notin$ set A

6. Variable attention costs can generate variable action sets

7. Joint representations of multiple states may be sparse

Fix (1978); Rose (1994); Berger (2003) in info theory; Matějka (2016); Stevens (2020); Jung, Kim, Matějka & Sims (2019); Caplin, Dean & Leahy (2019) in econ

In short, RI choices are based on compressed, noisy representation of their environment

What about delayed and infrequent adjustment?

RI in Dynamic Decision Problems

- Chris Sims' motivation for bringing info theory to economics was in fact about **aggregate dynamics** (Sims, 1998)
- Central issue in macro: **delayed responses** to shocks
- Sims hypothesized info choice subject to cap on Shannon entropy reduction can **replace** the myriad of adjustment costs & wedges modern macro models use to fit sluggish aggregate adjustment dynamics (for both prices & quantities)
- Hence, he proposed a plausible friction and an elegant way to operationalize it, and also promised a **big payoff**

RI: Dynamic Setup

- Recall our static problem:

$$\max_{q,A} E_{x,a}\left[u(a,x) - \theta \log \frac{q(a|x)}{\bar{q}(a)}\right]$$

$$\sum_{a\in A} q(a|x) = 1 \text{ for all } x \in \mathcal{X}$$

$$q(a|x) \geq 0 \text{ for all } x \in \mathcal{X}, a \in A$$

$$\bar{q}(a) = \sum_{x\in\mathcal{X}} p(x)q(a|x) > 0 \text{ for all } a \in A$$

where expectations are over the joint distribution $p(x)q(a|x)$

- Extend to a setting where the choice of the optimal representation becomes **forward-looking**

RI: Dynamic Setup

- Q: Can DM condition on past signals or time period freely?
- Suppose **yes**: the cap on/cost of uncertainty reduction applies to **new** information each period, and DM has **perfect recall** of the full history of past signals and actions
- Get **dampened, delayed, hump-shaped responses** to changes in the exogenous fundamental, as seen in the macro data Maćkowiak & Wiederholt (2009); Luo (2008); Acharya & Wee (2020); Steiner et al. (2017); Maćkowiak, Matějka & Wiederholt (2018)
- Suppose **no** and assume **equal cost** per unit of both new information and any information retrieved from memory
- Then can also obtain **infrequent adjustment**, as in micro Woodford (2009); Stevens (2020); Khaw et al. (2017); Morales-Jimenez & Stevens (2022)

RI Dynamic Setup - Version 1

- Consider a dynamic RI problem with costless recall:

$$\max_{\{q_t\}_{t\geq 0}} \left(u(a_t, x_t) - \theta \log \frac{q_t\left(a_t | x^t; a^{t-1}\right)}{\bar{q}_t\left(a_t | a^{t-1}\right)} \right)$$

- The history of actions forms a **prior**, which, combined with new info yields a posterior based on which DM makes decision $\rightarrow$ problem becomes dynamic

RI Dynamic Setup - Version 1

- Consider a dynamic RI problem with costless recall:

$$\max_{\{q_t\}_{t\geq 0}} E_{(x^t,a^t)} \left[\sum_{t=0}^{\infty} \beta^t \left(u(a_t, x_t) - \theta \log \frac{q_t\left(a_t | x^t; a^{t-1}\right)}{\bar{q}_t\left(a_t | a^{t-1}\right)} \right) \right]$$

$$\sum_{a_t \in A_t} q_t\left(a_t | x^t; a^{t-1}\right) = 1 \text{ for all reachable } x^t, a^{t-1}$$

$$q_t\left(a_t | x^t; a^{t-1}\right) \geq 0 \text{ for } -//- \text{ and all } a_t \in A_t$$

where $a^t = (a_0, ..., a_t)$, $\beta \in (0, 1)$ discounts future payoffs, and expectations are over the joint distribution of x^t, a^t

RI Dynamic Setup - Version 1

- Consider a dynamic RI problem with costless recall:

$$\max_{\{q_t\}_{t\geq 0}} E_{(x^t,a^t)}\left[\sum_{t=0}^{\infty}\beta^t\left(u(a_t,x_t)-\theta\log\frac{q_t\left(a_t|x^t;a^{t-1}\right)}{\bar{q}_t\left(a_t|a^{t-1}\right)}\right)\right]$$

$$\sum_{a_t\in A_t} q_t\left(a_t|x^t;a^{t-1}\right)=1 \text{ for all reachable } x^t, a^{t-1}$$

$$q_t\left(a_t|x^t;a^{t-1}\right)\geq 0 \text{ for } -//- \text{ and all } a_t\in A_t$$

Dynamic Setup with Free Recall

- The simplification that collapses signals to actions continues to apply in (most) dynamic settings (need cost linear in $\mathcal{I}$)
- Solution via two-step optimization
  (1) Fixing the marginal distribution Q_t, solve for the optimal conditional q_t to max net expected payoff
  (2) Given the solution for the conditional, solve for the optimal marginal $\bar{q}_t = \arg\min_Q \mathcal{I}(q_t, Q)$
- Problem becomes analogous to a dynamic control problem with an **optimal default distribution**, adapted to this problem, and optimal from the info-theoretic point of view

Dynamic Representation with Free Recall

$$(6) \quad \frac{q_t(a_t|x^t;a^{t-1})}{\bar{q}_t(a_t|a^{t-1})} = \frac{\exp\left\{\frac{U_t(a_t,x^t,a^{t-1})}{\theta}\right\}}{\sum_{\tilde{a}_t} \bar{q}_t(\tilde{a}_t|a^{t-1}) \exp\left\{\frac{U_t(\tilde{a}_t,x^t,a^{t-1})}{\theta}\right\}}$$

$$(7) \quad \bar{q}_t(a_t|a^{t-1}) = \sum_{x^t} Pr(x^t|a^{t-1}) q_t(a_t|x^t;a^{t-1}) > 0$$

$$U_t(a_t,x^t,a^{t-1}) = u(a_t,x_t) + \beta E\left[V_{t+1}(x^{t+1},a^t)|a_t,x^t,a^{t-1}\right]$$

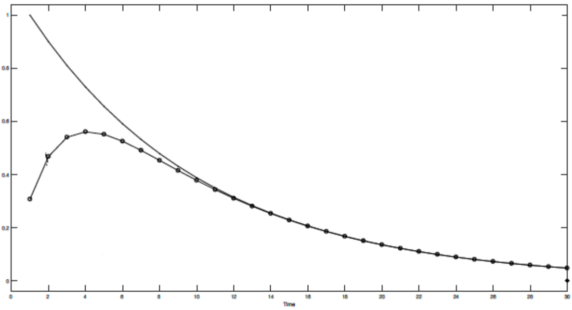
$$V_t(x^t,a^{t-1}) = \max_{q_t} E\left[u_t - \theta \log \frac{q_t}{\bar{q}_t} + \beta V_{t+1}(x^{t+1},a^t)|x^t,a^{t-1}\right]$$

Dynamic Representation with Free Recall

$$(8)\quad Z(a_t;\bar{q}_t)\begin{cases} = 1 & \text{for all } a_t \in A_t \\ \leq 1 & \text{for all } a_t \in \mathcal{A}, \quad \text{where} \end{cases}$$

$$Z(a_t;\bar{q}_t) = \sum_{x^t} Pr(x^t|a^{t-1}) \frac{\exp\left\{\frac{U\left(a_t,x^t,a^{t-1}\right)}{\theta}\right\}}{\sum\limits_{\tilde{a}_t\in A_t} \bar{q}_t(\tilde{a}_t|a^{t-1}) \exp\left\{\frac{U(\tilde{a}_t,x^t,a^{t-1})}{\theta}\right\}}$$

Example with AR(1) Fundamental



Maćkowiak et al. (2018)

dampened, hump-shaped response to x_t innovation

In short, RI choices are based on compressed, noisy representation of their environment and generate hump-shaped adjustment in dynamic settings

What about lumpy, infrequent adjustment?

RI: Dynamic Setup - Version 2

- Given discrete adjustment evidence, let us break the analysis of adjustment dynamics into two separate questions:
    - what determines **when adjustments occur?**
    - what determines **what action is taken** when adjustments occur?

- To generate sticky choices over time, let us also suppose that DM cannot condition on past signals, actions, time for free
    - instead faces **equal cost** when absorbing info regardless of source or type

Khaw et al. (2017) Model

- Choose

  (1) sequence of functions $\Lambda_t(x^t)$ specifying proba of adjustment in period t for each possible history

  (2) sequence of functions $\mu_t(x^t)$, specifying proba measure over possible new choices for each possible prior history, conditional on adjustment

- to maximize

$$\mathrm{E}\left\{\sum_{t=1}^{T}\left[r(x_t; a_t) - \psi_1\mathcal{I}_1 - \psi_2\Lambda_t\mathcal{I}_2\right]\right\}$$

- given costs $\psi_1, \psi_2 > 0$ of reducing uncertainty for the two decisions

Optimal adjustment hazard

$$\log \frac{\Lambda_t}{1-\Lambda_t} \;=\; \log \frac{\bar{\Lambda}}{1-\bar{\Lambda}} + \frac{\Delta_t}{\psi_1}$$

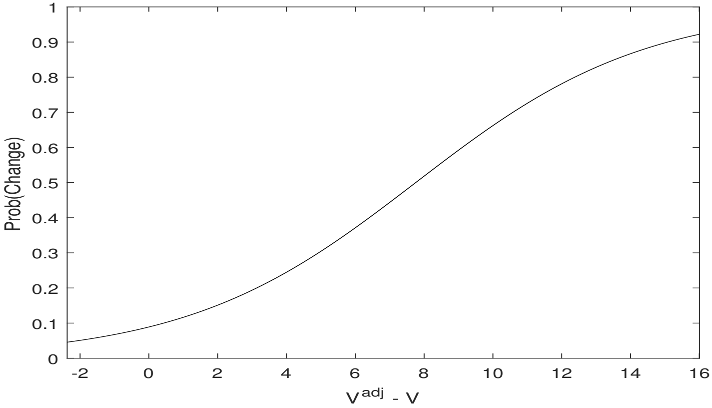
where the "RI value gap" $\Delta_t \equiv \int V_t(a)d\mu_t(a) - V_t(a_{t-1})$

$V_t(a)$ is the continuation value function

$\bar{\Lambda}$ is the expected frequency of adjustment over future states

$$\bar{\Lambda} = \frac{1}{T} E\left\{\sum_{t=1}^{T} \Lambda_t(x^t)\right\}$$

Adjustment Hazard for RI Value Gap ($\psi_1 = 3.34$)



RI hazard implied by best-fitting model

Optimal Adjustment Decision

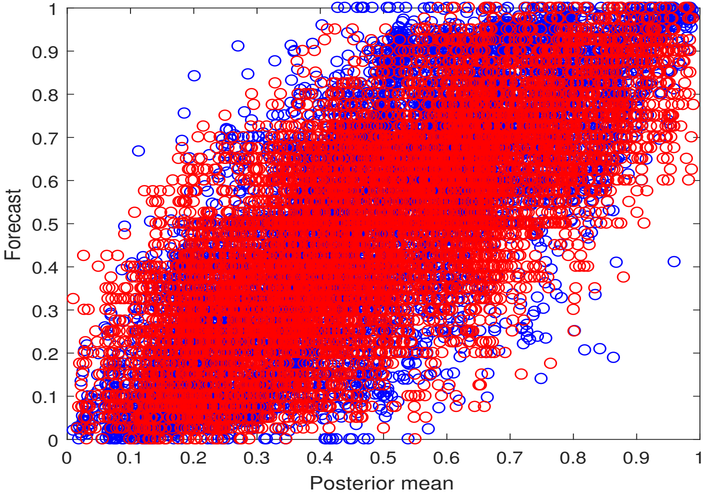
$$\mu_t(a) = \frac{\bar{\mu}(a)\exp\{V_t(a)/\psi_2\}}{\int d\bar{\mu}(\tilde{a})\exp\{V_t(\tilde{a})/\psi_2\}}$$

$V_t(p)$ is the continuation value function

$\bar{\mu}(a)$ is expected frequency of different actions, given the hazard function

$$\bar{\mu}(a) = \frac{E\left\{\sum_{t=1}^{T}\Lambda_t(x^t)\mu_t(x^t)\right\}}{E\left\{\sum_{t=1}^{T}\Lambda_t(x^t)\right\}}$$

Predicted Choices ($\psi_2 = 1.44$)



blue circles = data; red circles = prediction of RI model

Related Models of Stochastic Choice

- RI vs. soft max / control cost model for slider position choice

$$\mu_t(p) = \frac{\bar{\mu}(p) \exp\{V_t(p)/\psi_2\}}{\int d\bar{\mu}(\tilde{p}) \exp\{V_t(\tilde{p})/\psi_2\}} \quad \text{vs.} \quad \mu_t(p) = \frac{\exp\{V_t(p)/\kappa\}}{\int \exp\{V_t(\tilde{p})/\kappa\}}$$

$\rightarrow$ RI = soft max/control cost with a prior optimized to the context, rather than uniform

Related Models of Stochastic Choice

- Estimate best fitting reference distribution $\tilde{\mu}$ and proba $\tilde{\lambda}$

$$\mu_t(p) = \frac{\tilde{\mu}(p) \exp\{V_t(p)/\psi_2\}}{\int d\tilde{\mu}(\tilde{p}) \exp\{V_t(\tilde{p})/\psi_2\}} \quad \text{with} \quad \tilde{\mu}(p) = A\bar{\mu}(p)^\gamma$$

$\gamma = 1 \Rightarrow$ pure RI

$\gamma = 0 \Rightarrow$ pure soft max

$\tilde{\lambda} > \bar{\lambda} \Rightarrow$ preference for adjustment relative to pure RI

$\tilde{\lambda} < \bar{\lambda} \Rightarrow$ preference to adjust less frequently than pure RI

- Interpret $\gamma \in (0, 1), \tilde{\lambda} \neq \bar{\lambda}$ as generalized RI model with intrinsic preference for certain actions

- Estimate
  $\gamma = 0.45, \psi_1 = 0.64, \psi_2 = 1.23, \tilde{\lambda} = 11.4\% > \bar{\lambda} = 8.9\%$

Rationally Inattentive Behavior

- RI belongs to the “early noise” class of models of Bayes-optimal choices that condition on a noisy internal representation of the environment

- In RI models the nature of the noise is optimally adapted to a prior over possible situations

- RI decision-making
  - justifies discrete, stochastic choices as optimal way to save on information processing costs
  - justifies incomplete, delayed adjustment as optimal outcome when actions are based on imperfect information

References

Acharya, Sushant & Shu Lin Wee (2020), "Rational inattention in hiring decisions," *American Economic Journal: Macroeconomics* 12(1): 1–40.

Arimoto, Suguru (1972), "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory* 18(1): 14–20.

Berger, Toby (2003), "Rate-distortion theory," *Wiley Encyclopedia of Telecommunications* .

Blahut, Richard (1972), "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory* 18(4): 460–473.

Caplin, Andrew & Mark Dean (2013), "Behavioral implications of rational inattention with Shannon entropy," NBER worlking paper 19318.

Caplin, Andrew, Mark Dean & John Leahy (2019), "Rational inattention, optimal consideration sets, and stochastic choice," *The Review of Economic Studies* 86(3): 1061–1094.

Coibion, Olivier & Yuriy Gorodnichenko (2012), "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy* 120(1): 116–159.

Cover, Thomas M & Joy A Thomas (2006), *Elements of information theory 2nd Edition*, Wiley-Interscience.

Csiszár, Imre (1974), "On the computation of rate-distortion functions," *IEEE Transactions on Information Theory* 20(1): 122–124.

Fix, Stephen L (1978), "Rate distortion functions for squared error distortion measures," in *16th Annual Allerton Conference on Communication, Control and Computing, Monticello, Ill*, pp. 704–711.

Jung, Junehyuk, Jeong Ho Kim, Filip Matějka & Christopher A Sims (2019), "Discrete actions in information-constrained decision problems," *The Review of Economic Studies* 86(6): 2643–2667.

Khaw, Mel Win, Luminita Stevens & Michael Woodford (2017), "Discrete adjustment to a changing environment: Experimental evidence," *Journal of Monetary Economics* 91: 88–103.

Luo, Yulei (2008), "Consumption dynamics under information processing constraints," *Review of Economic dynamics* 11(2): 366–385.

Maćkowiak, Bartosz, Filip Matějka & Mirko Wiederholt (2018), "Dynamic rational inattention: Analytical results," *Journal of Economic Theory* 176: 650–692.

References (cont.)

Maćkowiak, Bartosz A. & Mirko Wiederholt (2009), "Optimal sticky prices under rational inattention," *The American Economic Review* 99(3): 769–803.

Matějka, Filip (2016), "Rationally inattentive seller: Sales and discrete pricing," *The Review of Economic Studies* 83(3): 1125–1155.

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

Morales-Jimenez, Camilo & Luminita Stevens (2022), "Nominal rigidities in U.S. business cycles," Working paper, University of Maryland.

Rose, Kenneth (1994), "A mapping approach to rate-distortion computation and analysis," *IEEE Transactions on Information Theory* 40(6): 1939–1952.

Shannon, Claude E. (1948), "A mathematical theory of communication," *The Bell System Technical Journal* 27(3): 379–423.

Shannon, Claude E. (1959), "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec* 4(142-163): 1.

Sims, Christopher A. (1998), "Stickiness," in *Carnegie-Rochester Conference Series on Public Policy*, vol. 49, pp. 317–356.

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

Steiner, Jakub, Colin Stewart & Filip Matějka (2017), "Rational inattention dynamics: Inertia and delay in decision-making," *Econometrica* 85(2): 521–553.

Stevens, Luminita (2020), "Coarse pricing policies," *The Review of Economic Studies* 87(1): 420–453.

Van Nieuwerburgh, Stijn & Laura Veldkamp (2010), "Information acquisition and under-diversification," *The Review of Economic Studies* 77(2): 779–805.

Woodford, Michael (2009), "Information-constrained state-dependent pricing," *Journal of Monetary Economics* 56(S): 100–124.