Introduction to Rational Inattention: motivation and applications

Filip Matějka

Sloan Nomis summer school

July, 2018

Motivation

- Imperfect and asymmetric information a foundation for many findings in economics.
- Available information is now more abundant than ever: beliefs not determined by what information is given to us, but by what we choose to attend to.
- We need a model of how we summarize, filter, and digest information,
    - Attention is scarce $\rightarrow$ mistakes.
    - We can choose what to attend to $\rightarrow$ optimal heuristics.
- Implications for: macro dynamics, labor matching, portfolio-allocation, consumer choices, politics, ...

Birth of rational inattention

Sims (1998,2003, 2006): sluggish dynamics in macro

MAIN BUILDING-BLOCKS:

- Available $\neq$ internalized information.
- Flexibility: agents select precision and nature of knowledge: form of mistakes subject to choice.
- Optimization subject to constraints on limited attention - a stretch, but natural benchmark - allows for discipline and policy analysis.

Action y, state x, utility $U(y, x)$:

If x observed: $y = \arg\max_{y'} U(y', x)$, i.e., $x \to y$
If form of noisy info given: $y = \arg\max_{y'} E[U(y', x)]$, i.e., $x \to f(y|x)$
RI, choice of info: choice of $f(y|x)$.

TWO STRANDS OF LITERATURE

1) Agents select precision of knowledge: linear-quadratic preferences, Gaussian uncertainty and signals.

Sims (1998,2003), Maćkowiak and Wiederholt(2009,2011), Van Nieuwerburgh and Veldkamp(2010), Mondria(2009,2010), Luo and Young(2009),...

2) Agents select nature of knowledge: general preferences and uncertainty.

Sims(2006), Woodford(2009,2015), Matějka(2010), Matějka and Sims(2010), Stevens (2014), Yang(2015), Matějka and McKay (2015), Caplin and Dean (2015), Ravid (2015), Caplin, Dean, and Leahy (2016),...

Simple example: amount of information 1/4

Objective, actions and uncertainty.

Manager sets price y to maximize profit

$$U(y, x) = -r\left(ax - y\right)^2 \tag{1}$$

less the cost of attention.

- x is unknown state (e.g., depends on random elasticity of demand or input cost),

$$x \sim N(0, \sigma_x^2),$$

- ax is the target price,
- parameter a is elasticity of the target to the shock x,
- parameter r scales stakes.

Simple example 2/4

Costly attention.

Refinement of knowledge about x: Gaussian signals,

$$s = x + \varepsilon, \text{ where } \varepsilon \sim N(0, \sigma_\varepsilon^2). \tag{2}$$

Precision is subject to choice.

More attention → more precise signal → more costly.

The cost function used in RI:

- $$cost = \lambda\kappa,$$

  $\lambda > 0$ is a parameter, and κ is the chosen amount of info.
- Amount κ: expected reduction of uncertainty (expressed by entropy) due to acquisition of signal s.

$$\kappa = H(x) - E\Big[H(x|s)\Big] = \frac{\log \sigma_x^2}{2\pi e} - \frac{\log \sigma_{x|s}^2}{2\pi e}.$$

Simple example 3/4

Decision problem.

The agent thus faces two choices in succession.
(i) How much attention to pay - the choice of posterior variance $\sigma^2_{x|s}$.
(ii) How to act upon posterior belief - the choice of optimal price y conditional on signal s.

The step (ii) is easy. $y = aE[x|s]$ maximizes the expectation of (1)

The choice of attention in (i) is given by:

$$\max_{\sigma^2_{x|s} \geq \sigma^2_x} E_x \left[E_s \left[-ra^2 (x - E[x|s])^2 \right] \right] - \lambda\kappa = \max_{\sigma^2_{x|s} \geq \sigma^2_x} \left(-ra^2 \sigma^2_{x|s} - \frac{\lambda}{2\pi e} \log \frac{\sigma^2_x}{\sigma^2_{x|s}} \right). \tag{3}$$

Simple example 4/4

Solution.

Bayesian updating:

$$E[x|s] = (1-\xi)\bar{x} + \xi s = \xi(x+\varepsilon),$$

$\xi \equiv \left(1 - \sigma^2_{x|s}/\sigma^2_x\right) \in [0,1]$ reflects the chosen level of attention.
Therefore,

$$y = (a\xi)x + \eta. \tag{4}$$

$\xi = 1$: attention, $y = ax$;
$\xi = 0$: no attention and no response to x.
The choice problem:

$$\max_{\xi\in[0,1]} \left(-ra^2(1-\xi)\sigma^2_x - \frac{\lambda}{2\pi e}\log(1-\xi)\right). \tag{5}$$

The solution is

$$\xi = max\left(0, 1 - \frac{\lambda}{ra^2\sigma^2_0}\right). \tag{6}$$

Basic implications of RI

$$y = \xi a x + \eta,$$
$$\xi = max\left(0, 1 - \frac{\lambda}{r a^2 \sigma_0^2}\right).$$

(i) Under-reaction: realized prices move less than optimal prices, $a\xi < a$.

(ii) Magnified relative elasticities: consider two products with elasticities $a_1 > a_2$; under RI the relative elasticities are $\frac{a_1 \xi(a_1)}{a_2 \xi(a_2)} > \frac{a_1}{a_2}$.

(iii) Uncertainty increases responses: ξ is increasing in the prior uncertainty σ_x^2.

(iv) Stakes and cost of information: higher stakes r and lower cost λ decrease responses.

MACRO: OPTIMAL STICKY PRICES 1/3

Mackowiak, Wiederholt (2009)

- Question: Individual prices change frequently and by large amounts. At the same time, the price level responds slowly to monetary policy shocks. How is this possible?

- Idea: Managers pay close attention to firm-specific conditions and thus prices respond quickly to idiosyncratic shocks, but managers pay less attention to aggregate conditions and therefore prices respond slowly to aggregate shocks.

- Furthermore: When prices are strategic complements, the fact that other firms pay limited attention to aggregate conditions reduces the incentive for an individual firm to pay attention to aggregate conditions.

- See also Woodford (2009): state-dependent pricing.

Macro: Optimal Sticky Prices 2/3

- Profit: $-\left(x_{it}-p_{it}\right)^2$
- Target price: $x_{it}=p_t+y_t+z_{it}$, aggregate price level: $p_t=\int_0^1 p_{it}di$
- Dynamics: $q_t=\rho_q q_{t-1}+\varepsilon_t^q, z_{it}=\rho_z z_{it-1}+\varepsilon_{it}^z$.
- Two signals: $s_{it}^q=q_t+\psi_{it}^q, s_{it}^z=z_{it}+\psi_{it}^z$
- Optimal allocation of attention:

$$\min_{\kappa_q,\kappa_z\geq 0}\left(\sigma_q^2 e^{-2\kappa_q}+\sigma_z^2 e^{-2\kappa_z}\right)$$

subject to

$$\kappa_q+\kappa_z\leq\kappa.$$

- Solution:

$$\kappa_q^*=\frac{1}{2}\kappa+\frac{1}{4}\ln\left(\frac{\sigma_q^2}{\sigma_z^2}\right)\quad \text{if } \frac{\sigma_q^2}{\sigma_z^2}\in\left[e^{-2\kappa},e^{2\kappa}\right].$$

Macro: Optimal Sticky Prices 3/3

- More volatile idiosyncratic conditions $\rightarrow$ firms pay more attention to idiosyncratic and prices react more quickly to them.
- Strategic complementarity in actions $\rightarrow$ strategic complementarity in information acquisition. Managers pay little attention to aggregates $\rightarrow$ other pay even less,...
- EVIDENCE; Mackowiak, Moench and Wiederholt (2009):
  - study U.S. sectoral price data,
  - prices respond faster to idiosyncratic shocks than to aggregate shocks,
  - prices respond more slowly to aggregate shocks in sectors with a higher volatility of idiosyncratic shocks.

Evidence: Expectations formation

Coibion, Gorodnichenko (2010, 2012)

- Data: U.S. Survey of Professional Forecasters (SPF), 1969-2010
- Great moderation: info rigidity declined in 70's until 83, and increasing afterwards (opp. to macro volatility).
- Differences in persistence explain 20-30% of differences in information rigidities.
- Endogeneity of rigidity: state vs time dependence
  - rigidity $\downarrow$ after a few quarters in a recession
  - rigidity $\downarrow\downarrow$ right after 9/11.

RI IN FINANCE 1/2

Van Nieuwerburgh, Veldkamp(2008): Under-Diversification

- Process information about returns of different assets

$$w = a_1 x_1 + a_2 x_2$$

- Risk management: learning and diversification.
- Under-diversification: agents process information about some assets only (and don't purchase the other ones).

Mondria(2009): Home bias, financial contagion:

- Fixed information constraint, shocks in two countries.
- Increased volatility in one country $\rightarrow$ more attention to that country, less to the other country $\rightarrow$ price decrease in both.
- Evidence: web search queries.

RI in Finance 2/2

Peng (2005)

- Correlated asset returns → info about common component;
- Attention to selected index;
- Co-movement of asset allocations.

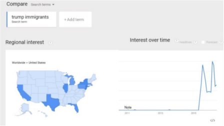
Evidence: Kacperczyk, Van Nieuwerburgh and Veldkamp (2016): stock-picking across business-cycle

- Increased volatility in recessions → more attention to aggregate.
- Data: covariance of each fund's portfolio holdings with the aggregate payoff shock indeed rises in recessions.

RI IN POLITICAL ECONOMY

Matejka, Tabellini (2016)

- Voters ignorant, but not in random, related to What is at Stake



How does selective attention interact with policy formation?
Magnified effect of stakes

- Small groups, extremists matter more (e.g., Stigler 1971);
- Minorities are more informed (Carpini & Keeter 1996);
- Public good underfunded; targeted tranfers/gifts;
- Policy fragmentation and info granularity: can be bad;
- Attention to divisive issues.

RI in Labor/Discrimination 1/3

Bartos, Bauer, Chytilova, and Matejka (2016)

- Statistical discrimination (Phelps 1972): weigh in group information.

$$E[q|s] = (1 - \xi)\bar{q} + \xi s,$$

  ξ: depends on precision of the individual-specific signal.
- Correspondence experiments (Bertrand and Mullainathan 2004)
  - minorities discriminated,
  - hypothesis: attention matters, implicit discrimination - minorities always paid less attention, automatic reaction.
- RI
  - statistical discrimination with endogenous precision (group signals affects precision of the individual signal),
  - correspondence field experiments: adaptable attention (minorities could be paid more attention).

RI in Labor/Discrimination 2/3

Experiment:

- Correspondence experiments + monitoring whether HR managers and landlords payed attention to applicants;
- Emails with hyperlink: "Dear Sir/Madam, I am writing because I am very interested in the Real Estate Agent job position advertised by your company. You can find my resume in this hyperlink: phanquyetnguyen1982.sweb.cz Best regards, Phan Quyet Nguyen."
- Additional treatment: "I have been looking for a job for two months [a year and a half], and I am writing because..."

Empirical findings:

- Minorities (Vietnamese, Roma) invited less.
- Minorities paid less attention on labor market, but more no rental market.
- Additional negative signal (unemployed) on the labor market decreases attention.

RI in Labor/Discrimination 3/3

Theory:

- Value of information increases when agent is less certain: accept vs reject.
- Labor market: very selective, cherry-picking (10% invitation rate), rental: thin, lemon-dropping (80% invitation rate)
- Labor market: more attention to better groups (higher prior), rental: to negatively stereotyped groups.

Implications:

- Attention selective, not implicit/fixed (econ vs psych - repeated decisions);
- RI $\rightarrow$ beliefs matter a lot (affect precision of all following signals) - name/sex-blind resumes? Long-term unemployment?
- Process data very useful.

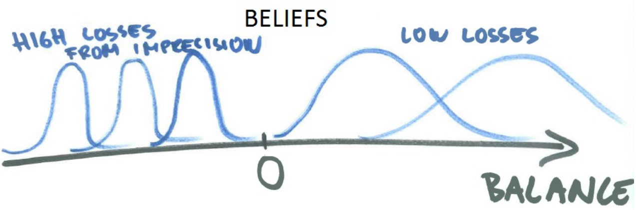
General RI: motivation 1/2

From how much information to what type of information.

Black-box of information acquisition (thinking)

- LQ loss: acquire info with equal precision across states,
- Unequel losses: unequal precision across states desirable.

  Example: information about credit-card balance

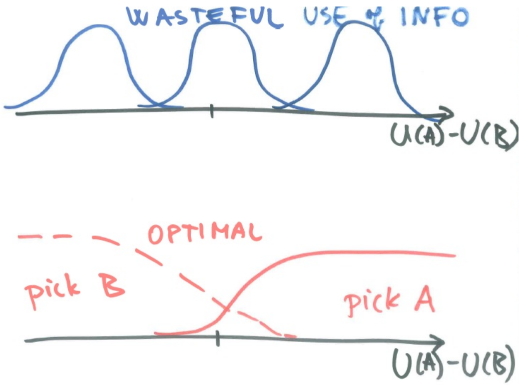


- Higher losses from imprecision when balance is negative,
- Procedure of information acquisition:
  1) Is the sign positive or negative (or black vs red color)?
  2) If negative, then read the balance in more detail.

GENERAL RI: MOTIVATION 2/2

- We simplify abundant information in all sorts of ways - RI provides a model (categorization, summarization,...)

  Example: binary choice: need to distinguish if $U(A) > U(B)$ only.



Rational Inattention: static model

here I follow Matějka, McKay (2015)

Indirect utility function $U(y,x)$;
Action y, unknown (random) state x, prior belief $g(x)$;

Agent's strategy $f(x,y)$: attention allocation and actions conditional on posterior beliefs.

$$\max_{f(\cdot,\cdot)} \int_x \int_y \quad U(x,y) f(x,y) dxdy - C(f), \tag{7}$$

$$\int_y f(x,y) dy = g(x) \qquad \forall x, \tag{8}$$

$$C(f) = \lambda \Big(H(x) - E[H(x|y)] \Big). \tag{9}$$

(1) maximization of expected utility,
(2) Bayesian consistency with prior $g(x)$,
(3) information constrain, $H[g(x)] = -\int g(x) \log g(x) dx$ is entropy.

Why joint distribution of states and actions?

$f(x, y)$ reflects the following choices (also Kamenica, Gentzkow 2011):

(1) What pieces of information (and how much),

- Choice of $f(s|x)$ defines signals,
- $f(x|s)$ is posterior belief,

(2) What action to take:

$$y = \arg\max_{y'} \int_x U(x, y') f(x|y \sim s) dx$$

$y \equiv s$: never two signals for one y (waste of information),

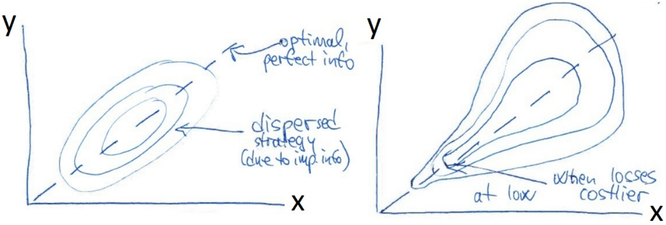


Figure: (a) $-(x-y)^2$, (b)higher losses for imprecision at lower x: $-\frac{(x-y)^2}{|x|}$

Important assumptions

Information is available in all forms

- Agents can design signals (and beliefs) in any way they like.
- As if agents could choose what questions to ask.

Entropy-based cost

- Reasonable: higher cost for higher precision,
- Axiomatic foundation: independence of sequential processing,
- Technology: agents receive blocks of signals,(information theory).

Optimized attention

- No cognitive frictions when choosing optimal strategy, which takes into account limited ability to process information,
- Costly re-optimization?

What RI model does not fit:

- If info is not fully available, e.g., if you need to walk to a store to find out if they sell a given product,
- If distance between states matters, e.g., if thermometer is used to find out about temperature.
- Some experiments where info is given in one form.

Solution: A generalized logit model

$$f(y|x) = \frac{e^{U(y,x)/\lambda+\alpha(y)}}{\int_z e^{U(z,x)/\lambda+\alpha(z)}dz}, \tag{10}$$

where $\alpha(y) = log(p(y))$.

The biases α describe the heuristics that the agent chooses to use, and reflect the agent's choice of attention.

- Bayesian updating + endogenous info structure;
- Summarize prior/choice set effects.

Main implications of RI 1/3

For simplicity: discrete choice, $i \in \{1..N\}$.

$$P(i|x) = \frac{e^{U(i,x)/\lambda+\alpha_i}}{\sum_j e^{U(j,x)/\lambda+\alpha_j}}, \quad \alpha_i = P(i) = E_x[P(i|x)]. \tag{11}$$

(i) Stochastic choice: RI agents make mistakes. Driven by cost function: $P(i) > 0 \Rightarrow P(i|x) > 0$;

(ii) Logistic choice and optimal heuristics (dimension $N-1$) are optimal for any preferences and prior beliefs.

- Tractable;
- Amenable to empirical work (McFadden 1974): effects of preferences and beliefs additively separable;
- Differences from logit: adjustable α's (dominated options never selected, IIA can be broken, ...);
- Dynamics: logit, too (Steiner et al. 2017).

Main implications of RI 2/3

$$P(i|x) = \frac{e^{U(i,x)/\lambda+\alpha_i}}{\sum_j e^{U(j,x)/\lambda+\alpha_j}}. \tag{12}$$

(iii) Linear-quadratic preferences and Gaussian prior uncertainty: Gaussian signals optimal.

(iv) Categorization, discreteness, and consideration sets. RI agents contemplate a low number of actions only, $P(i) > 0$. In continuous case: set of actions s.t. $p(y) > 0$ is discrete (Matějka and Sims 2010). An alternative explanation for adjustment cost: prices change between two levels only. See also Caplin, Dean, Leahy (2016)

Main implications of RI 3/3

(v) Violations of revealed preference. RI can imply choices that are seemingly irrational. If signals are endogenous to what options are presented to the agent, then regularity in choice can be violated (Woodford 2015, Matejka and McKay 2015)

(vi) Posterior invariance. If the number of possible states is no larger than the number of alternatives, then the set of possible posteriors the agents can acquire is independent of small changes to the prior (Caplin, Dean 2015);

(vii) Multi-dimensional simplification - indexation. If agents need to pay attention to several shocks and choose multiple actions, then RI models what simplified representation of the high-dimensional environment they use.

SOLVING FOR $\alpha_i = \log P_i^0$'S

$$\max_{\{\mathcal{P}_i^0\}_{i=1}^N} \int_{\mathbf{v}} \lambda \log \left(\sum_{i=1}^{N} \mathcal{P}_i^0 e^{v_i/\lambda} \right) G(d\mathbf{v}).$$

subject to $\forall i$: $\quad \mathcal{P}_i^0 \geq 0$ and $\sum_i \mathcal{P}_i^0 = 1$.
For all i such that $\mathcal{P}_i^0 > 0$:

$$\int_{\mathbf{v}} \frac{e^{v_i/\lambda}}{\sum_{j=1}^{N} \mathcal{P}_j^0 e^{v_j/\lambda}} G(d\mathbf{v}) = 1 \qquad (*).$$

see also Caplin, Dean, and Leahy (2016) who show: solution if and only if $LHS = 1$ in $(*)$ for i s.t. $\mathcal{P}_i^0 > 0$, and $LHS \leq 1$ if $\mathcal{P}_i^0 = 0$.

Effect of Co-movements 1/2

Logit: if and only if IIA(Independence from irrelevant alternatives): relative choice probabilities independent from choice set

IIA is counterintuitive, Debreu's (1960) example: train, yellow and red buses, choice probabilities should be $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$, not $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

RI: choice set is important, alters the form of logit
Problem:

- formalize "duplicate" actions as $Pr\left[v_i = v_j\right] = 1$
- then duplicate actions treated as a single action

$$\hat{\mathcal{P}}_i(\mathbf{u}) = \mathcal{P}_i(\mathbf{v}), \qquad \forall i < N \tag{13}$$

$$\hat{\mathcal{P}}_N(\mathbf{u}) + \hat{\mathcal{P}}_{N+1}(\mathbf{u}) = \mathcal{P}_N(\mathbf{v}), \tag{14}$$

- e.g. train = 1/2, yellow bus = 1/4, red bus = 1/4.

EFFECT OF CO-MOVEMENT 2/2, (EXAMPLE)

Problem:, payoffs: train=1/2, buses $\in \{0,1\}$, correlated (ρ),

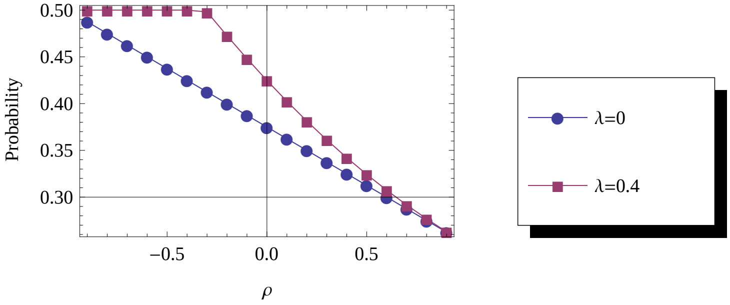


FIGURE: Unconditional probability of a bus.

- low ρ: agent compares buses only, does not compare relatively to train

THE WHOLE CHOICE SET MATTERS

- The model offers framework for studying many more perhaps interesting situations.
- Example: Regularity does not hold, RI-model cannot be represented by any RUM.

$P(A|A,B,C) > P(A|A,B)$, additional action helps an existing one

	state 1	state 2
action A	0	1
action B	1/2	1/2
action C	Y	$-Y$
prior prob.	$g(1)$	$g(2) = 1 - g(1)$

high $g(1)$ and Y

DISCRETENESS, CATEGORIZATION OF STATES

Jun, Kim, Matějka, and Sims (2013)

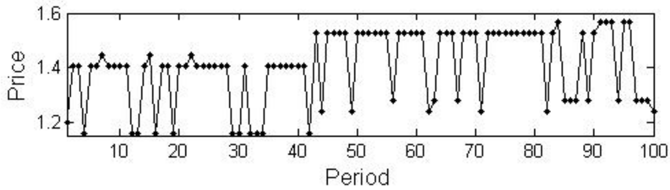
Result: action set is discrete even if continuous menu.

- if range of X, i.e., prior, is bounded,
- for tracking $U(x,y) = u(x-y)$: if $g(x) \neq \int f(y)e^{U(x,y)}dy$.

Intuition:

- discreteness allows for minimization of large errors,
- continuity: perfect balance between info about small and large scales.

Implications: categorization and inaction (as if adjustment cost)



See also Matějka (2016) and Stevens (2017)

Strategic setup: Coordination with RI 1/3

Yang (2015); see also Hellwig, Veldkamp (2009)

	invest	not invest
invest	θ, θ	$\theta - r, \theta$
not invest	$\theta, \theta - r$	0,0

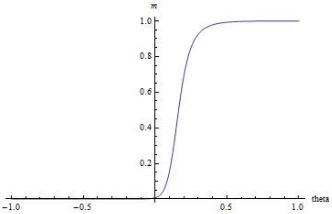
- 2 players, common prior, acting simultaneously
- players choose:
  - how to process information (about what θ)
  - how to act upon posterior knowledge
- players know strategy of the other players
- result: multiple Nash equilibria for low information cost, i.e. for precise signals (opposite to results in global games)

BINARY CHOICE WITH STRATEGIC INTERACTION 2/3

Payoff when the other player plays $P_{-1}(I|\theta)$:

	invest	not invest
payoffs:	$\theta + r(P_{-1}(I|\theta) - 1)$	0

$$P(I|\theta) = \frac{e^{\theta + r(P_{-1}(I|\theta)-1)/\lambda + K}}{e^{\theta + r(P_{-1}(I|\theta)-1)/\lambda + K} + 1}$$



- processes info about augmented value (logit in $\theta + r(P_{-1}(I|\theta) - 1)$)
- lower μ: higher strategic complementarity.

Multiplicity in RI 3/3

Uniqueness in global games: strategy given by threshold for exogenous signals (1D): invest right of the threshold (and contagion argument).

Multiplicity in RI:

- Strategy is given by a form of signals (2D): probability of investing for each θ, additional dimension on which to coordinate;
- 2 types of posteriors only (far apart);
- Lower λ: higher strategic complementarity, i.e. stronger responses of others to my actions (since they distinguish my actions better)

Example:

- As if the agent asked: "Is θ above θ_0?";
- Multiple Nash equilibria: $\theta_0 \in [0, r]$;
- Coordination on the form/position of the threshold-signal;
- With exogenous signal: not possible.

Summarization/indexation: Mental accounting 1/3

Matějka, Koszegi (2018)

Thaler (1985, 1999): mental accounts, like business accounts, are systems to record and summarize transactions, a way of aggregating large amounts of data to facilitate decision making by reducing the information load (but Prospect theory used instead).

Use RI, a model of optimal information aggregation:

- Two consumption goods that are relatively good substitutes,
- Agent uncertain about marginal utilities μ_1, μ_2, chooses consumptions,
- The question: what to get info (think) about?

$$U(c_1, c_2, \mu_1, \mu_2) = \left(\mu_1 c_1^{r_1} + \mu_2 c_2^{r_1}\right)^{r_2//r_1} - a(p_1 c_1 + p_2 c_2), \qquad (15)$$

Another example: Mental accounting 2/3

- The agent processes information about approx $\mu_1 - \mu_2$ only,
- For more goods and more complex preference structure: more accounts; very tight for low information, and broader for higher.

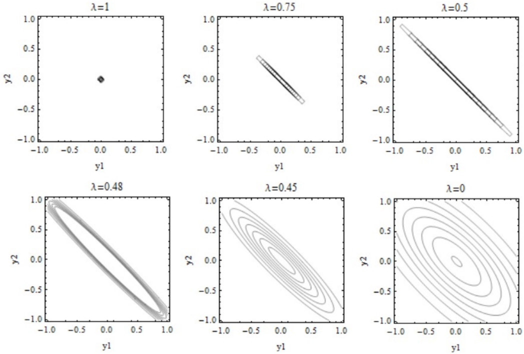


FIGURE: Joint distributions of consumption of 1 and 2, various costs of info.

Another example: Mental accounting 3/3

Intuition:

- Thinking about both μ_1 and μ_2 is a hard problem, e.g., how particular two types of education relate to utility from a vacation,
- Technically: uncertainty is reduced in the most important way only; the agent knows that substituting between c_1 and c_2 is most likely,
- RI magnifies elasticities,
- Optimal, endogenous, heuristic.

Methods: water-filling (engineering).

ESTIMATION

Joo (2017): can estimate U and α separately

- A.C.Nilsen supermarket data on prices, volumes sold, but also display;
- Assumption: display affects prior α, but not U;
- Result: 40% of quantity surcharges due to information, rest is preference.

Information Theory / Measuring information

Shannon (1948), Cover and Thomas (2006)
Consider sending series of 0/1.

- How many symbols do you need to describe a number in $\{1..2^N\}$? N.
- How many to describe one in $\{1..M\}$. On average: $\log_2(M)$.
- What if you could use decimal digits instead? then: $\log_{10}(M) = \xi \log_2(M) = -\log(p)$, were $p = 1/M$
- How many symbols would you on average need to describe a sequence of 0's and 1's, where 0's are more likely - encode (just like Morse's alphabet)
- you need a number of symbols equal to entropy

$$H = -E[\log(p)] = -\sum_i p_i \log(p_i),$$

where p_i is probability of state i, 0 or 1

Measuring information: Entropy

Entropy measures information content: how many symbols you on average need to send/receive to find out what the state is

- Independent of what the symbols are (0/1, digits, letters, pictures)
- Noise can be present - then you need more symbols, but entropy still proportional to the number. Entropy of a binary channel

$$H(p) = -p\log(p) - (1-p)\log(1-p).$$

- Entropy reduction measures information flow
  - random variable to learn about: X, prior $f(X)$
  - signal Y (sequence of symbols): posterior $f(X|Y)$
  - mutual information (proportional to the number of symbols needed):

$$I(X,Y) = H(X) - E[H(X|Y)]$$

ENTROPY

A good measure of cost when information is available

- when agent can choose signals / questions to ask
- when the information processing is repeated
- when there are "small" signals available to be pieced together

Example: not good when metric matters, e.g., when finding out temperature using a thermometer - given exogenous noise.

Shannon (1948): let there be N states, with pdf $p(i)$. How much do I learn when I find our that the state is i? Let $H(p)$ be the quantity. Three axioms imply that $H(p)$ is entropy.

- Continuity of $H(p)$,
- Monotonicity: for uniform p, $H(p)$ is increasing in N,
- Independence from intermediate steps, i.e., first you find which subgroup the state belongs to, and then wich state within the group.

See also: Csiszár (2008), de Oliveira (2014), and Hébert, Woodford (2016), Caplin, Dean, and Leahy (2018).

Differences from related approaches

(i) Signal-extraction, Lucas (1973): assumes a particular set of available signals, comparative statics are different.

(ii) Sticky-information, Manwik, Reis (2001): agents get all or no info. A very useful and tractable model. RI can, in contrast, explain differential responses to different shocks. Decision-making on the individual level is completely different.

(iii) Salience Bordalo et al.(2012): behavioral assumptions of what draws attention. Less related. Useful and simple.

(iv) Focusing Kőszegi, Szeidl (2013): assumes that agents put more weight on attributes that differ across various alternatives more; can be microfounded by RI. Useful and simple.

(v)Sparsity Gabaix (2014): agents choose costly loads of responses; RI builds on a connection to uncertainty, not as a behavioral assumption of low loads only. RI is also more general as it goes beyond linear loads, i.e., the logit model or discreteness. The assumption of sparsity can be derived from the quadratic-Gaussian version of RI. Useful and

SUMMARY

- Theory:
  - Flexible and selective (available) info acquisition;
  - LQ: under-reactions, magnified relative elasticity, effects of stakes and volatility;
  - General RI: many behavioral features.
- Empirical evidence (more is needed!):
  - Action: volatility and stakes (prices and portfolio allocations);
  - Expectations: effect of volatility (surveys);
  - Monitoring attention: effect of stakes (search data, audit study);
  - Lab experiments: detailed properties.
- For policy: realistic, endogenous, includes costs.

Still barely only half way there...