Introduction to Bounded Rationality and Limited Information

Mark Dean

Behavioral Economics G6943
Fall 2022

- In the first few lectures we are going to be focusing on the topics of bounded rationality
- And, in particular, limited attention
- Here I am going to offer an introduction to (how I understand) both topics

What is Bounded Rationality?

- Start with a 'standard' economic model
  - e.g. expected utility maximization

$$C(A) = \arg\max_{p \in A} \sum_{x \in X} p(x)u(x)$$

- If the model is wrong how can we adjust it?
- Two 'minimal' adjustments we could make
  1. Modify objective
  2. Modify constraints
- Most of behavioral economics concerned with approach 1
  - Loss aversion
  - Ambiguity aversion
  - etc
- Bounded rationality concerned with approach 2
  - Optimal behavior within some additional costs/constraints

What is Bounded Rationality?

- Costs to acquiring or processing information
  - E.g. Simon [1955], Stigler [1961], Sims [2003]
- Limits on reasoning
  - E.g. Camerer [2004], Crawford [2005]
- Thinking Aversion
  - E.g. Ergin and Sarver [2010], Ortoleva [2013]
- Bounded memory
  - E.g. Wilson [2014]
- Semi-Rational Models
  - E.g. Gabaix et al. [2008], Esponda [2008], Rabin and Vayanos [2010], Gabaix [2013],
- Heuristics
  - Tversky and Kahneman [1974], Gigerenzer [2000]

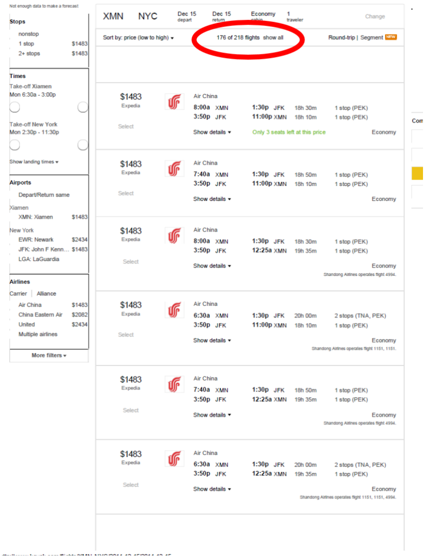
Advantages and Disadvantages of Bounded Rationality

- Advantage:
  - Intuitive plausibility
    - Evolution equipped us to optimize within constraints
  - Can 'microfound' behavioral models
  - Leads to new predictions: how behavioral phenomena can change with the environment
- Disadvantages:
  - May be wrong!
  - What is correct constraint?
  - Regress issue

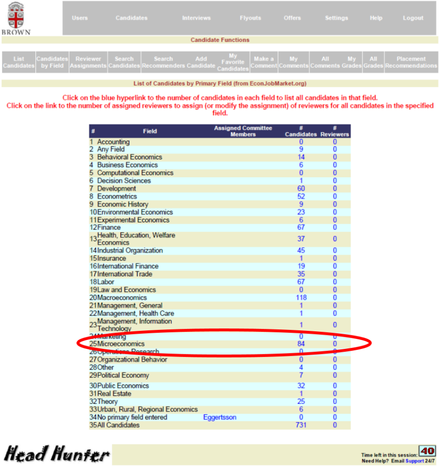
- For this course I am going to focus on one particular constraint on decision making:
    - Understanding the world is hard!
- More specifically, there is an enormous about of information out there that may be relevant for our choices
- It can be hard/impossible to process all of it
    - Even if it is 'freely' available
- This means there is likely to be a gap between the 'true' state of the world and that perceived by the decision maker

- This is
  - Fairly obvious through introspection

Am I fully Informed?



Am I fully informed?



BROWN

Users | Candidates | Interviews | Flyouts | Offers | Settings | Help | Logout

Candidate Functions

List Candidates | Candidates by Field | Reviewer Assignments | Search Candidates | Search Recommenders | Add Candidate | My Favorite Candidates | Make a Comment | My Comments | All Comments | My Grades | All Grades | Placement Recommendations

List of Candidates by Primary Field (from EconJobMarket.org)

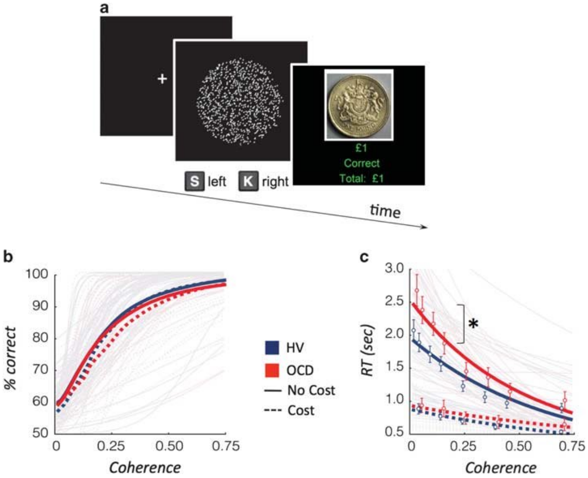
Click on the blue hyperlink to the number of candidates in each field to list all candidates in that field.
Click on the link to the number of assigned reviewers to assign (or modify the assignment) of reviewers for all candidates in the specified field.

#	Field	Assigned Committee Members	# Candidates	# Reviewers
1	Accounting		0	0
2	Any Field		9	0
3	Behavioral Economics		14	0
4	Business Economics		6	0
5	Computational Economics		0	0
6	Decision Sciences		1	0
7	Development		60	0
8	Econometrics		52	0
9	Economic History		9	0
10	Environmental Economics		23	0
11	Experimental Economics		6	0
12	Finance		67	0
13	Health, Education, Welfare Economics		37	0
14	Industrial Organization		45	0
15	Insurance		1	0
16	International Finance		19	0
17	International Trade		35	0
18	Labor		67	0
19	Law and Economics		0	0
20	Macroeconomics		118	0
21	Management, General		1	0
22	Management, Health Care		1	0
23	Management, Information Technology		1	0
24	Marketing		0	0
25	Microeconomics		84	0
26	Operations Research		0	0
27	Organizational Behavior		0	0
28	Other		4	0
29	Political Economy		7	0
30	Public Economics		32	0
31	Real Estate		1	0
32	Theory		25	0
33	Urban, Rural, Regional Economics		6	0
34	No primary field entered	Eggertsson	0	0
35	All Candidates		731	0

Head Hunter

Time left in this session: 40
Need Help? Email Support 24/7

- This is
  - Fairly obvious through introspection
  - Well documented in psychology experiments


a
+
S left
K right
£1
Correct
Total: £1
time
b
% correct
100
90
80
70
60
50
0
0.25
0.50
0.75
Coherence
HV
OCD
No Cost
Cost
c
RT (sec)
3.0
2.5
2.0
1.5
1.0
0.5
*
0
0.25
0.50
0.75
Coherence

- This is
  - Fairly obvious through introspection
  - Well documented in psychology experiments
  - Documented in economics experiments

Caplin Dean and Martin [2012]

4	2
3	13
20	11
15	8
8	10

Caplin Dean and Martin [2012]

4+6+10-11-23+9	2+3+6-11-14+9+10
3+9-17-99+102-6+15	6+18-19-55+70
20-27+7-19+2+3-5	11+2-5+7-8-9+10
15-5-5+6+16+17-20-9	8+9+10-11+8+2+6-32
8+8+9-13-9-6+7	10-9+17-23+10+2+15

Choice Objects

- Subjects choose between 'sums'

four plus eight minus four

- Value of option is the value of the sum
- 'Full information' ranking obvious, but uncovering value takes effort
- 6 treatments
  - 2 x complexity (3 and 7 operations)
  - 3 x choice set size (10, 20 and 40 options)
- No time limit

Size 20, Complexity 7

- ( ) zero
- ( ) seven minus four minus two minus four minus two plus eleven minus four
- ( ) six plus five minus eight plus two minus nine plus one plus four
- ( ) seven minus two minus four plus three plus four minus three minus three
- ( ) seven plus five minus two minus two minus three plus zero minus two
- ( ) six plus seven plus six minus two minus six minus eight plus four
- ( ) six plus two plus five minus four minus two minus seven plus three
- ( ) six minus four minus one minus one plus five plus three minus six
- (•) two plus six plus seven minus two minus four minus two plus zero
- ( ) two minus three minus five plus nine minus one plus five minus three
- ( ) three plus zero plus two plus zero plus one minus three minus one
- ( ) four plus three plus zero minus two plus three plus four minus ten
- ( ) seven plus two plus seven minus seven plus three minus two minus two
- ( ) three plus three minus two plus zero plus zero minus four plus five
- ( ) two minus two plus zero plus nine minus two minus one minus one
- ( ) three plus four minus three plus three minus four plus three minus four
- ( ) three plus five plus seven plus five minus two minus seven minus ten
- ( ) three plus six minus eight plus one plus two minus two plus zero
- ( ) three plus five plus zero plus four plus three minus four minus two
- ( ) eight minus one plus one minus four minus four minus five plus six
- ( ) four minus five plus four minus one minus four plus zero plus four

Finished

Failure rates (%) (22 subjects, 657 choices)

Failure rate		
	Complexity	
Set size	3	7
10	7%	24%
20	22%	56%
40	29%	65%

Average Loss ($)		
	Complexity	
Set size	3	7
10	0.41	1.69
20	1.10	4.00
40	2.30	7.12

- This is
  - Fairly obvious through introspection
  - Well documented in psychology experiments
  - Documented in economics experiments
  - The most straightforward explanation for many 'mistakes'

- Abaluck and Gruber: "Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program" [2011]

"Our findings are striking: along three dimensions, elders are making choices which are inconsistent with optimization under full information. First, elders place much more weight on plan premiums than they do on the expected out of pocket costs that they will incur under the plan. Second, they substantially under-value variance reducing aspects of alternative plans. Finally, consumers appear to value plan financial characteristics far beyond any impacts on their own financial expenses or risk. These findings are robust to a variety of specifications and econometric approaches."

Examples

- Chetty et al: "Salience and Taxation" [2009]
  - Prices are usually posted net of sales tax
  - Price is added a register
  - Adding a tag that includes the post tax price should be an 'inconsequential' change in the product
  - Does it affect choice?
- Experiment
  - Take 1 large supermarket
    - 30% of products have sales tax of 7.375% added at register
  - Take three 'impulse purchase' product categories
    - Cosmetics, hair care accessories, deodorants
    - 750 products in total
  - Add tags which displayed post tax price (as well as pre tax price)
    - Experiment lasted 3 weeks

Examples

TABLE 3
Effect of Posting Tax-Inclusive Prices: DDD Analysis of Mean Quantity Sold

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

Notes: Each cell shows mean quantity sold per category per week, for various subsets of the

Examples

- Basic message of these first two papers is that 'people screw up' relative to full information benchmark
- Other examples include:
  - Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment." American Economic Review, 105 (11): 3489-3529.
  - Saurabh Bhargava, George Loewenstein, and Justin Sydnor. Choose to lose: Health plan choices from a menu with dominated option. The Quarterly Journal of Economics, 132(3):1319|1372, 2017.
  - Benjamin R Handel and Jonathan T Kolstad. Health insurance for" humans": Information frictions, plan choice, and consumer welfare. American Economic Review, 105(8):2449{2500, 2015.
  - Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel. 2012. "Comparison friction: Experimental evidence from Medicare drug plans." The Quarterly Journal of Economics 127, no. 1: 199-235.

- This is
  - Fairly obvious through introspection
  - Well documented in psychology experiments
  - Documented in economics experiments
  - The most straightforward explanation for many 'mistakes'
  - Also the most straightforward explanation for ‘sluggishness’

- In most standard models, variables that are easy to adjust should jump immediately in response to news
- E.g.
    - Prices
    - Consumption
- A long literature has identified that this often doesn't happen
- For example Boivin, Jean, Marc P. Giannoni, and Ilian Mihov. "Sticky prices and monetary policy: Evidence from disaggregated US data." American economic review 99.1 (2009): 350-84
    - Look at responses of consumer (PCI) and producer (PPI) prices to different types of shock

Sluggish Price Responses (Boivin et al. 2009)

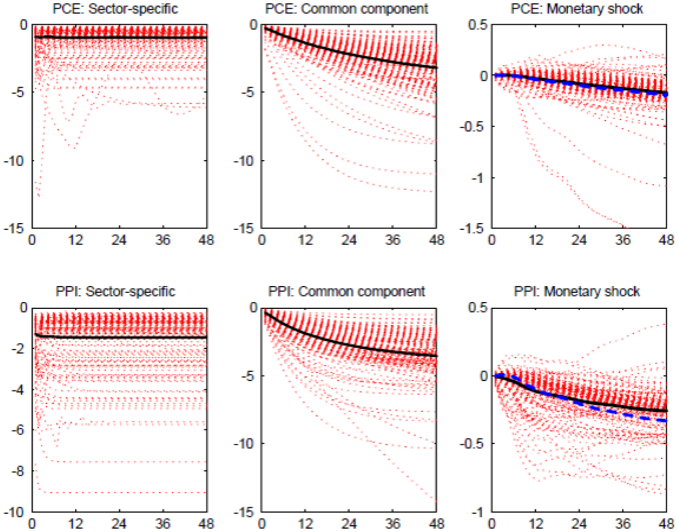


Figure 2. Sectoral price responses to various shocks

Introduction

- This is
  - Fairly obvious through introspection
  - Well documented in psychology experiments
  - Documented in economics experiments
  - The most straightforward explanation for many important economic behaviors
  - The most straightforward explanation for many 'mistakes'
  - Also the most straightforward explanation for 'sluggishness'
  - A possible unifying factor for many behavioral economic phenomena

Behavioral Economics as Limited Attention

- As you will see from this week's reading, if you squint you can use inattention to explain
  - Existence of shrouded attributes
  - Inattention to taxes
  - Nominal Illusion
  - Hyperbolic discounting
  - Prospect theory
  - Projection bias
  - Base rate neglect
  - Correlation neglect
  - Overconfidence
  - Left digit bias.....

My Take

- Limited attention is absolutely ubiquitous
- It is always the case that there is more potentially relevant information than we can (or should) process
- We are always making decisions based on a restricted data set
- The data set a decision maker uses is not (easily) observable to the outside researcher

My Take

- This leads to a number of first order important questions

1. How is the information that people use determined?
   - Do they selected it rationally?
   - Is it determined by features of the environment such as salience?
   - Do they use simplifying heuristics?
2. How should we adjust our economic models to take limited attention into account?
   - This question could be asked in pretty much an field you care to imagine
   - Currently mainly done in macro and a bit in IO
3. What are the normative implications?
   - Choice no longer equals preference
   - If attention is costly this should be taken into account
   - Are more options always better?

A Model of Bayesian Decision Noise

- Over the next few weeks you are going to see a wide variety of different ways of thinking about limited attention
- One simple workhorse model is that of a Bayesian decision maker who receives a noisy signal
- This all becomes very easy if everything is normal
- Gives rise to a functional form which you will come across a lot
    - Including in the Gabaix reading from this week

- Consider a DM who is trying to make inference about the value of a stimulus x
- Prior to receving any information they believe

$$x \sim N(\mu, \sigma_p^2)$$

- They then receive a signal s which they believe to be equal to

$$\begin{aligned} s &= x + \varepsilon \\ \text{where } \varepsilon &\sim N(0, \sigma_\varepsilon^2) \end{aligned}$$

- What should the beliefs of the DM be after observing s?

The Solution

- Recall from Bayes rule that

$$f(x|s) = \frac{f(s|x)f(x)}{f(s)}$$

Where

$$f(s|x) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(\frac{-(s-x)^2}{2\sigma_\varepsilon^2}\right)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma_p^2}\right)$$

and $f(s)$ is a constant

- Note that nothing outside the exponent is a function of x and so are normalizing constants

- Concentrating on the exponent bits of $f(s|x)f(x)$ gives us

$$
\begin{aligned}
& \exp\left(\frac{-(s-x)^2}{2\sigma_\varepsilon^2}+\frac{-(x-\mu)^2}{2\sigma_p^2}\right) \\
= \; & \exp\left(-\frac{s^2}{2\sigma_\varepsilon^2}-\frac{\mu^2}{2\sigma_p^2}\right)\exp\left(\frac{2xs-x^2}{2\sigma_\varepsilon^2}+\frac{2\mu s-x^2}{2\sigma_p^2}\right)
\end{aligned}
$$

- Concentrating on the term within the second exponential we have

$$
\begin{aligned}
& \frac{1}{2}\left(\frac{\sigma_p^2 2xs+2\sigma_\varepsilon^2 x\mu-(\sigma_p^2+\sigma_\varepsilon^2)x^2}{\sigma_\varepsilon^2\sigma_p^2}\right) \\
= \; & \frac{1}{2}\left(\frac{\frac{\sigma_p^2 s+\sigma_\varepsilon^2\mu}{(\sigma_p^2+\sigma_\varepsilon^2)}2x-x^2}{\frac{\sigma_\varepsilon^2\sigma_p^2}{(\sigma_p^2+\sigma_\varepsilon^2)}}\right)
\end{aligned}
$$

$$\frac{1}{2}\left(\frac{\frac{\sigma_p^2 s+\sigma_\varepsilon^2\mu}{(\sigma_p^2+\sigma_\varepsilon^2)}2x-x^2}{\frac{\sigma_\varepsilon^2\sigma_p^2}{(\sigma_p^2+\sigma_\varepsilon^2)}}\right)$$

$$=\quad\frac{1}{2}\frac{-\left(\frac{\sigma_p^2 s+\sigma_\varepsilon^2\mu}{(\sigma_p^2+\sigma_\varepsilon^2)}-x\right)^2+\left(\frac{\sigma_p^2 s+\sigma_\varepsilon^2\mu}{(\sigma_p^2+\sigma_\varepsilon^2)}\right)^2}{\frac{\sigma_\varepsilon^2\sigma_p^2}{(\sigma_p^2+\sigma_\varepsilon^2)}}$$

- As the latter term is again constant in x we can factor it out of the exponent, and we end up with

$$f(x|s)\propto\exp\left(\frac{-\left(\frac{\sigma_p^2 s+\sigma_\varepsilon^2\mu}{(\sigma_p^2+\sigma_\varepsilon^2)}-x\right)^2}{\frac{\sigma_\varepsilon^2\sigma_p^2}{(\sigma_p^2+\sigma_\varepsilon^2)}}\right)$$

- All of which gives the TL:DR that, conditional on observing s, the DM's posteriors will be distributed normally
- With a mean given by

$$\beta s + (1 - \beta)\mu$$
$$\text{for } \beta = \frac{\sigma_p^2}{(\sigma_p^2 + \sigma_\varepsilon^2)}$$

and variance given by

$$\frac{\sigma_\varepsilon^2 \sigma_p^2}{(\sigma_p^2 + \sigma_\varepsilon^2)}$$

- While specific to the normal case, this formula has a lot of nice features
  - People partially respond to new information
  - But on average a population will be bias towards prior mean
    - As $E(s) = x$, $E(E(x|s)) = \beta x + (1 - \beta)\mu$
  - Size of the bias depends on the perceived noise in the signal and in the prior
  - The more informed people are (as in lower σ_ε^2) the less bias they become
- Partly because of these features, this set up crops up a lot in the Gabaix paper
  - And other examples we will see

Risk Aversion through Noisy Encoding

- One nice application of similar ideas is in Khaw et al [2020]
  - Featuring our very own Mike Woodford
- Basic idea: Noisy encoding of numeric values can lead to risk aversion
  - Even if the objective is to choose the alternative with the highest expected value

Risk Aversion through Noisy Encoding

- Key assumption: DM receives a normal signal about the **log** value of the prize
- Why log?
- Because there is a lot of evidence that the absolute degree of error increases with the size of the signal
- Roughly speaking, the ability to differentiate between two signals n_1, n_2 is constant in $\frac{n_1}{n_2}$
    - Weber's law

Risk Aversion through Noisy Encoding

- If we assume that the DM receives signals r about a quantity n distributed according to

$$r \sim N(\log n, v^2)$$

- Then, if the DM receives signals r_1, r_2 about n_1, n_2, The probability that r_2 is higher is given by the probability

$$r_2 - r_1 > 0$$

- Which is distributed according to $N(\log n_2/n_1, 2v^2)$
  - Only n_2/n_2 matters
- Practically, using a log scale may be a way to help getting wide ranges of variables onto a bounded scale

Risk Aversion through Noisy Encoding

- Assume that, when faced with a lottery, the DM gets a signal about the value of each non zero prize x distributed according to

$$r \sim N(\log x, v^2)$$

- This is then combined with a prior belief distributed according to which x is distributed according to

$$\log x \sim N(\mu, \sigma^2)$$

- Using arguments similar to those we went through above

$$\log x | r \sim N(\bar{\mu}, \bar{\sigma}^2)$$

where

$$\begin{aligned} \bar{\mu} &= \beta r + (1-\beta)\mu \\ \beta &= \frac{\sigma^2}{\sigma^2 + v^2} \end{aligned}$$

Risk Aversion through Noisy Encoding

- Now imagine a DM choosing between C for sure and X with probability p
  - Assume that prior beliefs about X and C are the same
  - And that p is encoded without noise
- Then, for a given pair of signals r_X and r_C we have

$$
\begin{aligned}
E(X|r_X) &= e^{(1-\beta)\mu} e^{\beta r_x} \\
E(C|r_C) &= e^{(1-\beta)\mu} e^{\beta r_c}
\end{aligned}
$$

- If we assume that the DM is an expected value maximizer, then they will choose the risky prospect if

$$\log p + \beta r_x > \beta r_c$$

Risk Aversion through Noisy Encoding

- The risky prospect will be chosen if

$$r_x - r_c > \frac{\log p^{-1}}{\beta}$$

- We know that $r_x - r_c$ is distributed according to $N(\log(X/C), 2v^2)$
- So x will be chosen with 50% probability when

$$\begin{aligned} \log(X/C) &= \log p^{-\frac{1}{\beta}} \\ p^{\frac{1}{\beta}} X &= C \end{aligned}$$

- As $p \in (0,1)$ and $\beta \in (0,1)$ $p^{\frac{1}{\beta}} < p$ and so $p^{\frac{1}{\beta}} X < pX \rightarrow$ risk aversion

Risk Aversion through Noisy Encoding

$$p^{\frac{1}{\beta}} X = C$$

- As $\beta \longrightarrow 1$ risk neutrality
- As $\beta \longrightarrow 0$ very high risk aversion
- What is β?

  $$\frac{\sigma^2}{\sigma^2 + v^2}$$

  Where v is the noise in the signal
- So high v implies both more risk aversion and more stochastic choice
  - Which is what Khaw et al. find.