Rational Inattention Lecture 5

Mark Dean

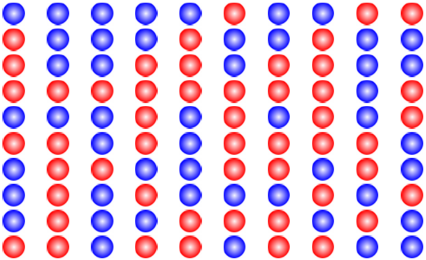
Behavioral Economics G6943
Fall 2022

- We have now described the mechanics behind the rational inattention model
- We are not going to talk through some experimental evidence
  - General model
  - Shannon model
- And some applications
  - Attention to quality
  - Discrimination
  - Elections
  - Dynamic Rational Inattention

- Introduce an experimental interface that can be used to collect state dependent stochastic choice data
- Will perform some specific tests based on the theory we have just seen
- But basic framework has (i think) a lot of mileage left in it
  - More taxing tests of the RI model
  - Test of any other model of limited attention you can think of

- We will perform some tests of both the general and Shannon models

1. Spillovers
    - RI vs EUM
2. Change in payoffs
    - RI vs Signal Extraction
    - Test ILR of Shannon model
3. Change in priors
    - Locally Invariant Posteriors
4. Many States
    - Test Invariance under Compression



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
- Prizes paid in probability points

An Aside: Testing Axioms with Stochastic Data

- Much of the following is going to come down to testing axioms of the following form

$$P(a|1) \geq P(a|2)$$

- These are conditions on the **population** probabilities
- We don't observe these, instead we observe **sample** estimates $\bar{P}(a|1)$ and $\bar{P}(a|2)$
- What to do?

An Aside: Testing Axioms with Stochastic Data

- We can make **statistical statements** about the validity of the axioms
- But there are two was to do this
  1. Can we reject a **violation** of the axiom
     - i.e., is it the case that $\bar{P}(a|1) > \bar{P}(a|2)$ and we can reject the hypothesis that $P(a|1) = P(a|2)$ at (say) the 5% level
  2. Can we find a **significant** violation of the axiom
     - i.e. is it the case that $\bar{P}(a|1) < \bar{P}(a|2)$ and we can reject the hypothesis that $P(a|1) = P(a|2)$ at (say) the 5% level
- (1) Is clearly a much tougher test that (2)
- If we have low power we will never be able to do (1)

- Recall that RUM implies monotonicity
  - For any two decision problems $\{A, A \cup b\}$, $a \in A$ and $b \notin A$

$$P_A(a|\omega) \geq P_{A\cup b}(a|\omega)$$

- Rational Inattention can lead to violations of monotonicity

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

- Does this happen in practice?

Table 1: Experiment 1						
	Payoffs					
DP	$U(a,1)$	$U(a,2)$	$U(b,1)$	$U(b,2)$	$U(c,1)$	$U(c,2)$
1	50	50	b_1	b_2	n/a	n/a
2	50	50	b_1	b_2	100	0

Table 2: Treatments for Exp. 1		
Treatment	Payoffs	
	b_1	b_2
1	40	55
2	40	52
3	30	55
4	30	52

Table 8: Results of Experiment 1

		$P(b\|1)$			$P(b\|2)$		
Treat	N	$\{a, b\}$	$\{a, b, c\}$	Prob	$\{a, b\}$	$\{a, b, c\}$	Prob
1	7	2.9	6.8	0.52	50.6	59.8	0.54
2	7	5.7	14.7	0.29	21.1	63.1	0.05
3	7	9.5	5.0	0.35	21.4	46.6	0.06
4	7	1.1	0.8	0.76	19.9	51.7	0.02
Total	28	4.8	6.6	0.52	28.3	55.6	$<$0.01

Expansion

- How does information gathering change with incentives?
- Simplest possible design: two states and two acts
- Change the value of choosing the correct act
- Can test
  - NIAS
  - NIAC
  - ILR

Experiment 2				
Decision Problem	Payoffs			
	$U(a,1)$	$U(a,2)$	$U(b,1)$	$U(b,2)$
1	5	0	0	5
2	40	0	0	40
3	70	0	0	70
4	95	0	0	95

- States equally likely
- Increase the value of making the correct choice
    - Payment in probability points
- 52 subjects

- In the symmetric 2x2 case, NIAS and NIAC have specific forms
- NIAS:

$$P_A(a|\omega_1) \geq \max\left\{\alpha P_A(a|\omega_2), \alpha P_A(a|\omega_2) + \beta\right\}, \qquad (1)$$

where

$$\alpha = \frac{u(b(\omega_2)) - u(a(\omega_2))}{u(a(\omega_1)) - u(b(\omega_1))}$$

$$\beta = \frac{u(a(\omega_1)) + u(a(\omega_2)) - u(b(\omega_1)) - u(b(\omega_2))}{(a(\omega_1)) - u(b(\omega_1))}$$

- In this case boils down to

$$P(a|\omega_1) \geq P(a|\omega_2)$$

- NIAC:

$$
\begin{aligned}
& \Delta P(a|\omega_1)\left(\Delta\left(u(a(\omega_1)) - u(b(\omega_1))\right)\right) + && (2)\\
& \Delta P(b|\omega_2)\left(\Delta\left(u(b(\omega_2)) - u(a(\omega_2))\right)\right) && (3)\\
\geq\ & 0 && (4)
\end{aligned}
$$

- In this case boils down to

$$
\begin{aligned}
& P_1(a|\omega_1) + P_1(b|\omega_2)\\
\leq\ & P_2(a|\omega_1) + P_2(b|\omega_2)\\
\leq\ & P_3(a|\omega_1) + P_3(b|\omega_2)\\
\leq\ & P_4(a|\omega_1) + P_4(b|\omega_2)
\end{aligned}
$$

Do People Optimally Adjust Attention?

- Alternative model: Choose optimally conditional on fixed signal
  - e.g. Signal detection theory [Green and Swets 1966]
- In general, choices can vary with incentives
  - Changes optimal choice in posterior state
- But not in this case
  - Optimal to choose *a* if $\gamma_1 > 0.5$, regardless of prize
- Change in choice between decision problems rules out Signal Detection Theory
  - Also rational inattention with fixed entropy

Testing NIAS: Experiment 1

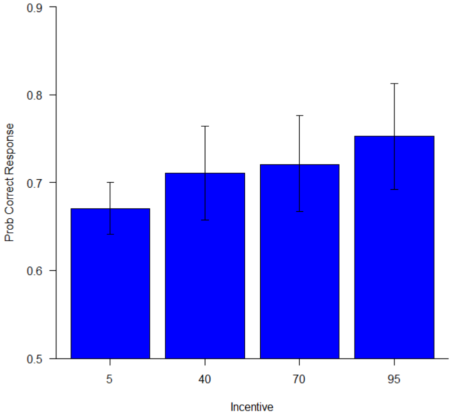
- NIAS test: For each decision problem

$$P(a|1) \geq P(a|2)$$

- From the aggregate data

Table 2: NIAS Test			
DP	$P_j(a\|1)$	$P_j(a\|2)$	Prob
1	0.74	0.40	0.000
2	0.76	0.34	0.000
3	0.78	0.34	0.000
4	0.78	0.27	0.000

Testing NIAC: Experiment 1

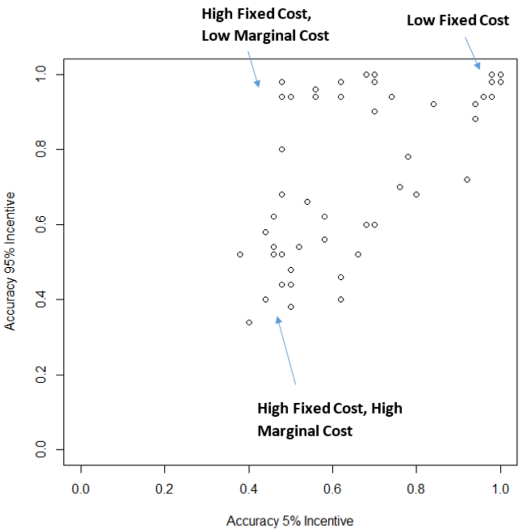


NIAC And NIAS: Individual Level

Violate	%
NIAS Only	2
NIAC Only	17
Both	0
Neither	81

- Counting only statistically significant violations

Recovering Costs - Individual Level



Invariant Likelihood Ratio and Responses to Incentives

- We can also use the same data to test a key implication of the Shannon model
  - Invariant Likelihood Ratio
- For chosen actions our condition implies

$$\frac{u(a(\omega)) - u(b(\omega))}{\ln \bar{\gamma}^a(\omega) - \ln \bar{\gamma}^b(\omega)} = \lambda$$

- Constrains how DM responds to changes in incentives

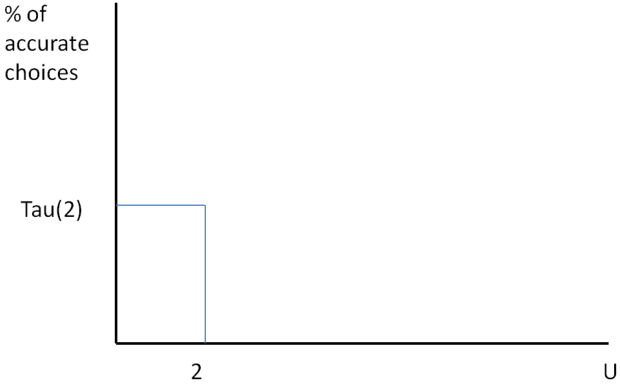
Invariant Likelihood Ratio - Example

Experiment 2				
Decision Problem	Payoffs			
	$U(a,1)$	$U(a,2)$	$U(b,1)$	$U(b,2)$
1	5	0	0	5
2	40	0	0	40
3	70	0	0	70
4	95	0	0	95

$$\frac{5}{\ln \bar{\gamma}^a(5) - \ln \bar{\gamma}^b(5)} = \frac{40}{\ln \bar{\gamma}^a(40) - \ln \bar{\gamma}^b(40)} = ... = \lambda$$

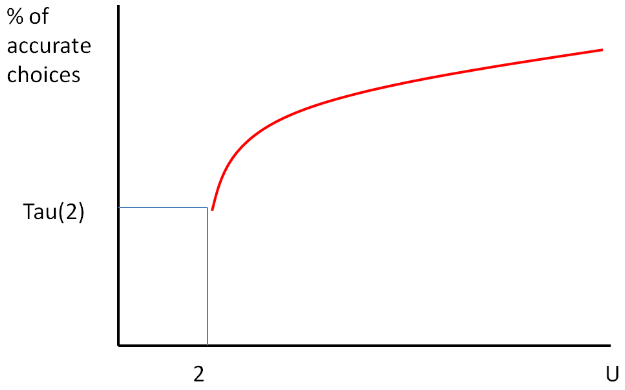
- One observation pins down λ
- Determines behavior in all other treatments

Invariant Likelihood Ratio - Example



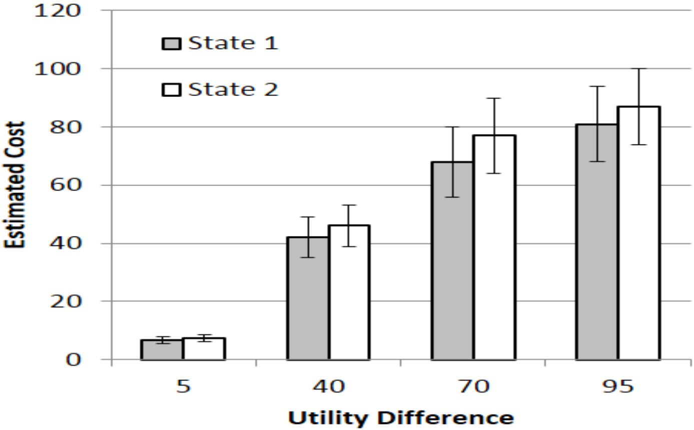
- Observation of choice accuracy for $x = 5$ pins down λ

Invariant Likelihood Ratio - Example

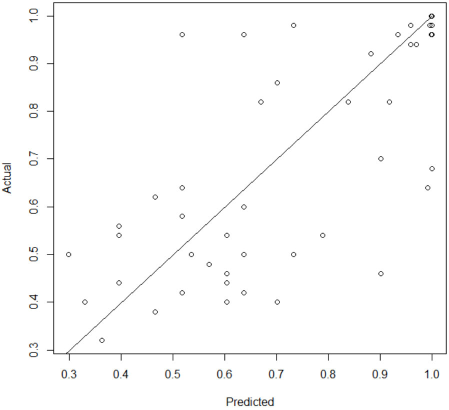


- Implies expansion path for all other values of x
- This does not hold in our experimental data

Invariant Likelihood Ratio - An Experimental Test



Individual Level Data



- Predicted vs Actual behavior in DP 4 given behavior in DP 1
- 44% of subjects adjust significantly more slowly than Shannon
- 19% significantly more quickly

Changing Priors

- How does information gathering change with prior beliefs?
- Simplest possible design: two states two acts
- Change the relative prior probability of the states

Experiment 3						
Decision Problem	$\mu(1)$	Payoffs $U(a(1))$	$U(a(2))$	$U(b(1))$	$U(b(2))$
1	0.50	10	0	0	10
2	0.60	10	0	0	10
3	0.75	10	0	0	10
4	0.85	10	0	0	10

- Two **unequally** likely states
- Two actions (a and b)
- 54 subjects

Questions

1. Are people rational?
   - i.e. do they respect NIAS
2. Do costs look like they are Posterior Separable
   - i.e. do they obey Locally Invariant Posteriors

- NIAS test: For each decision problem

$$P(a|1) \geq \frac{2\mu_1 - 1}{\mu_1} + \frac{1 - \mu_1}{\mu_1} P(a|2)$$

- From the aggregate data

DP	$P_j(a\|2)$	Constraint on $P_j(a\|1)$	$P_j(a\|1)$	Prob
5	0.29	0.29	0.77	0.000
6	0.38	0.39	0.88	0.000
7	0.40	0.80	0.90	0.045
8	0.51	0.91	0.91	0.538

- NIAS test: For each decision problem

$$P(a|1) \geq \frac{2\mu_1 - 1}{\mu_1} + \frac{1 - \mu_1}{\mu_1} P(a|2)$$

- Individual level data

Prior	0.5	0.6	0.75	0.85
% Significant Violations	0	2	2	11

Locally Invariant Posteriors

- Each subject has ‘threshold belief’
  - Determined by information costs
- If prior is within those beliefs
  - Both actions used
  - Learning takes place
  - Same posteriors always used
- If prior is outside these beliefs
  - No learning takes place
  - Only one action used

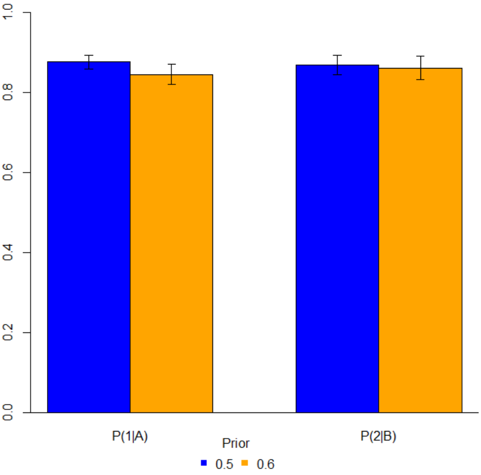
- Distribution of thresholds for 54 subjects

Posterior Range	N	%
[0.5,0.6)	14	25
[0.6,0.75)	12	22
[0.75,0.85)	12	22
[0.85,1]	16	29

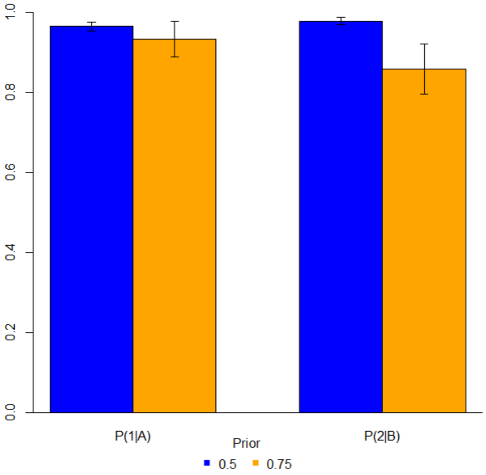
- Fraction of subjects who gather no information and always choose *a*

Table 10: Testing the 'No Learning' Prediction:				
Fraction of subjects who never choose *b*				
		$\mu(1)$		
		DP8	DP9	DP10
		0.6	0.75	0.85
Significant differences	$\gamma_7^a(1) < \mu_i(1)$	33%	46%	41%
	$\gamma_7^a(1) \geq \mu_i(1)$	3%	10%	14%

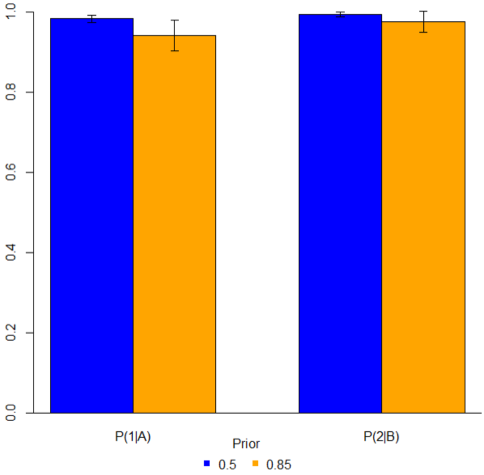
Results - Threshold Greater than 0.6



Results - Threshold Greater than 0.75



Results - Threshold Greater than 0.85



Symmetry

- Compression axiom: distinguishes Shannon from the more general posterior separable model
- Optimal revealed posteriors depend **only** on the relative value of acts in that state
- Implies that there is no concept of ‘perceptual distance’

A Simple Example

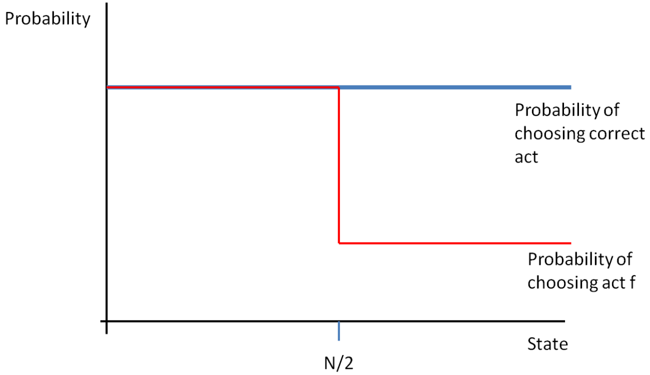
- N equally likely **states of the world** $\{1, 2....., N\}$
- Two **actions**

	Payoffs	
States	$1, ... \frac{N}{2}$	$\frac{N}{2} + 1, .., N$
action f	10	0
action g	0	10

- Mutual Information predicts a *quantized* information structure
  - Optimal information structure has **2 signals**
  - Probability of making correct choice is **independent of state**

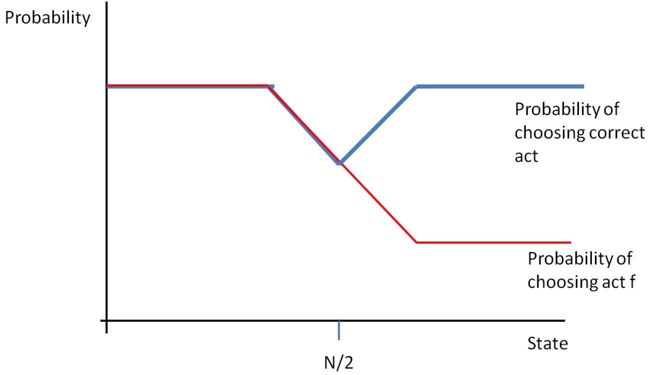
$$\frac{\exp\left(\frac{u(10)}{\kappa}\right)}{1 + \exp\left(\frac{u(10)}{\kappa}\right)}$$

Predictions for the Simple Problem - Shannon

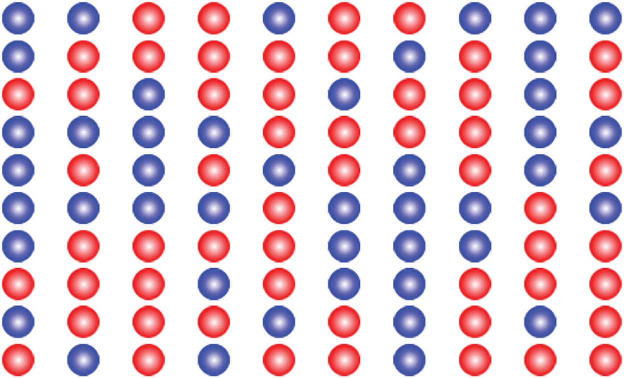


- Probability of correct choice does not go down near threshold

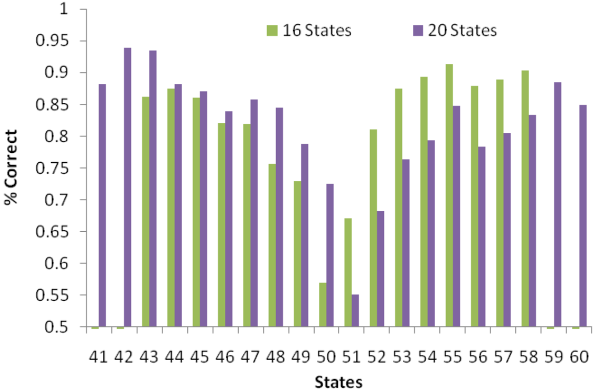
Predictions for the Simple Problem - Shannon



- Not true of other information structures (e.g. uniform signals)



Action	Payoff $\leq$ 50 Red	Payoff $>$ 50 Red
f	10	0
g	0	10



- Probability of correct choice significantly correlated with distance from threshold ($p<0.001$)

Can we Improve on Shannon?

- These experiments tested three key properties of Shannon
  - Locally Invariant Posteriors
  - Invariant Likelihood Ratio
  - Invariance Under Compression (and in particular symmetry)
- LIP did okay(ish), the others did pretty badly
  - Expansion path problem
  - Symmetry problem
- Can we modify the Shannon model to better fit this data?
  - And in doing so do we provide a **quantitatively** better fit of the data?

Expansion Path

- To fix the expansion path problem there are two obvious routes

1. Posterior Separable cost functions

$$K(\mu, \pi) = \sum_{\Gamma} Q(\gamma) T(\gamma) - T(\mu)$$

   - e.g. we could use Generalized Entropy

$$T_{\rho}^{Gen}(\gamma) = \begin{cases} \left(\frac{1}{(\rho-2)(\rho-1)|\Gamma|} \sum_{\Gamma} \hat{\gamma}^{2-\rho} - 1 \right) & \text{if } \rho \neq 1 \text{ and } \rho \neq 2; \\ \frac{1}{|\Gamma|} \left(\sum_{\Gamma} \hat{\gamma} \ln \hat{\gamma} \right) & \text{if } \rho = 1; \\ -\frac{1}{|\Gamma|} \left(\sum_{\Gamma} \ln \hat{\gamma} \right) & \text{if } \rho = 2. \end{cases}$$

2. Drop the assumption that costs are linear is Shannon mutual information

$$K(\mu, \pi) = \kappa \left(\sum_{\gamma \in \Gamma(\pi)} \pi(\gamma) \left[-H(\gamma) \right] - \left[-H(\mu) \right] \right)^{\sigma}$$

Symmetry

- It is fairly obvious why symmetry fails
  - Nearby states are harder to distinguish than those further away
  - Shannon cannot take this into account
- Hebert and Woodford [2017] propose a solution
  - Divide the state space into I overlapping 'neighborhoods' $X_1 \ldots X_I$
  - An information structure is assigned a cost for each neighborhood based on the prior and posteriors conditional on being in that neighborhood
  - Total costs is the sum across all neighborhoods

$$\sum_{i=1}^{I} \mu(X_i) \sum_{\gamma} Q(\gamma|X_i) \left[-H(\gamma|X_i)\right] - \left[-H(\mu|X_i)\right]$$

- Has a number of attractive features
  - Introduces perceptual distance to Shannon-like models
  - Qualitatively fits data from psychometric experiments
  - Can be 'microfounded' as resulting from a process of sequential information acquisition

Symmetry

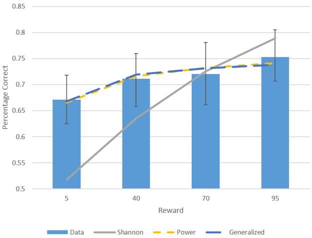
- Has a number of attractive features
  - Introduces perceptual distance to Shannon-like models
  - Qualitatively fits data from psychometric experiments
  - Can be ‘microfounded’ as resulting from a process of sequential information acquisition
- However, not the only option
  - See for example "The Cost of Information" by Strack et al [2022]

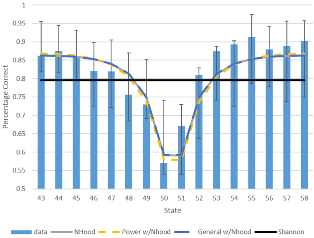
Applying Alternative Cost Functions

- We can combine these ideas to come up with a family of cost function to estimate

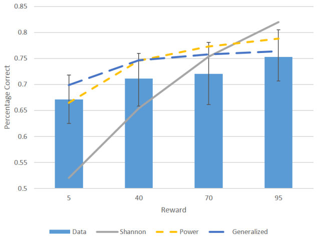
1. Linear mutual information with neighborhoods
   - Assume one global neighborhood, plus one neighborhood for each sequential pair of states
   - Cost within each neighborhood based on mutual information
   - Two parameters:
     - κ_g: marginal cost of information for the global neighborhood
     - κ_l: marginal cost of information for each of the local neighborhoods
2. Non-linear mutual information with neighborhoods
   - As (1), but costs raised to a power
   - Introduces one new parameter σ
3. General mutual information with neighborhoods
   - As (1) but mutual information replaced with expected change in generalized entropy
   - Introduces one new parameter ρ

Fitted Values (Estimated Separately on Each Experiment)





Fitted Values (Estimated Jointly)



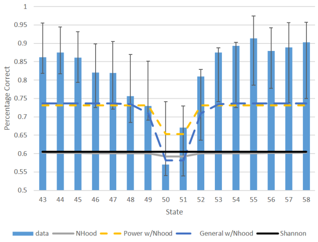


Table 12: Parameter Estimates - Aggregate Data						
Model	κ_g	κ_l	σ	ρ	BIC	AIC
Experiment 2 Only						
NHood	28.82	-	-	-	379	372
Power	7728.00	-	4.23	-	55	41
Generalized	0.16	-	-	13.41	56	42
Experiment 4 Only						
Shannon	7.38	-	-	-	485	479
NHood	5.40	5.04	-	-	326	313
Power w/NHood	4.98	5.63	0.94	-	334	315
Generalized w/NHood	5.36	4.99	-	1.05	334	315

Table 12: Parameter Estimates - Aggregate Data

Model	κ_g	κ_l	σ	ρ	BIC	AIC
Experiment 2 and 4						
Shannon	23.49	-	-	-	1689	1681
NHood	3.24	23.93	-	-	738	722
Power w/NHood	14.31	88.90	1.99	-	447	423
Generalized w/NHood	0.01	1.80	-	8.41	421	397

Columns 2-5 report parameter estimates from the models described in table 11
Columns 6 and 7 report the Akaike information criterion and Bayes information criterion respectively

- Dewan, Ambuj, and Nathaniel Neligh. "Estimating information cost functions in models of rational inattention." Journal of Economic Theory 187 (2020): 105011.
- Caplin, Andrew, et al. "Rational inattention, competitive supply, and psychometrics." The Quarterly Journal of Economics 135.3 (2020): 1681-1724.
- Cheremukhin, Anton, Anna Popova, and Antonella Tutino. "A theory of discrete choice with information costs." Journal of Economic Behavior & Organization 113 (2015): 34-50.
- Bronchetti, E. T., Kessler, J. B., Magenheim, E. B., Taubinsky, D., & Zwick, E. (2020). Is Attention Produced Rationally?.

Application

- There are many 'classic' applications or rational inattention
  - Slow adjustment in macro models (e.g. Sims [2003], Mackowiak and Wiederholdt [2015])
  - Pricing (e.g. Mackowiak and Wiederholdt [2009], Matejka [2015, 2016])
  - Portfolio selection (e.g. Van Nieuwerburg and Veldkamp (2009), Mondria (2010))
- I am not going to concentrate on these
  - More the domain of the Macro Rational Inattention foks
  - See for a nice discussion Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt.. "Rational Inattention: A Review." Journal of Economic Literature..
- Instead cover some more recent, esoteric applications
  - Rational inattention to quality
  - Discrimination
  - Elections
  - Dynamic Rational Inattention

- Other applications we don't have time for
  - Jakobsen, Alexander M. 2020. "A Model of Complex Contracts." American Economic Review, 110 (5): 1243-73
  - Gaglianone, W. P., Giacomini, R., Issler, J. V., & Skreta, V. (2020). Incentive-driven inattention. Journal of Econometrics.
  - Bhattacharya V, and Howard G, 2020, "Rational Inattention in the Infield."
  - Ambuehl, Sandro and Ockenfels, Axel and Stewart, Colin, Who Opts In? (May 20, 2020). Rotman School of Management Working Paper No. 3154197
  - Kroft, Kory, et al. Salience and Taxation with Imperfect Competition. No. w27409. National Bureau of Economic Research, 2020.
  - Heidhues, Paul, Johannes Johnen, and Botond Kőszegi. "Browsing versus studying: A pro-market case for regulation." The Review of Economic Studies 88.2 (2021): 708-729.

Application: Price Setting with Rationally Inattentive Consumers

- Consider buying a car
- The price of the car is easy to observe
- But quality is difficult to observe
- How much effort do consumers put into finding out quality?
- How does this affect the prices that firms charge?
- This application comes from Martin [2017]

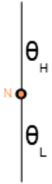
Application: Price Setting with Rationally Inattentive Consumers

- Model this as a simple game
  1. Quality of the car can be either high or low
  2. Firm decides what price to set depending on the quality
  3. Consumer observes price, then decides how much information to gather
  4. Decides whether or not to buy depending on their resulting signal
  5. Assume that consumer wants to buy low quality product at low price, but not at high price
- Key point: prices may convey information about quality
- And so may effect how much effort buyer puts into determining quality

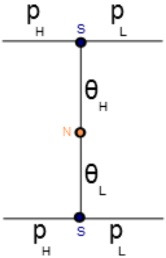
Market Setting

- One off sales encounter
    - One buyer, one seller, one product

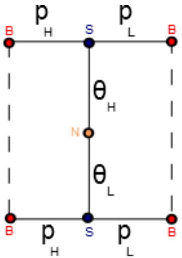
- Nature determines quality $\theta \in \{\theta_L, \theta_H\}$
  - Prior $\mu = \Pr(\omega_H)$



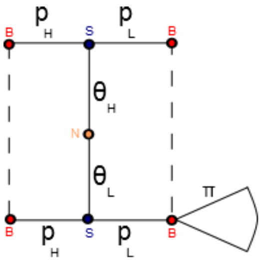
- Seller learns quality, sets price $p \in \{p_L, p_H\}$



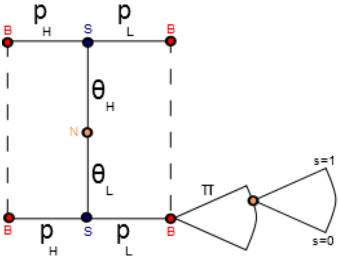
- Buyer learns p, forms interim belief μ_p (probability of high quality given price)
  - Based on prior μ and seller strategies



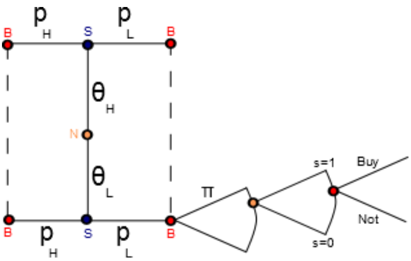
- Choose attention strategy contingent on price $\{\pi^H, \pi^L\}$
  - Costs based on Shannon mutual information



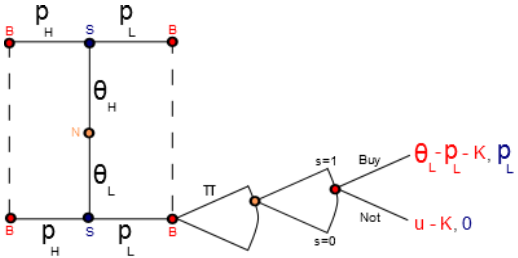
- Nature determines a signal
    - Posterior belief about product being high quality



- Decides whether to buy or not
  - Just a unit of the good



- Standard utility and profit functions (risk neutral EU)
    - $u \in \mathbb{R}_+$ is outside option, $K \in \mathbb{R}_+$ is Shannon cost



Equilibrium

- How do we make predictions in this setting?
- We need to find
  - A pricing strategy for low and high quality firms
  - An attention strategy for the consumer upon seeing low and high prices
  - A buying strategy for the consumers
- Such that
  - Firms are optimizing profits given the behavior of the customers
  - Consumers are maximizing utility given the behavior of the firms

Equilibrium

- There is **no** equilibrium in which low quality firm charges p_L and high quality firm charges p_H
- Why?
- If this were the case, the consumer would be completely inattentive with probability 1 at both prices
  - Price conveys all information
- Incentive for the low quality firm to cheat and charge the high price
- Would sell with probability 1
- This highlights a recurring issue with RI in games
  - With pure strategies there is nothing to learn

Equilibrium

- Always exists "Pooling low" Equilibrium
  - High quality sellers charge a *low price* with probability 1
  - Low quality sellers charge a *low price* with probability 1
  - Buyer believes that high price is a signal of low quality
- However, this is not a 'sensible' equilibrium:
  - Perverse beliefs on behalf of the buyer:
  - High price implies low quality
  - Allowed because beliefs never tested in equilibrium

Theorem

For every cost κ, there exists an equilibrium ("mimic high") where high quality sellers price high with probability 1 and low quality sellers price high with a unique probability $\eta \in [0, 1]$.

Explaining the Equilibrium

- How do rationally inattentive consumers behave?
- If prices are low, do not pay attention
- If prices are high, choose to have two signals
  - ‘bad signal’ - with high probability good is of low quality
  - ‘good signal’ - with high probability good is of high quality
- Buy item only after good signal

Explaining the Equilibrium

- Give rise to two posteriors (prob of high quality):
  - $\gamma^0_{p_H}$ (bad signal)
  - $\gamma^1_{p_H}$ (good signal)
- We showed that these optimal posterior beliefs are determined by the relative rewards of buying and not buying in each state

$$\ln\left(\frac{\gamma^1_{p_H}}{\gamma^0_{p_H}}\right) = \frac{(\theta_H - p_H) - u}{\kappa}$$

$$\ln\left(\frac{1-\gamma^1_{p_H}}{1-\gamma^0_{p_H}}\right) = \frac{(\theta_L - p_H) - u}{\kappa}$$

- LIP tells us that these posteriors do not vary with the prior

- Let $\mu_{p_H}(H)$ be the prior probability that the good is of high quality given that it is of high price
- Let $d^{\theta_L}_{p_H}$ be the probability of buying a good if it is actually low quality if the price is high:
  - i.e $\pi_{p_H}(\gamma^1_{p_H}|\theta_L)$
- Using Bayes rule, we can show:

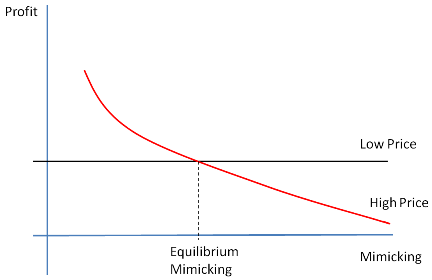
$$d^{\theta_L}_{p_H} = \frac{\left(\frac{1-\gamma^1_{p_H}}{\gamma^1_{p_H}-\gamma^0_{p_H}}\right)\left(\mu_{p_H}(H) - \gamma^0_{p_H}\right)}{\left(1 - \mu_{p_H}(H)\right)}$$

- Conditional demand is
  - Strictly increasing in interim beliefs μ_{p_H}
  - So strictly decreasing in 'mimicking' η

- What about firm behavior?
- If the low quality firm sometimes prices high and sometimes prices low, we need them to be indifferent between the two

$$d_{p_H}^{\theta_L} \times p_H = p_L \Rightarrow d_{p_H}^{\theta_L} = \frac{p_L}{p_H}$$

- As low quality firms become more likely to mimic, it decreases the probability that the low quality car will be bought
- And so reduces the value of setting the high price


Profit
Low Price
High Price
Equilibrium
Mimicking
Mimicking

- What is the unique value of η when $\eta \in (0,1)$?

$$\eta = \frac{\kappa}{1-\kappa} \frac{\left(1-\gamma^0_{p_H}\right)\left(1-\gamma^1_{p_H}\right)}{\gamma^0_{p_H}\left(1-\gamma^1_{p_H}\right) + \frac{p_L}{p_H}\left(\gamma^1_{p_H} - \gamma^0_{p_H}\right)}$$

- We can use a model of rational inattention to solve for
  - Consumer demand
  - Pricing strategies
- Can use the model to make predictions about how these change with parameters of the model
  - E.g as $\kappa \to 0$, $\eta \to 0$

Discrimination [Bartos et al 2016]

- A second recent application of the rational inattention model has been to study discrimination
- Imagine you are a firm looking to recruit someone for a job
- You see the name of the applicant at the top of the CV
- This gives you a clue to which 'group' an applicant belongs to
    - e.g. British vs American
- You have some prior belief about the abilities of these groups
    - e.g. British people are worse than Americans
- Do you spend more time looking at the CVs of Brits or Americans?

A Formal Version of the Model

- You are considering an applicant for a position
  - Hiring for a job
  - Looking for someone to rent your flat
- An applicant is of quality q, which you do not observe
- If you hire the applicant you get payoff q
- Otherwise you get 0

- Initially you get to observe which group the applicant comes from
  - Brits (B) or Americans (A)
- Your prior beliefs depend on this group
- If the person is British you believe

$$q \sim N(q_B, \sigma^2)$$

- American

$$q \sim N(q_A, \sigma^2)$$

  with $q_B > q_A$
- This is your 'bias'

Information

- Before deciding whether to hire the applicant you receive a normal signal

$$y = q + \varepsilon$$

  Where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$
- You get to choose the **precision** of the signal
  - i.e. get to choose σ_ε^2
- Pay a cost based on the precision of the signal
  - $M(\sigma_\varepsilon^2)$
- Note, it doesn't have to be the case that costs are equal to Shannon
  - Only assume that lower variance gives higher costs

- What are the benefits of information?
- What do you believe after seeing signal if variance is σ_ε^2?

$$q' = \alpha y + (1-\alpha) q_G$$

Where q_G is the beliefs given the group (i.e. q_B or q_A)

$$\alpha = \frac{\sigma^2}{\sigma^2 + \sigma_\varepsilon^2}$$

- As signal gets more precise (i.e σ_ε^2 falls) then
  - More weight is put on the signal
  - Less weight put on the bias
- If information was free then bias wouldn't matter

- If you got signal y, what would you choose?
- If

$$q' = \alpha y + (1 - \alpha) q_G > 0$$

- Will hire the person
- Otherwise will not

- Value of the information structure is the value of the choice for each y

$$\max\left\{\alpha y + (1-\alpha) q_G, 0\right\}$$

- Integrated over all possible values of y

$$G(\sigma_\varepsilon^2) = \int_{-\frac{(1-\alpha)}{\alpha} q_G}^{\infty} \left[\alpha y + (1-\alpha) q_G\right] f(y) dy$$

- So the optimal strategy is to

1. Choose the precision of the signal σ_ε^2 to maximize

$$G(\sigma_\varepsilon^2) - M(\sigma_\varepsilon^2)$$

2. Hire the worker if and only if

$$\alpha y + (1-\alpha) q_G > 0$$

or

$$\varepsilon > q + \frac{(1+\alpha)}{\alpha} q_G$$

Questions

- What type of question can we answer with this model?

1. Do Brits or Americans recieve more attention
2. Does 'Rational Inattention' help or hurt the group that discriminated against?
    - i.e. would Brits do better or worse if σ_ε^2 had to be the same for both groups?

Cherry Picking or Lemon Dropping

- It turns out the answer depends on whether we are in a 'Cherry Picking' or 'Lemon Dropping' market
- Cherry Picking: would not hire the 'average' candidate from either group
  - i.e. $q_A < q_B < 0$
  - Only candidates for which good signals are received are hired
  - e.g. hiring for a job
- Lemon Dropping: would hire the 'average' candidate from either group
  - i.e. $0 < q_A < q_B$
  - Only candidates for which bad signals are received are not hired
  - e.g. looking for people to rent an apartment

Theorem

In Cherry Picking markets, the 'worse' group gets less attention, and rational attention hurts the 'worse' group

Theorem

In Lemon Dropping markets, the 'worse' group gets more attention, and rational attention hurts the 'worse' group

- 'Hurts' in this case means relative to a situation in which the 'worse' group had to be given the same attention as the 'better' group
- Discriminated against group get screwed either way!

- Intuition:

1. Attention is more valuable to the hirer the closer a group is from the threshold on average
   - If you are far away from the threshold, less likely information will make a difference to my choice
   - In the cherry picking market the 'worse' group is further away from the threshold, and so get less attention
   - In the lemon dropping market the worse group is closer to the threshold and gets more attention
2. Attention is more likely to get you hired in the cherry picking market, less likely to get you hired in the lemon dropping market
   - In the first case only hired if there is high quality evidence that you are good
   - In the latter case hired unless there is high quality evidence that you are bad

Experimental Evidence

- Market 1: Lemon Dropping - Housing Applications
- Market 2: Cherry Picking - Job Applications
- Experiment run in Czech Republic
- In each case used dummy applicants with different 'types' of name
  - White
  - Asian
  - Roma

Housing Market

Table 1—Czech Rental Housing Market: Invitation Rates and Information Acquisition by Ethnicity, Comparison of Means

	White majority name (W) (1)	Pooled Asian and Roma minority name (E) (2)	Percentage point difference: W − E, (*p*-value) (3)	Asian minority name (A) (4)	Percentage point difference: W − A, (*p*-value) (5)	Roma minority name (R) (6)	Percentage point difference: W − R, (*p*-value) (7)	Percentage point difference: R − A, (*p*-value) (8)
Panel A. Invitation for a flat visit								
No Information Treatment ($n = 451$)	0.78	0.41	37 (0.00)	0.39	39 (0.00)	0.43	36 (0.00)	3 (0.57)
Monitored Information Treatment ($n = 762$)	0.72	0.49	23 (0.00)	0.49	23 (0.00)	0.49	23 (0.00)	0 (0.92)
Monitored Information Treatment[a] ($n = 293$)	0.84	0.66	18 (0.00)	0.71	13 (0.00)	0.62	21 (0.00)	−9 (0.20)
Monitored Information Treatment[b] ($n = 469$)	0.66	0.37	29 (0.00)	0.35	31 (0.00)	0.39	27 (0.00)	4 (0.51)
Treatment with additional text in the e-mail ($n = 587$)	0.78	0.52	26 (0.00)	0.49	29 (0.00)	0.55	23 (0.00)	5 (0.29)
Panel B. Information acquisition in the Monitored Information Treatment								
Opening applicant's personal website	0.33	0.41	−8 (0.03)	0.38	−5 (0.24)	0.44	−11 (0.01)	6 (0.15)
Number of pieces of information acquired	1.29	1.75	−0.46 (0.01)	1.61	−0.32 (0.09)	1.88	−0.59 (0.00)	0.27 (0.17)
At least one piece of information acquired	0.30	0.40	−10 (0.01)	0.37	−7 (0.12)	0.44	−13 (0.00)	7 (0.12)
All pieces of information acquired	0.19	0.26	−8 (0.02)	0.24	−6 (0.12)	0.28	−10 (0.01)	4 (0.33)
Number of pieces of information acquired[a]	3.91	4.24	−0.33 (0.06)	4.23	−0.32 (0.15)	4.25	−0.34 (0.09)	0.02 (0.90)
At least one piece of information acquired[a]	0.92	0.98	−6 (0.02)	0.97	−5 (0.15)	0.98	−7 (0.03)	2 (0.47)
All pieces of information acquired[a]	0.56	0.64	−7 (0.23)	0.64	−8 (0.30)	0.64	−7 (0.30)	−0 (0.96)

- This is a lemon dropping market and there is discrimination
- More information acquired on minority candidates
- Landlords more responsive to info for minority candidates

Table 4—Czech Labor Market: Invitation Rates and Information Acquisition by Ethnicity, Comparison of Means

	White majority name (W) (1)	Pooled Asian and Roma minority name (E) (2)	Percentage point difference: W − E, (p-value) (3)	Asian minority name (A) (4)	Percentage point difference: W − A, (p-value) (5)	Roma minority name (R) (6)	Percentage point difference: W − R, (p-value) (7)	Percentage point difference: R − A, (p-value) (8)
Panel A. Employer's response								
Callback	0.43	0.20	23 (0.00)	0.17	26 (0.00)	0.25	18 (0.01)	8 (0.22)
Invitation for a job interview	0.14	0.06	8 (0.03)	0.05	9 (0.03)	0.08	6 (0.18)	3 (0.46)
Invitation for a job interview[a]	0.19	0.09	10 (0.06)	0.09	10 (0.12)	0.10	9 (0.16)	1 (0.83)
Panel B. Information acquisition								
Opening applicant's resume	0.63	0.56	7 (0.22)	0.47	16 (0.03)	0.66	−3 (0.69)	19 (0.01)
Acquiring more information about qualification[a]	0.16	0.10	6 (0.27)	0.06	10 (0.12)	0.14	2 (0.73)	8 (0.24)
Acquiring more information about other characteristics[a]	0.18	0.18	0 (0.92)	0.19	−1 (0.85)	0.18	0 (0.99)	1 (0.85)

- There is discrimination and it is a cherry picking market
- Attention discrimination agains Asian candidates

Voting

- Voters are typically not very well informed
- However, the spread of information is not uniform or random
- Which voters choose to get informed about which issue?
- How does this impact the formation of policies
- These issues are discussed in Matejka and Tabellini [2021]

- Two candidates A and B
- Pick policy platform: vector q_C in order to maximize prob of winning an election
- N groups of voters
    - Each group contains a continuum of voters of mass m^J
- Utility of voter v in group J if each candidate wins is

$$
\begin{aligned}
U_A^{v,J} &= U^J(q_A) \\
U_B^{v,J} &= U^J(q_B) + x^v \\
x^v &= \hat{x} + \hat{x}^V
\end{aligned}
$$

Rational Inattention in Games

- This is going to be a game between the candidates and the voters
- Applying rational inattention to game theory is hard
  - In equilibrium, strategy of other players is 'known'
  - What to learn about?
- Typically it is assumed that learning is about some exogenous state
- Though even here there is complications
  - e.g. would like my learning to be correlated
- For discussions see
  - Denti "Unrestricted Information Acquisition", 2019
  - Morris and Yang "Coordination and Continuous Stochastic Choice 2019
  - Afrouzi, Hassan. "Strategic inattention, inflation dynamics and the non-neutrality of money." 2017
  - Martin, Daniel, and Edwin Muñoz-Rodriguez. "Misperceiving Mechanisms: Imperfect Perception and the Failure to Recognize Dominant Strategies." 2019

- Assume that there is some irreducible noise around the candidate's platform
  - Candidate chooses $\hat{q}_c$, actual platform

$$q_{C,i} = \hat{q}_{C,i} + \varepsilon_{C,i} \text{ with } \varepsilon_{C,i} \sim N(\bar{q}, \sigma^2_{C,i})$$

- Voters recieve a normal signal

$$s^{v,J}_{C,i} = q_{C,i} + \varepsilon^{v,J}_{C,i} \text{ with } \varepsilon^{v,J}_{C,i} \sim N(0, \gamma^J_{C,i})$$

  - Define $\zeta^J_{C,i} = \frac{\sigma^2_{C,i}}{\sigma^2_{C,i} + \gamma^J_{C,i}}$
- Choose variance optimally
  - Costs based on entropy
  - Benefits?

- Sequence of events
  1. Voters form priors and choose attention strategies
  2. Candidates choose platforms
  3. Voters observe signal
  4. x^v is realized and election is held
- Voters vote for candidate A if

$$E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] > x^v$$

- In equilibrium
  - Voter priors correct given candidate strategies
  - Voter information acquisition optimal given these priors
  - Candidates strategies optimal given strategies of voters

- If information costs are zero this boils down to a standard voting model
- Probability of each candidate winning is increasing in their social welfare
- A's probability of winning is

$$p_A = \frac{1}{2} + \phi \left[\sum_J m^J \left(U^J(q_A) - U^J(q_B) \right) \right]$$

  - where ϕ is a constant
- If attention is costly, this gets replaced by

$$p_A = \frac{1}{2} + \phi E^J_{\varepsilon, q_A, q_B} \left[\sum_J m^J (E(U^J(q_A | s_A^{v,J}) - E(U^J(q_B | s_B^{v,J})) \right]$$

  - The perceived social welfare function

- Each candidate will try to maximize their perceived social welfare
- If information is free then the weight of each group is just its size m^J
- If attention is costly, then differential attention can play a role
- Indeed, if we can use a quadratic approximation for utility around $\bar{q}_C$ then

$$E(U^J(q_C|s_C^{v,J}) = U^J(\bar{q}) + u_C^J(\bar{q})\left(E(q|s_C^{v,J}) - \bar{q}\right)$$

where $u_{C,i}^J = \frac{\partial u^J(q_{C,i})}{\partial q_{J,i}}$

- But

$$E(q|s_C^{v,J}) = \zeta_C^J s_C^{v,J} + \left(1 - \zeta_C^J\right)\bar{q}$$

and, as $E_{\varepsilon,q_A,q_B}^J\left(s_C^{v,J}\right) = \hat{q}_C$ the objective function becomes

$$\frac{1}{2} + \phi\sum_J m^J\left[U^J(\bar{q}) + u_C^J(\bar{q})\left(\zeta_C^J\left(\hat{q}_C - \bar{q}\right)\right)\right]$$

- And the FOC

- Candidates that pay more attention will get a higher weight
- Policy will be skewed towards those groups that pay higher attention
- Who pays more attention?
- Under the same approximation, the benefits of attention are given by

$$\sum_C \sum_i \zeta^J_{C,i} \left(u^J_{C,i}\right)^2 \sigma^2_{C,i}$$

  - This is the variance of the difference in expected utility between candidates
- Attention will be
  - Increasing in the prior variance
  - Decreasing in the cost of attention
  - Increasing in the derivative of the utility function

- To see how these forces play out, consider the case in which there is only one dimension
  - Bliss point of group J is t^J
  - Cost of attention for group J is Λ^J
  - $U^J(q) = U(q - t^J)$ where U is concave and symmetric
- Note: Voters pay more attention for $\bar{q}_C$ further away from t^J as marginal utility is higher
  - Voters with extreme preferences have higher stakes
- With only two voters we have

$$\frac{u^1(q_C)}{u^2(q_C)} = -\frac{m^2 \zeta_C^2}{m^1 \zeta_C^1}$$

- Relative to fixed attention
  - Groups with more extreme views will pay more attention and so be given higher weight
  - Smaller groups will be less pandered to, and so be more 'extreme' and pay more attention
  - Attention offsets group size

TABLE 1. Regressing *Political Attention* on dummy variables for being an extremist on each of the following policy dimensions: desired size of government spending, globalization, and civil rights.

	Dependent Variable: Political Attention					
	(1)	(2)	(3)	(4)	(5)	(6)
Extremist on redistribution	0.101*** (0.012)			0.101*** (0.013)		
Extremist on globalisation	0.063*** (0.010)			0.062*** (0.010)		
Extremist on civil rights	0.032*** (0.012)			0.027** (0.011)		
Extremist on red. and glo.		0.122*** (0.019)			0.119*** (0.019)	
Extremist on all three issues			0.117*** (0.036)			0.106*** (0.036)
Democrat				0.040*** (0.011)	0.044*** (0.009)	0.043*** (0.009)
Republican				0.041*** (0.011)	0.053*** (0.009)	0.054*** (0.009)
Mean dep.	0.61	0.61	0.61	0.61	0.61	0.61
Observations	5,720	8,245	8,245	5,709	8,222	8,222
Adjusted (R^2)	0.16	0.15	0.15	0.17	0.16	0.15
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable *Political Attention* ranges from 0 to 1, and 1 indicates full attention of the respondent to what is going on in government and politics. Additional controls include age, age squared, gender, education, and race. Standard errors are robust to heteroskedasticity. Significance levels: *0.10, **0.05, and ***0.01.

- Other results
  - RI amplifies the effect of preference intensity and dampens effect of group size
  - Groups with lower attention cost get higher weight (possibly larger groups)?
  - More general predictions depend on the distribution of bliss points
    - If distribution is asymmetric, those in longer tail pay more attention
  - In general RI must lower social welfare, distorting towards more informed groups
  - If candidates have different costs, higher cost candidate will pander to the more extreme voters

Dynamic Rational Inattention

- So far we have dealt exclusively with static rational inattention problems
- Of course many interesting problems have a dynamic aspect
- A recent literature has addressed these issues

Dynamic Rational Inattention

- Two branches
  1. 'Stopping problems': Dynamic accrual of information prior to making a choice
     - Hébert, Benjamin, and Michael Woodford. Rational inattention and sequential information sampling. No. w23787. National Bureau of Economic Research, 2017.
     - Zhong, Weijie. "Optimal dynamic information acquisition." 2017.
     - Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki. "Speed, accuracy, and the optimal timing of choices." American Economic Review 108.12 (2018): 3651-84.
  2. 'Dynamic problems': Make a choice in every period
     - Steiner, Jakub, Colin Stewart, and Filip Matějka. "Rational Inattention Dynamics: Inertia and Delay in Decision-Making." Econometrica 85.2 (2017): 521-553.
     - Miao, Jianjun, and Hao Xing. Dynamic Rationally Inattentive Discrete Choice: A Posterior-Based Approach. 2019.
     - Afrouzi, Hassan, and Choongryul Yang. "Dynamic rational inattention and the Phillips curve." Available at SSRN 3770462 (2019).

Dynamic Rational Inattention

- Steiner, Stewart and Matejka (SSM) write down conditions for optimality in a dynamic RI problem
  - Costs linear in mutual information
- First observation: if costs are linear in mutual information then actions are sufficient statistics for signals
  - So we can model choice of actions directly
- This is obvious in the static case
- Less obvious in the dynamic case
  - Maybe want to gather information earlier than needed to smooth information costs
- But linear mutual information costs have no such smoothing motive
  - See also Afrouzi and Yang [2019]

- Second observation: Dynamic problem can be reduced to a sequence of static problems
- Let p be a dynamic choice strategy (i.e stochastic mapping from Θ^t to $\Delta(A)$ for every t
- p is an interior optimum if, at every history z it solves the static RI problem with
  - State space Θ^t
  - Prior $\mu(\theta^t) = \pi^p(\theta^{t-1}|z^{t-1})\pi(\theta^t|\theta^{t-1})$
  - And utility function

$$
\begin{aligned}
\hat{u}(a, \theta^t, z^{t-1}) &= \hat{u}(a, \theta^t) + \delta E V_{t+1}\left(\theta^{t+1}\right)|a_t, \theta^t, z^{t-1}) \\
V_{t+1}\left(\theta^{t+1}\right) &= \ln \sum_{a_t} p(a_t|z^{t-1}) \exp \hat{u}(a, \theta^t, z^{t-1})
\end{aligned}
$$

  - where z^t is the history of actions and exogenous signals
- This solution can still be very cumbersome
  - Miao offer an alternative using posteriors as states

- This is another case in which the LQG framework has lead to more tractability
  - Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. "Dynamic rational inattention: Analytical results." Journal of Economic Theory 176 (2018): 650-692.
  - Miao, Jianjun, Jieran Wu, and Eric Young. Multivariate Rational Inattention. No. WP2019-07. Boston University-Department of Economics, 2019.
  - Afrouzi, Hassan, and Choongryul Yang. "Dynamic Rational Inattention and the Phillips Curve." Available at SSRN 3465793 (2019).
- This last paper in particular comes with a really fast solution algorithm that has been packaged with easy to use software
- And Hassan is at Columbia!