Rational Inattention Lecture 2

Mark Dean

Behavioral Economics G6943
Fall 2016

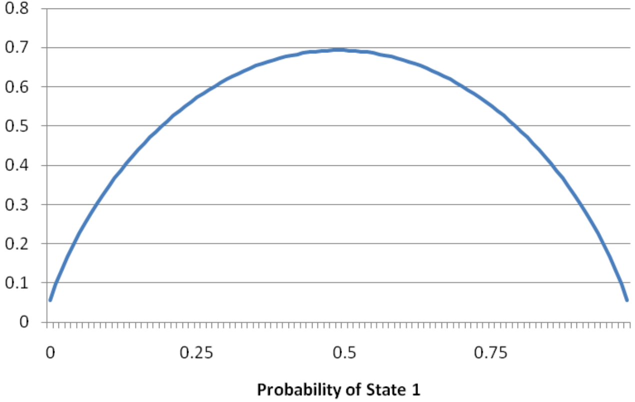
Rational Inattention and Shannon Information Costs

- We have so far considered what we can say when we are agnostic about information costs
- We now move consider behavior under a specific assumed cost for information
- Based on the concept of Shannon Entropy
  - Extremely popular in the applied literature
  - Consider this the 'Cobb Douglas' case to last week's 'revealed preference' treatment
- Long history of research in information theory
  - Quite a lot is known about how these costs behave
  - Cover and Thomas is a great resource

- Shannon Entropy is a measure of how much 'missing information' there is in a probability distribution
- In other words - how much we do not know, or how much we would learn from resolving the uncertainty
- For a random variable X that takes the value x_i with probability $p(x_i)$ for $i = 1...n$, defined as

$$\begin{aligned} H(X) &= E(-\ln(p(x_i))) \\ &= -\sum_i p(x_i)\ln(p_i) \end{aligned}$$

Shannon Entropy



- Can think of it as how much we learn from result of experiment

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
  - $H(X) = H(p)$

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
  - $\max_{p \in \Delta^M} H(p) = H\left(\left\{\frac{1}{M}, \frac{1}{M}, ..., \frac{1}{M}\right\}\right)$

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
- Unaffected by adding zero probability state
  - $H(\{p_1....p_M\}) = H(\{p_1....p_M, 0\})$

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
- Unaffected by adding zero probability state
- Additive
  - $H(X, Y) = H(X) + \sum_x p(x)H(Y|x)$
  - How much you learn from observing X, plus how much you additionally learn from observing Y
  - Implies that the entropy of two independent variables is just $H(X) + H(Y)$
  - 'Constant returns to scale' assumption
  - (Most 'controversial' - other entropies relax this assumption)

Justification for Shannon Entropy

- Say we want our measure of entropy to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
- Unaffected by adding zero probability state
- Additive
- Then Entropy must be of the form (Khinchin 1957)

$$H(X) = -\sum_i p(x_i) \ln(p_i)$$

Entropy and Information Costs

- Related to the notion of entropy is the notion of Mutual Information

$$I(X, Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

- Measure of how much information one variable tells you about another
- Note that $I(X, Y) = 0$ if X and Y are independent

- Note also that mutual information can be rewritten in the following way

$$
\begin{aligned}
I(X, Y) &= \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x, y) \log \frac{p(x|y)}{p(x)} \\
&= \sum_y \sum_x p(x, y) \ln P(x|y) - \sum_x \sum_y p(x, y) \ln p(x) \\
&= \sum_y p(y) \sum_x p(x|y) \ln P(x|y) - \sum_y p(x) \ln p(x) \\
&= H(X) - E(H(X|Y))
\end{aligned}
$$

- Difference between entropy of X and the expected entropy of X once Y is known

Mutual Information and Information Costs

- Mutual Information between prior and posteriors often used to model information costs

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

- For convenience use γ to refer to the posterior beliefs generated by signal γ

Mutual Information and Information Costs

- Can be justified by information theory
- Say you are going to observe n repetitions of the state Ω (let ω^n be a typical element)
- You are allowed to send a message consisting of nR bits (R is the rate)
- Decoded in order to generate n repetitions of the signal space Γ (let γ^n be a typical element)
- Define $d(\omega, \gamma)$ be the loss associated with receiving signal γ in state ω, and $\hat{d}(\omega^n, \gamma^n) = \frac{1}{n} \sum d(\omega_i^n, \gamma_i^n)$

- Rate Distortion Theorem: Let $R(D)$ be the minimal rate needed to generate loss D as $n \to \infty$, then

$$R(D) = \min_{\pi \in \Pi} I(\Omega, \Gamma) \text{ s.t. } \sum_{(\gamma,\omega)} \mu(x)\pi(\gamma|x)d(\omega,\gamma) \leq D$$

- Implies (assuming strict monotonicity)

$$\min \sum_{(\gamma,\omega)} \mu(x)\pi(\gamma|x)d(\omega,\gamma) \text{ s.t. } I(\Omega,\Gamma) \leq R(D)$$

- is equivalent to

$$\min \sum_{(\gamma,\omega)} \mu(x)\pi(\gamma|x)d(\omega,\gamma) \text{ s.t. } R \leq R(D)$$

- See Cover and Thomas Chapter 10.

- Key feature: Entropy is strictly *concave*
- So negative of entropy is strictly convex
- Say we choose a signal structure with two posteriors γ and γ'
- It must be that

$$P(\gamma)\gamma + P(\gamma')\gamma' = \mu$$

- so

$$\begin{aligned} P(\gamma)H(\gamma) + P(\gamma')H(\gamma') &< H(P(\gamma)\gamma + p(\gamma')\gamma') \\ &= H(\mu) \end{aligned}$$

- So the cost of ‘learning something’ is always positive

Solving Rational Inattention Models

- Solving the Shannon model can be difficult analytically
  - Though easier than many other models
- General approach - ignore choice of information structure, instead focus on joint distribution of choice variable and state
  - i.e. choose state dependent stochastic choice directly
  - Can do this because optimal strategy will always be 'well behaved'
  - Each action taken in at most one state
- Example (Matejka and McKay 2015) - continuous state space, finite action space
- We will talk about analytical approaches
  - Alternative, algorithmic approaches
  - e.g. Blahut-Arimotio algorithm
  - See Cover and Thomas (page 191)

- $\mathcal{P}$ set of all state contingent stochastic choice functions for some state space Ω and set of acts A
- Remember $P(a|\omega)$ is the probability of choosing a in state ω
- Remember that, for $P \in \mathcal{P}$, the mutual information between choices a and objective state ω is given by

$$I(A, \Omega) = H(A) - H(A|\Omega)$$

Solving Rational Inattention Models

- Decision problem of agent is to choose $P \in \mathcal{P}$ to maximize

$$\sum_{a \in A} \int_{\omega} u(a(\omega)) P(a|\omega) \mu(d\omega)$$
$$-\lambda \left[\sum_{a \in A} \int_{\omega} P(a|\omega) \ln P(a|\omega) \mu(d\omega) + \sum_{a \in A} P(a) \ln P(a) \right]$$

- Subject to

$$\sum_{a \in A} P(a|\omega) = 1 \text{ Almost surely}$$

- Where $P(a)$ is the unconditional probability of choosing a
- Note another constraint which we will ignore for now

$$P(a|\omega) \geq 0 \ \forall \ a, \omega$$

The Lagrangian Function

$$\sum_{a\in A}\int_{\omega} u(a(\omega))P(a|\omega)\mu(d\omega)$$
$$-\lambda\left[\sum_{a\in A}\int_{\omega} P(a|\omega)\ln P(a|\omega)\mu(d\omega)+\sum_{a\in A}P(a)\ln P(a)\right]$$
$$-\int_{\omega}\rho(\omega)\left[\sum_{a\in A}P(a|\omega)-1\right]\mu(d\omega)$$

- $\rho(\omega)$ Lagrangian multiplier on the condition that $\sum_{a\in A} P(a|\omega)=1$
- FOC WRT $P(a|\omega)$ (assuming $>$0)

$$u(a(\omega))-\rho(\omega)+\lambda[\ln P(a)+1-\ln P(a|\omega)-1]=0$$

- Note that this is a convex problem

Solution

- FOC WRT $P(a|\omega)$ (assuming >0)

$$u(a(\omega)) - \rho(\omega) + \lambda[\ln P(a) + 1 - \ln P(a|\omega) - 1] = 0$$

- Which gives

$$P(a|\omega) = P(a) \exp^{\frac{u(a(\omega)) - \rho(\omega)}{\lambda}}$$

- Plug this into

$$\begin{aligned} \sum_{a' \in A} P(a'|\omega) &= 1 \\ \Rightarrow \exp^{\frac{\rho(\omega)}{\lambda}} &= \sum_{a' \in A} P(a') \exp^{\frac{u(a'(\omega))}{\lambda}} \end{aligned}$$

- Which in turn gives...

Comments

$$P(a|\omega) = \frac{P(a)\exp^{\frac{u(a(\omega))}{\lambda}}}{\sum_{c\in A} P(c)\exp^{\frac{u(c(\omega))}{\lambda}}}$$

- Similar in form to logistic random choice
- If alternatives are ex ante identical, this *is* logistic choice
- Otherwise choice probabilities are 'warped' by $P(a)$ - which contains information on the prior value of each option
- As costs go to zero, deterministically pick best option in that state
- As costs go to infinity, deterministically pick the best option ex ante

Comments

- The MM conditions ignore the constraint

$$P(a|\omega) \geq 0 \ \forall \ a, \omega$$

- Need to know which acts will be chosen with positive probability
- Typically there will be many acts not chosen at the optimum (Matejka and Sims 2010)
- There will be many solutions to the necessary conditions
- Ideally, would like necessary and sufficient conditions

Necessary and Sufficient Conditions

- Let $z(a(\omega))$ be 'normalized utilities'

$$z(a,\omega) = \exp\left\{\frac{u(a,\omega)}{\lambda}\right\}$$

- Note that the MM conditions are

$$P(a|\omega) = \frac{P(a)z(a,\omega)}{\sum_{c\in A} z(c,\omega)}$$

Necessary and Sufficient Conditions

Theorem

P is consistent with rational inattention with mutual information costs **if and only if**

$$\sum_{\omega}\left[\frac{\mu(\omega)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}\right] \leq 1 \text{ all } a\in A$$

$$\sum_{\omega}\left[\frac{\mu(\omega)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}\right] = 1 \text{ all } a \text{ s.t. } P(a)>0$$

and

$$P(a|\omega)=\frac{P(a)z(a,\omega)}{\sum_{c\in A}P(c)z(c,\omega)}$$

1. Identify correct **unconditional** choice probabilities
   - Equality condition for chosen actions
   - Check inequality condition for unchosen actions
2. Read off **conditional** choice probabilities using MM conditions

Example: Finding the Good Act

- Choose from a set of goods $A = \{a_1, ..., a_N\}$
- Only one of these goods is of high quality
  - u_h utility of the high quality good
  - u_l utility of the low quality good
  - μ_i prior probability that good i is the high quality good
  - WLOG assume $\mu_1 \geq \mu_2 .... \geq \mu_N$
- Common set up in many psychology experiments

Solution

- Cutoff strategy in prior probabilities: Exists c such that
  - $\mu_i > c \Rightarrow i$ chosen with positive probability
  - $\mu_i < c \Rightarrow i$ never chosen and nothing is learned about their quality
- Endogenously form a 'consideration set'
  -
- Let $\delta = \frac{\exp(\frac{u_h}{\lambda})}{\exp(\frac{u_l}{\lambda})} - 1$: 'additional' utility from high act
- Search the best K alternatives, where K solves

$$\mu_K > \frac{\sum_{k=1}^{K} \mu_k}{K + \delta} \geq \mu_{K+1}.$$

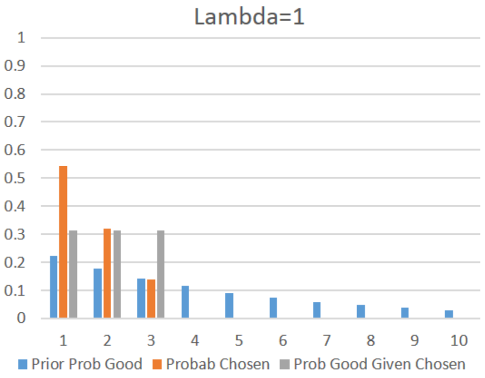
- Can use equality constraints to solve for unconditional choice probabilities

$$P(a_i) = \frac{\mu(\omega_i)(K+\delta) - \sum_{k=1}^{K} \mu(\omega_k)}{\delta \sum_{k=1}^{K} \mu(\omega_k)}$$

- MM conditions to solve for conditional choice probabilities

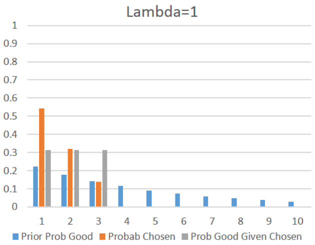
$$P(b|b = u_h) = \frac{P(b)\delta}{\sum_{c \in A} P(c)}$$

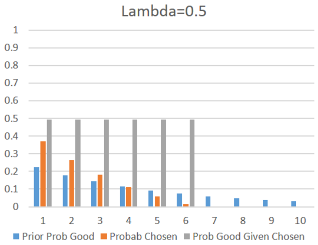
Choice Probabilities - Example

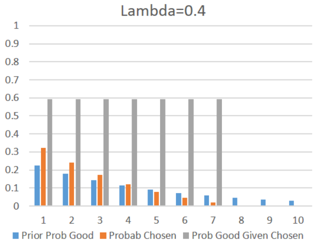


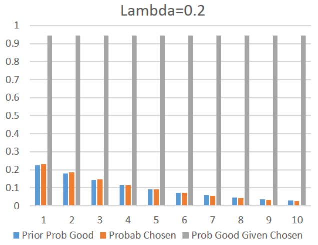
- Exponential priors
- $u_h = 1, \; u_l = 0$

Choice Probabilities - Example









Importance of Sufficient Conditions

- The MM necessary conditions could be solved for many possible ‘consideration sets’
  - Choosing any option with probability 1 will solve the necessary conditions
  - For any set C with worst alternative $\mu_{\bar{C}}$ there is a solution to the necessary conditions if

$$\frac{\mu_{\bar{C}}}{\sum_{k \in C} \mu_k} > \frac{1}{|C| + \delta}.$$

- Do no reference unchosen actions
- Do not determine whether higher utility could be obtained with a different consideration sets
- This is the advantage of the sufficient conditions

The Linear Quadratic Gaussian Case

- One case in which this problem becomes more tractable is if the input and output signal are both normal
- The entropy of a normal variable $X \sim N(\mu, \sigma_x^2)$ is given by

$$H(Y) = \frac{1}{2}\ln(2\pi e \sigma_x^2)$$

- If Y and X are both normal, then

$$E(H(Y|X)) = \int_x f(x) \int_y f(y|x) \ln(y|x) d(y) d(x)$$

- As $y|x$ is distributed normally with variance $(1-\rho^2)\sigma_y^2$, this becomes

$$\begin{aligned} E(H(Y|X)) &= \int_x f(x) \frac{1}{2} \ln(2\pi e \sigma_{y|x}^2) d(x) \\ &= \frac{1}{2}\ln(2\pi e(1-\rho^2)\sigma_y^2) \end{aligned}$$

The Linear Quadratic Gaussian Case

- As mutual information is given by

$$\begin{aligned} & H(Y) - E(H(Y|X)) \\ = \quad & \frac{1}{2}\ln(2\pi e\sigma_y^2) - \frac{1}{2}\ln(2\pi e(1-\rho^2)\sigma_y^2) \end{aligned}$$

- In this case, the mutual information is given by

$$\frac{1}{2}\ln(1-\rho^2)$$

- So information costs depend only on the covariance of the two signals!
- It turns out that joint normality is optimal if the utility function is quadratic in the relationship between the objective and subjective state
    - Choice of variance on some normally distributed error term
- However, note that some papers *assume* normality (this is bad)

A Posterior Based Approach

- Rather than think of the problem as choosing *posteriors* rather than choosing *state dependent stochastic choice*
- Can rewrite the objective function as

$$\sum_{\gamma\in\Gamma(\pi)} P(\gamma)W(\gamma) - \lambda\left[\sum_{\gamma\in\Gamma(\pi)} P(\gamma)H(\gamma) + \lambda H(\mu)\right]$$
$$\sum_{\gamma\in\Gamma(\pi)} P(\gamma)\left(W(\gamma) - \lambda H(\gamma)\right) + \lambda H(\mu)$$

- Where
  - $P(\gamma)$ is the unconditional probability of posterior γ
  - $W(\gamma) = \sum_{\omega\in\Omega}\gamma(\omega)u(a^*(\omega))$ be the expected utility of a^*, optimal choice at posterior γ
  - $H(\gamma)$ is the entropy associated with γ

Implications

- For each posterior we can define the net utility

$$N(\gamma) = W(\gamma) - \lambda H(\gamma)$$

- Optimal strategy: Choose posteriors to maximize the weighted average of $N(\gamma)$, subject to

$$\sum_{\gamma \in \Gamma(\pi)} P(\gamma)\gamma = \mu$$

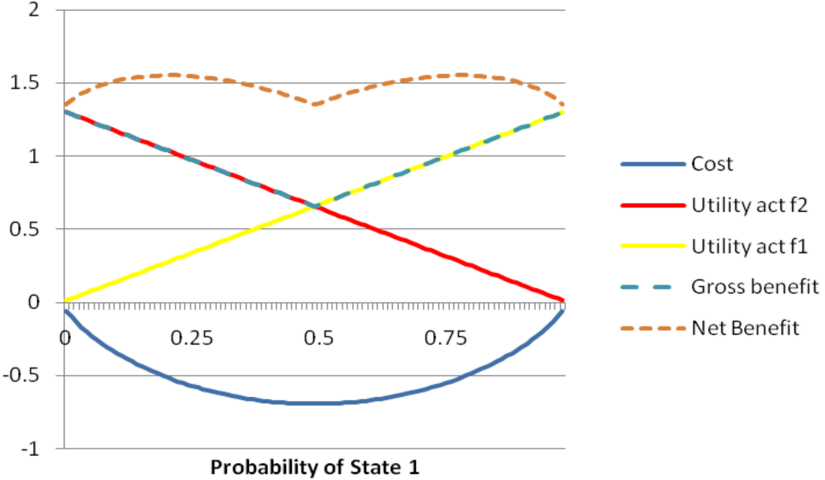
- If same number of posteriors as states this pins down $P(\gamma)$ once posteriors have been chosen

An Example of the Posterior Based Approach

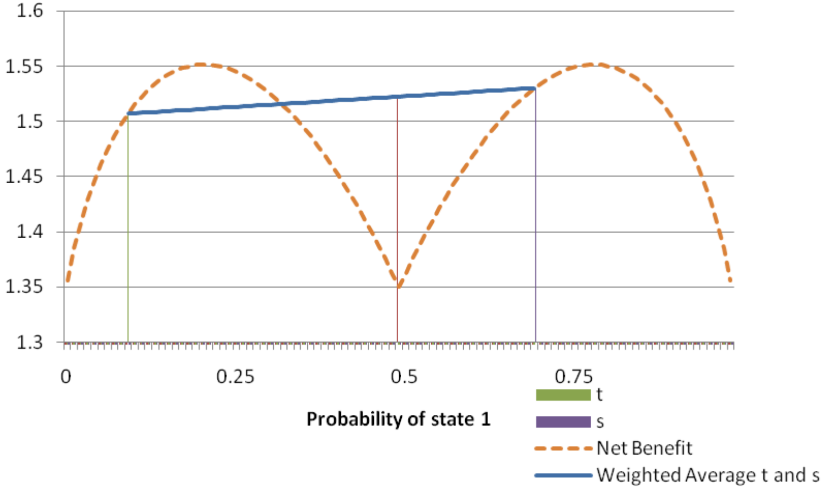
- Example: 2 states, 2 actions

Action	**Payoff in state 1**	**Payoff in state 2**
$\mathbf{f}^1$	x	0
$\mathbf{f}^2$	0	x

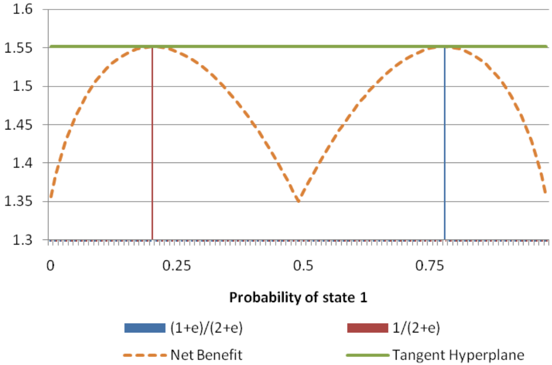
Constructing the Net Utility Function



Value as a Weighted Average of Net Utility



Finding the Optimal Strategy



- Optimal posteriors identified by hyperplane that supports the set of feasible net utilities.

Theorem

Given decision problem $(\mu, A) \in \Gamma \times \mathcal{F}$ a set of posteriors are rationally inattentive if and only if:

1. **Invariant Likelihood Ratio (ILR) Equations for Chosen Acts**: *given $a, b \in B$, and $\omega \in \Omega$,*

$$\frac{\gamma^a(\omega)}{z(a(\omega))} = \frac{\gamma^b(\omega)}{z(b(\omega))}$$

2. **Likelihood Ratio Inequalities for Unchosen Acts**: *given act a chosen with positive probability and $b \in A$,*

$$\sum_{\omega \in \Omega} \left[\frac{\gamma^a(\omega)}{z(a(\omega))} \right] z(b(\omega)) \leq 1.$$

Behavioral Properties

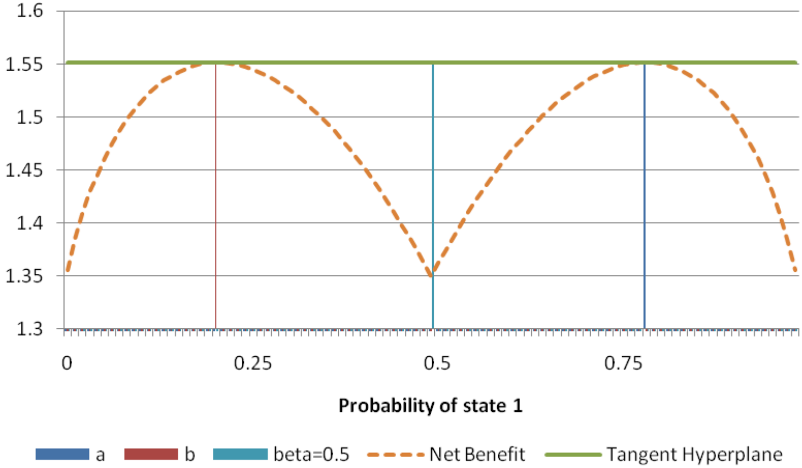
- Locally Invariant Posteriors
- Invariant Likelihood Ratio and Response to Incentives
- Symmetry

- Locally Invariant Posteriors
- Invariant Likelihood Ratio and Response to Incentives
- Symmetry

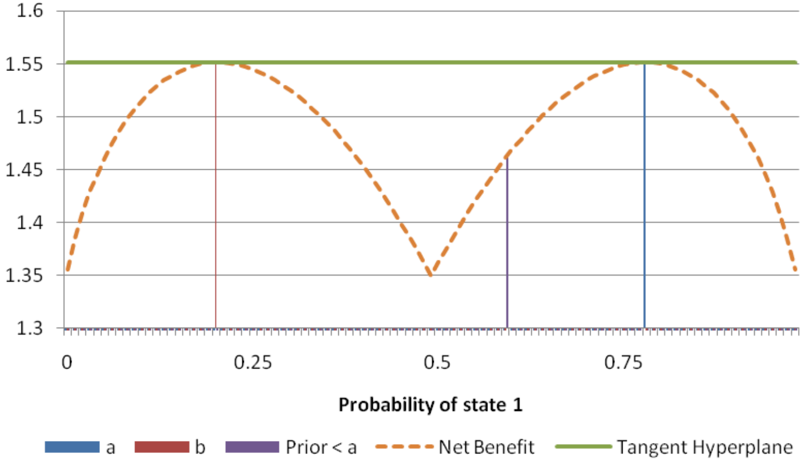
- Example: 2 states, 2 actions

Action	**Payoff in state 1**	**Payoff in state 2**
$\mathbf{f}^1$	x	0
$\mathbf{f}^2$	0	x

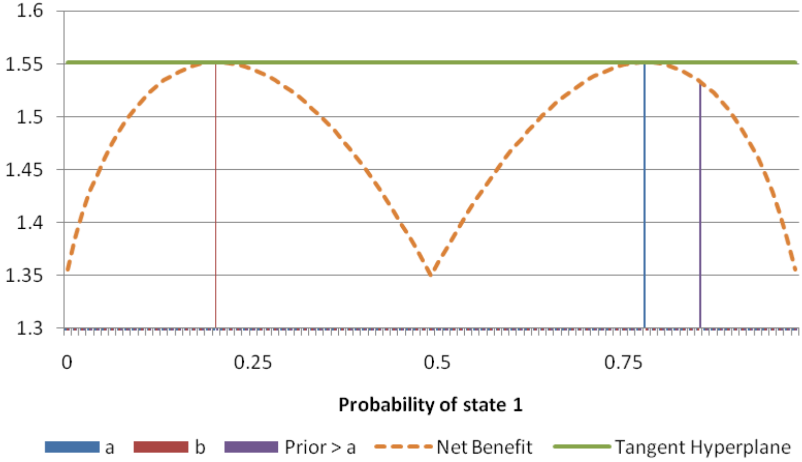
Behavior at 0.5 Prior



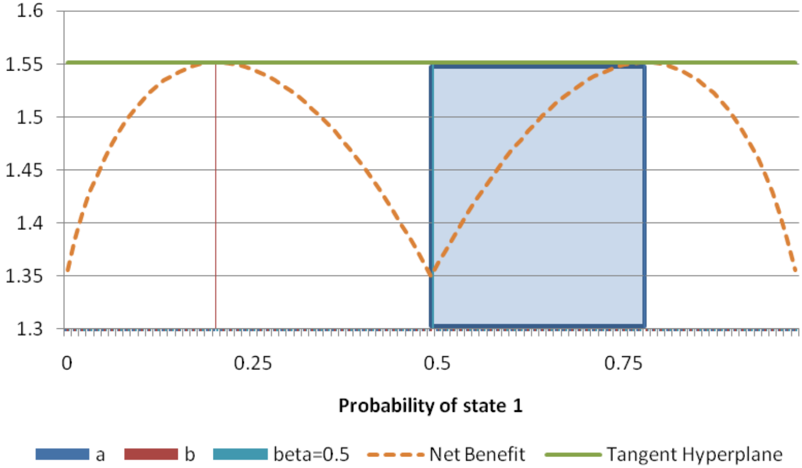
Behavior for prior<a



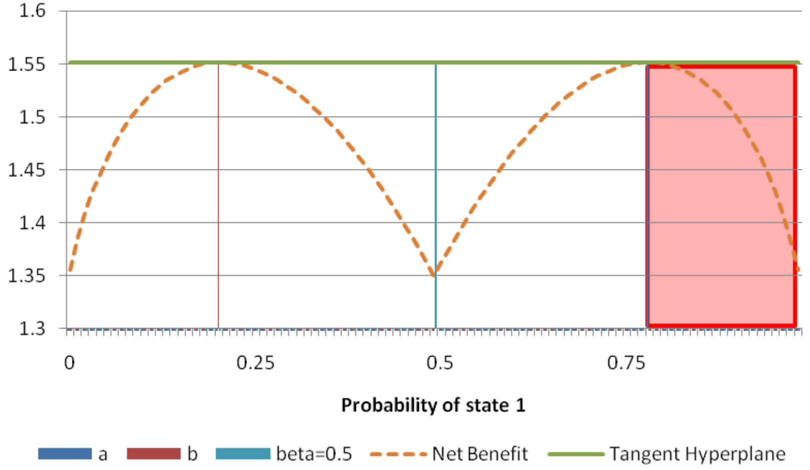
Behavior for prior>a



Same Posteriors as for 0.5 prior



No Information Gathered



Theorem (Locally Invariant Posteriors)

If a set of posteriors $\{\gamma^a\}_{a \in A}$ are optimal for decision problem $\{\mu, A\}$ and are also feasible for $\{\mu', A\}$ then they are also optimal for that decision problem

- Choice probabilities move ‘mechanically’ with prior to maintain posteriors
- Useful in, for example, models in which consumers are rationally inattentive to quality
  - As the prior distribution of quality changes, posterior beliefs do not
  - See Martin [2014]

Table 1: Experiment

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)
- 23 subjects

Prediction

- 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 23 subjects

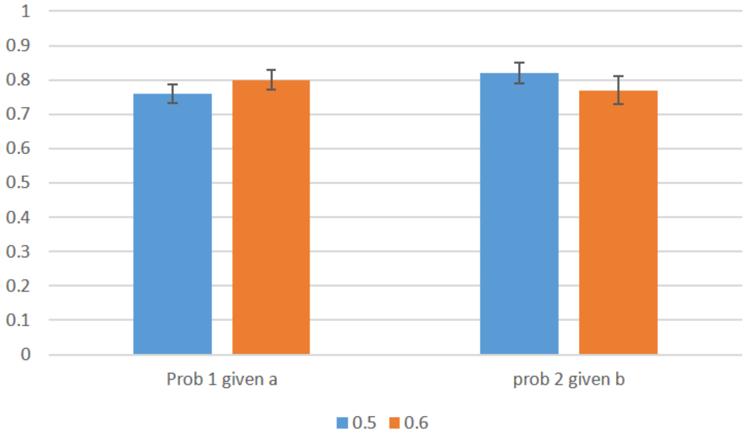
Threshold	%
[0.5,0.6)	30
[0.6,0.75)	30
[0.75,0.85)	26
[0.85,1]	13

- Fraction of subjects who gather no information

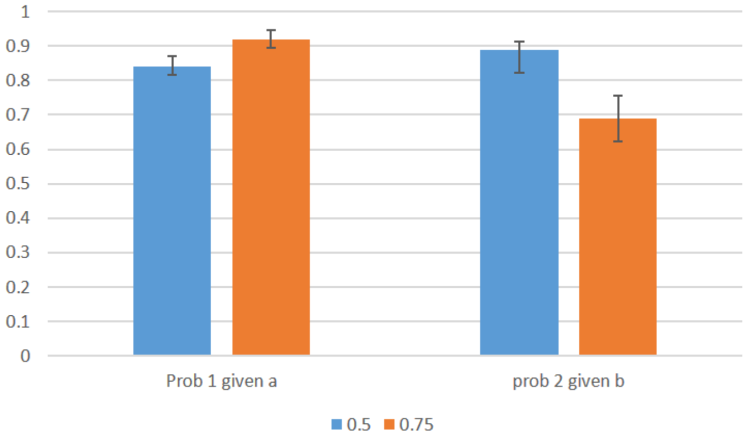
	μ		
	0.6	0.75	0.85
Threshold below μ	29%	29%	35%
Threshold above μ	0%	0%	0%

- Fraction of subjects who gather statistically insignificant amounts of information

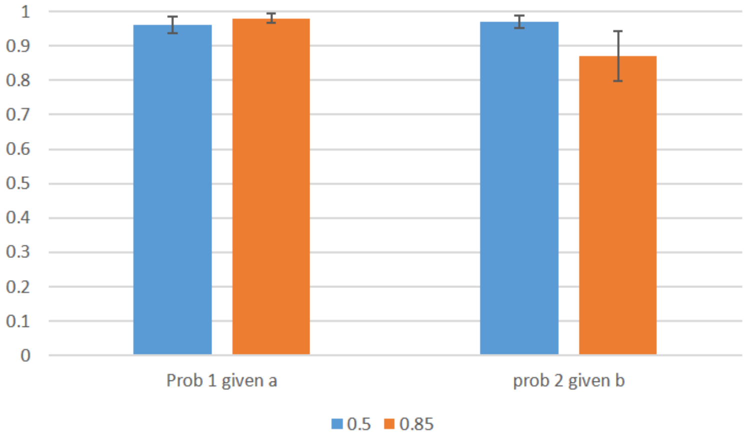
	μ		
	0.6	0.75	0.85
Threshold below μ	71%	35%	50%
Threshold above μ	19%	0%	0%


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Prob 1 given a
prob 2 given b
0.5
0.6

Results



Results



Further Prior Invariance

- Another implication of Shannon Mutual Information is that one only pays expected information costs
- Mutual information can also be written as

$$-\sum_{\omega \in \Omega} \mu(\omega) H(A|\omega) + H(\Omega)$$

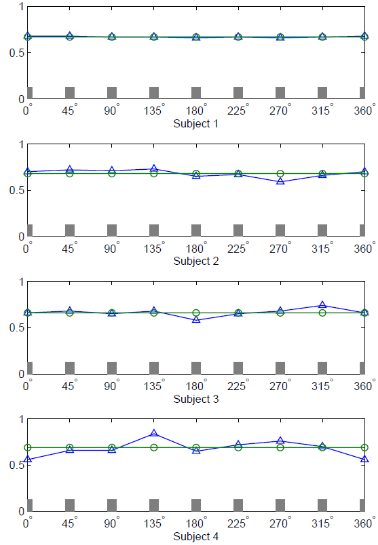
- It is cheap to be accurate in rare states.
- In fact, accuracy doesn't depend on prior distribution, as we can see from the ILR condition

$$\frac{\gamma^a(\omega)}{z(a(\omega))} = \frac{\gamma^b(\omega)}{z(b(\omega))}$$

Further Prior Invariance

- Does this work out?
- Not always - Shaw and Shaw [1977]
- Have to recognize which of three letters has appeard
- Letter can appear at any of 8 points in a circle
- Each appearance point equally likely

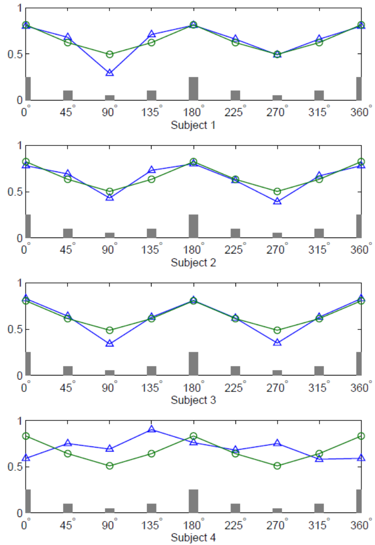
Further Prior Invariance



Further Prior Invariance

- Now make it more likely that letter appears at 'Due North' or 'Due South'
- Changes priors, but not payoffs
- Should not affect behavior

Further Prior Invariance



-

- Locally Invariant Posteriors
- Invariant Likelihood Ratio and Response to Incentives
- Symmetry

Invariant Likelihood Ratio and Responses to Incentives

- 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

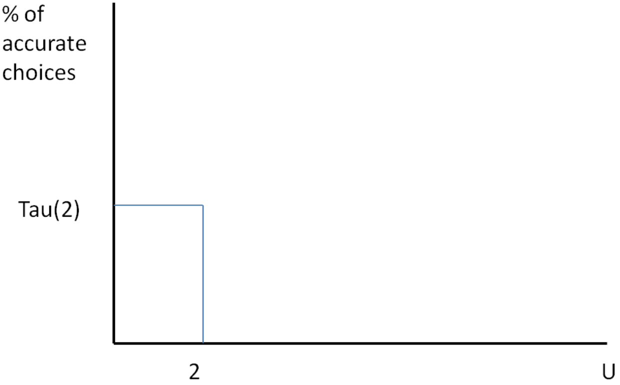
Invariant Likelihood Ratio - Example

Table 2: Experiment

Decision Problem	Payoffs $u(a(1))$	$u(a(2))$	$u(b(1))$	$u(b(2))$
1	2	0	0	2
2	10	0	0	10
3	20	0	0	20
4	30	0	0	30

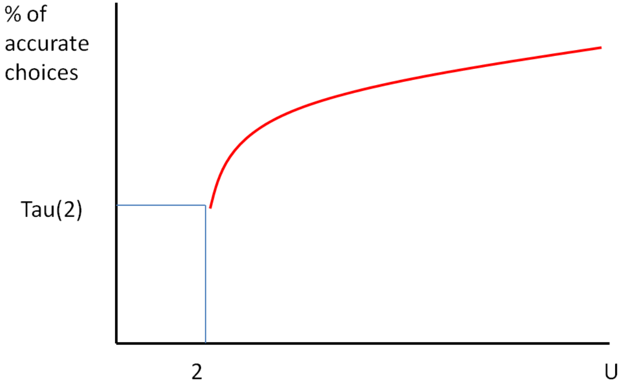
$$\frac{2}{\ln \bar{\gamma}^a(2) - \ln \bar{\gamma}^b(2)} = \frac{10}{\ln \bar{\gamma}^a(10) - \ln \bar{\gamma}^b(10)} = ... = \lambda$$

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



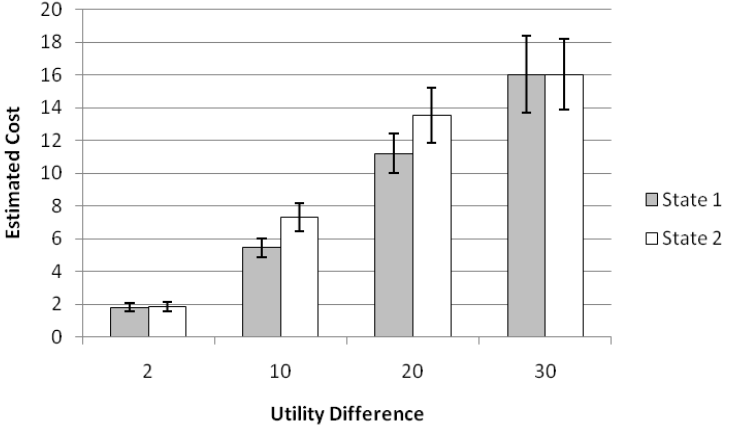
- Observation of choice accuracy for $x = 2$ pins down λ

Invariant Likelihood Ratio - Example

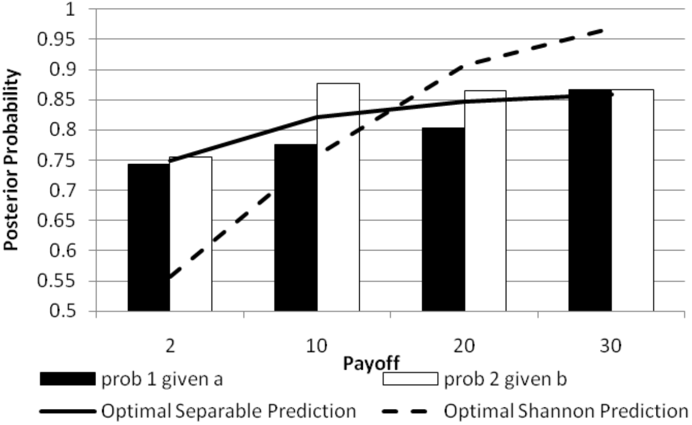


- Implies expansion path for all other values of x
- This does not hold in our experimental data

Invariant Likelihood Ratio - An Experimental Test



Fitting the Data



- Locally Invariant Posteriors
- Invariant Likelihood Ratio and Response to Incentives
- Symmetry

Symmetry

- Shannon Mutual Information has the property of **symmetry**
- Behavior invariant to the labelling of states

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

- Optimal beliefs depend **only** on the relative value of actions in that state
- Implies that there is no concept of 'perceptual distance'

A Simple Example

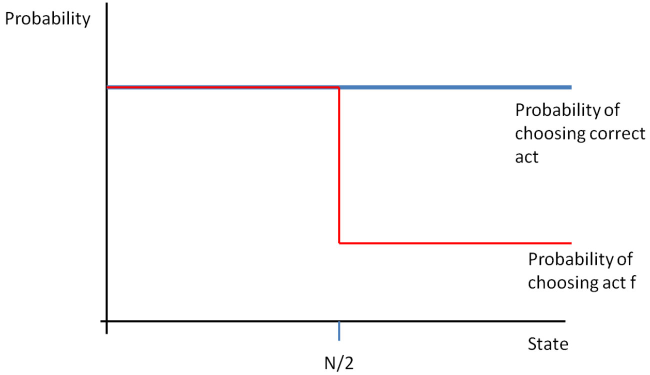
- N equally likely **states of the world** $\{1, 2....., N\}$
- Two **actions**

	Payoffs	
States	$1, \ldots \frac{N}{2}$	$\frac{N}{2} + 1, .., N$
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**

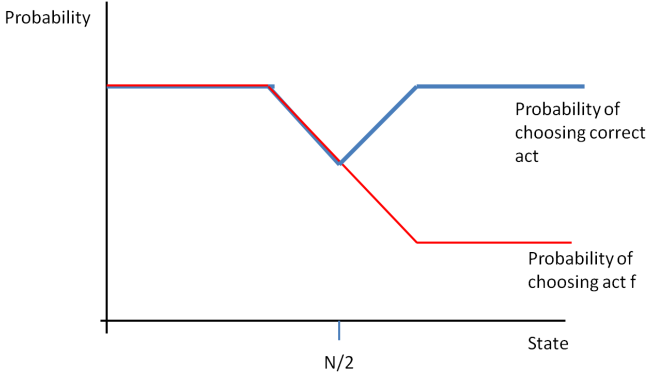
$$\frac{\exp\left(\frac{u(10)}{\lambda}\right)}{1 + \exp\left(\frac{u(10)}{\lambda}\right)}$$

Predictions for the Simple Problem - Shannon



- Probability of correct choice does not go down near threshold

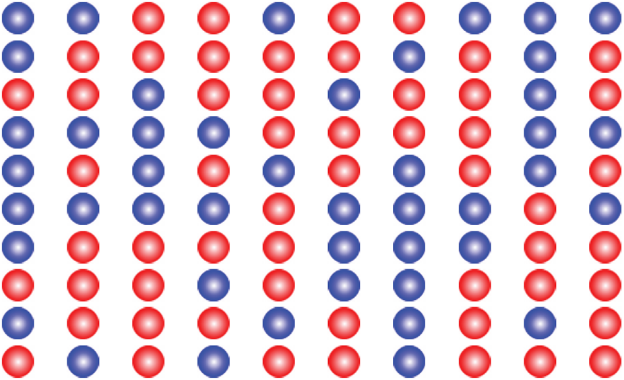
Predictions for the Simple Problem - Shannon



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

Symmetry

- Shannon Model makes strong predictions for the simple problem
  - Accuracy not affected by closeness to threshold
  - In contrast to (e.g.) uniform signals
- Which model is correct?
  - It may depend on the **perceptual environment**
- Test prediction in two different environments



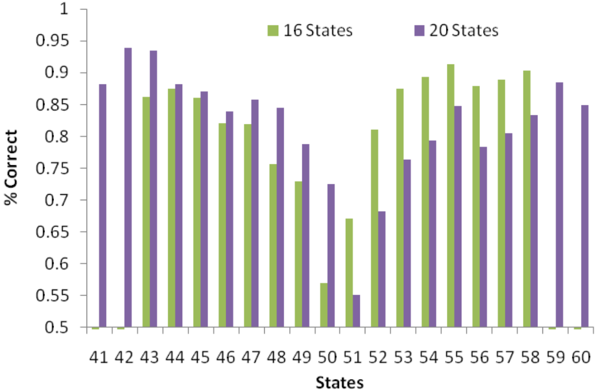
Action	**Payoff $\leq$ 50 Red**	**Payoff $>$ 50 Red**
f	10	0
g	0	10

Environment 2 (Letters)

J	P	P	J	J	L
P	N	K	N	K	M
J	Q	M	O	L	O
O	M	L	N	Q	J
Q	K	J			

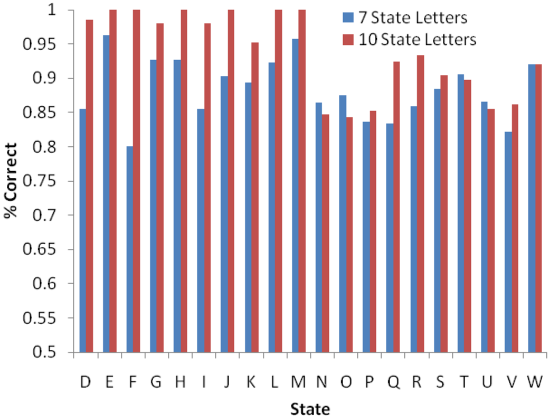
Action	**Payoff state letter $<$ N**	**Payoff state letter $\geq$ N**
f	10	0
g	0	10

- 2 treatments
- ‘Balls’ Experiment
  - 23 subjects
  - Vary the number of states
- ‘Letters’ Experiment
  - 24 subjects
  - Vary the relative frequency of the state letter
- Test whether probability of correct choice is lower nearer the threshold



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

Letters Experiment



- Probability of correct choice does vary between states
- But is not correlated with distance from threshold ($p=0.694$)

Alternative Cost Functions

- So far we have seen three potential failures of the Shannon Cost function
  - Prior invariance
  - Expansion path
  - Lack of percepual distance
- How can we address these shortcomings?

Alternative Cost Functions

- So far we have seen three potential failures of the Shannon Cost function
  - Prior invariance
  - Expansion path
  - Lack of percepual distance
- How can we address these shortcomings?

Prior Invariance

- Shannon implies a lack of response to priors which may be unrealistic
- We could replace costs based on Shannon mutual information with those based on Shannon **capacity**

$$C(\Omega, \Gamma) = \max_{\mu \in \Delta(\Omega)} I(\Omega, \Gamma)$$

- The maximial possible entropies across all prior distributions
- Intuitively, means that it is no longer cheap to be accurate in unlikely states
- See Woodford [2012]

- So far we have seen three potential failures of the Shannon Cost function
  - Prior invariance
  - Expansion path
  - Lack of percepual distance
- How can we address these shortcomings?

- We saw that people were less responsive to information than implied by Shannon
- Shannon Cost function:

$$K(\pi, \mu) = \lambda \left[-H(\mu) + \sum_{\gamma \in \Gamma(\pi)} P(\gamma) H(\gamma) \right].$$

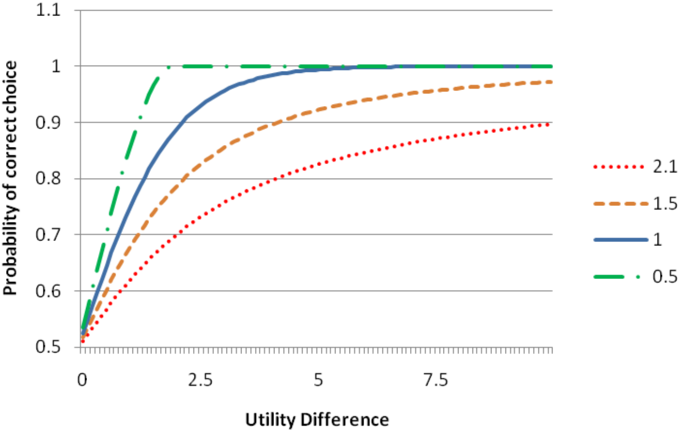
- Posterior- Separable cost functions:

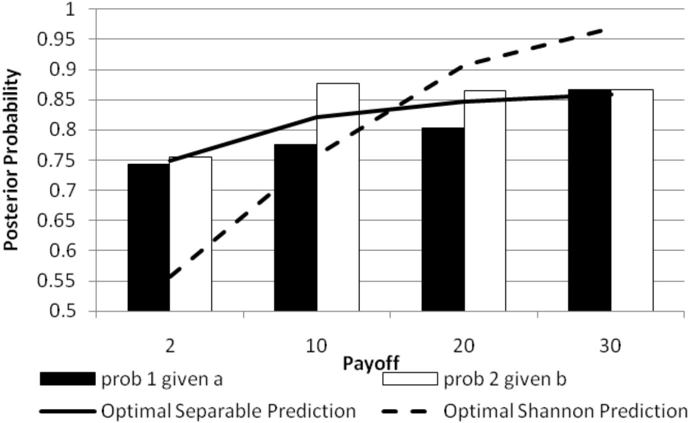
$$K(\pi, \mu) = \lambda \left[-L(\mu) + \sum_{\gamma \in \Gamma(\pi)} P(\gamma) L(\gamma) \right].$$

- where

$$L_{\{\rho,\lambda\}}(\gamma) = \begin{cases} -\lambda \left(\sum_{\Omega} \gamma(\omega) \left[\frac{\gamma(\omega)^{1-\rho}}{(\rho-1)(\rho-2)} \right] \right) \text{ if } \rho \neq 1 \text{ and } \rho \neq 2; \\ -\lambda \left(\sum_{\Omega} \gamma(\omega) \ln \gamma(\omega) \right) \text{ if } \rho = 1. \\ -\lambda \left(\sum_{\Omega} \gamma(\omega) \frac{\ln \gamma(\omega)}{\gamma(\omega)} \right) \text{ if } \rho = 2. \end{cases},$$

Response to Incentives: Posterior Separable Cost Functions




1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Posterior Probability
2
10
20
30
Payoff
prob 1 given a
prob 2 given b
Optimal Separable Prediction
Optimal Shannon Prediction

- So far we have seen three potential failures of the Shannon Cost function
  - Prior invariance
  - Expansion path
  - Expansion path
- How can we address these shortcomings?

- One promising approach being developed by Hubert and Woodford
- "Rational Inattention with Sequential Information Sampling"
- Watch this space....

Summary

- Introduced Shannon Mutual Information as a potential cost function
  - Popular in the literature
  - 'Cobb Douglas' vs 'Revealed Preference'
- Introduced some analytical tools to help solve the Shannon model
  - MM - necessary conditions
  - Necessary + Sufficient Conditions
  - Posterior-based approach
- Shown that the Shannon model can give rise to endogenous consideration set formation
- Discussed the experimental evidence for other behavioral implications
  - LIP
  - ILP
  - Symmetry