Rational Inattention

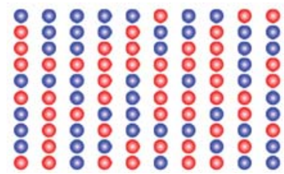
Mark Dean

ECON 1820 - Behavioral Economics

The Story So Far.....

- (Hopefully) convinced you that attention costs are important
- Introduced the 'satisficing' model of search and choice
- But, this model seems quite restrictive:
  - Sequential Search
  - 'All or nothing' understanding of alternatives
- Seems like a good model for choice over a large number of simple alternatives
- Not for a small number of complex alternatives

A Non-Satisficing Situation



Act	Payoff 47 red dots	Payoff 53 red dots
$\mathbf{f}_2$	20	0
$\mathbf{g}_2$	0	10

Rational Inattention

- An alternative model of information gathering
- The world can be in one of a number of different states
  - 47 or 53 balls on a screen
  - Demand for your product can be high or low
  - Quality of a used car can be good or bad
- Initially have some beliefs about the likelihood of different states of the world (prior)
- By exerting effort, we can learn more
  - Count some of the balls
  - Run a customer survey
  - Ask a mechanic to look at the car
- But this learning comes with costs
  - Time, cognitive effort, money, etc.

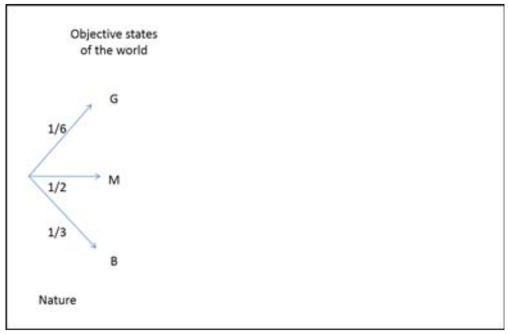
The Choice Problem

- The specifics of the process of information acquisition may be very complex
- We model the choice of information in an *abstract* way
- The decision maker chooses an *information structure*
  - Set of signals to receive
  - Probability of receiving each signal in each state of the world
- Then choose what action to take based only on the signal.
  - Value of the action depends on the state of the world
- More informative information structures are more costly, but lead to better decisions
  - Sets up a trade off

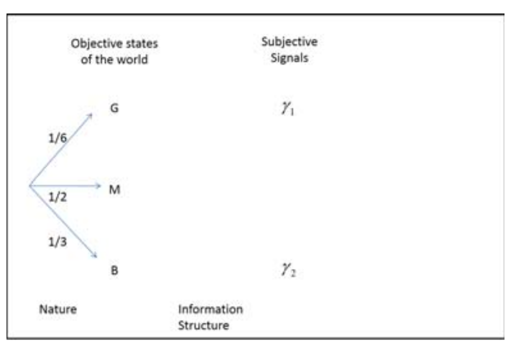
Set Up

- Objective states of the world
  - e.g. Demand could be 'good', 'medium' or 'bad'
- Decision maker chooses an action
  - e.g. Set price to be high, average, or low
- Gross payoff depends on action and state
  - e.g. Quantity sold depends on price and demand
- Decision maker get to learn something about the state before choosing action
  - e.g. Could do market research, focus groups, etc.

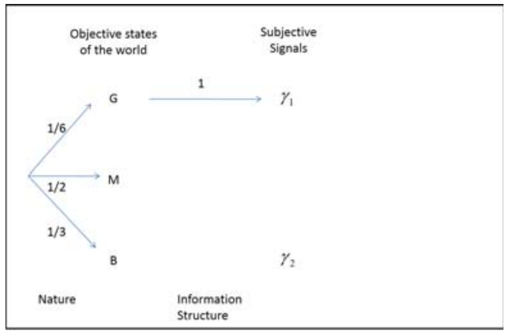
The Choice Problem



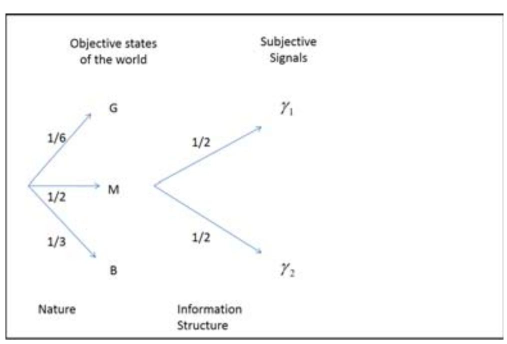
The Choice Problem



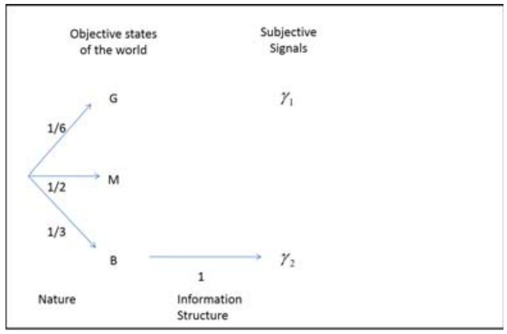
The Choice Problem



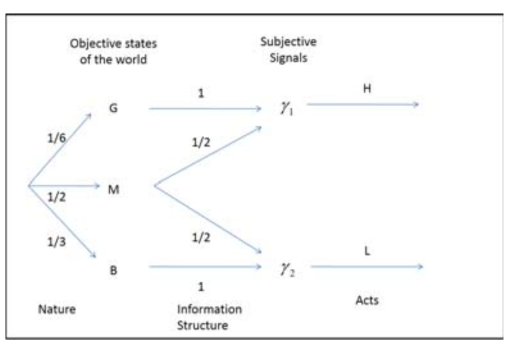
The Choice Problem



The Choice Problem



The Choice Problem



Describing an Information Structure

- $\Omega = \{\omega_1, ...\omega_M\}$: States of the world (number of balls, quality of the car, etc)
  - with prior probabilities μ
- Information structure defined by:
  - Set of signals: $\Gamma(\pi)$
  - Probability of receiving each signal γ from each state $\omega : \pi(\gamma|\omega)$
- In previous example

	Signal (Γ)	
State (Ω)	R	S
G	1	0
M	$\frac{1}{2}$	$\frac{1}{2}$
B	0	1

What Information Structure to Choose?

- Better information will lead to better choices
- But will cost more
  - Time, effort, money etc
- How to decide what information structure to choose?
- Trade off
  - Benefit of information (easy to measure)
  - Cost of information (hard to measure)
- Assume that this trade off is done *optimally*

The Value of An Information Structure

- What is the value of an information structure?
- In the end you will have to choose an *action*
  - Defined by the outcome it gives in each state of the world
- In previous example, could choose three actions
  - set price H, A or L
- The following table could describe the profits each price gives at each demand level

	Price		
State	H	A	L
G	10	3	1
M	1	2	1
B	-10	-3	-1

- Let $u(a(\omega))$ be the utility (profit) that action a gives in state ω

The Value of An Information Structure

- What would you choose if you gathered no information?
  - i.e. if you had your prior beliefs
  - Use μ to describe the prior

$$\mu(G) = \frac{1}{6}, \mu(M) = \frac{1}{2}, \mu(B) = \frac{1}{3}$$

- Calculate the expected utility for each act

$$\begin{aligned} \frac{1}{6}u(H(G)) + \frac{1}{2}u(H(M)) + \frac{1}{3}u((H(B)) &= \frac{-7}{6} \\ \frac{1}{6}u(A(G)) + \frac{1}{2}u(A(M)) + \frac{1}{3}u((A(B)) &= \frac{1}{2} \\ \frac{1}{6}u(L(G)) + \frac{1}{2}u(L(M)) + \frac{1}{3}u((L(B)) &= \frac{1}{3} \end{aligned}$$

- Choose A
- Get utility $\frac{1}{2}$

The Value of An Information Structure

- What would you choose upon receiving signal R?
- Depends on beliefs conditional on receiving that signal
- Luckily we can calculate this using Bayes Rule

$$\begin{aligned} P(G|R) &= \frac{P(G \cap R)}{P(R)} \\ &= \frac{\mu(G)\pi(R|G)}{\mu(G)\pi(R|G) + \mu(M)\pi(R|M) + \mu(B)\pi(R|B)} \\ &= \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{4} + 0} = \frac{2}{5} \end{aligned}$$

The Value of An Information Structure

- We can therefore calculate posterior beliefs conditional on signal R

$$\begin{aligned} P(G|R) &= \frac{2}{5} = \gamma^R(G) \\ P(M|R) &= \frac{3}{5} = \gamma^R(M) \\ P(B|R) &= 0 = \gamma^R(B) \end{aligned}$$

- And calculate the value of choosing each act given these beliefs

$$\begin{aligned} \frac{2}{5}u(H(G)) + \frac{3}{5}u(H(M)) &= \frac{23}{5} \\ \frac{2}{5}u(A(G)) + \frac{3}{5}u(A(M)) &= \frac{12}{5} \\ \frac{2}{5}u(L(G)) + \frac{3}{5}u(L(M)) &= \frac{2}{5} \end{aligned}$$

The Value of An Information Structure

- If received signal R, would choose H and receive $\frac{23}{5}$
- By similar process, can calculate that if received signal S
  - Choose L and receive $-\frac{1}{7}$
- Can calculate the value of the information structure as

$$\begin{aligned} P(R)\frac{23}{5} + P(S)\frac{-1}{7} &= \\ \frac{5}{12}\frac{23}{5} + \frac{7}{12}\frac{-1}{7} &= \frac{11}{6} \end{aligned}$$

- How much would you pay for this information structure?

The Value of An Information Structure

- Value of this information structure is $\frac{11}{6}$
- Value of being uninformed is $\frac{1}{2}$
- Would prefer this information structure to being uninformed if cost is below $\frac{8}{6}$
- Note that the value of an information structure depends on the acts available

$$\begin{aligned} G(\pi, A) &= \sum_{\gamma \in \Gamma(\pi)} P(\gamma) g(\gamma, A) \\ g(\gamma, A) &= \max_{a \in A} \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \end{aligned}$$

- $g(\gamma, A)$ value of receiving signal γ if available actions are A
  - Highest utility achievable given the resulting posterior beliefs

The Choice of Information Structure

- What information structure would you choose?
- In general, more information means better choices, and higher values
- Without further constraints, would choose to be fully informed
- To make the problem interesting and realistic, need to introduce a **cost to information** K
- The 'net value' of an information structure π in choice set A is

$$G(\pi, A) - K(\pi)$$

What is the cost of information?

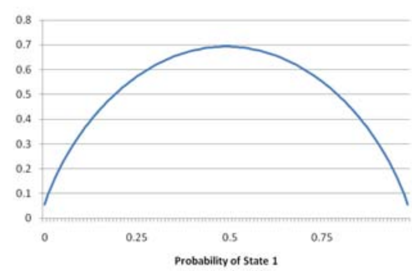
- What form should information costs K take?
- Good question!
- Many alternatives have been considered in the literature
  - Normal Signals
  - All or nothing
  - Partitions
- We will focus on 'Shannon mutual information' (Sims 2003)
  - A way of measuring how much information is gained by using an information structure

Shannon Entropy

- 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 $i = 1 \ldots n$, 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
  - i.e. actually determining what x is

Justification for Shannon Entropy

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

Justification for Shannon Entropy

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

Justification for Shannon Entropy

- Say we want our measure of information to have the following features
- Depends only on the probability distribution
- Maximized at a uniform probability distribution
- Unaffected by adding zero probability state
- Additive
  - $H(X, Y) = H(X) + \sum_x P(x) H(Y|x)$

Justification for Shannon Entropy

- Say we want our measure of information 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 it must be of the form (Khinchin 1957)

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

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
- Can be rewritten as

$$\begin{aligned} & \sum_y p(y) \sum_x p(x|y) \ln p(x|y) - \sum_y p(x) \ln p(x) \\ = \quad & H(X) - \sum_y P(y) H(X|y) \end{aligned}$$

- The expected reduction in entropy about variable x from observing y

Mutual Information and Information Costs

- Mutual Information measures the expected reduction in entropy from observing a signal
- We can use it as a measure of information costs

$$K(\pi,\mu) = -\lambda\,[\text{ expected entropy of signals - entropy of prior}]$$

$$= -\lambda \left[\sum_{\gamma \in \Gamma(\pi)} P(\gamma) \sum_{\omega \in \Omega} \gamma(\omega) \ln \gamma(\omega) - \sum_{\omega \in \Omega} \mu(\omega) \ln \mu(\omega) \right]$$

- Can be justified by information theory
  - Consider a signal which consists of a sequence of n ones and zeros (an information channel)
  - An information structure can be achieved by an information channel if and only if the expected decrease in the entropy is less than the amount of information processed
  - Proportional to n

Working with Rational Inattention

- Now we have defined information costs, the optimization problem is well defined
- For any set of alternatives A, choose π to maximize

$$G(\pi, A) - K(\pi)$$

- What does this tell us about behavior?

A Simple Example

- Consider the case of two state and two acts

	ω_1	ω_2
a	$U(a(\omega_1))$	$U(a(\omega_2))$
b	$U(b(\omega_1))$	$U(b(\omega_2))$

- It is easy to show that decision maker will never choose more than 2 signals
  - Why?
- Assume $\mu(1) = \mu(2) = 0.5$
- Assume that they do choose two signals
  - γ^a, after which a is chosen
  - γ^b, after which b is chosen

Solving for Optimal Behavior

- Choose
  - $P(\gamma^a)$: Probability of signal γ^a
  - $\gamma^a(\omega_1)$: Posterior probability of state ω_1 following γ^a
  - $\gamma^b(\omega_1)$: Posterior probability of state ω_1 following γ^b
- To maximize

$$P(\gamma^a)\left[\gamma^a(\omega_1)u(a(\omega_1)) + (1-\gamma^a(\omega_1))u(a(\omega_2))\right] +$$
$$(1-P(\gamma^a))\left[\gamma^b(\omega_1)u(b(\omega_1)) + (1-\gamma^b(\omega_1))u(b(\omega_2))\right]$$

$$-\lambda\left[\begin{array}{c} P(\gamma^a)\left(\begin{array}{c}\gamma^a(\omega_1)\ln\gamma^a(\omega_1)+\\ (1-\gamma^a(\omega_1))\ln(1-\gamma^a(\omega_1))\end{array}\right) \\ (1-P(\gamma^a))\left(\begin{array}{c}\gamma^b(\omega_1)\ln\gamma^b(\omega_1)\\ (1-\gamma^b(\omega_1))\ln(1-\gamma^b(\omega_1))\end{array}\right)\end{array}\right]$$

- subject to

$$P(\gamma^a)\gamma^a(\omega_1) + (1-P(\gamma^a))\gamma^b(\omega_1) = \mu(\omega_1)$$

Implies

- You will show

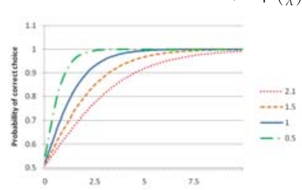
$$\frac{\gamma^a(\omega_1)}{\gamma^b(\omega_1)} = \exp\left(\frac{u(a(\omega_1)) - u(b(\omega_1))}{\lambda}\right)$$
$$\frac{\gamma^a(\omega_2)}{\gamma^b(\omega_2)} = \exp\left(\frac{u(a(\omega_2)) - u(b(\omega_2))}{\kappa}\right)$$

- Ratio of beliefs in each states depends only on the 'cost of mistakes' in that state
- Posterior beliefs do not depend on priors

Implies

- We can use these formula to calculate how probability of correct choice changes with reward.
- Assume
  - $u(a(\omega_1)) = u(b(\omega_2)) = c$, $u(a(\omega_2)) = u(b(\omega_2)) = 0$,
- Imples that

$$\pi(\gamma^a|\omega_1) = \pi(\gamma^b(\omega_2)) = \frac{\exp\left(\frac{c}{\lambda}\right)}{1+\exp\left(\frac{c}{\lambda}\right)}$$



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?

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

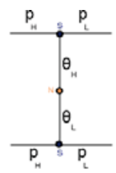
- Once off sales encounter
  - One buyer, one seller, one product

Market Setting

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

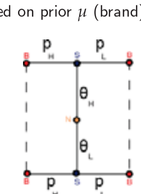
Market Setting

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



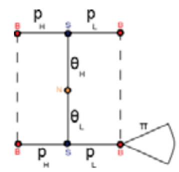
Market Setting

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



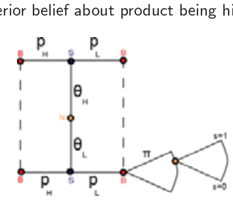
Market Setting

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



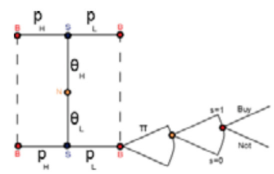
Market Setting

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



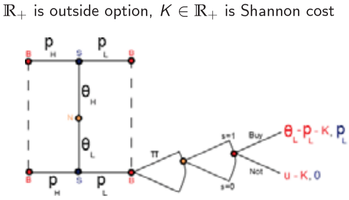
Market Setting

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



Market Setting

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



Equilibrium

- An equilibrium of the model is
  - 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 cannot make any more profit by changing their strategy
  - Consumers cannot increase their utility by changing their strategy
  - Beliefs are correct

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

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
  - Strategic ignorance: Buyers never attend, strong beliefs
- 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

Equilibrium

Theorem

For every cost λ, 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

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

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

Explaining the Equilibrium

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

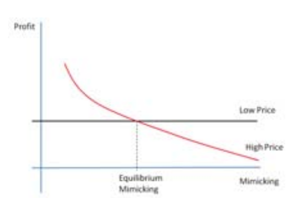
$$d_{p_H}^{\theta_L} = \frac{\left(\frac{1-\gamma_{p_H}^1}{\gamma_{p_H}^1-\gamma_{p_H}^0}\right)\left(\mu_{p_H}(H) - \gamma_{p_H}^0\right)}{\left(1-\mu_{p_H}(H)\right)}$$

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

Firm Behavior

- What about firm behavior?
- When $\eta \in (0,1)$, need low quality seller indifference:

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



Equilibrium

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

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

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

Other Applications

- Consumption and Savings [Sims 2003]
  - Standard permanent income hypothesis: consumption responds immediately and fully to changes in income
  - Rational Inattention: consumption responses occur gradually over time
  - Fits stylized facts in the macro literature
- Discrete Pricing [Matejka 2010]
  - Standard model: Firms prices should respond continuously to cost shocks
  - Rational Inattention: Firms will 'jump' between a small number of discrete prices
  - In line with observed date
- Home Bias [Van Nieuwerburgh and Veldkamp 2009]
  - Standard model: investors should diversify portfolio internationally
  - Rational Inattention: investors should specialize in assets they know more about
  - Leads to 'Home Bias' in investment

Summary

- Rational Inattention provides a way of modelling how people choose to learn about the state of the world
  - Applicable in cases in which satisficing is not appropriate
- Assumes people choose information to maximize value net of costs
  - Value depends on the choices to be made
  - Costs generally based on Shannon Entropy
- We can make predictions about learning and choice based on the rewards available in the environment
- Can be used to address a number of 'puzzles'