Choice Set Effects

Mark Dean

Behavioral Economics G6943
Autumn 2018

Context Effects

- We will now think about context effects that are more general than simple reference dependence
- Standard Model: Adding options to a choice set can only affect choice in a very specific way
  - Either a new option is chosen or it isn't
  - Independence of Irrelevant Alternatives
- Work from economics, neuroscience and psychology suggest a different channel
  - Changing the choice set changes the *context* of choice
  - Context affects preferences
  - Can lead to violations of IIA

Observing Context Effects

- We are going to consider three examples

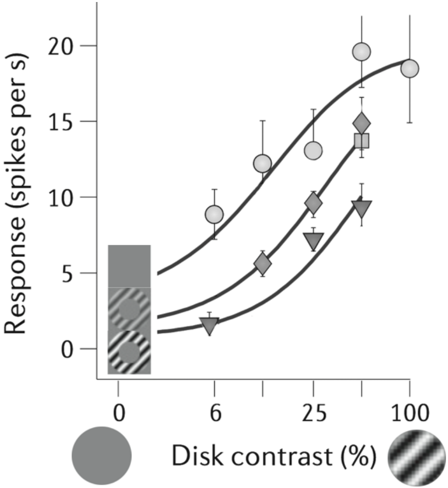
1. Stochastic Choice
   - Divisive Normalization: Louie, Khaw and Glimcher [2013]
2. Choice between multidimensional alternatives
   - Relative Thinking: Bushong, Rabin and Schwartzstein [2015]
   - Salience: Bordalo, Gennaioli and Shleifer [2012]

- Examples of models in which the choice set affects how distances are perceived

- The brain needs some way of representing (or encoding) stimuli
  - Brightness of visual stimuli
  - Loudness of auditory stimuli
  - Temperature etc.
- Typically, a given brain region will have the task of encoding a particular stimuli at a particular point in space and time
  - e.g. the brightness of a light at a particular point in the visual field
- How is this encoding done?
- A 'naive' mode: neural activity encodes the absolute value of the stimuli

$$\mu_i = KV_i$$

  - μ_i: neural activity in a particular region
  - V_i: The value of the related stimuli



- Encoding depends not only on the value of the stimuli, but also on the context [Carandini 2004]

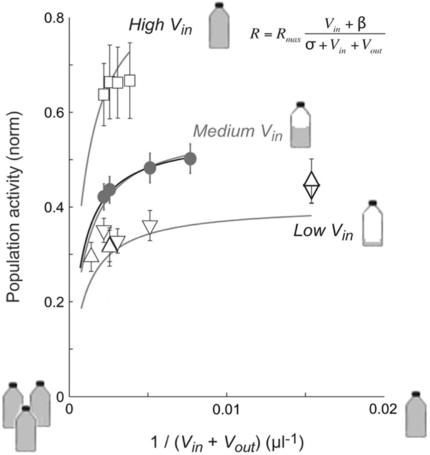
- Divisive Normalization:

$$\mu_i = K \frac{V_i}{\sigma_H + \sum_j w_j V_j}$$

  - σ_H: Normalizing constant (semi-saturation)
  - w_j: Weight of comparison stimuli j
  - V_j: Value of comparison stimuli j

- Why would the brain do this?
  - Efficient use of neural resources [Carandini and Heager 2011]
  - Neurons can only fire over a finite range
  - Want the same system to work (for example) in very bright and very dark conditions
  - Absolute value encoding is inefficient
    - In dark environments, everything encoded at the bottom of the scale
    - In light environments, everything encoded at the top of the scale
  - Normalization encodes relative to the mean of the available options
  - Encodes things near the middle of the scale.
- Allows stimuli to be encoded further apart from each other
  - Reduces errors that occur due to noise

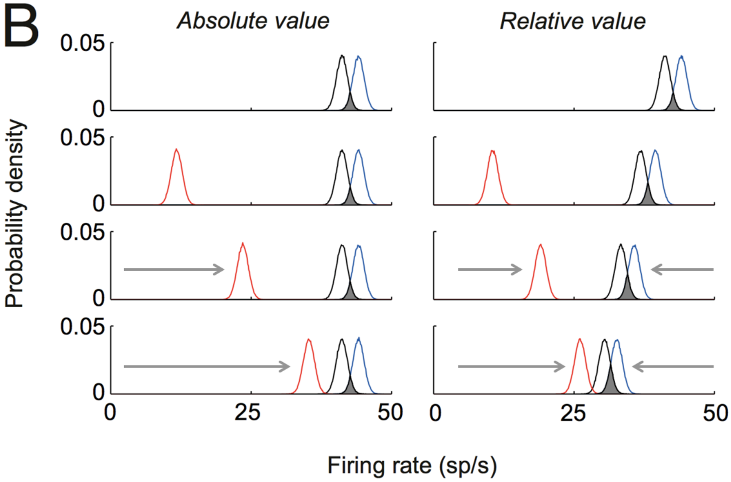
Divisive Normalization and Choice

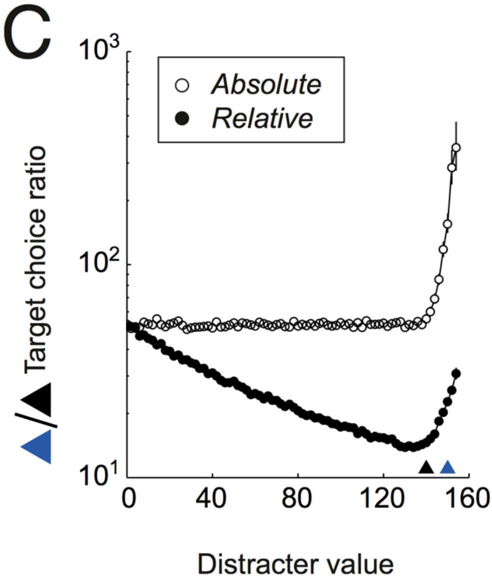


- There is also evidence that the value of choice alternatives is normalized [Louie et. al. 2011]

Divisive Normalization and Choice

- Why should normalization matter for choice?
- Does not change the ordering of the valuation of alternatives, so why should it change choice?
- Because choice is *stochastic*
- The above describes *mean* firing rates
- Choice will be determined by a draw from a random distribution around that mean
  - Claim that such stochasticity is an irreducible fact of neurological systems
- Probability of choice depends on the difference between the encoded value of each option
- Utility has a cardinal interpretation, not just an ordinal one


B
Absolute value
Relative value
Probability density
0.05
0
0.05
0
0.05
0
0.05
0
0
25
50
0
25
50
Firing rate (sp/s)


C
10^3
10^2
10^1
Absolute
Relative
Target choice ratio
0
40
80
120
160
Distracter value

- How do these predictions vary from standard random utility model?
- Luce model:

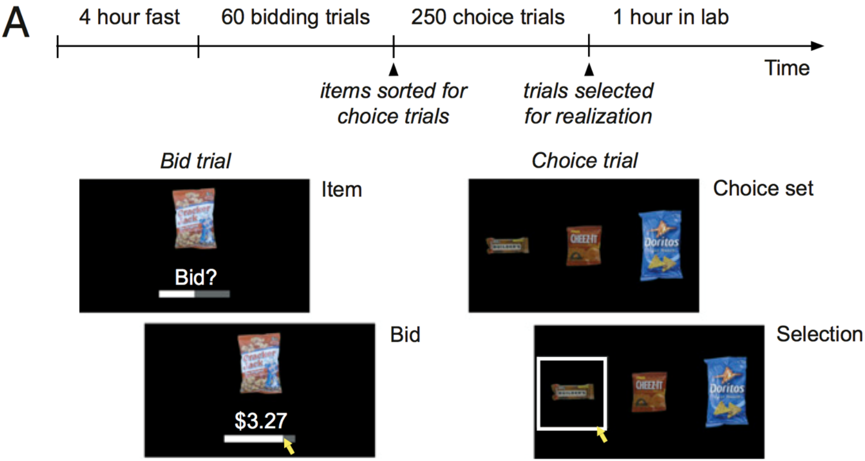
$$p(a|A) = \frac{u(a)}{\sum_{b \in A} u(b)}$$

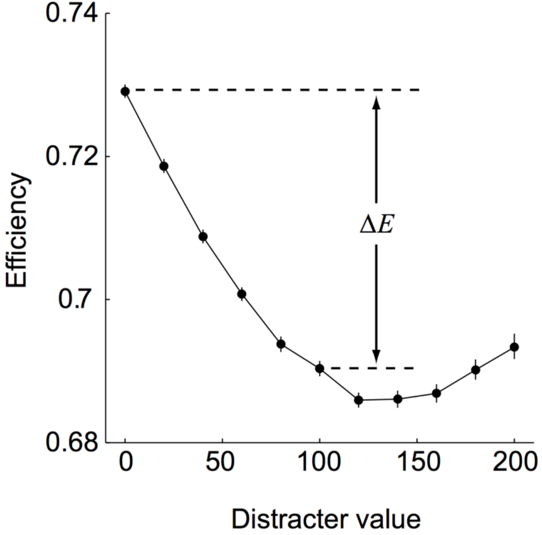
  - Implies that the relative likelihood of picking a over b is independent of the other available alternatives
  - Stochastic IIA
- More general RUM
  - Adding an alternative c can affect the relative likelihood of choosing a and b
  - But only because c itself is chosen
  - Can 'take away' probability from a or b
  - The amount c is chosen bounds the effect it can affect the choice of a or b

Experimental Evidence

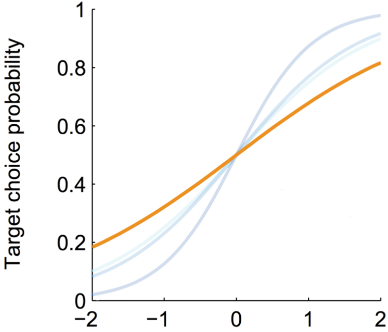
- Subjects (40) took part in two tasks involving snack foods

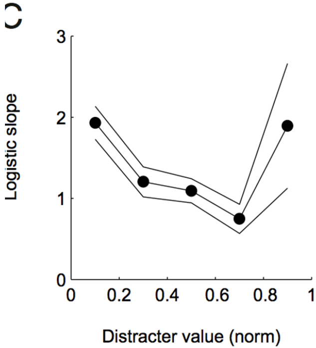
1. Asked to *bid* on each of 30 different snack foods to elicit valuation
    - BDM procedure used to make things incentive compatible
2. Asked to make a choice from three alternatives
    - Target, alternative and distractor
    - 'True' value of each alternative assumed to be derived from the bidding stage


A
4 hour fast
60 bidding trials
250 choice trials
1 hour in lab
Time
items sorted for choice trials
trials selected for realization
Bid trial
Item
Bid?
Bid
$3.27
Choice trial
Choice set
Selection


0.74
0.72
0.7
0.68
Efficiency
0
50
100
150
200
Distracter value
ΔE

Experimental Evidence





- But see "A neural mechanism underlying failure of optimal choice with multiple alternatives" by Chau et al. Nat Neurosci. 2014 Mar; 17(3): 463–470.

Choice with Multidimensional Alternatives

- In the model we just saw, adding a third ‘distractor’ changed the ‘distance’ between the value of two targets
    - Context changed apparent magnitude of the difference
- This could not be seen in 'standard' choice data
    - Is observable in stochastic choice

Choice with Multidimensional Alternatives

- Another data set in which such effects could be observed is choice over goods defined over multiple attributes
  - $c = \{c_1, ..., c_K\}$
- Utility is assumed additive,

$$U(c|A) = \sum_{k=1}^{K} w_k^A u_k(c_k)$$

  - $u_k(.)$ the true (context independent) utility on dimension k
  - w_k^A is a context dependent weight on dimension k
- Utility also assumed to be observable
  - Koszegi and Szeidl [2013] suggest how this can be done
- Context can change the distance between values on one dimension
  - Change the trade off relative to other dimensions

Choice with Multidimensional Alternatives

- Many recent papers make use of this framework
  - Bordalo, Gennaioli and Shleifer [2012, 2013]: Salience
  - Soltani, De Martino and Camerer [2012]: Range Normalization
  - Cunningham [2013]: Comparisons
  - Koszegi and Szeidl [2013]: Focussing
  - Bushong, Rabin and Schwartzstein [2015]: Relative thinking
  - Landry and Webb [2018]: Pairwise Divisive Normalization

- Many recent papers make use of this framework
  - Bordalo, Gennaioli and Shleifer [2012, 2013]: Salience
  - Soltani, De Martino and Camerer [2012]: Range Normalization
  - Cunningham [2013]: Comparisons
  - Koszegi and Szeidl [2013]: Focussing
  - Bushong, Rabin and Schwartzstein [2015]: Relative thinking
  - Landry and Webb [2018]: Pairwise Divisive Normalization
- We will consider these two

- In the Louie et al. [2013] paper, normalization was relative to the *mean* of the value of the available options
- There is also a long psychology literature which suggests that *range* can play an important role in normalization
- A given absolute difference will seem *smaller* if the total range under consideration seems *larger*
- Bushong et al. [2015] suggest conditions on the weights w_k^A to capture this effect .

$$U(c|A) = \sum_{k=1}^{K} w_k^A u_k(c_k)$$

Relative Thinking: Assumptions

1. $w_k^A = w(\Delta_k(A))$ where

$$\Delta_k(A) = \max_{a \in A} u_k(a_k) - \min_{a \in A} u_k(a_k)$$

   - The weight given to dimension k depends on the range of values in this dimension

2. $w_k^A(\Delta)$ is diffable and decreasing in Δ
   - A given absolute difference receives less weight as the range increases

3. $w_k^A(\Delta)\Delta$ is strictly increasing, with $w(0)0 = 0$
   - The change in weight cannot fully offset a change in absolute difference

4. $\lim_{\Delta \to \infty} w(\Delta) > 0$
   - Absolute differences still matter even as the range goes to infinity

- An example of such a function

$$w_k^A(\Delta) = (1-\rho) + \rho \frac{1}{\Delta^\alpha}$$

- Bushong et al. [2015] do not fully characterize the behavioral implications of their model
  - Potentially interesting avenue for future research
- However, some of the implications are made clear in the following examples

Example 1

$$c = \left\{ \begin{array}{c} 2 \\ 3 \\ 0 \end{array} \right\}, \; c' = \left\{ \begin{array}{c} 0 \\ 0 \\ 5 \end{array} \right\}$$

- Assume these payoffs are in utility units
- What will the DM choose?
- They would choose c, despite the fact that the 'unweighted' utility of the two options is the same

$$2w(2) + 3w(3) > 5w(3) > 5w(5)$$

- DM favors benefits spread over a large number of dimensions

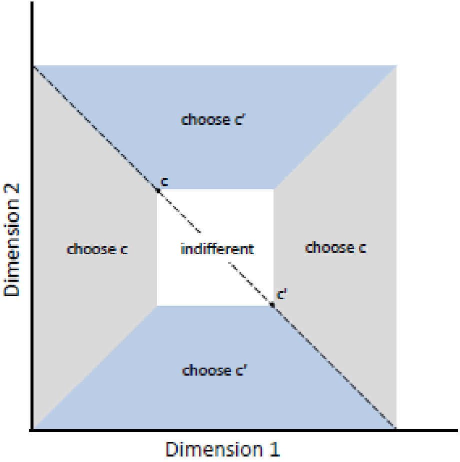
$$c = \left\{ \begin{array}{c} 2 \\ 1 \end{array} \right\}, \; c' = \left\{ \begin{array}{c} 1 \\ 2 \end{array} \right\}$$

- Assume utility is linear
- Say that, in the choice set $\{c, c'\}$ the DM is indifferent between the two.
- What would they choose from

$$c = \left\{ \begin{array}{c} 2 \\ 1 \end{array} \right\}, \; c' = \left\{ \begin{array}{c} 1 \\ 2 \end{array} \right\}, \; c'' = \left\{ \begin{array}{c} 2 \\ 0 \end{array} \right\}$$

- They would choose c
- The introduction of c'' increases the range of dimension 2, but not dimension 1
- Reduces the weight on the dimension in which c' has the advantage
- This is an example of the asymmetric dominance effect

Example 2


choose c'
c
choose c
indifferent
choose c
c'
choose c'
Dimension 2
Dimension 1

Salience Theory

- Basic Idea: Attention is not spread evenly across the environment
- Some things draw our attention whether we like it or not
    - Bright lights
    - Loud noises
    - Funky dancing
- The things that draw our attention are likely to have more weight in our final decision
- Notice here that attention allocation is *exogenous* not *endogenous*
    - Potentially could be thought of as a reduced form for some endogenous information gathering strategy

- Bordalo, Gennaioli and Shleifer [2013] formulate salience in the following way

$$\begin{aligned} U(c|A) &= \sum_{k=1}^{K} w_{k,c}^{A} u_k(c_k) \\ &= \sum_{k=1}^{K} w_{k,c}^{A} \theta_k c_k \end{aligned}$$

  - θ_k is the 'true' utility of dimension k
  - $w_{k,c}^{A}$ is the 'salience' weight of dimension k for alternative c
- Notice that the weight that dimension k receives may be different for different alternatives

Determining Salience

- How are the weights determined?
- First define a 'Salience Function'

$$\sigma(c_k, \bar{c}_k)$$

  - $\bar{c}_k$ is the reference value for dimension k (usually, but not always, the mean value of dimension k across all alternatives)
  - $\sigma(c_k, \bar{c}_k)$ is the salience of alternative c on dimension k

- Properties of the Salience function
  1. Ordering: $[\min(c_k, \bar{c}_k), \max(c_k, \bar{c}_k)] \supset [\min(c'_k, \bar{c}'_k), \max(c'_k, \bar{c}'_k)] \Rightarrow \sigma(c'_k, \bar{c}'_k) \leq \sigma(c_k, \bar{c}_k)$
  2. Diminishing Sensitivity: $\sigma(c_k + \varepsilon, \bar{c}_k + \varepsilon) < \sigma(c_k, \bar{c}_k)$
  3. Reflection: $\sigma(c'_k, \bar{c}'_k) > \sigma(c_k, \bar{c}_k) \Rightarrow \sigma(-c'_k, -\bar{c}'_k) > \sigma(-c_k, -\bar{c}_k)$

- An example of a salience function

$$\sigma(c_k, \bar{c}_k) = \frac{|c_k - \bar{c}_k|}{|c_k| + |\bar{c}_k|}$$

- Note:
  - Shares some features with both the previous approaches we have seen
  - Normalization by the mean
  - Diminishing sensitivity (but relative to zero, rather than the range)
  - The precise differences in the behavioral implications between these different models is somewhat murky

From Salience to Decision Weights

- Use $\sigma(c_k, \bar{c}_k)$ to rank the salience of different dimensions for good c
    - $r_{k,c}$ is the salience rank of dimension k (1 is most salient)
- Assign weight $w^A_{k,c}$ as

$$\frac{\delta^{r_{k,c}}}{\sum_j \theta_j \delta^{r_{j,c}}}$$

- Then plug into

$$\sum_{k=1}^{K} w^A_{k,c} \theta_k c_k$$

- More salient alternatives get a higher decision weight
- δ indexes degree to which subject is affected by salience
    - lower δ, more affected by salience

Application: Choice Under Risk

- Bordalo et al [2012] apply the salience model to choice under risk
- Choice objects are lotteries on a subjective state space
- Dimensions are states of the world
  - c_k is the utility provided by lottery c in state of the world k
  - θ_k is the objective probability of state of the world k
- Someone who does not have salience effects maximizes expected utility
- Salience leads to probability weighting
- Note: in binary choices, assume that each alternative has the same salience for each state
- e.g.

$$\sigma(c_k, c'_k) = \frac{|c_k - c'_k|}{|c_k| + |c'_k| + \lambda}$$

- Example: Salience and the Allais Paradox
- Allais Paradox: Consider the following pairs of choices:

$$\begin{aligned} c &= (0.33 : 2500; 0.01 : 0; 0.66 : 2400) \\ \text{or } c' &= (0.34 : 2400; 0.66 : 2400) \end{aligned}$$

$$\begin{aligned} \bar{c} &= (0.33 : 2500; 0.01 : 0; 0.66 : 0) \\ \text{or } \bar{c}' &= (0.34 : 2400; 0.66 : 0) \end{aligned}$$

- Typical choice is c' over c but $\bar{c}$ over $\bar{c}'$
- Inconsistent with expected utility theory
- Can be explained by salience

- Consider choice 1

$$\begin{aligned} c &= (0.33 : 2500; 0.01 : 0; 0.66 : 2400) \\ \text{or } c' &= (0.34 : 2400; 0.66 : 2400) \end{aligned}$$

- Represent by the following state space:

State	c	c'
s_1	2500	2400
s_2	0	2400
s_3	2400	2400

- State s_2 is the most salient state, receives most weight
- c' chosen if

$$\delta 0.33 \times 100 < 0.01 \times 2400$$

- More susceptible to salience, the more likely to choose c'

- Consider choice 2

$$\begin{aligned} \bar{c} &= (0.33 : 2500; 0.01 : 0; 0.66 : 0) \\ \text{or } \bar{c}' &= (0.34 : 2400; 0.66 : 0) \end{aligned}$$

- *Assume independence* and represent by the following state space:

State	$\bar{c}$	$\bar{c}'$
s_1	2500	2400
s_2	2500	0
s_3	0	2400
s_4	0	0

- Salience ranking is s_2, then s_3, then s_1
- Now the upside of $\bar{c}$ is most salient
- $\bar{c}'$ chosen if

$$0.33 \times 0.66 \times 2500 - \delta 0.67 \times 0.34 \times 2400 + \delta^2 0.33 \times 0.34 \times 100 < 0$$

- Which is never true for $\delta \geq 0$

Summary

- There is a large body of evidence which suggests that context effects are important in economic choice
- This is a violation of the standard model (via IIA)
- A new class of models have tried to explain these effects via the channel of 'normalization'
  - The context of a choice affects whether a given difference is seen as big or small
- Many open questions in this literature
  - Type of normalization
  - What is the 'context'?
  - How do we behaviorally differentiate between classes of models?