Efficient Perceptual Coding and Reference-Dependent Valuations

Michael Woodford

Columbia University

7th Conference on Monetary Economics
Banco de Portugal
June 15-16, 2012

Reference-Dependent Valuations

- Experimentally observed choices not always consistent with the existence of consistent preferences over final outcomes (independent of path by which they are reached)

  — a key feature of the prospect theory of Kahneman and Tversky (1979)

Kahneman-Tversky (1979)

Problem

In addition to whatever you own, you have been given 1000. You are now asked to choose between (a) winning an additional 500 with certainty, or (b) a gamble with a 50 percent chance of winning 1000 and a 50 percent chance of winning nothing.

Kahneman-Tversky (1979)

Problem

In addition to whatever you own, you have been given 1000. You are now asked to choose between (a) winning an additional 500 with certainty, or (b) a gamble with a 50 percent chance of winning 1000 and a 50 percent chance of winning nothing.

Problem

In addition to whatever you own, you have been given 2000. You are now asked to choose between (a) losing 500 with certainty, and (b) a gamble with a 50 percent chance of losing 1000 and a 50 percent chance of losing nothing.

Kahneman-Tversky (1979)

Problem

In addition to whatever you own, you have been given 1000. You are now asked to choose between (a) winning an additional 500 with certainty, or (b) a gamble with a 50 percent chance of winning 1000 and a 50 percent chance of winning nothing.

Majority of subjects choose (a)

Problem

In addition to whatever you own, you have been given 2000. You are now asked to choose between (a) losing 500 with certainty, and (b) a gamble with a 50 percent chance of losing 1000 and a 50 percent chance of losing nothing.

Majority of subjects choose (b)

Kahneman-Tversky (1979)

- Problem for standard EU theory: in both cases, subjects are choosing between the same probability distributions over final wealth levels:

(a) initial wealth + 1500 with certainty

VS

(b) 50 percent chance of initial wealth + 1000,
50 percent chance of initial wealth + 2000

Kahneman-Tversky (1979)

- Problem for standard EU theory: in both cases, subjects are choosing between the same probability distributions over final wealth levels:

  (a) initial wealth + 1500 with certainty

  VS

  (b) 50 percent chance of initial wealth + 1000,
  50 percent chance of initial wealth + 2000

- Explanation proposed by prospect theory: people don't evaluate only distribution over final situations, they care about distribution of gains or losses relative to "reference point" (wealth prior to the choice)

Explaining Reference-Dependence

- Kahneman-Tversky (1979), Koszegi-Rabin (2006) seek to incorporate ref-dependence into economic theory through non-standard preferences, maximized under full information

Explaining Reference-Dependence

- Kahneman-Tversky (1979), Koszegi-Rabin (2006) seek to incorporate ref-dependence into economic theory through non-standard preferences, maximized under full information
- Alternative proposal here: standard preferences, but choice based on imperfect perception of available options

Explaining Reference-Dependence

- Kahneman-Tversky (1979), Koszegi-Rabin (2006) seek to incorporate ref-dependence into economic theory through non-standard preferences, maximized under full information
- Alternative proposal here: standard preferences, but choice based on imperfect perception of available options
- Idea: reference-dependence a feature of the way objective characteristics are mapped into subjective representations
  - this can be optimal, due to constraints on information-processing capacity

Explaining Reference-Dependence

- Strategy: motivate a particular model of perceptual coding — and the account of information constraints under which it can be judged to be constrained-optimal — by reference to documented imperfections in visual perception

Explaining Reference-Dependence

- Strategy: motivate a particular model of perceptual coding — and the account of information constraints under which it can be judged to be constrained-optimal — by reference to documented imperfections in visual perception

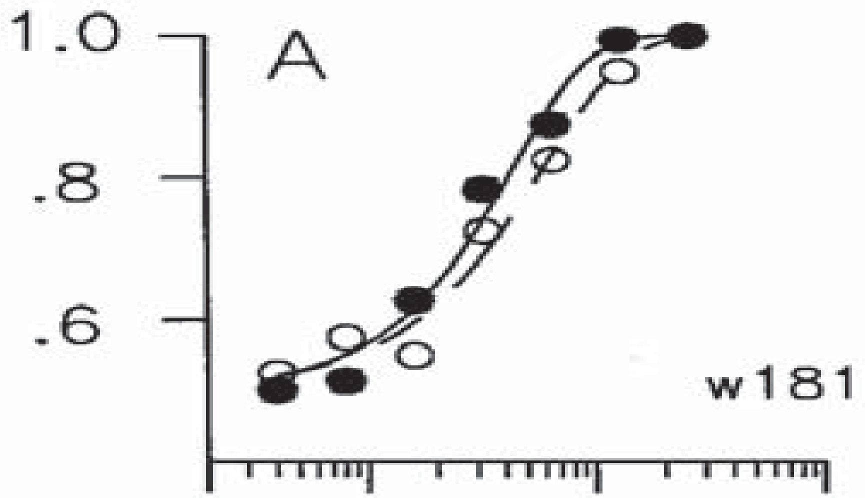
  - similar patterns observed in many different sensory domains

    — suggesting common processing constraints may be responsible for usefulness of similar computational strategies in multiple contexts

Imperfect Sensory Perception

- Long experimental literature in psychophysics shows that discrimination between stimuli is both imprecise and probabilistic:
    - probability of correct discrimination of relative brightness, direction of motion, etc., increasing function of objective difference ("psychometric function")

Motion Discrimination: Britten *et al.* (1992)



“Psychometric” and “neurometric” functions compared.

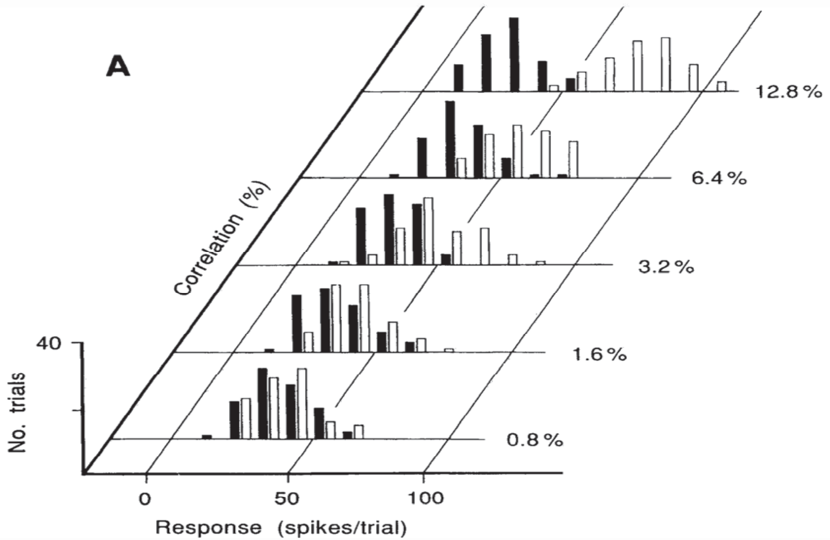
Imperfect Sensory Perception

- Long experimental literature in psychophysics shows that discrimination between stimuli is both imprecise and probabilistic:
    - probability of correct discrimination of relative brightness, direction of motion, etc., increasing function of objective difference ("psychometric function")
- Interpretation: relation between objective characteristics of stimulus and subjective representation is stochastic

Imperfect Sensory Perception

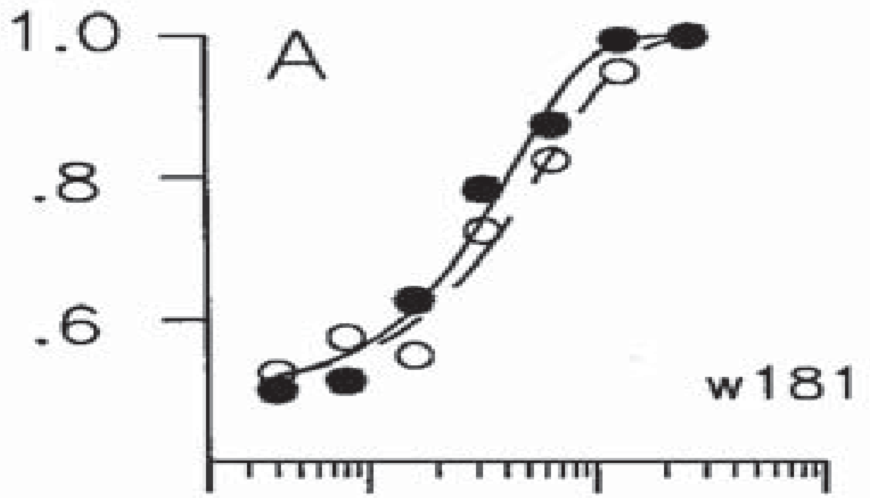
- Long experimental literature in psychophysics shows that discrimination between stimuli is both imprecise and probabilistic:
  - probability of correct discrimination of relative brightness, direction of motion, etc., increasing function of objective difference ("psychometric function")
- Interpretation: relation between objective characteristics of stimulus and subjective representation is stochastic
  - in animals, the stochastic relationship between stimulus and neural coding can be directly observed
  - e.g., Britten *et al.* (1992) study of ability of rhesus monkeys to discriminate the direction of motion of moving dots

Motion Discrimination: Britten *et al.* (1992)



Histograms of number of spikes in MT neuron.

Motion Discrimination: Britten *et al.* (1992)



"Psychometric" and "neurometric" functions compared.

Sensory Adaptation

Two features of adaptation observed in multiple contexts (Clifford *et al.*, 2007):

1. adjustment of the range of stimuli over which finer discriminations are possible

Sensory Adaptation

Two features of adaptation observed in multiple contexts (Clifford *et al.*, 2007):

1. adjustment of the range of stimuli over which finer discriminations are possible

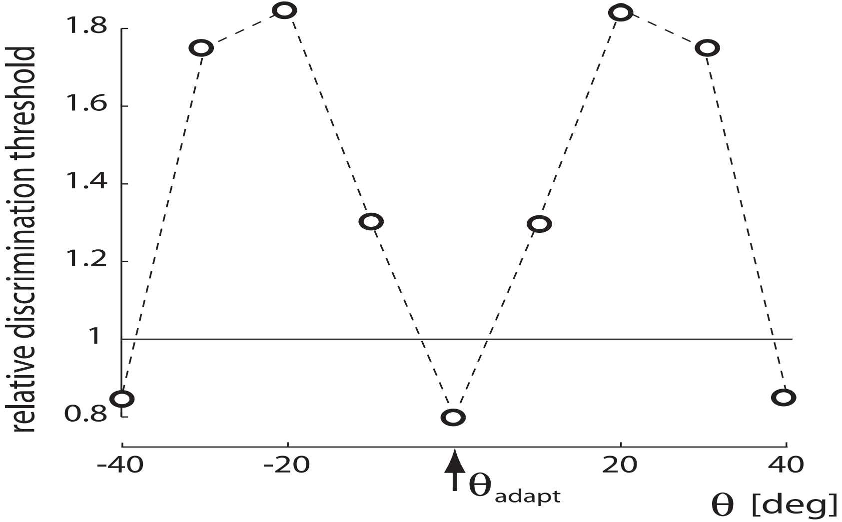
    - common experience: adjustment of vision to dark room, or bright sunlight

Sensory Adaptation

Two features of adaptation observed in multiple contexts (Clifford *et al.*, 2007):

1. adjustment of the range of stimuli over which finer discriminations are possible
   - common experience: adjustment of vision to dark room, or bright sunlight
   - experiment: prolonged exposure to a particular stimulus leads to increased discriminability of stimuli similar to the adaptor, but reduced discriminability of stimuli farther from it

Post-Adaptation Discrimination Thresholds



Discrimination thresholds for direction of motion
[Data: Phinney *et al.*, 1997]

Sensory Adaptation

Two features of adaptation observed in multiple contexts (Clifford *et al.*, 2007):

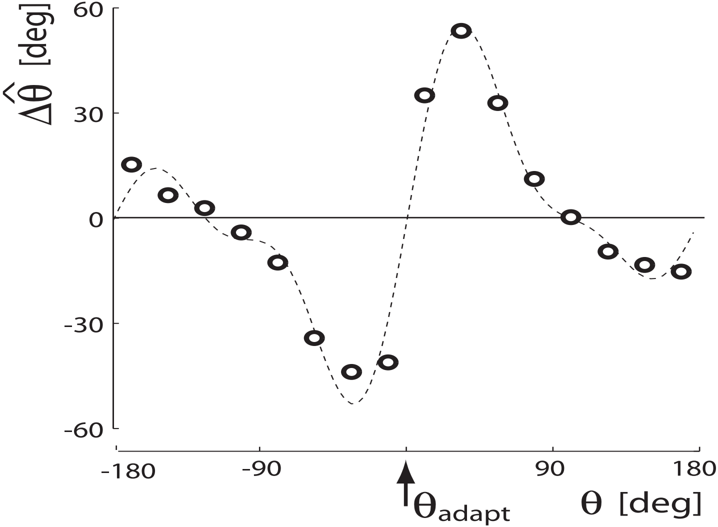
1. adjustment of the range of stimuli over which finer discriminations are possible
2. repulsion effects: stimuli different from what visual system has come to "expect" are perceived as even more different than they really are

Sensory Adaptation

Two features of adaptation observed in multiple contexts (Clifford *et al.*, 2007):

1. adjustment of the range of stimuli over which finer discriminations are possible
2. repulsion effects: stimuli different from what visual system has come to "expect" are perceived as even more different than they really are
   - after-effects from prolonged exposure to an intense stimulus: brightness, color, tilt of lines, direction of motion, etc.

Repulsion Effect: Direction of Motion



Bias in post-adaptation perceived direction of motion
[Data: Schrater and Simoncelli, 1998]

Sensory Adaptation

Two features of adaptation observed in multiple contexts (Clifford *et al.*, 2007):

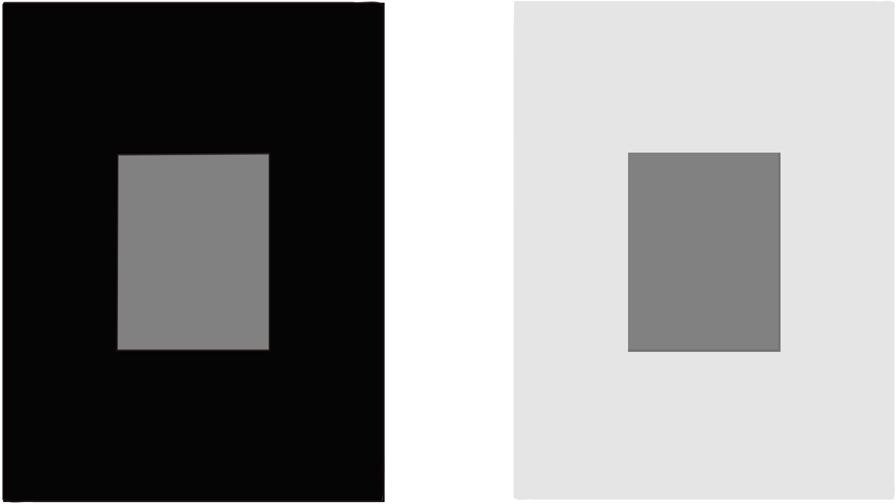
1. adjustment of the range of stimuli over which finer discriminations are possible

2. repulsion effects: stimuli different from what visual system has come to "expect" are perceived as even more different than they really are

    - after-effects from prolonged exposure to an intense stimulus: brightness, color, tilt of lines, direction of motion, etc.

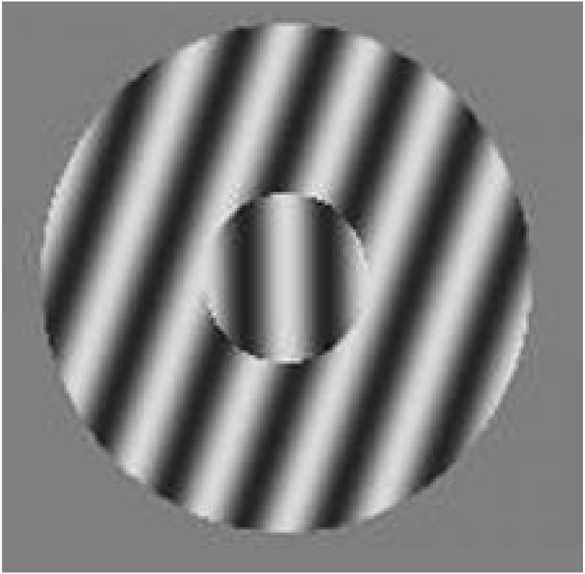
    - similar repulsion effects from other elements of visual field: many common visual illusions

Perceived Brightness Depends on Contrast



The central squares reflect equal amounts of light.

Perceived Orientation Depends on Contrast



The central bars are actually vertical.

Explaining Sensory Adaptation

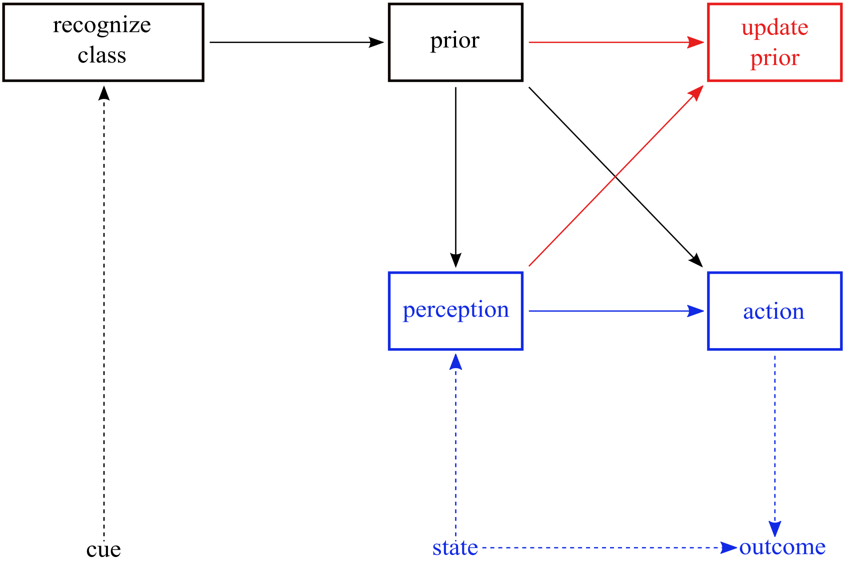
- Proposal: perceptual coding adjusts to the (apparent) frequency distribution of stimuli in a given environment (class of situations)

Explaining Sensory Adaptation

- Proposal: perceptual coding adjusts to the (apparent) frequency distribution of stimuli in a given environment (class of situations)

- It does this so as to make efficient use of limited information-processing capacity of the channel over which observations of the world must be transmitted to the nervous system

Levels of Processing



Efficient Perception

- Concern here: how a subjective representation r of the situation is produced, as function of the actual state s, and the prior [distribution $f(s)$] for this class of situations

Efficient Perception

- Concern here: how a subjective representation r of the situation is produced, as function of the actual state s, and the prior [distribution $f(s)$] for this class of situations
  - note: the mapping from s to r will be stochastic

Efficient Perception

- Concern here: how a subjective representation r of the situation is produced, as function of the actual state s, and the prior [distribution $f(s)$] for this class of situations
  - note: the mapping from s to r will be stochastic

- Hypothesis: the perceptual mapping is efficient, given
  - the decision problem [defined by $U(s, a)$, a to be a function of r]
  - the prior $f(s)$
  - capacity of the channel for transmission of information about s used to produce r

Processing Constraints

- The bottleneck: a communication channel with possible inputs X, possible outputs R, produces output y with conditional probability $p(r|x)$

Processing Constraints

- The bottleneck: a communication channel with possible inputs X, possible outputs R, produces output y with conditional probability $p(r|x)$
- Given these properties, the (Shannon) capacity of the channel is defined as

$$C(p) \equiv \max_{\pi} I_{\pi}(X, R)$$

where for any frequency distribution $\pi(x)$ over inputs x, I_{π} is the mutual information between random variables X and R when joint distribution is $p(x, r) = \pi(x)p(r|x)$

Efficient Perception

- Complete perception/action circuit:
  - encoding: $x = g(s)$
  - transmission [perception]: r produced with prob $p(r|x)$
  - decoding [action]: $a = h(r)$

Efficient Perception

- Complete perception/action circuit:
  - encoding: $x = g(s)$
  - transmission [perception]: r produced with prob $p(r|x)$
  - decoding [action]: $a = h(r)$

- Relevant measure of accuracy of perception: given joint distribution for (s, r) implied by prior f, encoding g, and channel p,

$$W(g, p; f) \equiv \mathrm{E}_{s,r}\left[\max_{h(r)} \mathrm{E}_{s'}[U(h(r), s')|r]\right]$$

Efficient Perception

Two (nested) efficiency hypotheses:

1. Efficient Coding Hypothesis: the encoding function g is optimized for the prior f and channel p

$$\hat{W}(p; f) \equiv \max_{g} W(g, p; f)$$

   - implies adaptation of g (hence joint dist'n $p(s, r)$) to changing environment f

Efficient Perception

Two (nested) efficiency hypotheses:

1. Efficient Coding Hypothesis: the encoding function g is optimized for the prior f and channel p

$$\hat{W}(p; f) \equiv \max_g W(g, p; f)$$

   - implies adaptation of g (hence joint dist'n $p(s, r)$) to changing environment f

2. Efficient Channel Hypothesis: the channel p is also optimized, for some prior over states, subject to a bound on possible channel capacity

$$\max_p \hat{W}(p; f) \quad \text{s.t.} \quad C(p) \leq \bar{C}$$

Efficient Perception

- Adaptation of channel may occur more slowly than adaptation of coding

Efficient Perception

- Adaptation of channel may occur more slowly than adaptation of coding
  - hence p may be optimal for typical past environment, but not current f

Efficient Perception

- Adaptation of channel may occur more slowly than adaptation of coding

  - hence p may be optimal for typical past environment, but not current f

  - or p may be optimized for a class of possible environments $\{f_\theta\}$, given prior over parameters θ

$$\max_p \mathrm{E}_\theta \hat{W}(p; f_\theta)$$

Perceived Direction of Motion

Illustrative application:

- State θ = direction of motion of stimulus (angle on the circle)
- Action: estimated direction $\hat{\theta}$ (also an angle on the circle)
- Objective: maximization of expected accuracy

$$\mathrm{E}[\cos(\hat{\theta} - \theta)]$$

  - approximately minimization of MSE, when accuracy is high: but a periodic function

Perceived Direction of Motion

- Channel: assumed to be optimized for a uniform prior

$$f(\theta) = 1/2\pi \quad \forall\theta$$

Perceived Direction of Motion

- Channel: assumed to be optimized for a uniform prior

$$f(\theta) = 1/2\pi \quad \forall\theta$$

- Optimal channel:
  - input space = angles $\hat{\phi}$ on the circle
  - optimal encoding [wlog]: $\hat{\phi}(\theta) = \theta$
  - output space = angles ϕ on the circle
  - distribution of outputs [subjective perceptions of direction] given input:

    $$p(\phi|\hat{\phi}) = A\, e^{\lambda^{-1}\cos(\phi-\hat{\phi})} \quad \forall\phi$$

    where $\lambda > 0$ varies inversely with capacity $\bar{C}$
  - optimal decoding: $\hat{\theta}(\phi) = \phi$

Perceived Direction of Motion

- Adaptation: suppose subject exposed repeatedly to a single direction of motion [wlog: $\theta_{adapt} = 0$]

Perceived Direction of Motion

- Adaptation: suppose subject exposed repeatedly to a single direction of motion [wlog: $\theta_{adapt} = 0$]

- Hypothesis: no change in channel, but encoding $\hat{\phi}(\theta)$ adjusts, to be optimal for a new perceived environment, described by a new prior on the circle, with mode at $\theta = 0$:

$$f(\theta) = B\ e^{\beta \cos \theta}$$

where $\beta > 0$ indicates degree of habituation to adaptor [$\beta = 0$ means unchanged prior]

Perceived Direction of Motion

- Adaptation: suppose subject exposed repeatedly to a single direction of motion [wlog: $\theta_{adapt} = 0$]

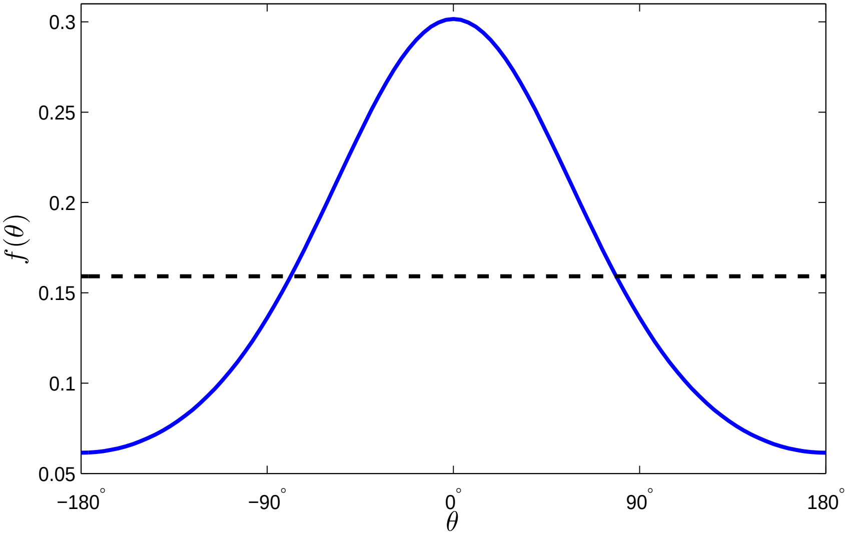
- Hypothesis: no change in channel, but encoding $\hat{\phi}(\theta)$ adjusts, to be optimal for a new perceived environment, described by a new prior on the circle, with mode at $\theta = 0$:

$$f(\theta) = B\ e^{\beta \cos \theta}$$

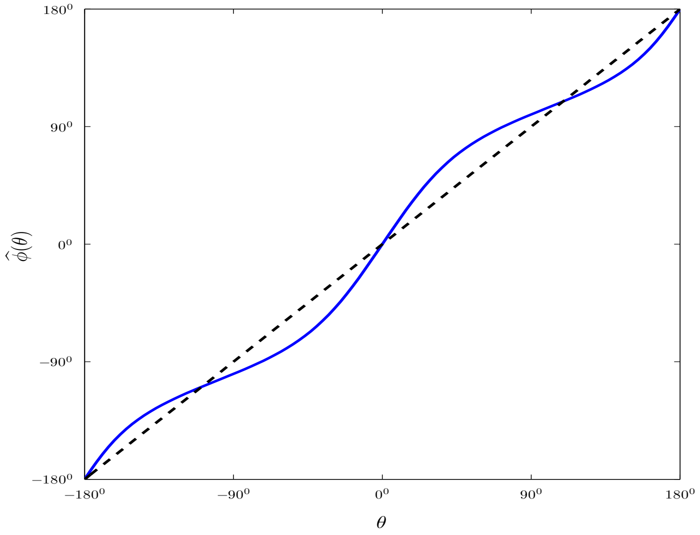
  where $\beta > 0$ indicates degree of habituation to adaptor [$\beta = 0$ means unchanged prior]

- Numerical example: $\lambda = 1, \beta = 0.8$

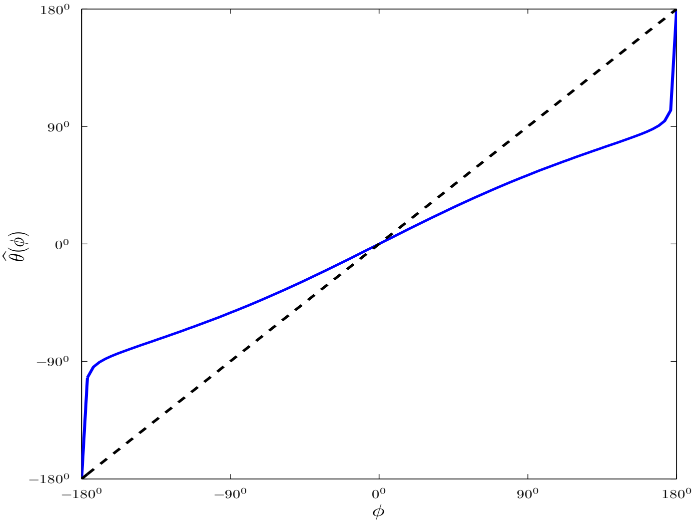
Post-Adaptation Prior



Post-Adaptation Optimal Coding



Post-Adaptation Optimal Decoding



Perceived Direction of Motion

Predictions for experimental data:

1. Effect of adaptation on perceived direction of motion:
   - define the average perceived direction $\bar{\phi}$ (given distribution of representations ϕ) as the angle that max's $\mathrm{E}_\phi[\cos(\phi - \bar{\phi})]$
   - then measure bias by

$$\text{bias}(\theta) \equiv \bar{\phi}(\theta) - \theta$$

Perceived Direction of Motion

Predictions for experimental data:

1. Effect of adaptation on perceived direction of motion:
   - define the average perceived direction $\bar{\phi}$ (given distribution of representations ϕ) as the angle that max's $\mathrm{E}_\phi[\cos(\phi - \bar{\phi})]$
   - then measure bias by

$$\text{bias}(\theta) \equiv \bar{\phi}(\theta) - \theta$$

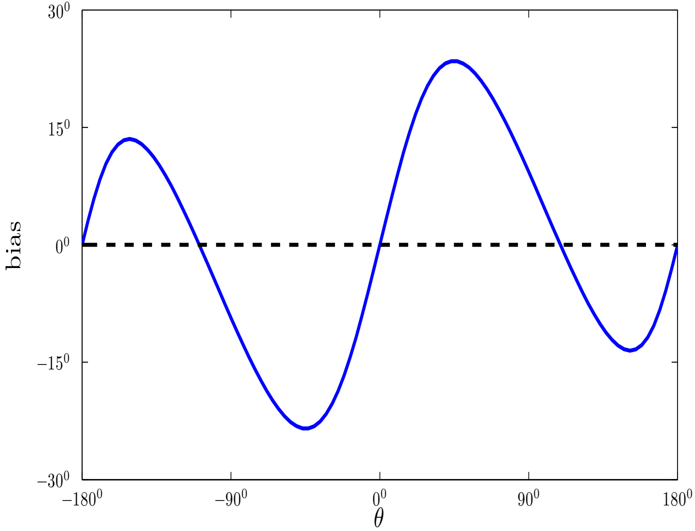
2. Effect on sensitivity to differences in direction of motion:
   - measure the variability of perceived direction by

$$\sigma_\phi^2 \equiv 2\{1 - \mathrm{E}[\cos(\phi - \bar{\phi})]\}$$

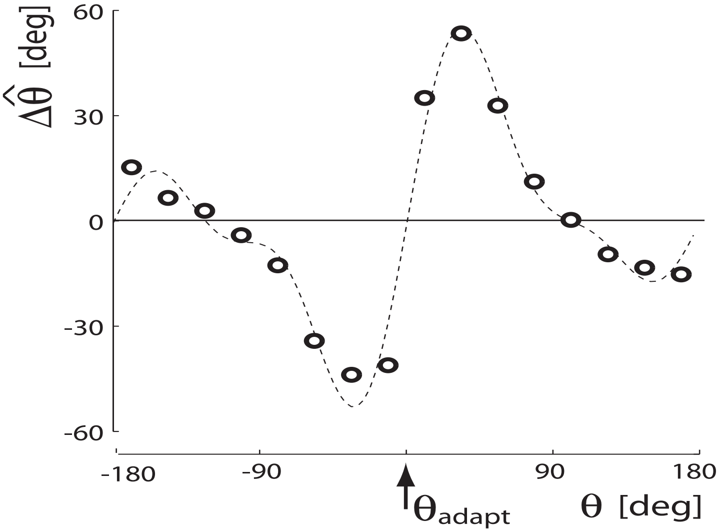
   - then define the discrimination threshold function by

$$\tau(\theta) \equiv \frac{\sigma_\phi(\theta)}{\bar{\phi}'(\theta)}$$

Prediction: Repulsion Effect

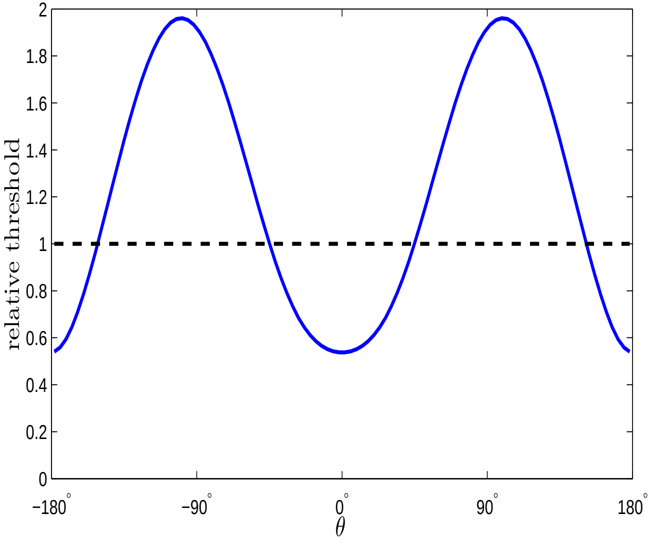


Evidence: Repulsion Effect

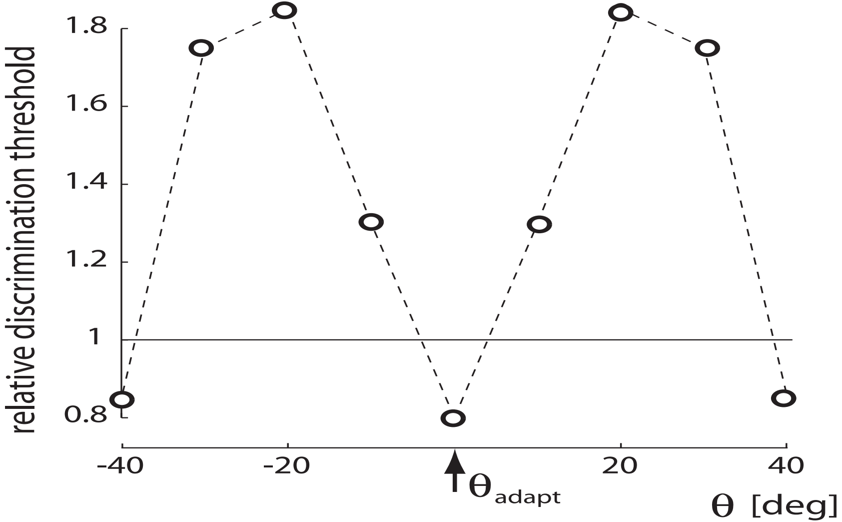


Bias in post-adaptation perceived direction of motion
[Data: Schrater and Simoncelli, 1998]

Prediction: Sensitivity Effect



Evidence: Sensitivity Effect



Discrimination thresholds for direction of motion
[Data: Phinney *et al.*, 1997]

Perceptions of Value

- Here hypothesize that imperfect perceptions of value in economic decisions are also optimal, subject to a similar info-processing constraint

Perceptions of Value

- Here hypothesize that imperfect perceptions of value in economic decisions are also optimal, subject to a similar info-processing constraint

- Model: suppose that true utility value of each option is a sum

$$u = \sum_a s_a$$

of the values of each of a number of attributes, each of which must be perceived separately

Perceptions of Value

- Assumed form of channel:
    - a subjective perception r_a for each attribute
    - parallel processing: conditional probabilities $p_a(r_a|x_a)$ for each attribute are independent of other attributes, where x_a is a function only of s_a

Perceptions of Value

- Assumed form of channel:
    - a subjective perception r_a for each attribute
    - parallel processing: conditional probabilities $p_a(r_a|x_a)$ for each attribute are independent of other attributes, where x_a is a function only of s_a
- Assumed objective: minimize MSE of subjective estimate of u

Perceptions of Value

- Assumed form of channel:
  - a subjective perception r_a for each attribute
  - parallel processing: conditional probabilities $p_a(r_a|x_a)$ for each attribute are independent of other attributes, where x_a is a function only of s_a
- Assumed objective: minimize MSE of subjective estimate of u
- Assume priors are independent for each attribute

$$f(\vec{s}) = \prod f_a(s_a)$$

Optimal Perception of Value

- Optimal decoding of each attribute depends only on prior, coding, and channel for that attribute:

$$\hat{s}_a = E[s_a | r_a]$$

Optimal Perception of Value

- Optimal decoding of each attribute depends only on prior, coding, and channel for that attribute:

$$\hat{s}_a = E[s_a | r_a]$$

- Optimal coding of attribute a depends only on f_a, p_a

Optimal Perception of Value

- Optimal decoding of each attribute depends only on prior, coding, and channel for that attribute:

$$\hat{s}_a = E[s_a | r_a]$$

- Optimal coding of attribute a depends only on f_a, p_a

- Optimal channel for attribute a depends only on f_a and shadow value λ of additional capacity (same for all attributes)

Implications of Optimal Coding

- Suppose different possible classes θ of choice situations are associated with different priors f^θ over possible values of the attributes; but suppose that prior for attribute a is of the same form

$$f_a^\theta(s_a) = \frac{1}{\sigma_a}\,\hat{f}_a\left(\frac{s_a - \mu_a^\theta}{\sigma_a^\theta}\right)$$

for all θ

— example: prior always belongs to Gaussian family

Implications of Optimal Coding

- Suppose different possible classes θ of choice situations are associated with different priors f^θ over possible values of the attributes; but suppose that prior for attribute a is of the same form

  $$f_a^\theta(s_a) = \frac{1}{\sigma_a}\,\hat{f}_a\left(\frac{s_a - \mu_a^\theta}{\sigma_a^\theta}\right)$$

  for all θ

  — example: prior always belongs to Gaussian family

- Then given channel p_a, optimal coding will be of same form

  $$g_a^\theta(s_a) = \hat{g}_a\left(\frac{s_a - \mu_a^\theta}{\sigma_a^\theta}\right)$$

  for each class θ

Implications of Optimal Coding

$$g_a^\theta(s_a) = \hat{g}_a\left(\frac{s_a - \mu_a^\theta}{\sigma_a^\theta}\right)$$

- Hence coding of value will be reference-dependent: what is coded is not absolute value s_a, but s_a relative to the prior mean

Implications of Optimal Coding

$$g_a^\theta(s_a) = \hat{g}_a\left(\frac{s_a - \mu_a^\theta}{\sigma_a^\theta}\right)$$

- Hence coding of value will be reference-dependent: what is coded is not absolute value s_a, but s_a relative to the prior mean
    - the "reference point" is thus determined by expectations, as argued by Koszegi and Rabin (2006)
    - but in general, no significance of any single "reference point": coding is relative to the prior distribution

Character of Optimal Channel

- Outputs will generally be stochastic function of inputs
  - hence subjective representations will be stochastic function of true state
  - not deterministic quantization as proposed by Saint-Paul (2011), Gul *et al.* (2011)

Character of Optimal Channel

- Outputs will generally be stochastic function of inputs
  - hence subjective representations will be stochastic function of true state
  - not deterministic quantization as proposed by Saint-Paul (2011), Gul *et al.* (2011)
  - not because it is assumed that feasible channels are inevitably stochastic: it is actually a more efficient use of bounded capacity

Character of Optimal Channel

- Outputs will generally be stochastic function of inputs
  - hence subjective representations will be stochastic function of true state
  - not deterministic quantization as proposed by Saint-Paul (2011), Gul *et al.* (2011)
  - not because it is assumed that feasible channels are inevitably stochastic: it is actually a more efficient use of bounded capacity

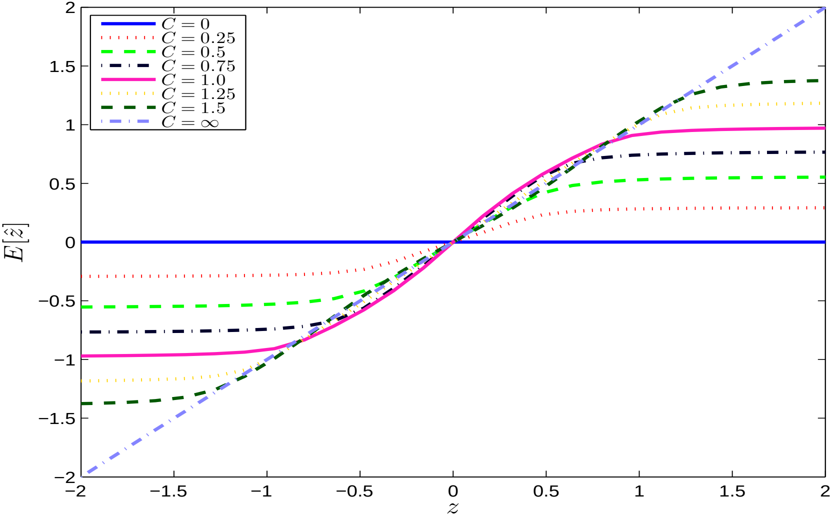
- Generally only a finite number of output states
  - hence little difference in the way that extreme states (relative to the prior) are coded

Character of Optimal Channel

- Outputs will generally be stochastic function of inputs
  - hence subjective representations will be stochastic function of true state
  - not deterministic quantization as proposed by Saint-Paul (2011), Gul *et al.* (2011)
  - not because it is assumed that feasible channels are inevitably stochastic: it is actually a more efficient use of bounded capacity

- Generally only a finite number of output states
  - hence little difference in the way that extreme states (relative to the prior) are coded
  - how far into the tails of prior states continue to be distinguished depends on capacity allocated to perception of the attribute in question

Mean Subjective Value vs. True Value



sensitivity increases with processing capacity C_a

Kahneman-Tversky (1979) Example

- Problem for standard EU theory: in both cases, subjects are choosing between the same probability distributions over final wealth levels:

  (a) initial wealth + 1500 with certainty

  VS

  (b) 50 percent chance of initial wealth + 1000,
  50 percent chance of initial wealth + 2000

Kahneman-Tversky (1979) Example

- Problem for standard EU theory: in both cases, subjects are choosing between the same probability distributions over final wealth levels:

  (a) initial wealth + 1500 with certainty

  VS

  (b) 50 percent chance of initial wealth + 1000,
  50 percent chance of initial wealth + 2000

- Explanation proposed by "prospect theory": outcomes evaluated relative to a different reference point in the two cases (wealth prior to the choice)

Applying the Theory

- Choice between two lotteries: the relevant attributes of each are its payoffs in the two equi-probable states

  (a) (1500, 1500)
  (b) (1000, 2000)

Applying the Theory

- Choice between two lotteries: the relevant attributes of each are its payoffs in the two equi-probable states

  (a) (1500, 1500)
  (b) (1000, 2000)

- Subjective perception of each of the two attributes depends on distribution of expected possible values for that attribute

Applying the Theory

- Suppose the class θ for which perceptual coding is optimized is "choice between two lotteries, starting from initial wealth θ"
  - prior f^{θ} is evoked by receipt of initial payment θ (the "cue"), before presentation of lotteries

Applying the Theory

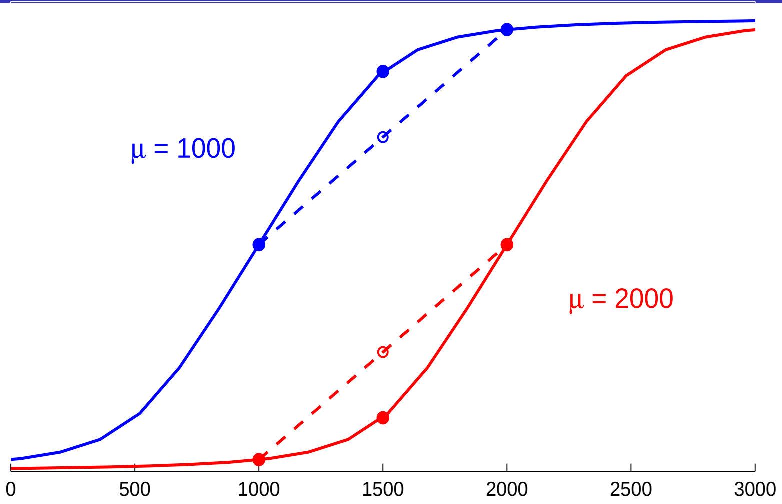
- Suppose the class θ for which perceptual coding is optimized is "choice between two lotteries, starting from initial wealth θ"
  - prior f^{θ} is evoked by receipt of initial payment θ (the "cue"), before presentation of lotteries

- For concreteness, suppose the prior f^{θ} is $\mathcal{N}(\theta, 1000^2)$
  - distribution of possible lottery payoffs in each state is expected to be independent of initial wealth

Applying the Theory

- Suppose the class θ for which perceptual coding is optimized is "choice between two lotteries, starting from initial wealth θ"
  - prior f^θ is evoked by receipt of initial payment θ (the "cue"), before presentation of lotteries
- For concreteness, suppose the prior f^θ is $\mathcal{N}(\theta, 1000^2)$
  - distribution of possible lottery payoffs in each state is expected to be independent of initial wealth
- Efficiency hypothesis:
  - channel is optimized for the entire class of possible priors $\{f^\theta\}$ (prior over θ doesn't matter!)
  - encoding optimally adapts to the particular prior f^θ after initial wealth θ is learned

Mean Subjective Valuations of Options



(a) has higher MNSV when $\mu^\theta = 1000$, but (b) higher when $\mu^\theta = 2000$

Comparison with Prospect Theory

Explanation by a sigmoid-shaped "value function" formally the same, though here the nonlinear transformation interpreted as reflecting perception rather then preferences. Also:

Comparison with Prospect Theory

Explanation by a sigmoid-shaped "value function" formally the same, though here the nonlinear transformation interpreted as reflecting perception rather then preferences. Also:

- Deeper explanation for shape of the value function proposed

Comparison with Prospect Theory

Explanation by a sigmoid-shaped "value function" formally the same, though here the nonlinear transformation interpreted as reflecting perception rather then preferences. Also:

- Deeper explanation for shape of the value function proposed
- Provides a theory of location of "reference point"

  — and not always status quo

Comparison with Prospect Theory

Explanation by a sigmoid-shaped "value function" formally the same, though here the nonlinear transformation interpreted as reflecting perception rather then preferences. Also:

- Deeper explanation for shape of the value function proposed
- Provides a theory of location of "reference point"

  — and not always status quo
- Predicts that both reference-dependence, and departures from risk-neutrality for small gambles, will be greater in contexts where stakes are small enough for low-capacity channel to be optimal

Comparison with Sims's "Rational Inattention"

- RI assumes joint optimization of coding and channel (hence no need to discuss separate encoding stage)

Comparison with Sims's "Rational Inattention"

- RI assumes joint optimization of coding and channel (hence no need to discuss separate encoding stage)
- RI assumes joint coding of all aspects of choice situation

Comparison with Sims's "Rational Inattention"

- RI assumes joint optimization of coding and channel (hence no need to discuss separate encoding stage)
- RI assumes joint coding of all aspects of choice situation
  - in above case, would be optimal to code only mean outcome over states $\Rightarrow$ (a) and (b) equivalent

Comparison with Sims's "Rational Inattention"

- RI assumes joint optimization of coding and channel (hence no need to discuss separate encoding stage)

- RI assumes joint coding of all aspects of choice situation
  - in above case, would be optimal to code only mean outcome over states $\Rightarrow$ (a) and (b) equivalent

- RI assumes upper bound on mutual information rather than channel capacity

Comparison with Sims's "Rational Inattention"

- RI assumes joint optimization of coding and channel (hence no need to discuss separate encoding stage)

- RI assumes joint coding of all aspects of choice situation
  - in above case, would be optimal to code only mean outcome over states $\Rightarrow$ (a) and (b) equivalent

- RI assumes upper bound on mutual information rather than channel capacity
  - in above case, optimal coding would make $\mathrm{E}[\hat{z}|z]$ a linear function $\Rightarrow$ (a) and (b) have same MNSV

Further Implications of the Theory

- The efficient coding hypothesis simultaneously offers candidate explanation for other behavioral anomalies:
  - stochasticity of choice
    — implication of optimal channel, even if not optimized for same prior as the encoding

Further Implications of the Theory

- The efficient coding hypothesis simultaneously offers candidate explanation for other behavioral anomalies:
  - stochasticity of choice
    — implication of optimal channel, even if not optimized for same prior as the encoding
  - "focusing effects"
    — optimal channel may transmit no information at all about some utility-relevant attributes (corner solution)

Further Implications of the Theory

- The efficient coding hypothesis simultaneously offers candidate explanation for other behavioral anomalies:
  - stochasticity of choice
    — implication of optimal channel, even if not optimized for same prior as the encoding
  - "focusing effects"
    — optimal channel may transmit no information at all about some utility-relevant attributes (corner solution)
  - "menu effects"
    — if prior implies correlation of attributes of choices in a given choice set

Further Implications of the Theory

- A variety of modifications of standard choice theory have been proposed to allow for each of these anomalies:
  - random utility models (McFadden, 1974)
  - prospect theory (Kahneman-Tversky, 1979)
  - reference-dependent preferences (Koszegi-Rabin, 2006)
  - preference for sparse representations (Gabaix, 2010)
  - range-dependent weights on attributes (Koszegi-Szeidl, 2011)
  - "local thinking" (Bordalo *et al.*, 2010, 2011)
  - preferences depend on "comparison set" (Cunningham, 2011)

Further Implications of the Theory

- A variety of modifications of standard choice theory have been proposed to allow for each of these anomalies:
    - random utility models (McFadden, 1974)
    - prospect theory (Kahneman-Tversky, 1979)
    - reference-dependent preferences (Koszegi-Rabin, 2006)
    - preference for sparse representations (Gabaix, 2010)
    - range-dependent weights on attributes (Koszegi-Szeidl, 2011)
    - "local thinking" (Bordalo *et al.*, 2010, 2011)
    - preferences depend on "comparison set" (Cunningham, 2011)

- But the present proposal offers a unified explanation of all of these phenomena

  — in a way that is also consistent with observations about perception in other contexts