Preference reversal or limited sampling? Maybe túngara frogs are rational after all.

Paulo Natenzon
Washington University in Saint Louis

Representation of Value and Economic Choice Workshop
Columbia University — May 6th, 2016

Motivation

Datasets with context-dependent choice behavior

- B chosen more than 50% of trials from $\{A, B\}$;
- A chosen more than 50% of trials from $\{A, B, C\}$

Choice reversal seems incompatible with maximizing utility:
$A \succ B$? $B \succ A$?

Adding an error term to utility does not help:

$$U_A = u_A + \varepsilon_A$$
$$U_B = u_B + \varepsilon_B$$
$$U_C = u_C + \varepsilon_C$$

$$\rho(A, ABC) = \mathbb{P}\{U_A > U_B \text{ and } U_A > U_C\} \leq \mathbb{P}\{U_A > U_B\} = \rho(A, AB)$$

A common interpretation of such data: irrational behavior

RESEARCH | *REPORTS*

SEXUAL SELECTION

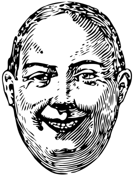
Irrationality in mate choice revealed by túngara frogs

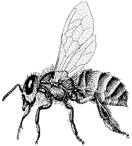
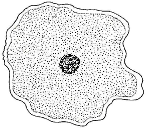
Amanda M. Lea[1*] and Michael J. Ryan[1,2]

Mate choice models derive from traditional microeconomic decision theory and assume that individuals maximize their Darwinian fitness by making economically rational decisions. Rational choices exhibit regularity, whereby the relative strength of preferences between options remains stable when additional options are presented. We tested female frogs with three simulated males who differed in relative call attractiveness and call rate. In binary choice tests, females' preferences favored stimulus caller B over caller A; however, with the addition of an inferior "decoy" C, females reversed their preferences and chose A over B. These results show that the relative valuation of mates is not independent of inferior alternatives in the choice set and therefore cannot be explained with the rational choice models currently used in sexual selection theory.

Evolutionary theory is tightly linked to traditional decision theory, which predicts consumer behavior by assuming individuals' decisions will lead to outcomes that max- rational mate choices. The preference function concept, central to sexual selection theory (*9*), assumes that mate choice rules obey formal rationality (*2*, *7*). The results of the scant empirical

Systematic context-dependent choice: evidence



Decoy effects found in people, but also:
Rhesus macaques (Parrish, Evans, Beran 2015)
Gray jays (Shafir, Waite, Smith 2002)
Honeybees (Shafir, Waite, Smith 2002)
Slime mold (Latty, Beekman 2010)

Testing choice theories using data

deterministic theory + noise model = random choice model

The most common noise models are random utility models: Logit, Probit, Nested Logit, Mixed Logit, etc.

- Fluctuation in tastes
- Hand trembling mistakes
- Taste heterogeneity

Example: Hey and Orme (*Econometrica*, 1994)

$$\left.\begin{array}{r}\text{Expected Utility}\\ \text{Disappointment Aversion Utility}\\ \text{Rank-dependent Utility}\end{array}\right\} + \text{Gaussian}\,\varepsilon = \text{Probit}$$

Contribution: a different noise model

Context-dependent data is incompatible with random utility:

$$\text{deterministic theory} + \text{noise model} = \text{random choice model}$$

One possible approach is to relax the deterministic theory:
Axiom 1: $\succsim$ is complete and transitive
Axiom 2: $\cdots$

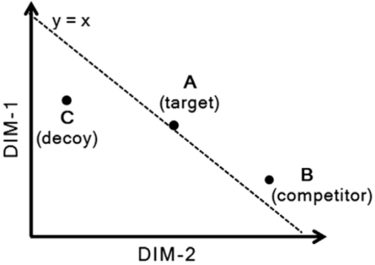
Instead, we introduce a more realistic noise model:

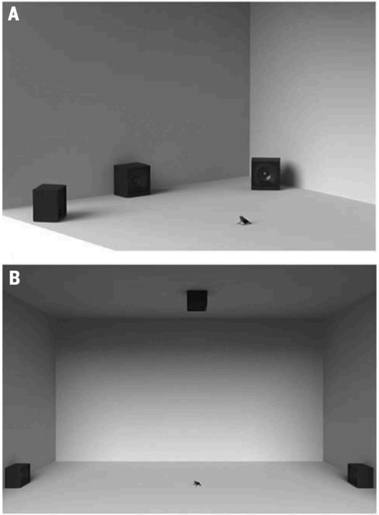
- $\Rightarrow$ better fit and out-of-sample prediction
- $\Rightarrow$ able to identify underlying preferences
- $\Rightarrow$ makes standard welfare analysis possible again

Maybe túngara frogs are rational after all...

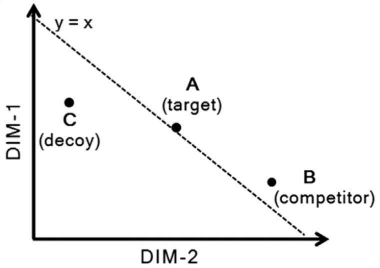
Choice data: frog mating selection

Physalaemus pustulosus





Choice data: frog mating selection



Menu	A	B	C
$\{A, B\}$	.37	.63	—
$\{B, C\}$	—	.69	.31
$\{A, C\}$	.84	—	.16
$\{A, B, C\}$	.55	.28	.17
$\{A, B, \not{C}\}$	.61	.39	—

Source: Lea and Ryan (2015), supplemental materials online.

Our new explanation for context-dependent choice

- Limited Sampling $\Rightarrow$ mistakes
- Two lessons from experimental psychophysics:
  - Mistakes and Values
  - Mistakes and Similarity
- Bayesian updating: favors options that are easier to compare
- Parametric Model: Bayesian probit
  - t Limited Sampling
  - μ Preference
  - σ Similarity
- Better data fit and better out-of-sample prediction
- Makes welfare analysis possible again

Limited Sampling: frog mating choice

Experimental data from Lea and Ryan (2015)

Female túngara frogs choose a mate based on its call.

Where does the noise in the data come from?

Decision maker obtains imperfect information about the value of the alternatives before making a choice:

- Choices made in dynamic social environments
- Potential mates have complex traits
- Limited cognitive resources
- Limited perceptual systems
- Time is costly: predator risk, lost mating opportunities

Two lessons from cognitive choice tasks

Examples of cognitive choice tasks:

- Which triangle is larger?
- Which star has more points?
- Which building is taller?
- Which object is heavier?
- Which sound is louder?

Special feature: analyst knows the utility function.

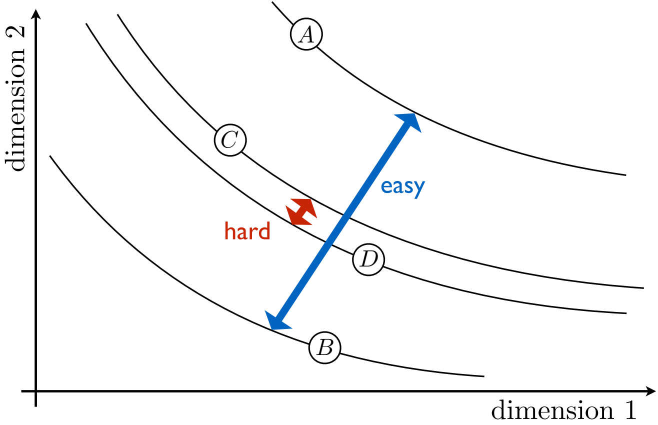
$\Longrightarrow$ Easy to identify and analyze **rate of mistakes**.

Two lessons about mistakes:
the effect of preference and the effect of similarity.

Lesson 1: The effect of preference

A is much better than B, while C is a little better than D.

Mistakes more likely in $\{C, D\}$ than in $\{A, B\}$.



Lesson 2: The effect of similarity

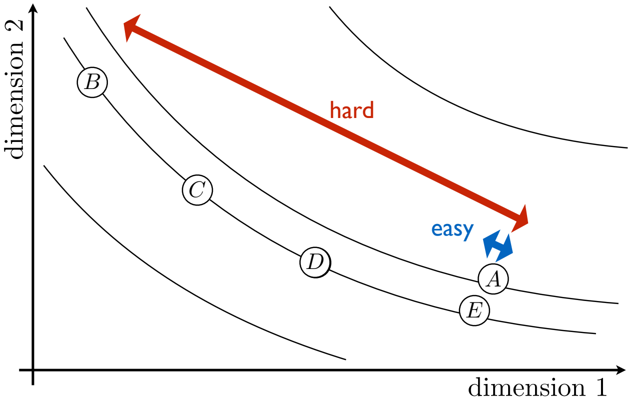
Tverksy and Russo (1969):

> *"it has been hypothesized that for a fixed difference between the psychological scale values, the more similar the stimuli, the easier the comparison or the discrimination between them."*

Lesson 2: The effect of similarity

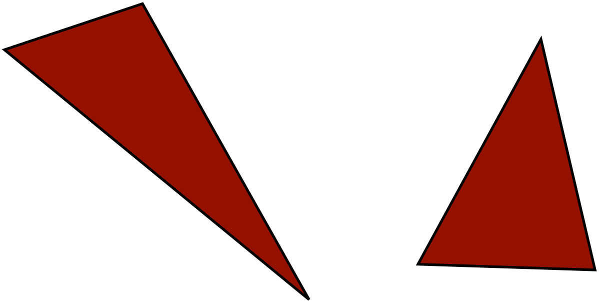
A is better than B, C, D, E.

Mistakes more likely in $\{A, B\}$ than in $\{A, E\}$.



Similarity example: triangle areas

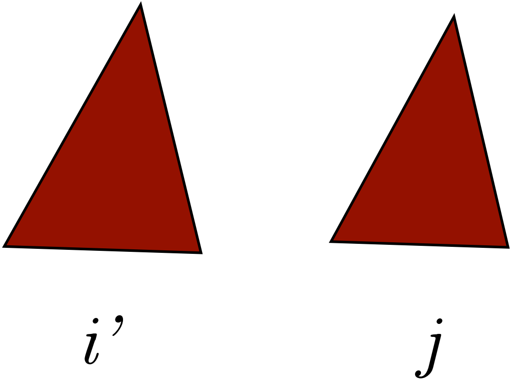
Which object in this pair has the largest area?



i j

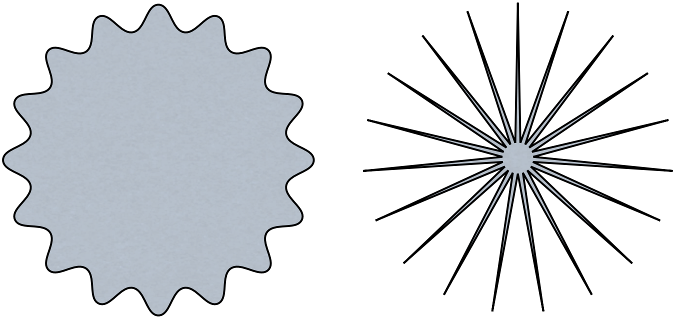
Similarity example: triangle areas

Which object in this pair has the largest area? And now?



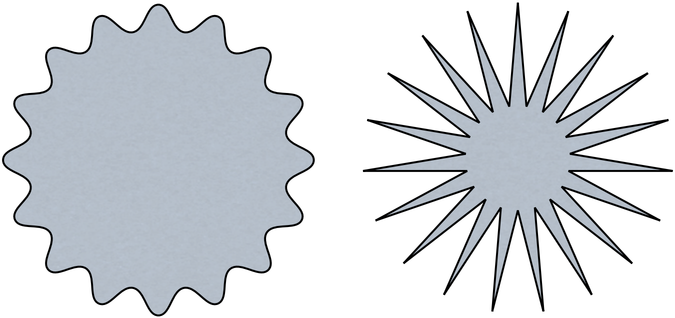
Similarity example: star points

Which star has more points?



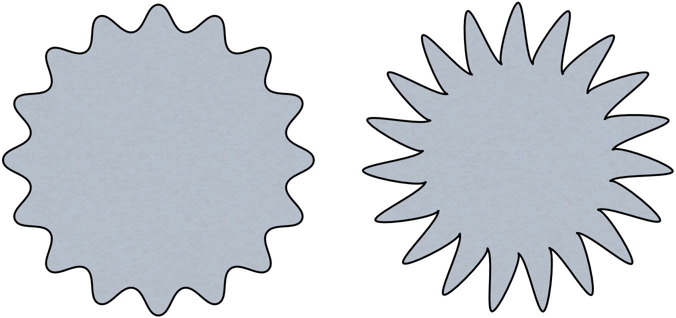
Similarity example: star points

Which star has more points?



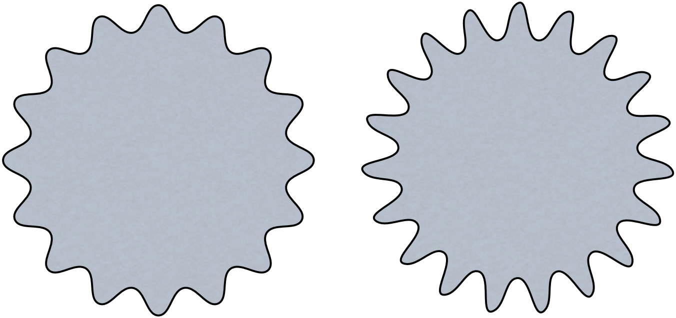
Similarity example: star points

Which star has more points?



Similarity example: star points

Which star has more points? And now?



Summary: lessons from experimental psychophysics

Comparing two options A and B:

Lesson 1: A, B easier to compare the greater the distance in value.
Mistakes more likely when $V_A - V_B$ is small.

Lesson 2: *Keeping values V_A and V_B fixed,*
A, B easier to compare when they are more similar.
Mistakes more likely when A and B are very different.

Easy to compare + Bayesian updating

Take a random draw of three frogs from the same population:

$$\begin{matrix} B & A & A & B & C & C \\ A & B & C & C & B & A \\ C & C & B & A & A & B \end{matrix}$$

Easy to compare + Bayesian updating

Take a random draw of three frogs from the same population:

B	A	A	B	C	C
A	B	C	C	B	A
C	C	B	A	A	B

If the only reliable comparison shows that

$(A \succ C)$

B	A	A
A	B	C
C	C	B

or

$(C \succ A)$

B	C	C
C	B	A
A	A	B

then alternative B starts at a disadvantage.

Parametric model: Bayesian Probit (Natenzon, 2010)

A, B, C choice alternatives

μ_A, μ_B, μ_C utility values

Prior μ_i distributed iid $\mathcal{N}(m, s)$

Signals $X_i = \mu_i + \varepsilon_i$ for each available $i = A, B, C$

$\varepsilon_A, \varepsilon_B, \varepsilon_C \sim \mathcal{N}(0, \Sigma)$ joint normal

- $\mathbb{E}[\varepsilon_i] = 0$
- $\sigma_{ij} \in [0, 1]$ correlation of $\varepsilon_i, \varepsilon_j$
- $\mathrm{Var}[\varepsilon_i] = 1/p > 0$ equal precision for all i

Choice Alternative j that maximizes X_j (Thurstone, 1927)

Parametric model: Bayesian Probit (Natenzon, 2010)

A, B, C choice alternatives

μ_A, μ_B, μ_C utility values

Prior μ_i distributed iid $\mathcal{N}(m, s)$

Signals $X_i = \mu_i + \varepsilon_i$ for each available $i = A, B, C$

$\varepsilon_A, \varepsilon_B, \varepsilon_C \sim \mathcal{N}(0, \Sigma)$ joint normal

- $\mathbb{E}[\varepsilon_i] = 0$
- $\sigma_{ij} \in [0, 1]$ correlation of $\varepsilon_i, \varepsilon_j$
- $\text{Var}[\varepsilon_i] = 1/p > 0$ equal precision for all i

Choice ~~Alternative j that maximizes X_j (Thurstone, 1927)~~

Alternative j that maximizes $\mathbb{E}[\mu_j | (X_i)_{i \in B}]$

Identification

Prior iid $\mathcal{N}(m, s)$

Signals joint $\mathcal{N}\left(\mu, \frac{1}{p}\Sigma\right)$

Proposition

The Bayesian probit with parameters $(\tilde{m}, \tilde{s}, \tilde{\mu}, \tilde{p}, \Sigma)$
is observationally equivalent to
the Bayesian probit with parameters $(0, 1, \mu, p, \Sigma)$, *where*

$$\mu_i = \frac{1}{\sqrt{\tilde{s}}}(\tilde{\mu}_i - \tilde{m}) \quad \textit{and} \quad p = \frac{\tilde{p}}{(1/s)}.$$

Choice probabilities:

$$\rho_p^{\mu\sigma}(j, B) = \mathbb{P}\left\{\mathbb{E}[\mu_j|X] \geq \mathbb{E}[\mu_k|X], \forall k \in B\right\}$$

where

$$\mathbb{E}[\mu|X] = [I + (1/p)\Sigma]^{-1} X$$

Behavioral content of parameters p, μ, σ

Proposition

$$\rho_p^{\mu\sigma}(i, \{i, j\}) = \Phi\left(\frac{\sqrt{p}}{\sqrt{2}} \frac{(\mu_i - \mu_j)}{\sqrt{1 - \sigma_{ij}}}\right)$$

where Φ is the standard normal cdf

Empirical content of each parameter (Natenzon, 2010):

p	information precision
μ	revealed preference
σ	revealed similarity

Axiom (Moderate Stochastic Transitivity)

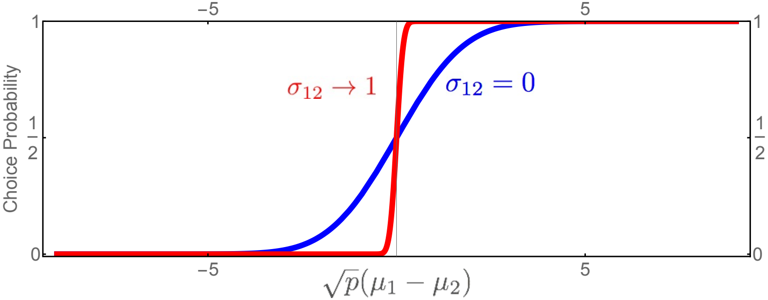
$\rho(i, j) > 1/2$ *and* $\rho(j, k) > 1/2 \Rightarrow \rho(i, k) > \min\{\rho(i, j), \rho(j, k)\}$.

Effect of correlation σ

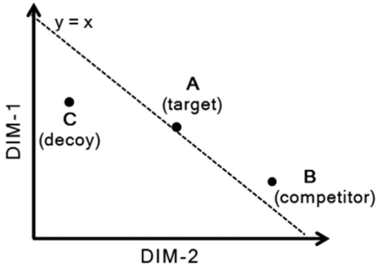
$$\left.\begin{aligned} X_i &= \mu_i + \varepsilon_i \\ X_j &= \mu_j + \varepsilon_j \end{aligned}\right\} \Longrightarrow X_i - X_j = \mu_i - \mu_j + (\varepsilon_i - \varepsilon_j)$$

$$\rho_p^{\mu\sigma}(i, \{i, j\}) = \Phi\left(\frac{\sqrt{p}}{\sqrt{2}} \frac{(\mu_i - \mu_j)}{\sqrt{1 - \sigma_{ij}}}\right)$$

- σ matters when $\sqrt{p}\,(\mu_i - \mu_j)$ is small



Closer look at the binary data



Menu	n	A	B	C
$\{A, B\}$	118♀	.37	.63	–
$\{B, C\}$	90♀	–	.69	.31
$\{A, C\}$	90♀	.84	–	.16

- $B \succ A \succ C$ rational benchmark
- More mistakes in B, C than in A, C
  $\Rightarrow A, C$ revealed more similar $\sigma_{AC} > \sigma_{BC}$
- Bayesian updating $\Rightarrow$ introducing C hurts B

Goodness-of-fit: Bayesian probit versus RUM

	data			BP			RUM		
Menu	A	B	C	A	B	C	A	B	C
$\{A, B\}$	.37	.63	–	.37	.63	–	.48	.52	–
$\{B, C\}$	–	.69	.31	–	.69	.31	–	.69	.31
$\{A, C\}$	.84	–	.16	.84	–	.16	.83	–	.17
$\{A, B, C\}$	.55	.28	.17	.57	.26	.17	.48	.35	.17
$\{A, B, \not{C}\}$	.61	.39	–	.60	.40	–	.48	.52	–

Log Likelihood: -265.2 (BP) and -271.0 (RUM).

$\Longrightarrow$ BP is at least $e^{5.886} \simeq 360$ times more likely to generate dataset than any RUM.

Akaike information criterion (AIC):
$BP \succ RUM \succ Logit \succ Probit$

Out-of-sample prediction

Estimate each model on restricted dataset:

	data		
Menu	A	B	C
$\{A, B\}$	.37	.63	–
$\{B, C\}$	–	.69	.31
$\{A, C\}$	.84	–	.16
$\{A, B, C\}$	.55	.28	.17

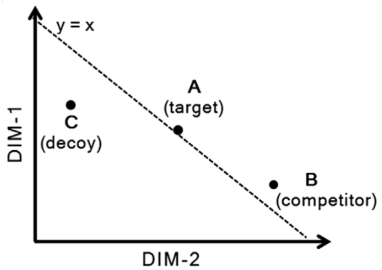
And predict choices for excluded menu $\{A, B, \not{C}\}$.

Model	A	B	$\not{C}$
data	.61	.39	–
BP	.61	.39	–
RUM	.42	.58	–
Probit	.46	.54	–
Logit	.49	.51	–

Discussion: are frogs irrational?

Estimated BP parameters:

$$\mu_B = 1.959\,(97.5\%), \qquad \sigma_{AB} = 0.154, \qquad t = 0.075$$
$$\mu_A = 0.454\,(67.5\%), \qquad \sigma_{AC} = 0.952$$
$$\mu_C = -0.566\,(28.6\%), \qquad \sigma_{BC} = 0.000$$



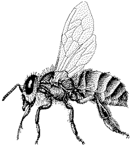
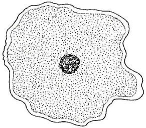
Menu	A	B	C
$\{A, B\}$	.37	.63	—
$\{B, C\}$	—	.69	.31
$\{A, C\}$	.84	—	.16
$\{A, B, C\}$	.55	.28	.17
$\{A, B, \not{C}\}$	.61	.39	—

Discussion: Bayesian updating, Monty Hall, evolution

Imagine frogs playing the Monty Hall game:

- Frogs that never switch doors win the prize 1/3 of the time
- Frogs that always switch doors win the prize 2/3 of the time

Any heuristic that behaves *as if* doing Bayesian updating could have an evolutionary advantage
$\Longrightarrow$ Possible evolutionary explanation for decoy effects

Conclusion

- Phenomenon: context-dependent choice behavior
- Limited Sampling $\Rightarrow$ mistakes
- Two lessons from psychophysics:
  - Mistakes and Value
  - Mistakes and Similarity
- Bayesian updating: favors options that are easier to compare
- Parametric Model: Bayesian probit

  t Limited Sampling
  μ Preference
  σ Similarity

- Better fit and out of sample prediction
  Allows standard welfare analysis

Thank you!

Louis Leon Thurstone (1887–1955)

‘Law’ of Comparative Judgement (Thurstone, 1927)

The ‘law’ is a **model** of binary comparisons:

Alternatives ordered in a **psychological continuum**

- gradations of gray, weight, excellence

The **discriminal process** for each alternative $X_i = \mu_i + \varepsilon_i$

ε_i discriminal deviation $\sim \mathcal{N}(0, 1/t)$

$1/t$ discriminal dispersion

$$\rho(1, \{1, 2\}) = \mathbb{P}\{X_1 > X_2\}$$

Comparison: Bayesian probit versus RUM

Multinomial Probit, Logit, Nested Logit, Cross-nested Logit, Mixed Logit are random utility models (RUMs).

Lemma (Block Marschak 1960)

Every RUM is equivalent to a probability measure over the $n!$ strict rankings of alternatives.

Let

$$
\begin{aligned}
p_{ABC} &= \mathbb{P}\{A \succ B \succ C\} \\
p_{ACB} &= \mathbb{P}\{A \succ C \succ B\} \\
p_{BAC} &= \mathbb{P}\{B \succ A \succ C\} \\
p_{BCA} &= \mathbb{P}\{B \succ C \succ A\} \\
p_{CAB} &= \mathbb{P}\{C \succ A \succ B\} \\
1 - p_{ABC} - p_{ACB} - p_{BAC} - p_{BCA} - p_{CAB} &= \mathbb{P}\{C \succ B \succ A\}
\end{aligned}
$$