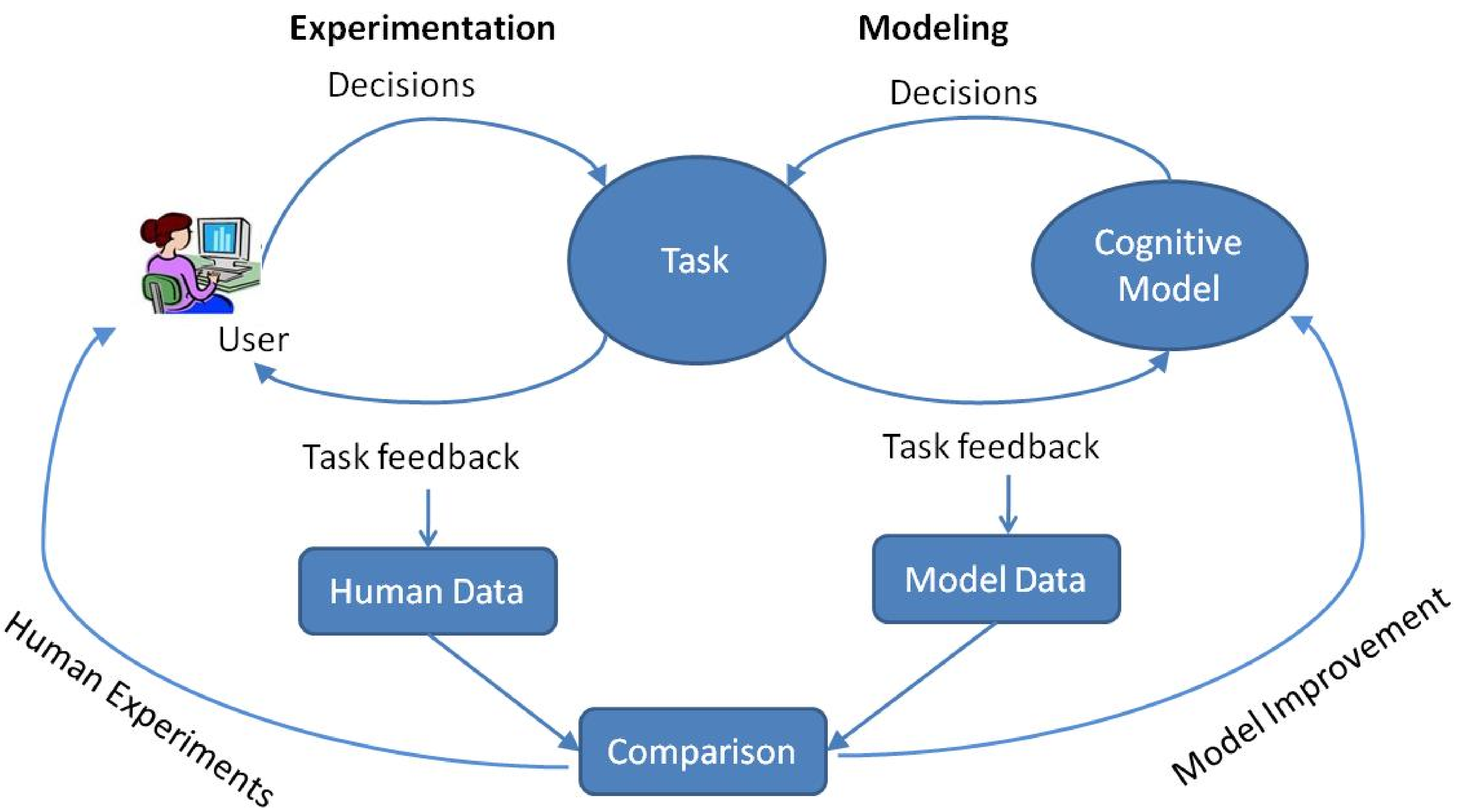
Dynamic Decision Making: Implications for Recommender System Design

Cleotilde (Coty) Gonzalez

Dynamic Decision Making Laboratory (www.cmu.edu/ddmlab)
Department of Social and Decision Sciences
Carnegie Mellon University

Research process and methods: Comparing Cognitive models against human data



Choice Explosion

Choice explosion in a cyber world

“A wealth of information

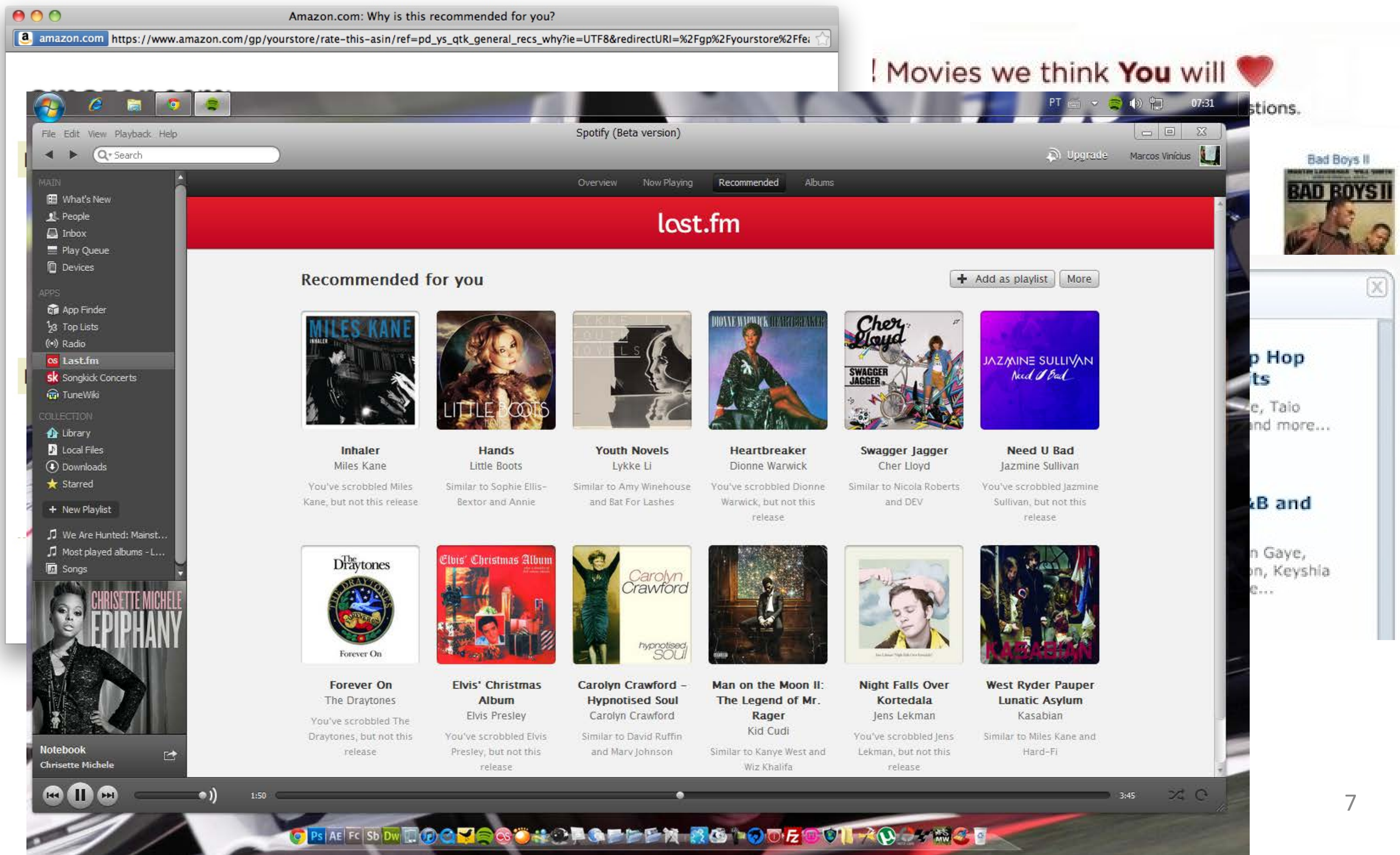
creates a poverty of attention

and a need to allocate it

Efficiently”

~Herb Simon (Nobel Prize Winner)

Recommender systems: many flavors



Human Decisions: Essence of Recommender systems

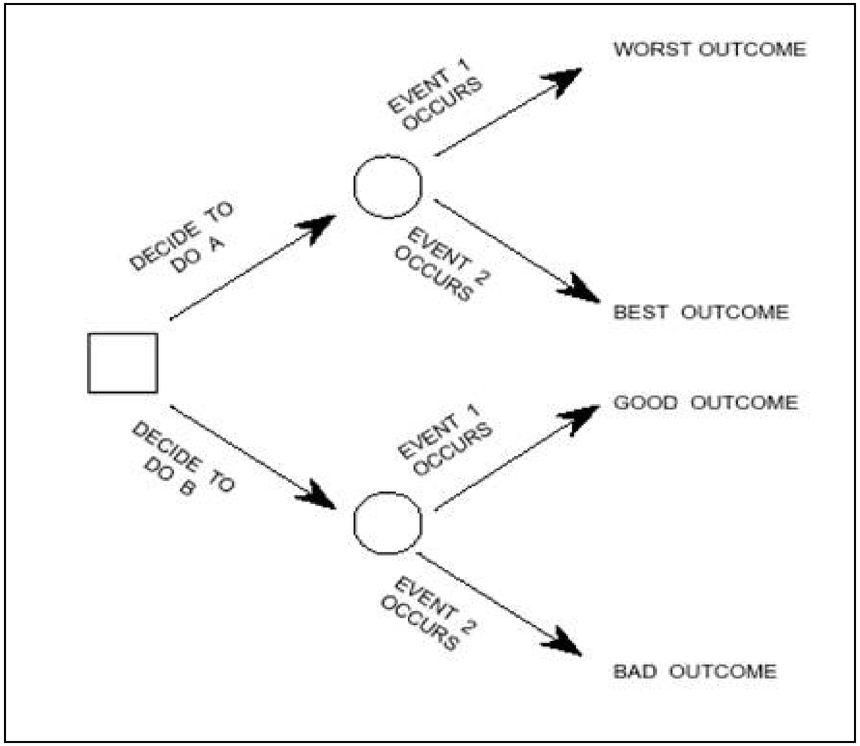
- Recommender systems aim at predicting preferences and ultimately human choice
- Human faced with a decision
  - Making a choice among a large set of alternatives
  - Relying on preferences:
    - Personal knowledge: preferences constructed through past experience (choices & outcomes experienced in the past)
    - Given knowledge: preferences constructed from information provided
- Human preferences are dynamic and contingent to the environment.

Premise: Dynamic decision making research may help to build recommender systems that learn and adapt recommendations dynamically to a particular user's experience to maximize benefits and overall utility from her choices

Outline:

- Offer a conceptual framework of decision making different from traditional choice: dynamic decision making
- Present main behavioral results obtained from experimental studies in dynamic situations
  - some initial findings on the dynamics of choice and trust on recommendations
- A theory (process and representations) and a computational model (algorithm) with demonstrated accuracy in predicting human choice

Static Decisions from Description



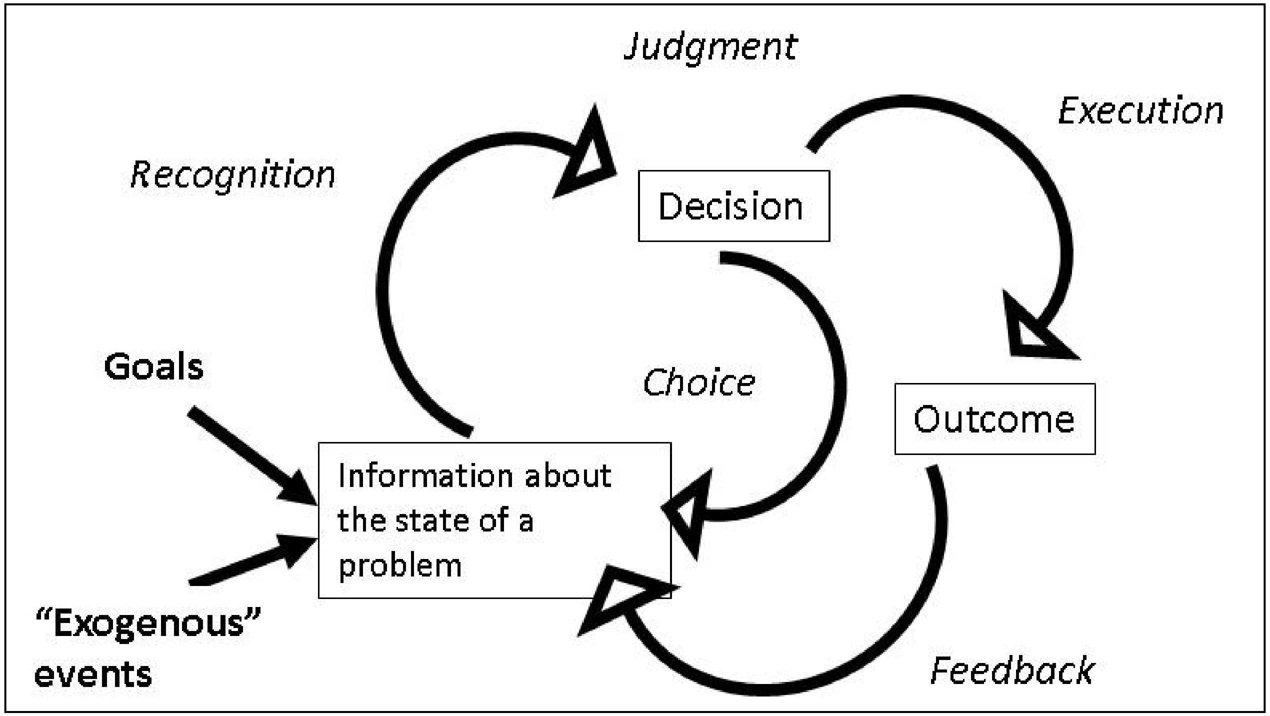
Which of the following would you prefer?

A: Get $4 with probability .8, $0 otherwise
B: Get $3 for sure

Assumptions:

1) Full information: options may be described by explicit outcomes and probabilities
2) Unlimited time and resources: No constraints in the decision making process
3) Stability: mapping between choice attributes and utility remain constant over time (and across individuals, and within a single individual).

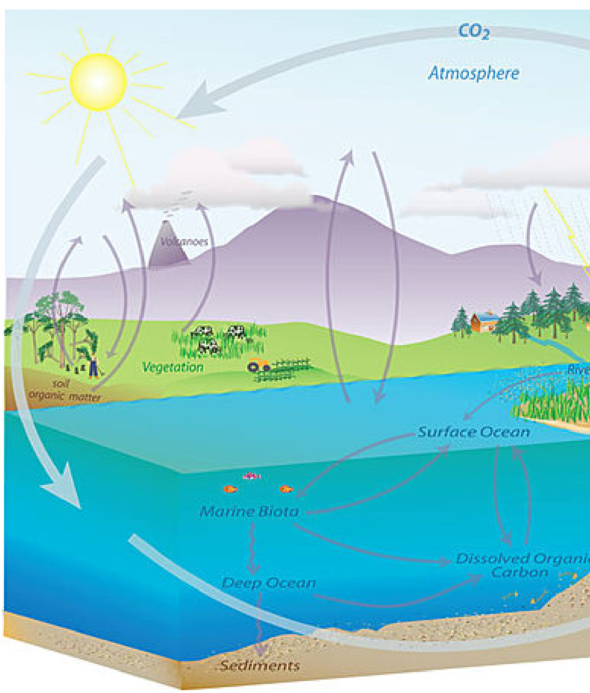
Dynamic Decisions from Experience



Dynamic Decision Making

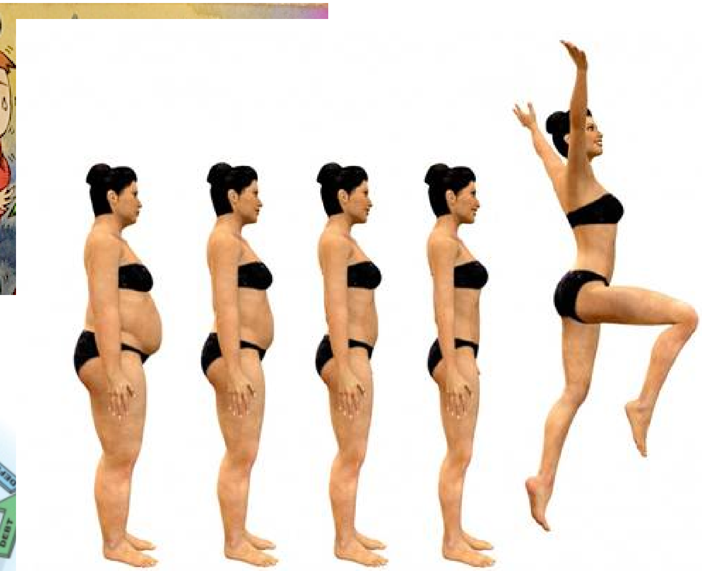
1. Series of Decisions
2. Decisions are interdependent: the output of one becomes the input of the future ones
3. Environment changes: either independently or dependently as a result of previous decisions
4. Utility of decisions is time-dependent (according to *when* they are made)
5. Resources and Time are limited


EXIT 129
NORTH
G. S. Parkway
N. J. Turnpike
EXIT ONLY


CO_2
Atmosphere
Carbon Cycle
Volcanoes
Vegetation
soil organic matter
Surface Ocean
Marine Biota
Dissolved Organic Carbon
Deep Ocean
Sediments


CREDIT CARDS


DEBT



Memory, Experience, Learning

A Continuum of "dynamics"
Only requirement: A sequence of decisions

Least Dynamic	Most Dynamic
No changes in the environment although the environment is probabilistic, probabilities and values don't change over the course of decisions	**Environment changes** (Independently and as a consequence of the actions of the decision maker)
Immediate feedback (Action-Outcome closest in time)	**Delayed feedback and Credit assignment problem** (Multiple actions and multiple outcomes separated in time)
Value is time independent (Time of the decision is determined by the decision maker, no penalty for waiting)	**Value is time-dependent** (Value decreases the farther away the decision is from the optimal time)
Simple	Complex

Complex dynamic environments: Microworld research

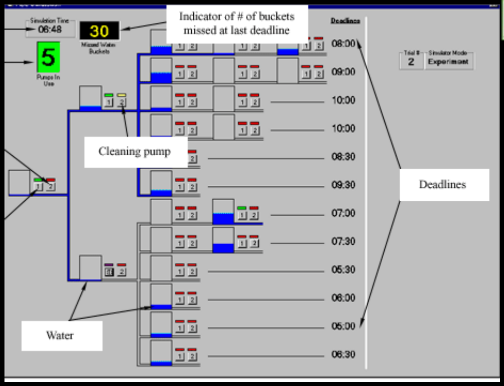
Gonzalez, Vanyukov & Martin, 2005

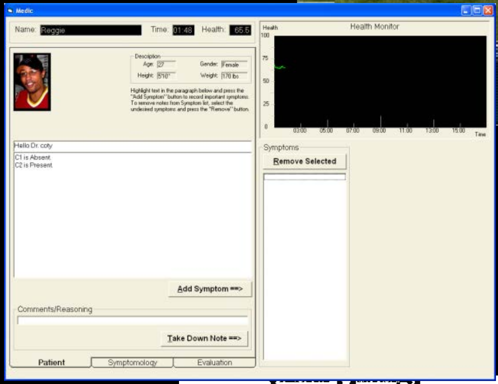
Dynamic Visual Detection

Conflict Resolution

Military Command and Control

Real-time resource allocation

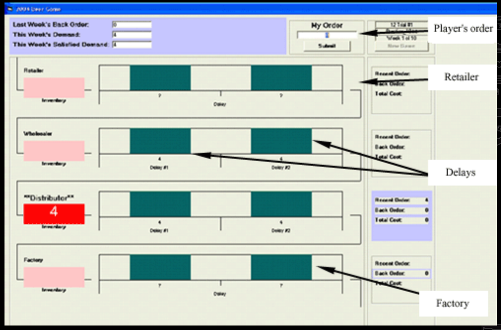


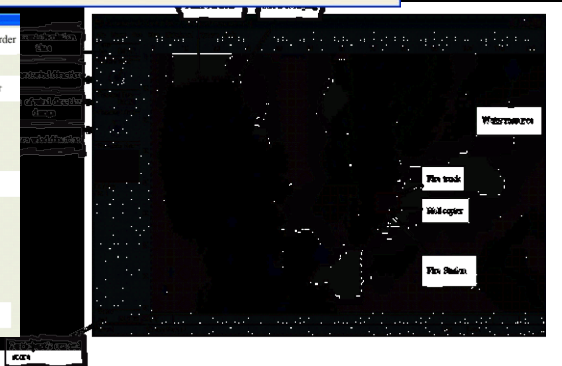


Medical Diagnosis

Climate Change

Supply-Chain Management





Fire Fighting

Main findings from my research with Microworlds (summarized in Gonzalez 2012)

- More "headroom" during training helps adaptation
  - Time constraints (Gonzalez, 2004): Slow pace training helps adaptation to high time constrains
  - High workload(Gonzalez, 2005): Low workload during training helps adaptation to high workload
- Heterogeneity of experiences helps adaptation
  - High diversity of experiences (Gonzalez & Quesada, 2003; Gonzalez & Thomas, 2008; Gonzalez & Madhavan, 2011; Brunstein and Gonzalez, 2011) helps detection of novel items
- Ability to "pattern-match" and see similarities is associated to better performance in DDM tasks (Gonzalez, Thomas and Vanyukov, 2005)
- Feedforward helps future performance of DDM tasks without feedback (Gonzalez, 2005)

A Continuum of "dynamics"
Only requirement: A sequence of decisions

Least Dynamic

No changes in the environment although the environment is probabilistic, probabilities and values don't change over the course of decisions

Immediate feedback (Action-Outcome closest in time)

Value is time independent (Time of the decision is determined by the decision maker, no penalty for waiting)

Simple

Most Dynamic

Environment changes (Independently and as a consequence of the actions of the decision maker)

Delayed feedback and Credit assignment problem (Multiple actions and multiple outcomes separated in time)

Value is time-dependent (Value decreases the farther away the decision is from the optimal time)

Complex

Choice: Abstract and simple experimental paradigms

Repeated choice Paradigm

(Barron & Erev, 2003)

4

4

3

… 0 …..

Fixed number of trials

Sampling Paradigm

(Hertwig et al. 2004)

Sampling

4

4

3

0

Make a choice:

4

Description-Experience Gap

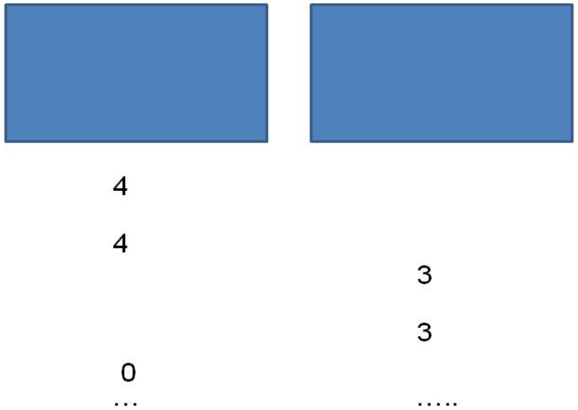
Barron & Erev (2003); Hertwig, Barron, Weber & Erev (2004)

- <u>Description:</u>

A: Get $4 with probability .8, $0 otherwise

B: Get $3 for sure

- <u>Experience</u>



Make a final choice:

DEGap: Pmax (A choices) = 36% - Pmax = 88% = 52

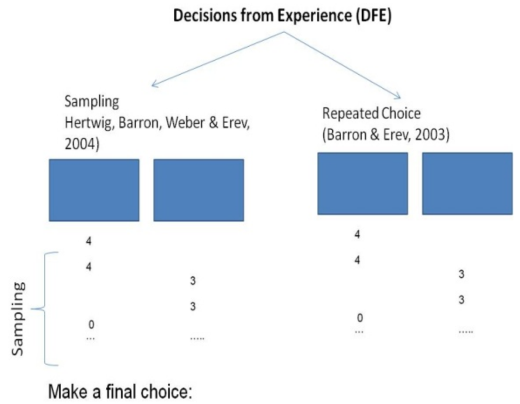
<u>Description</u>: According to Prospect Theory people **overweight** the probability of the rare event

<u>Experience</u>: as if people **underweight** the probability of the rare event

Exploration process: a theoretical divide?

DE-Gap is due to

Sampling		Repeated Choice
Reliance on small samples		Reliance on recent outcomes



Exploration transitions – A theoretical divide?

Exploration – Exploitation two distinct processes

Models often assume that sampling is random

Exploration - Exploitation tradeoff

Increase selection of best known option over time

Gonzalez & Dutt (2011)

- Demonstrate the behavioral regularities between sampling and consequential choice paradigms:
  - Similar Description-Experience(DE)-Gap
  - Gradual decrease of exploration over time
  - Maximization in choice
  - Prediction of choice from memory: Selection of option with the highest experienced expected outcome during past experience
- Demonstrate that people rely on remarkably similar cognitive processes in both paradigms:
  - People explore options aiming to get the best possible outcome
  - Rely on their (faulty) memories (frequency, recency and noise)
- A single cognitive model based on Instance-Based Learning Theory (IBLT; Gonzalez, Lerch, & Lebiere, 2003):
  - Explains the learning process and predicts choice better than models that were designed for one paradigm alone (e.g., the winners of the Technion Modeling competition - TPT)

Human data sets

	Description	Sampling	Repeated Choice
6 problems	Hertwig et al., 2004 N=50	Hertwig et al., 2004 N=50	Barron & Erev, 2003 N=144
Technion Prediction Tournament (TPT) Erev et al., 2011	N=100 60 problems Estimation set N=100 60 problems Competition set	N=100 60 problems Estimation set N=100 60 problems Competition set	N=100 60 problems Estimation set N=100 60 problems Competition set

Similar DEGap in Sampling and Consequential Choice paradigms

	Description	Sampling	Repeated Choice
6 problems	Hertwig et al., 2004 N=50	Hertwig et al., 2004 N=50	Barron & Erev, 2003 N=144
Technion Prediction Tournament (TPT) Erev et al., 2011	N=100 60 problems Estimation set N=100 60 problems Competition set	N=100 60 problems Estimation set N=100 60 problems Competition set	N=100 60 problems Estimation set N=100 60 problems Competition set

Significant gap for each of the 6 problems

r = .93, p =.01

r = –.53, p=.0004

r = .83, p =.0001

r = –.37, p =.004

Similar risky choices across DFE paradigms, but is exploration similar?

In TPT data sets

- P-risky choices (Estimation and Competition)
  - Sampling = 0.49 & 0.44
  - Repeated choice = 0.40 & 0.38

- Alternation rate (A-rate) is a measure of exploration. A-rate (Estimation and Competition)
  - Sampling = 0.34 & 0.29
  - Repeated choice = 0.14 & 0.13

- Alternation correlations between sampling and consequential choice over time
  - r =.93, p=.01 Estimation set
  - r =.89, p=.01 Competition set

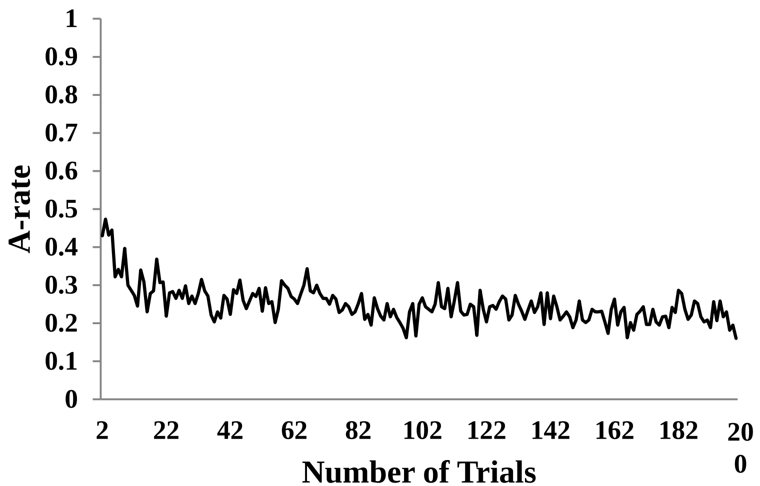
Exploration decreases over time

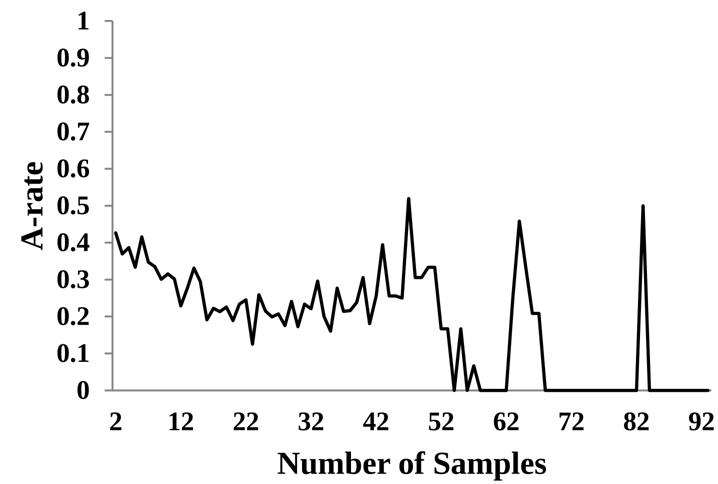
Gonzalez & Dutt, 2011

Repeated Choice

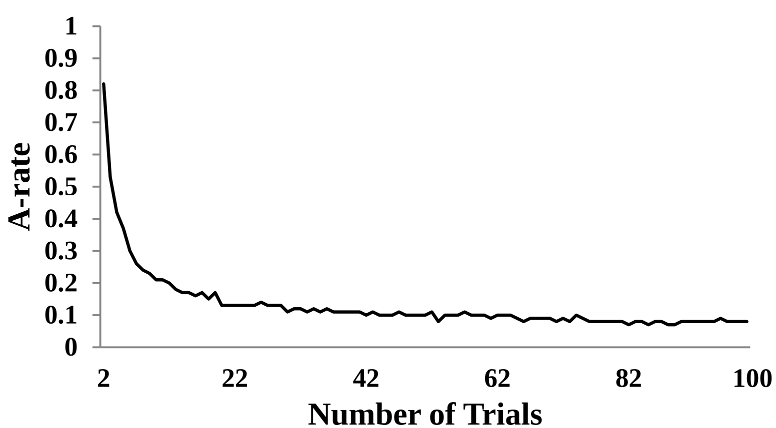
Sampling

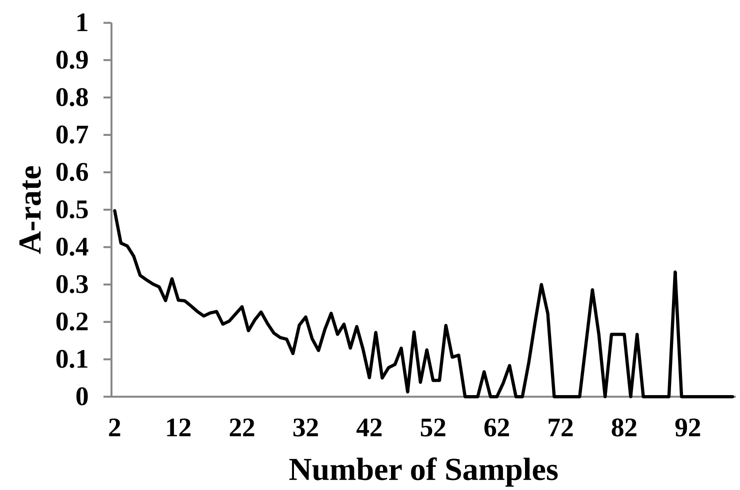
6 problems Hertwig et al., 2004





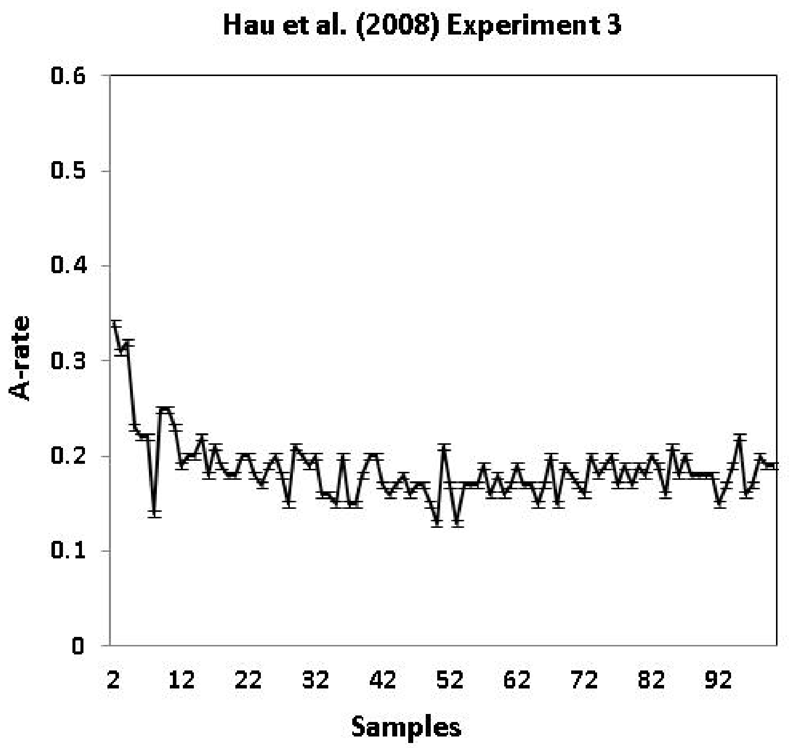
Technion Prediction Tournament (TPT) Erev et al., 2011

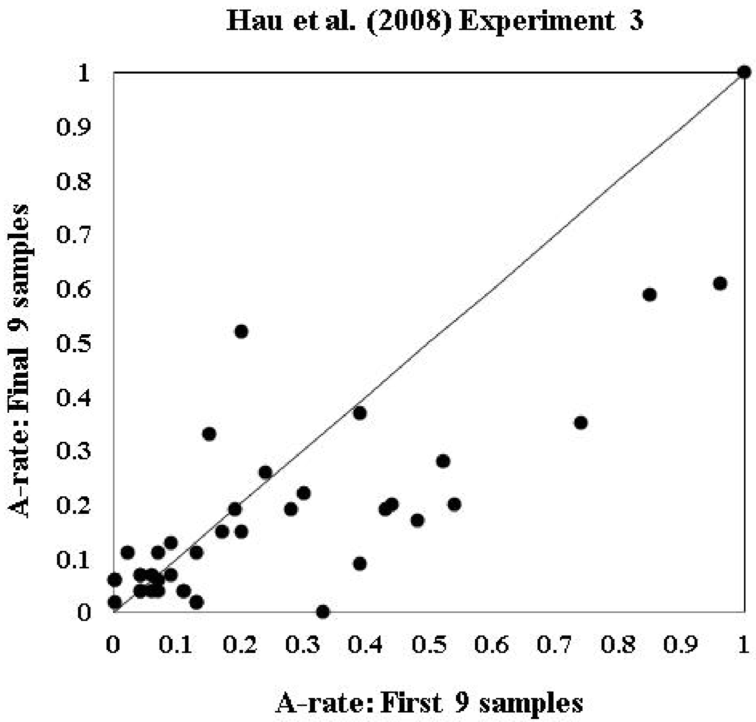




Decreased exploration over time occurs for most individuals

Gonzalez & Dutt, 2012



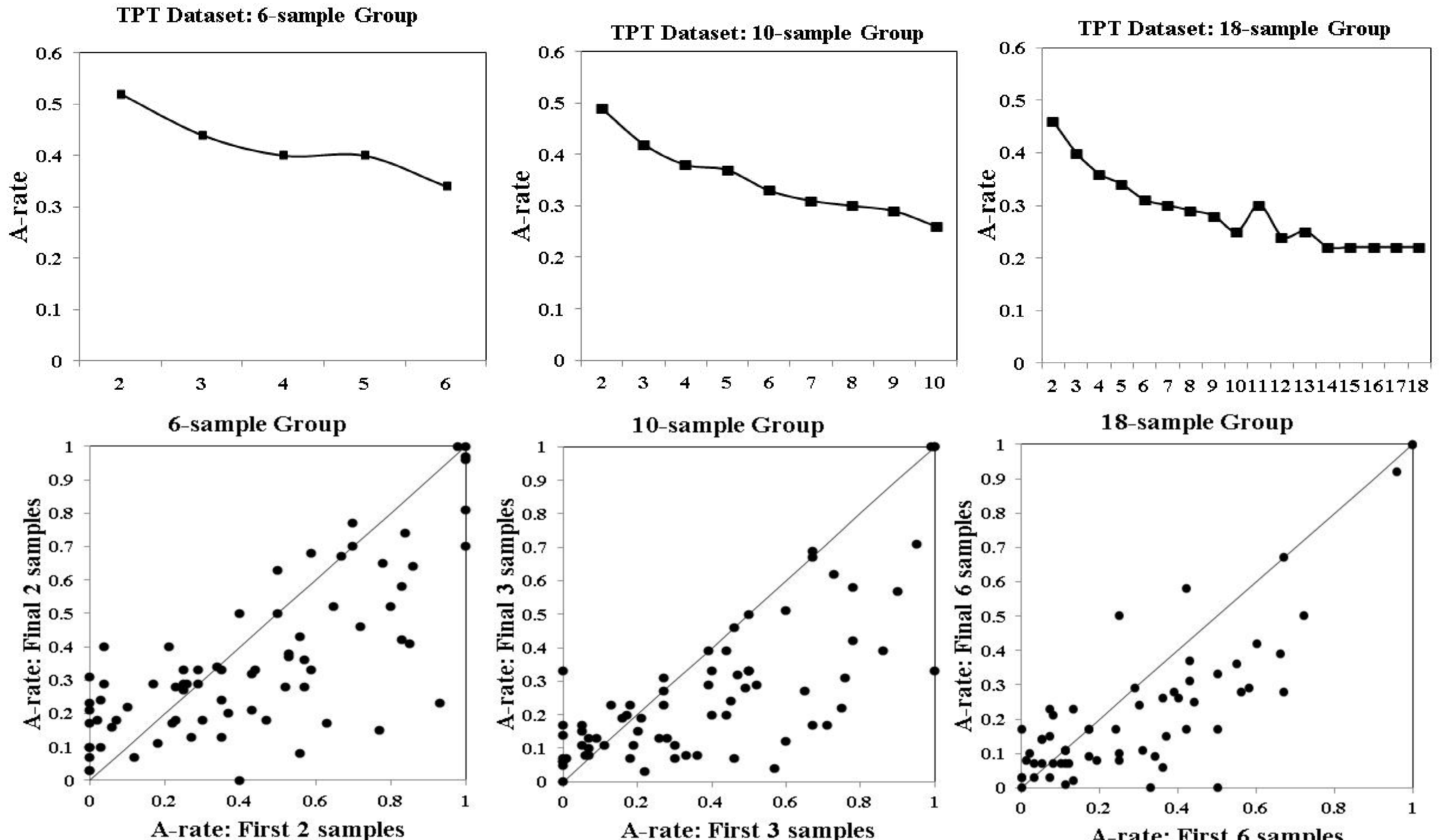


In first 11 trials A-rate falls 44% and then the curve flattens to about 19% ➔ remarkably similar to consequential choice

Initial and final A-rates at the individual level. 4/40 (10%) kept their initial and final A-rates constant; 12/40 (30%) increased A-rate; and 24/40 (60%) fell below the diagonal, decreased A-rate

The longer individuals sample, the more they decrease exploration

(Gonzalez & Dutt, 2012)

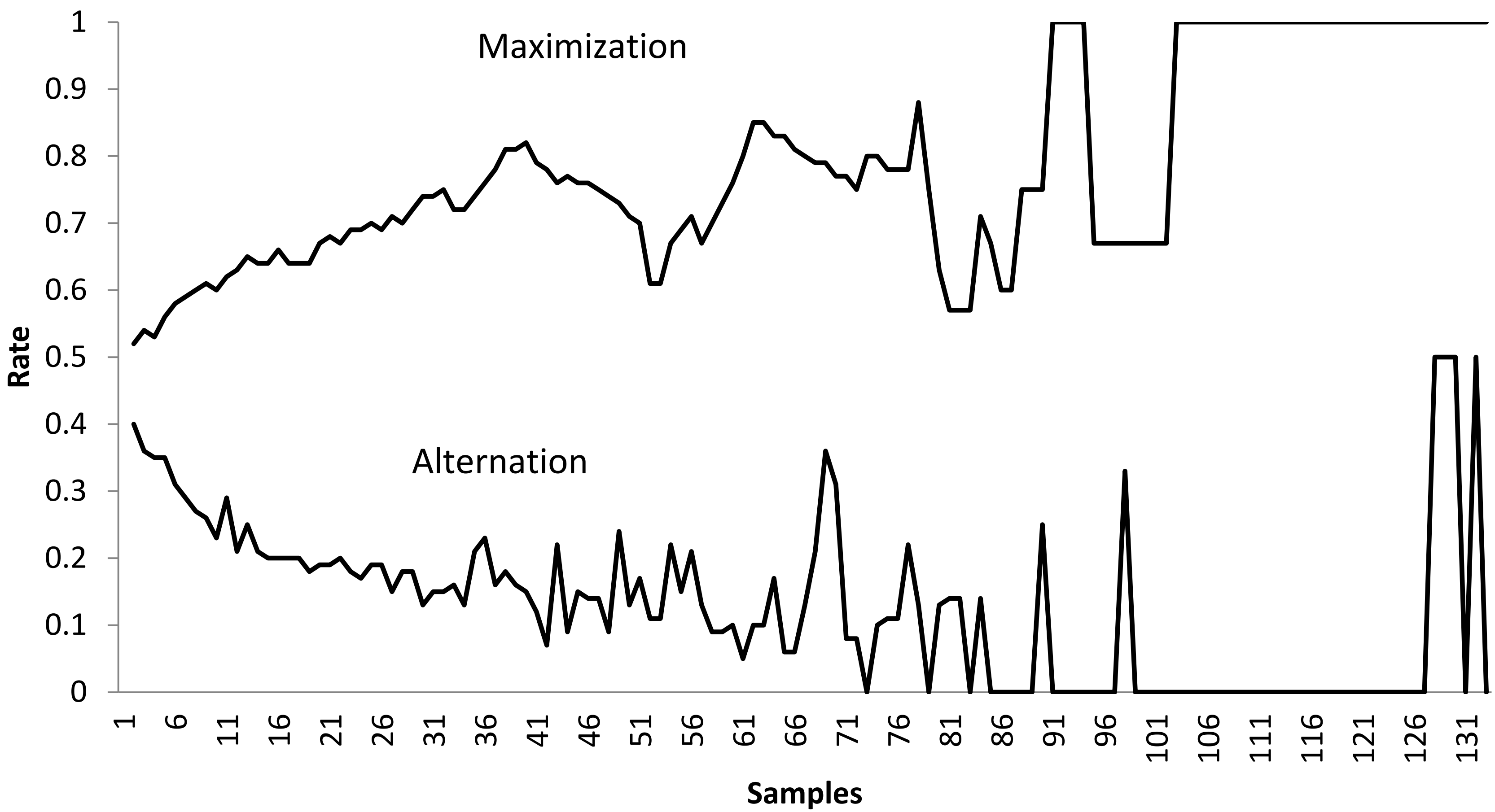


Choice is predicted by maximization from experience

Gonzalez & Dutt, 2011; Gonzalez & Dutt, 2012; Mehlhorn et al., 2014

- In Hau et al.'s data (2008)
  - Maximization during sampling & Maximization at choice ($r(38) = 0.36$, $p < .05$).
  - 60% of the choices predicted by maximizing option during sampling are consistent with final choices.
- In TPT sampling data set
  - A positive correlation of Maximization behavior in the three groups:
    - $r(73) = .26$, $p < .05$ for the 6-samples group
    - $r(70) = .34$, $p < .01$ for the 10-samples group
    - $r(60) = .40$, $p < .01$ for the 18-samples group
  - 84% of the choices predicted by the maximizing option during sampling are consistent with the final choices.

Concurrence of Exploration and Maximization in Decisions from Sampling (Gonzalez & Dutt, under review)



$r_s = -.48, p < .01$

Learning in imperfect recommendation systems
(Harman, Odonovan, Abdelzaher, Gonzalez, 2014: Recsys 2014)

Value obtained from choice

High/Low outcome from choice

Value obtained from choice

High/Low outcome from choice

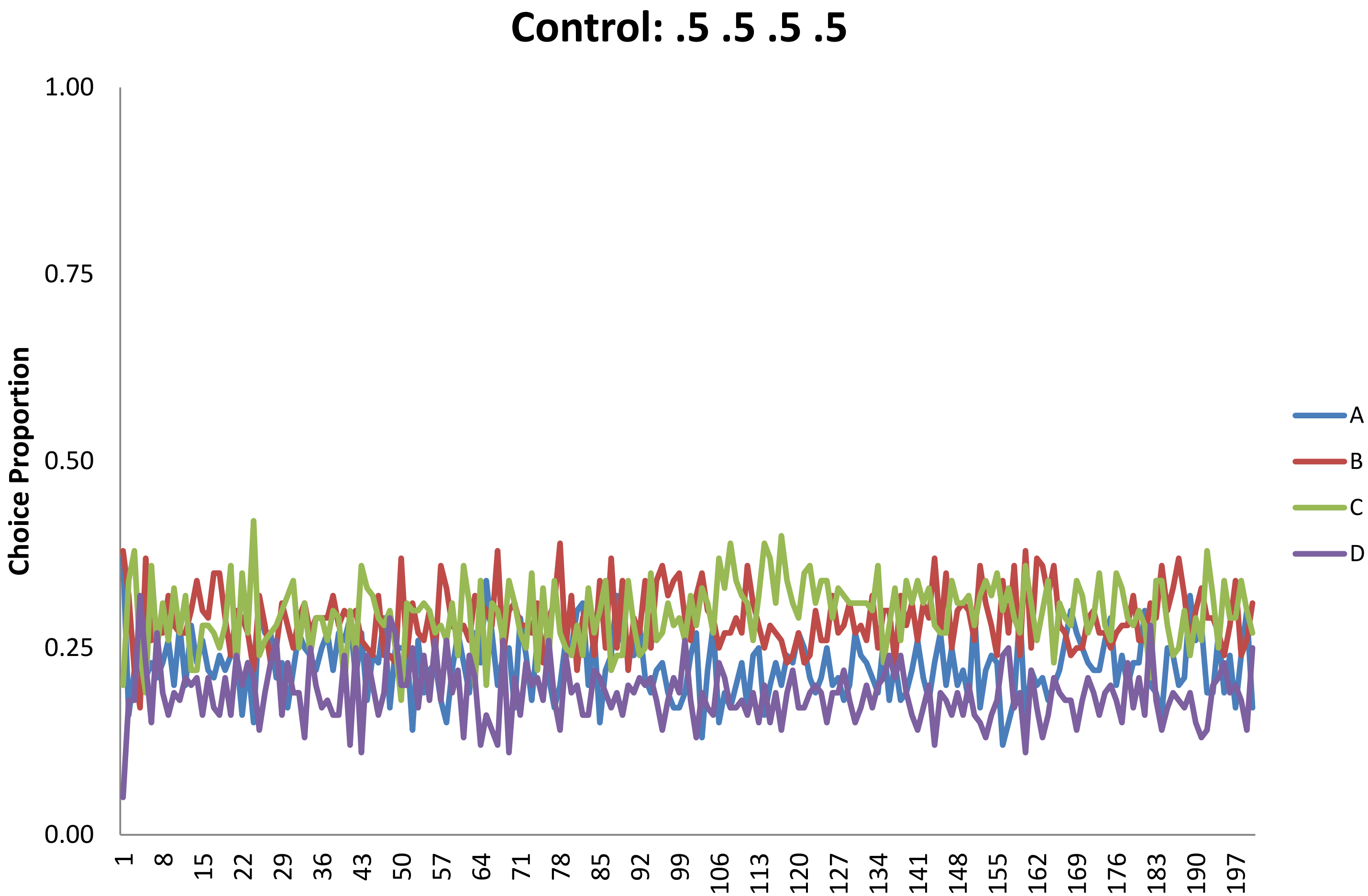
Accuracy of the recommender

High/Low accuracy

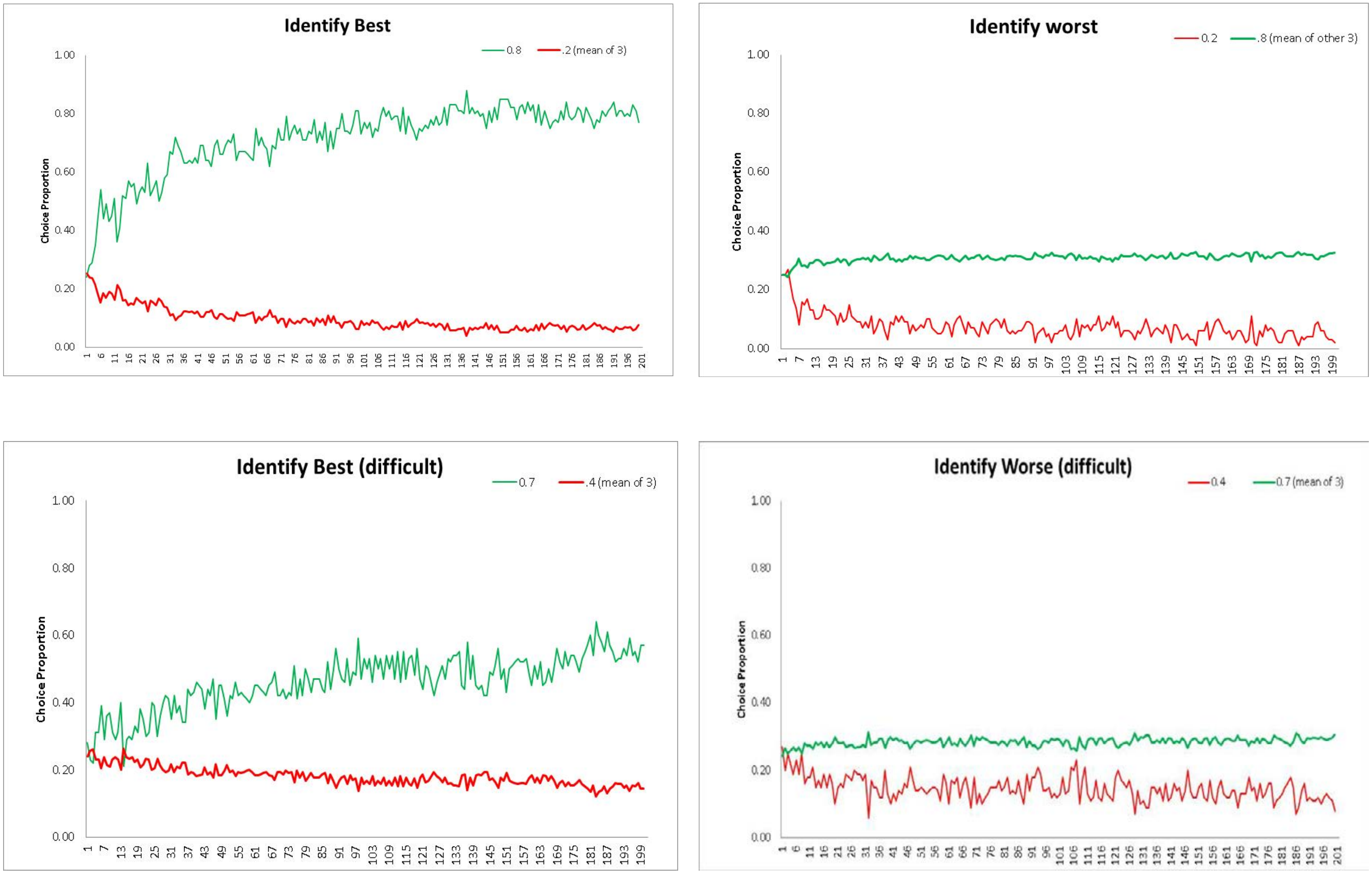
Experiments

- **Exp 1**: Learning value (over 200 trials) without recommendations. Each Condition 100 participants. Conditions represent the probability of obtaining a high (1) outcome.
  - Control condition: .5 .5 .5 .5
  - Identify best/worst value:
    - Easy: .8 .2 .2 .2/ .2 .8 .8 .8
    - Difficult: .7 .4 .4 .4/.4 .7 .7 .7
  - Identify best value among distinct/similar sources:
    - Distinct: .2 .4 .6 .8
    - Similar: .4 .5 .6 .7
- **Exp 2**: Learning value with recommendations. Same as Exp. 1, but with accurate (p=1) or inaccurate (.5) recommendations.

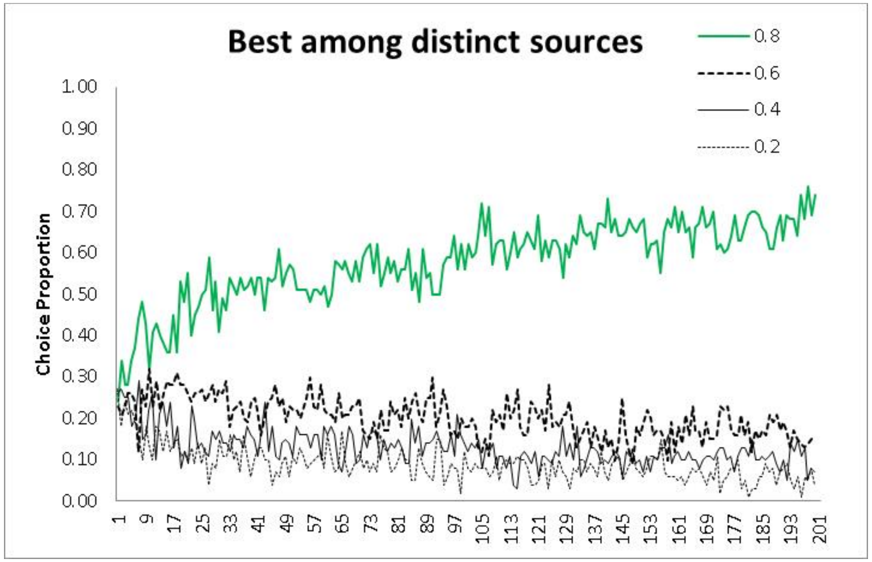
Exp. 1: Control Condition

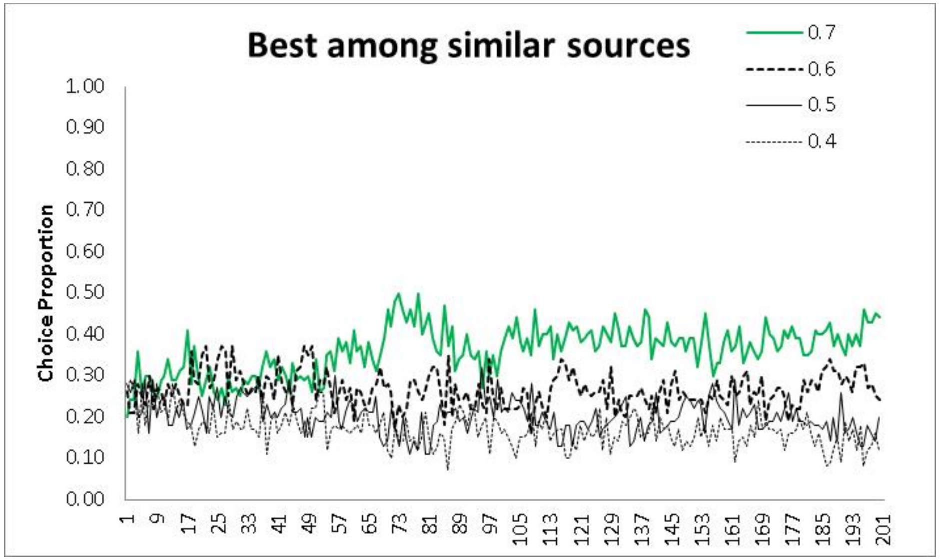


Exp. 1: Identify Best/Worst value

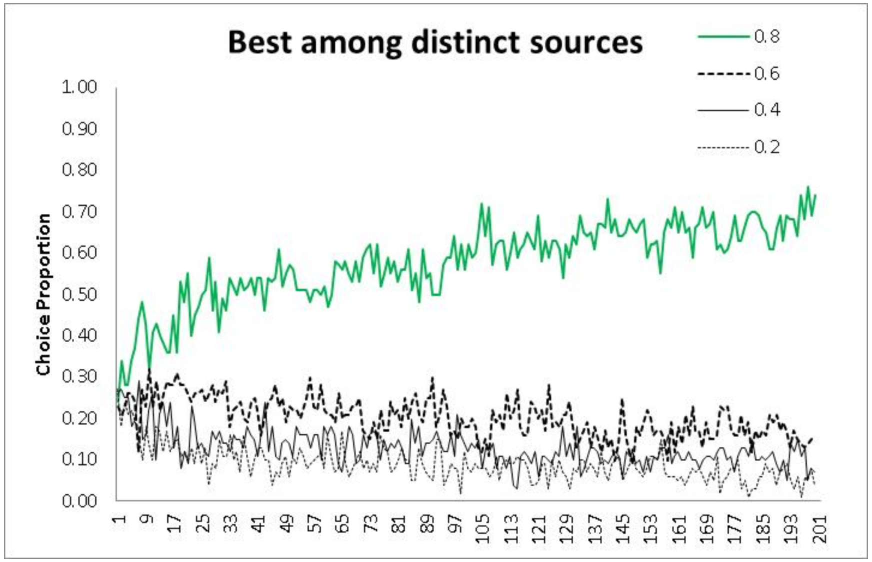


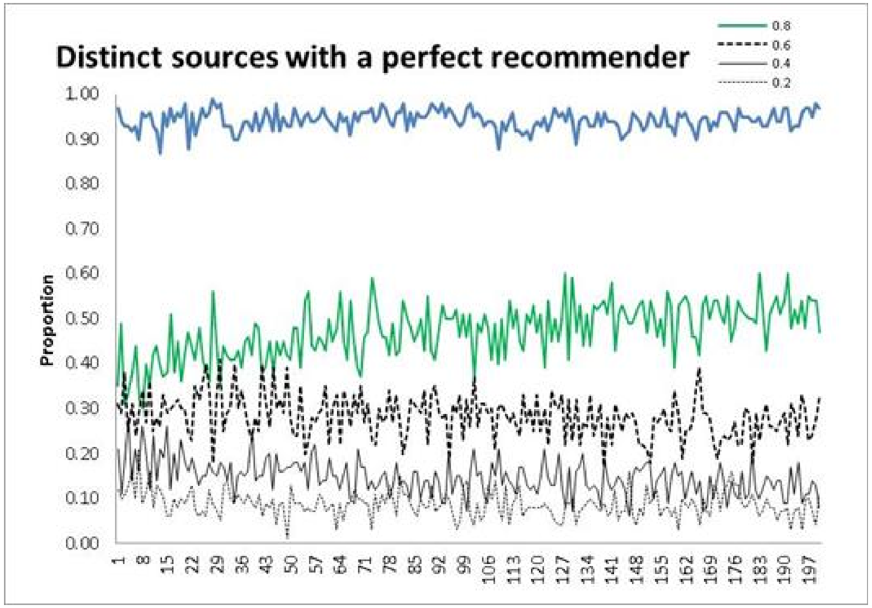
Exp. 1: Identify best among distinct/similar sources

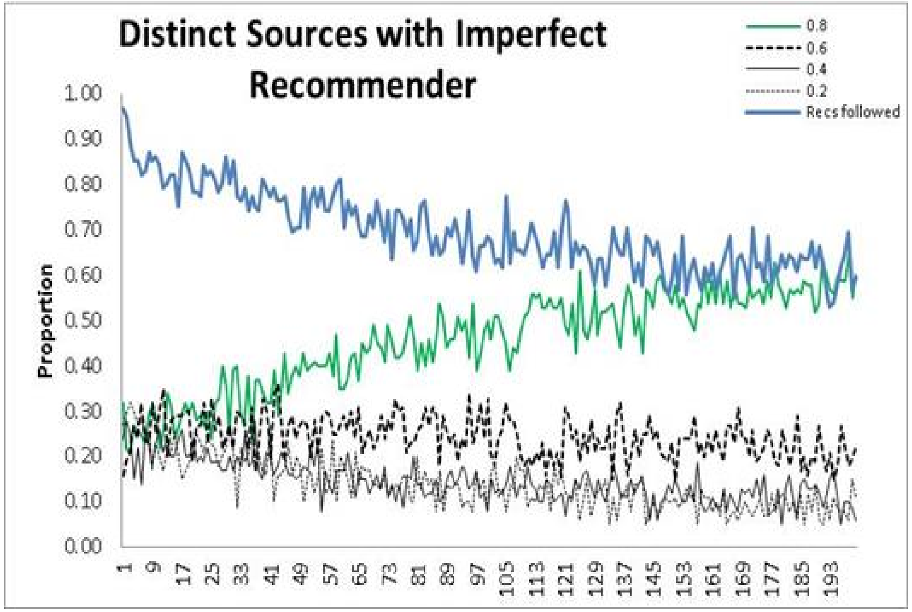




Exp. 2: Identify Best/Worst value among distinct/similar sources with a recommender help







Summary of behavioral phenomena

Conditional Reinforcement:

Increasingly select actions that led to best outcomes in similar past experiences

Reduced Exploration:

Decrease exploration of options over time in consistent environments

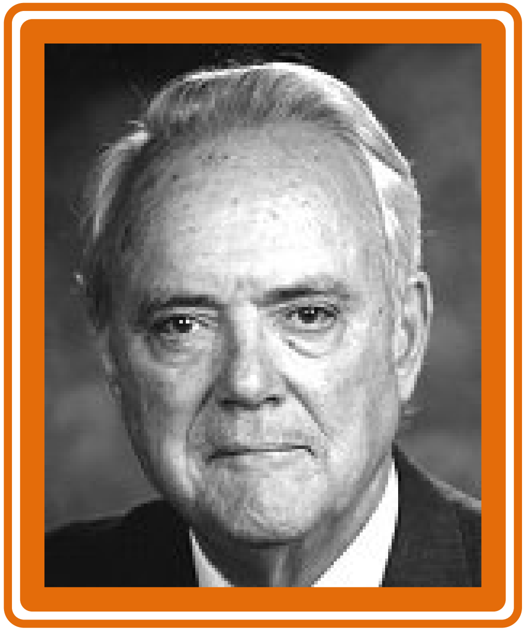
Recommender systems:

Recommenders may act as distractions for humans' own exploration and search for best value

Humans abandon imperfect recommenders


Decision theory
behaviour
human
rule
maximization
believing
evangelical
unknown
reinterpretation
regularities
fallacy
events
optimal
values
renewed
alternative
probability
argument
concerned
model
infinite
prescriptive
normative
departures
approximative
antecedent
presuppositions
systematic
theorems
reformulated
rates
subjective
unquestioning
procedure
probabilistic
axioms
outcome
outsized
illustrative
worshiped
distribution
maximizing
paper
expected
compute
situation
postulates
decision
people
hypothesis
fuzzy
identify
assumptions
outperforms
admissible
work
behavior
paradox
cargo
uncertain
estimation
predictions
deviates
descriptive
axiomatic
choice
rational
rationality
capricious
complexity
theory

"... static decision theories have only a limited future. Human beings learn, and probabilities and values change; these facts mean that the really applicable kinds of decision theories will be dynamic, not static" Edwards (1961, page 485).



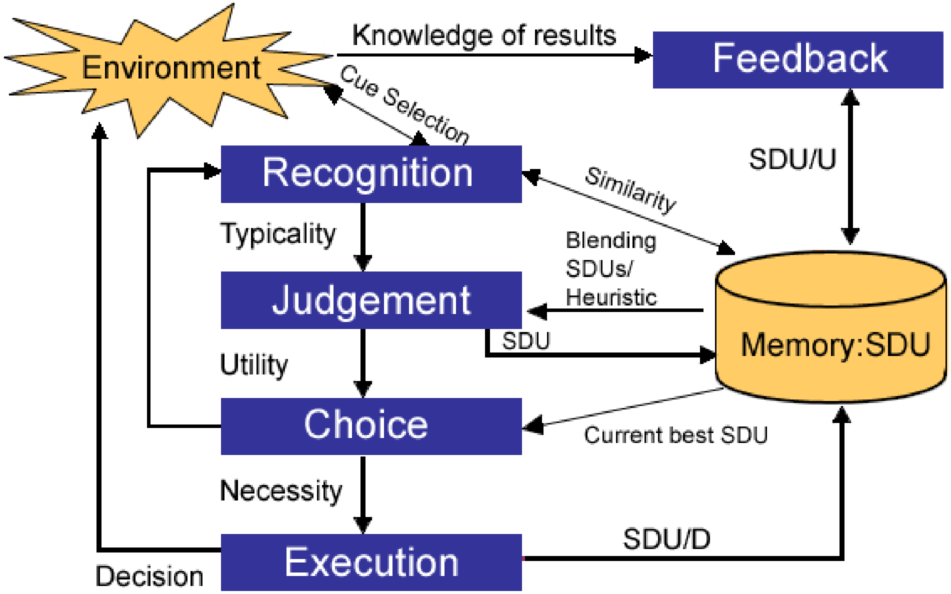
Ward Edwards (1927-2005)

Dynamic Decision Theories: Learning Theories

- Psychology is full of learning theories!
  - Toward an instance theory of automatization (Logan, 1988)
  - The use of specific instances to control dynamic systems (Dienes & Fahey, 1995)
  - Learning in Dynamic Decision Tasks (Gibson, Fichman & Plaut, 1997)
  - Case-Based Decision Theory (Gilboa & Shmedlier, 1995)
- Instance-Based Learning Theory (IBLT) (Gonzalez, Lerch and Lebiere, 2003)
  - Descriptive account of the cognitive structures and learning processes involved in human decision making in dynamic environments (Gonzalez et al., 2003)
  - IBLT characterizes learning in dynamic tasks by storing a sequence of instances, “Situation-Decision-Utility” triplets, produced by experienced events in memory.

Dynamic Decision Theory
Instance-Based Learning Theory (IBLT)
(Gonzalez, Lerch, & Lebiere, 2003)

- Proposes a generic DDM cognitive process:
  Recognition, Judgment, Choice, Execution, Feedback
- Formalizes representations:
  - Instance: tripled: Situation, Decision, Utility (SDU)
  - Relies on mathematical mechanisms proposed by ACT-R
- Represents processes computationally: to provide concrete predictions of human behavior in various task types



IBL model of choice

1. Each experience combination is created as an instance in memory (e.g. S-10; P-8; S-1; P-5; S-5) when the outcome is experienced
2. Each instance has a memory “activation” value based on frequency, recency, similarity, etc.
3. The probability of retrieving an instance from memory depends on activation
4. For each option, memory instances are “blended” to determine next choice by combining value and probability
5. Choose the option with the maximum blended value

10

10

8

1

5

5

...

.....

A formalization of an IBL model of binary-choice (Gonzalez & Dutt, 2011; Lejarraga et al., 2012)

1. **Each Instance** has an Activation: simplification of ACT-R's mechanism (Anderson & Lebiere, 1998):

$$A_{i,t} = ln\left(\sum_{t_i \in \{1,\ldots,t-1\}} (t - t_i)^{-d}\right) + \sigma \,.\, ln\left(\frac{1-\gamma_{i,t}}{\gamma_{i,t}}\right)$$

Frequency **Recency**

Free parameters: *d* : high d-> More recency *Noise:* σ : high s -> high variability

2. **Each Instance** has a probability of retrieval is a function of memory Activation (A) of that outcome relative to the activation of all the observed outcomes for that option given by:

$$P_{i,t} = \frac{e^{A_{i,t}/\tau}}{\sum_j e^{A_{j,t}/\tau}} \qquad \tau = \sigma \cdot \sqrt{2}$$

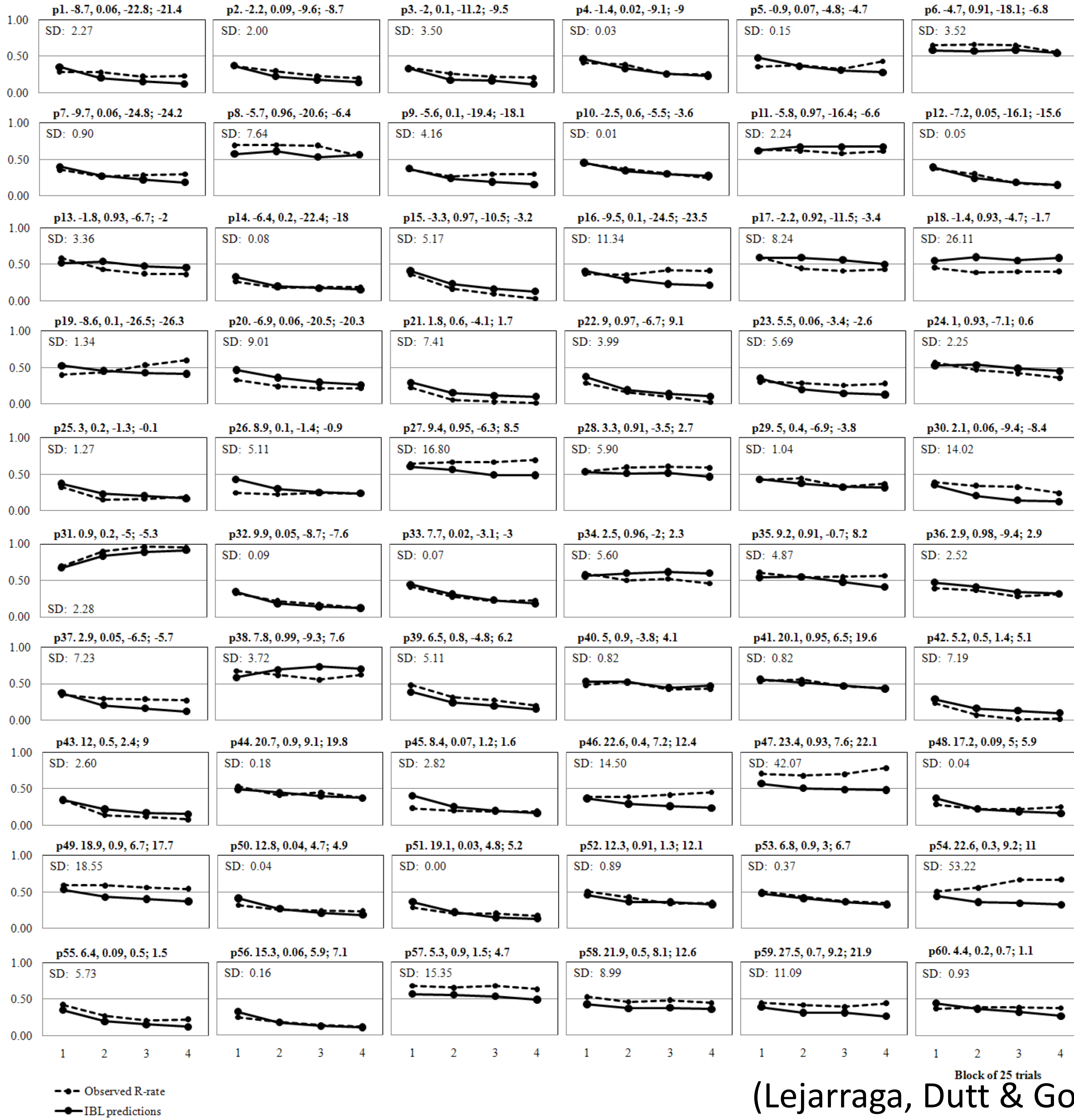
3. **Each Option** has a Blended Value that combines the probability of retrieval and outcome of the instances:

$$V_j = \sum_{i=1}^{n} p_i x_i$$

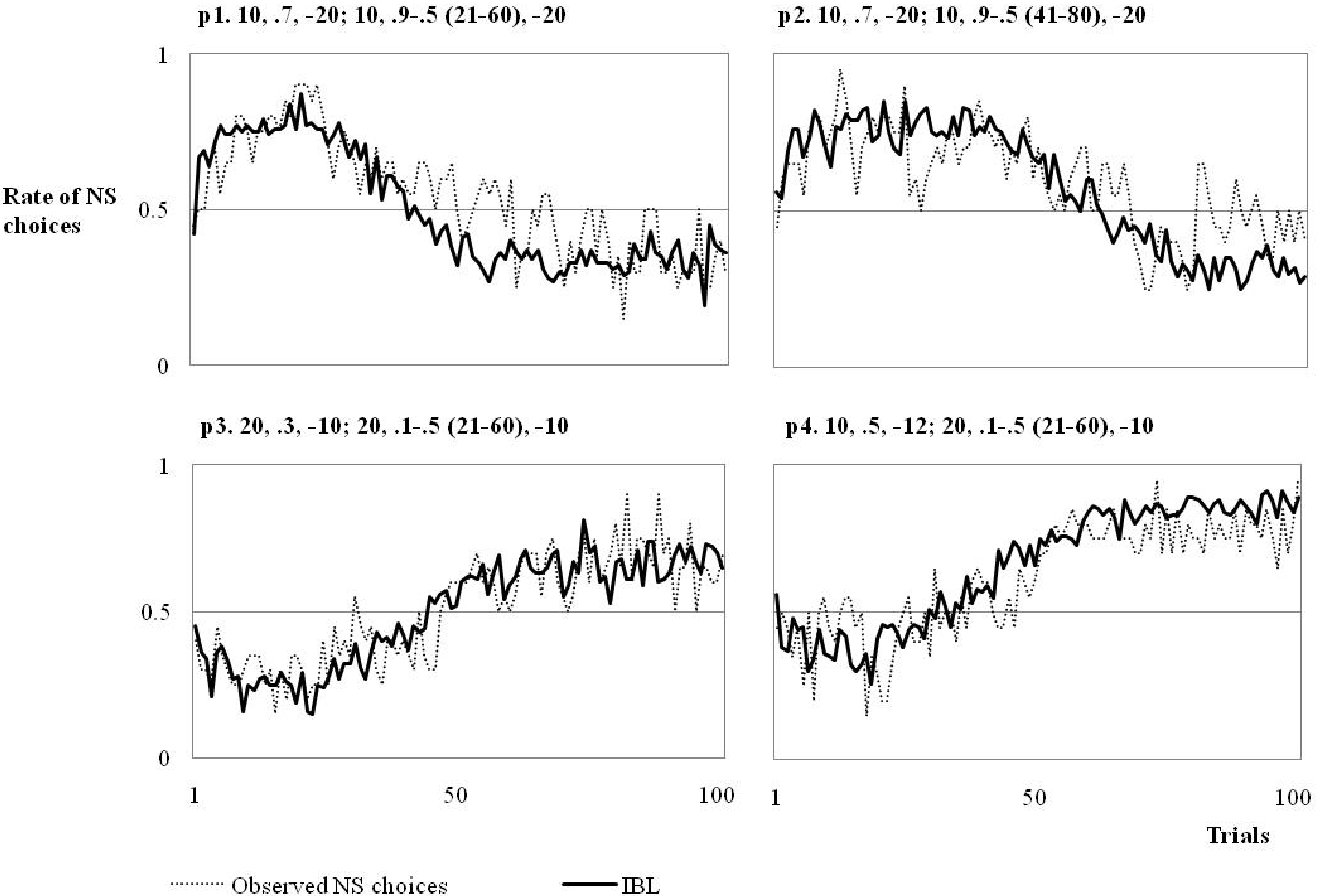
4. **Choose** the option with the highest experienced expected value ("blended" value)

Robustness of the IBL model's prediction

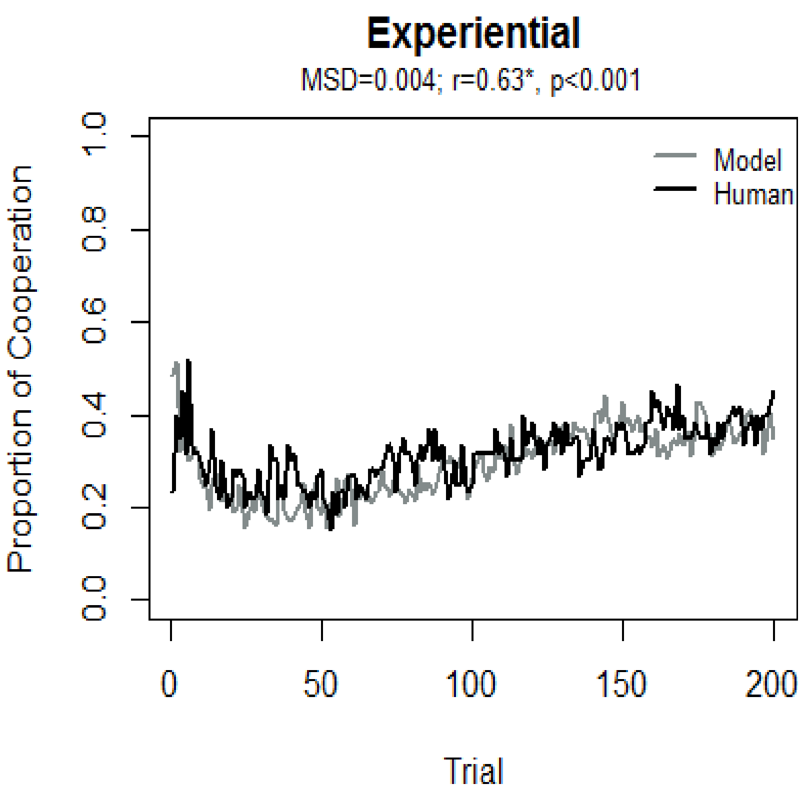
- In three different tasks: Repeated choice; Probability Learning; Repeated choice with non-stationary probabilities (Lejarraga et al., 2012)
- Across two different paradigms: sampling and repeated choice (Gonzalez & Dutt, 2011)
- In a market entry task (Gonzalez, Dutt & Lejarraga, 2011)
- To demonstrate how decision "biases" disappear when making decisions from experience (Hartman & Gonzalez, 2014; Mehlhorn et al., 2013; Gonzalez & Mehlhorn, 2014)
- To demonstrate the short and long-term dynamics of cooperation in the Prisoner's dilemma and other social dilemmas (Gonzalez, Ben-Asher, Martin & Gonzalez, 2014)
- Learning with imperfect recommendations (Harman, Abdelzaher, Gonzalez, in prep)

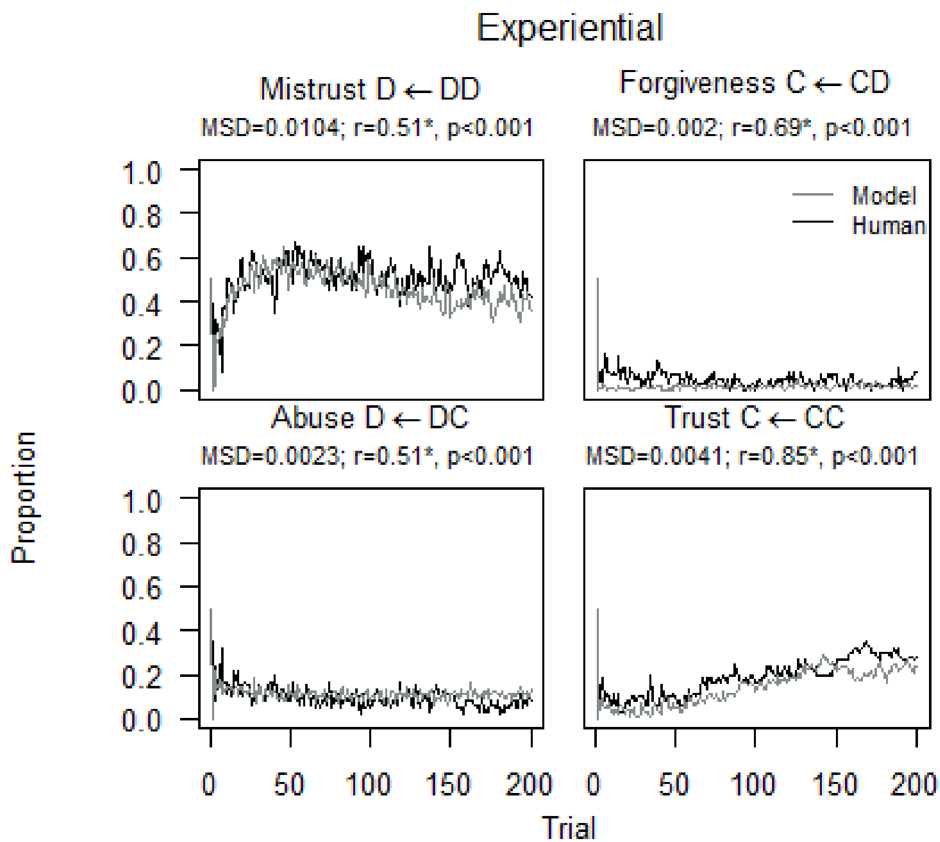


(Lejarraga, Dutt & Gonzalez, 2012)



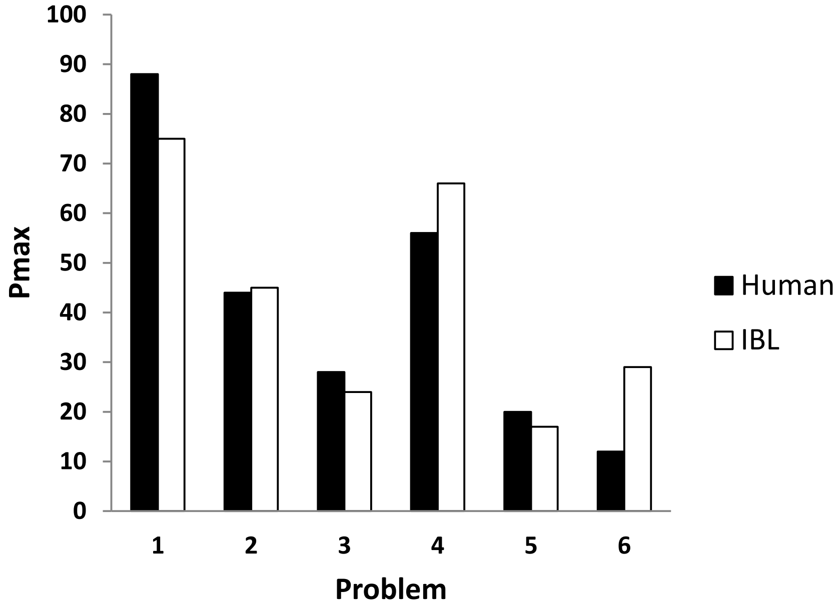
(Lejarraga, Dutt & Gonzalez, 2012)



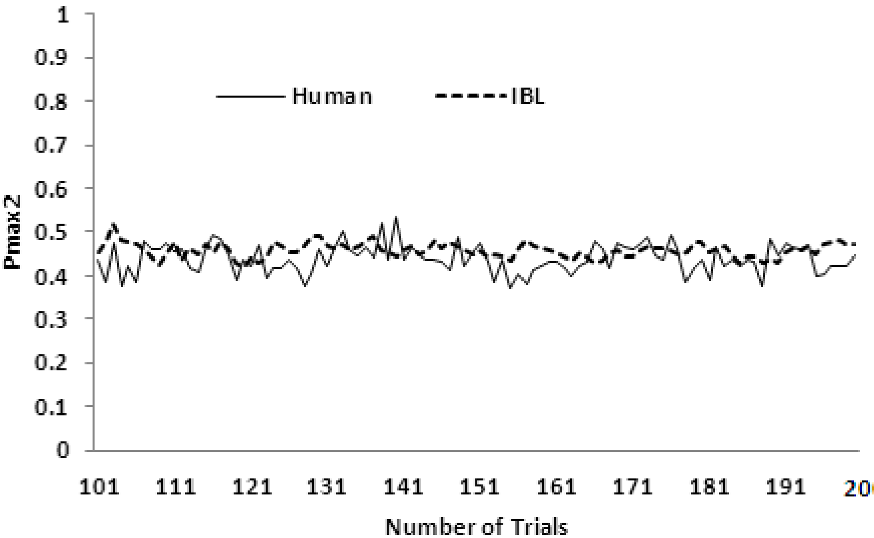


(Gonzalez, Ben-Asher, Martin & Dutt, 2014)

Pmax at final choice in sampling paradigm

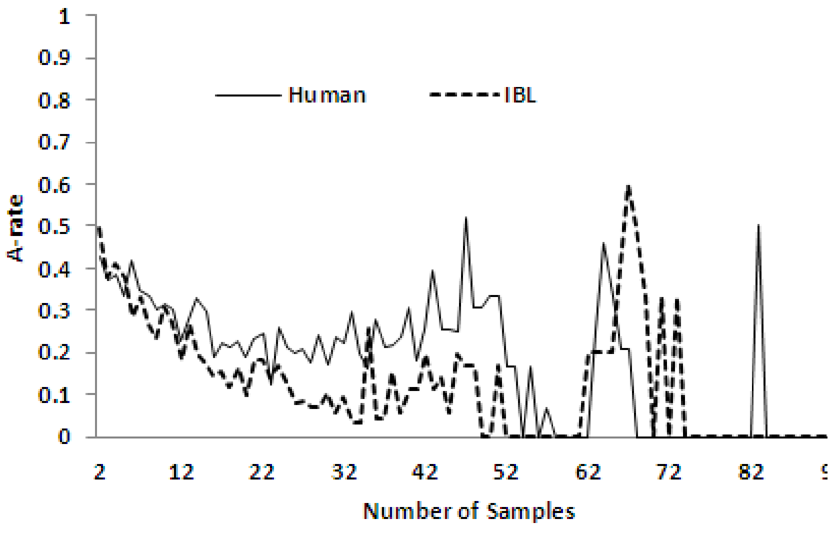


Pmax during repeated consequential choice

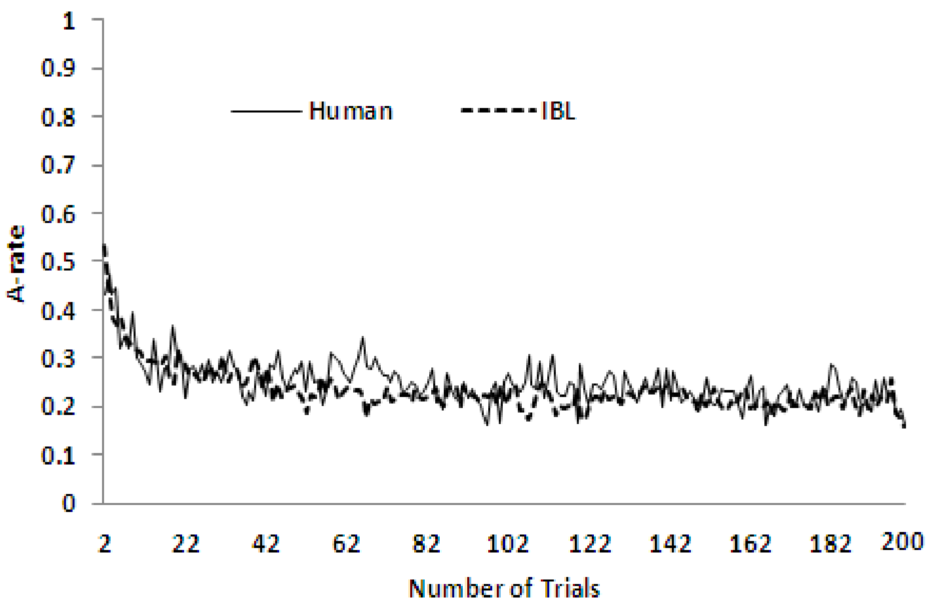


(Gonzalez & Dutt, 2011)

A-rate during sampling



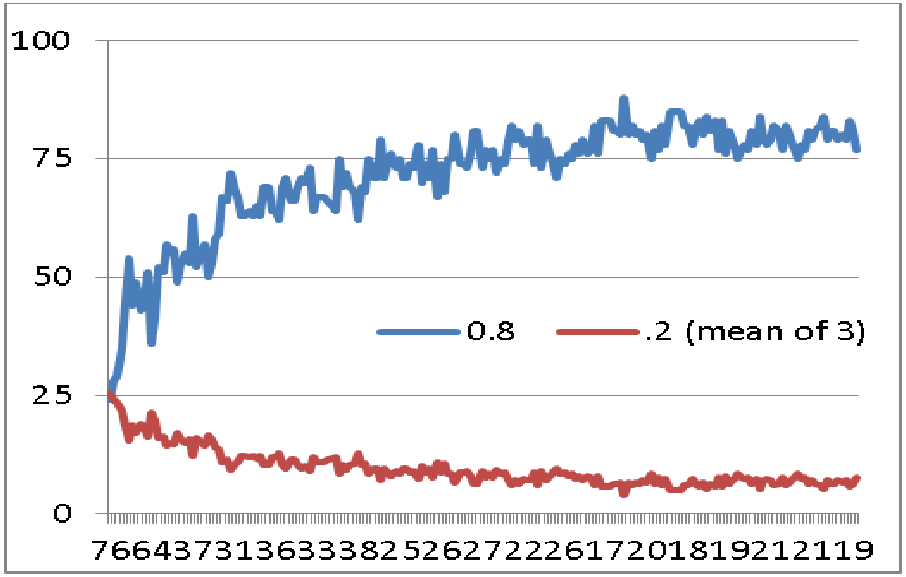
A-rate during repeated Choice

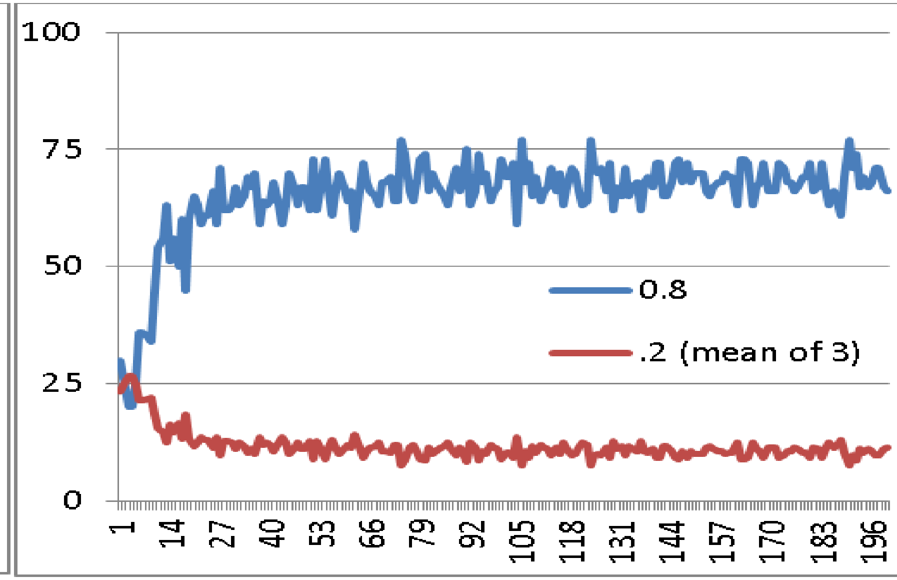


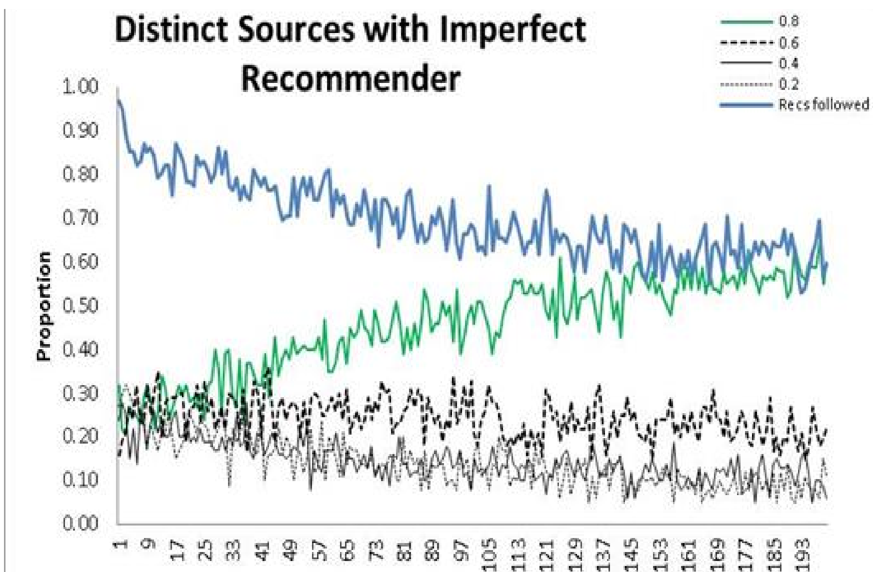
(Gonzalez & Dutt, 2011)

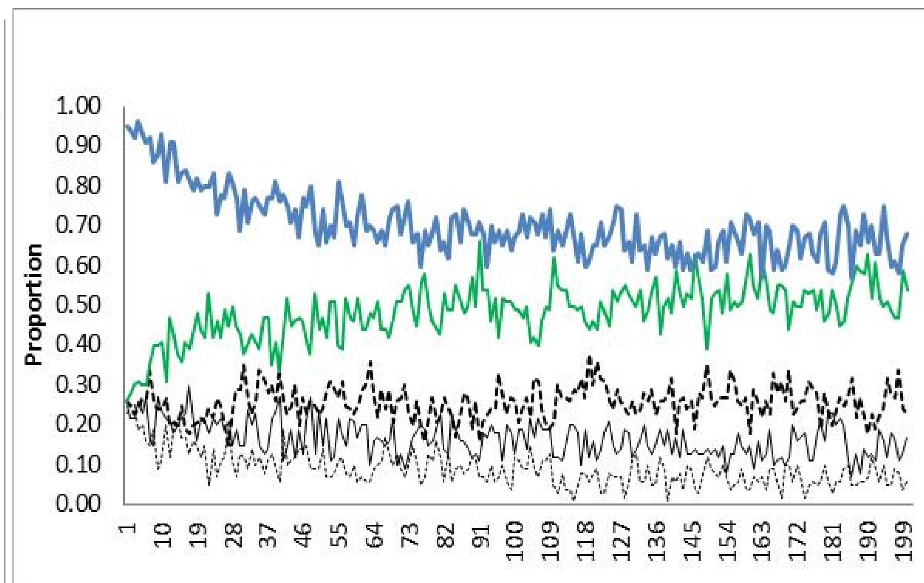
IBL Model predictions

Observed | **IBL MODEL**









(Harman et al., in prep)

Summary of behavioral phenomena

Conditional Reinforcement:

Increasingly select actions that led to best outcomes in similar past experiences

Reduced Exploration:

Decrease exploration of options over time in consistent environments

Recommender systems:

Recommenders may act as distractions for humans' own exploration and search for best value.

Humans abandon imperfect recommenders

IBL model captures human cognitive processes, but there are some challenges:

- Risk tolerance and sequential accumulation of information
- Complex interrelationships of events over time
- Complex similarities among objects
- Feedback delays: processing of cause-effect relationships
- The positive linear causality effect: positive correlations are easier to comprehend than their negative counterparts
- Credit assignment problem: one to one cause-effect relationships

Scaling up IBL models and Experimental Paradigms to increased dynamic complexity

Least Dynamic

Most Dynamic

No changes in the environment although the environment is probabilistic, probabilities and values don't change over the course of decisions

Immediate feedback (Action-Outcome closest in time)

Value is time independent (Time of the decision is determined by the decision maker, no penalty for waiting)

Environment changes (Independently and as a consequence of the actions of the decision maker)

Delayed feedback and Credit assignment problem (Multiple actions and multiple outcomes separated in time)

Value is time-dependent (Value decreases the farther away the decision is from the optimal time)

Simple

Complex

references

Gonzalez, C., et al. (2003). "Instance-based learning in dynamic decision making." Cognitive Science 27(4): 591-635.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors, 46*(3), 449–460.

Gonzalez, C. (2005). The relationship between task workload and cognitive abilities in dynamic decision making. *Human Factors, 47*(1), 92-101.

Gonzalez, C., & Quesada, J. (2003). Learning in dynamic decision making: The recognition process. *Computational and Mathematical Organization Theory, 9*(4), 287–304.

Gonzalez, C., Thomas, R., & Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision making. *Intelligence*, *33*(2), 169–186.

Gonzalez, C. (2005). Decision support for real-time dynamic decision making tasks. *Organizational Behavior and Human Decision Processes*, *96*, 142–154.

Gonzalez, C., et al. (2005). "The use of microworlds to study dynamic decision making." Computers in Human Behavior 21(2): 273-286.

Gonzalez, C., & Thomas, R. (2008). Effects of automatic detection on dynamic decision making. *Journal of Cognitive Engineering and Decision Making*. 2(4), 328-348.

Gonzalez, C. & Madhavan, P. (2011). Diversity during practice enhances detection of novel stimuli. *Journal of Cognitive Psychology.* 23(3), 342-350.

References (cont.)

Gonzalez, C. (2012). Training decisions from experience with decision making games. Adaptive technologies for training and education. P. Durlach and A. M. Lesgold. New York, Cambridge University Press: 167-178.

Gonzalez, C. and V. Dutt (2011). "Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms." Psychological Review 118(4): 523-551.

Gonzalez, C. and V. Dutt (2012). "Refuting data aggregation arguments and how the IBL model stands criticism: A reply to Hills and Hertwig (2012)." Psychological Review 119(4): 893-898.

Barron, G. and I. Erev (2003). "Small feedback-based decisions and their limited correspondence to description-based decisions." Journal of Behavioral Decision Making 16(3): 215-233.

Brunstein, A. & Gonzalez, C. (2011). Preparing for Novelty with Diverse Training. *Applied Cognitive Psychology*. 25(5), 682-691.

Erev, I., et al. (2010). "A choice prediction competition: Choices from experience and from description." Journal of Behavioral Decision Making 23(1): 15-47.

Harman, J., O'Donovan, J., Abdelzaher, T. & Gonzalez, C. (2014). Dynamics of Human Trust in Recommender Systems. The ACM Conference Series on Recommender Systems. RECSYS 2014. October 6th-10th. Foster City, Silicon Valley, USA.

Hertwig, R., et al. (2004). "Decisions from experience and the effect of rare events in risky choice." Psychological Science 15(8): 534-539.

Mehlhorn, K., Ben-Asher, N., Dutt, V. & Gonzalez, C. (In Press). Observed Variability and Values Matter: Towards a Better Understanding of Information Search and Decisions from Experience. *Journal of Behavioral Decision Making.*