A Case of Divergent Predictions Made by Delta and Decay Rule Learning Models

Darrell A. Worthy – *Texas A&M*

Ross Otto – *McGill University*

Astin C. Cornwall – *Texas A&M*

Hilary J. Don – *University of Sydney*

Tyler Davis – *Texas Tech*

Introduction

- Delta and Decay rules are used to update expected value representations (EVs) in learning models.
  - E.g. - what is the value of each option in predicting reward?
- Delta rule – learns **average** reward provided by each option.
- Decay rule – learns **cumulative** reward provided by each option.
- These learning rules usually lead to similar predictions.
- Can diverge if choice options are presented with different frequency.
  - More frequent option may be associated with higher cumulative, but lower average reward.
- Here, we test each model’s predictions in this type of situation.

Delta Rule

$$Ev_{j(t)} = Ev_{j(t-1)} + \alpha(outcome_{(t)} - Ev_{j(t-1)}) \,.\, I_j$$

where:

α is a **recency** parameter ($0 < a < 1$)

outcome: 1 if reward, 0 if no reward

I_j: 1 if option j chosen, 0 otherwise

Decay Rule

$$Ev_{j(t)} = Ev_{j(t-1)} \, . \, a + outcome_{(t)} \, . \, I_j$$

where:

a is a **decay** rate parameter ($0 < a < 1$)

outcome: 1 if reward, 0 if no reward

I_j: 1 if option j chosen, 0 otherwise

Choice rule

The predicted probability that option j will be chosen on trial t is calculated using a Softmax rule:

$$P|C_{jt}| = e^{\theta(t).Evj(t)} / (e^{\theta(t).Evj(t)} + e^{\theta(t).Ek(t)})$$

where:

$$\theta(t) = 3c - 1$$

c is an inverse temperature parameter that determines how consistently the option with higher EV is selected ($0 < c < 5$)

when c = 0, choices are random

when c = 5 choices are deterministic

Binary choice task

Similar to Estes (1976)

TRAINING:

	Frequent trials (x100)			Infrequent trials (x50)		
	A	vs	B	C	vs	D
p(reward):	**.65**		.35	**.75**		.25

TEST:

Critical transfer trial

	A	vs	C
p(reward):	LOWER		HIGHER
Freq(reward):	HIGHER		LOWER

MODEL PREDICTIONS

Delta Rule
Choose C
(higher probability of reward)

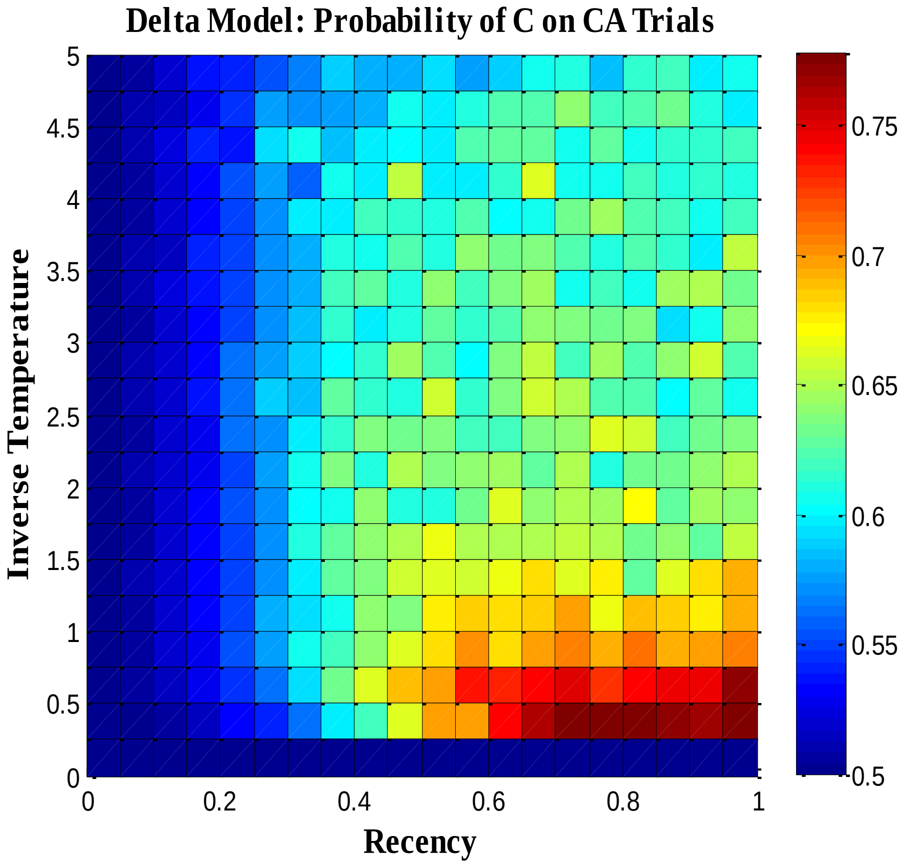
vs.

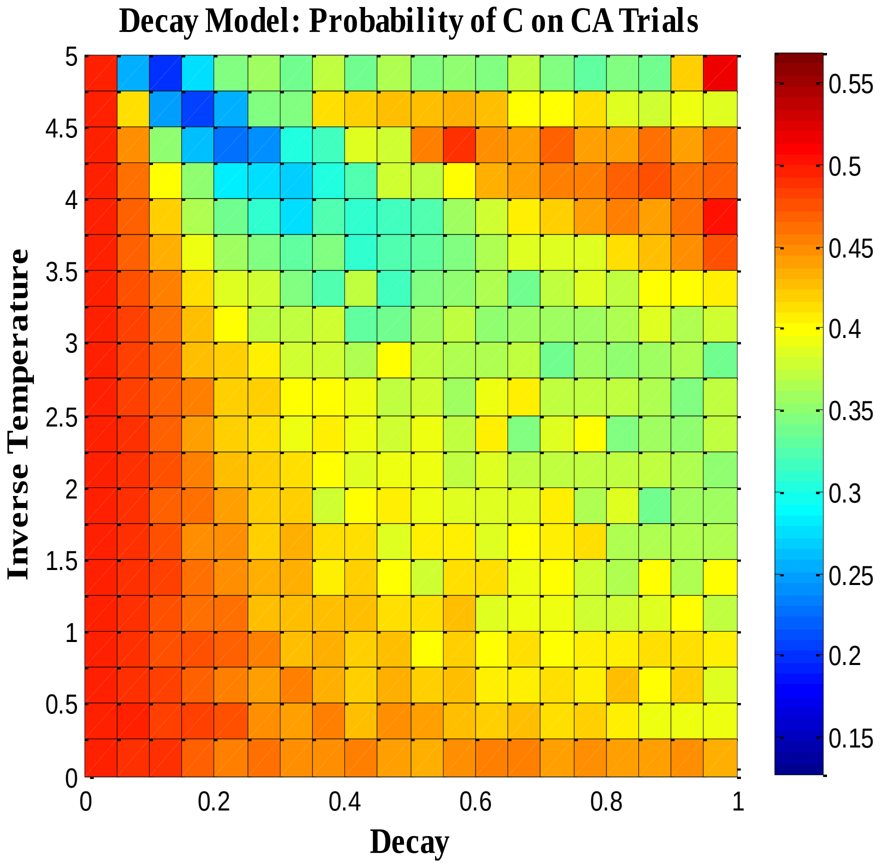
Decay Rule
Choose A
(more frequently rewarded)

Simulations

- To verify this we simulated this task with each model.
- 1000 simulations for over 400 pairs of parameters across the space.
- Parameter values for the recency and decay parameters ranged from 0 to 1 in increments of .05.
- Inverse temperature, c, had increments of .25 from 0 to 5.
- To obtain predictions for the test phase we computed the probability of selecting C on CA test trials following the training phase.

Simulation Results





-Delta model predicts between 50% and ~75% C choices

-Decay model predicts between 15% and ~55% C choices

Experiments

- We conducted three experiments with the same reward structure, but with slight differences (total N=133).

Experiment 1 (n=33)

Training

Test

Experiments 2 & 3 (n=50)

Training

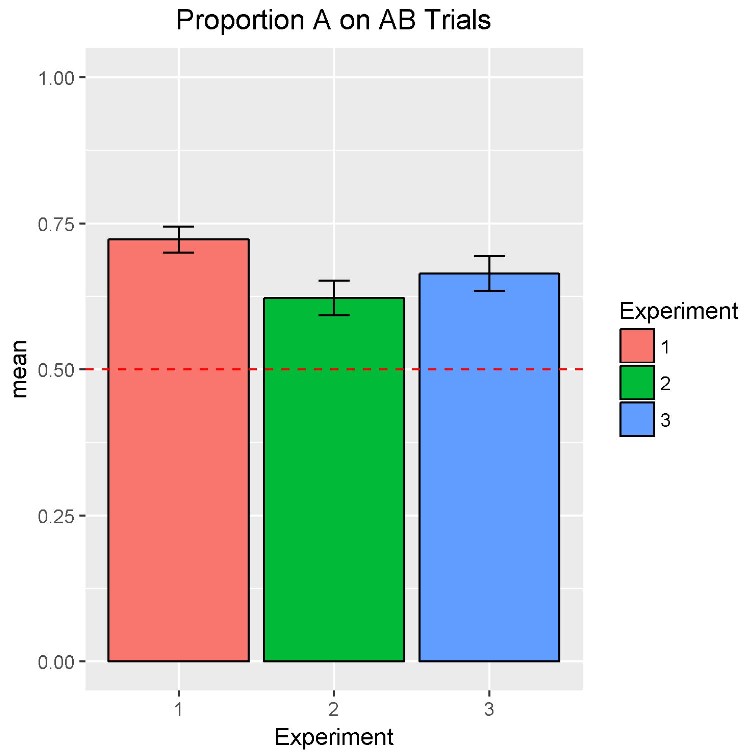
Test

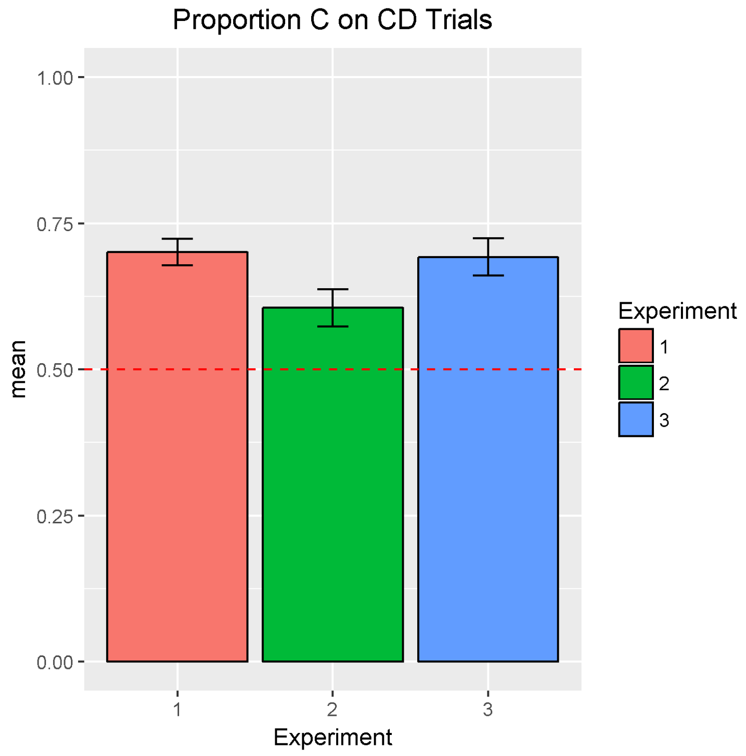
- Exps. 1 & 2 - feedback during test
- Exps. 2 & 3 - Ss received monetary payouts for test

Analysis

- Optimal choice proportions for AB and CD trials were compared against chance (.5) to evaluate whether participants learned.
- Choice proportions during the test phase were compared against the objective reward ratio between two options.
    - CA - .75/(.75+.65) = **.5357**
    - CB - .75/(.75+.35) = **.6818**
    - AD - .65/(.65+.25) = **.7222**
    - BD - .35/(.35+.25) = **.5833**
- Used Bayesian one sample t-tests with these values as the comparison metric.
- Fit Delta and Decay models to all data and compared using BIC.

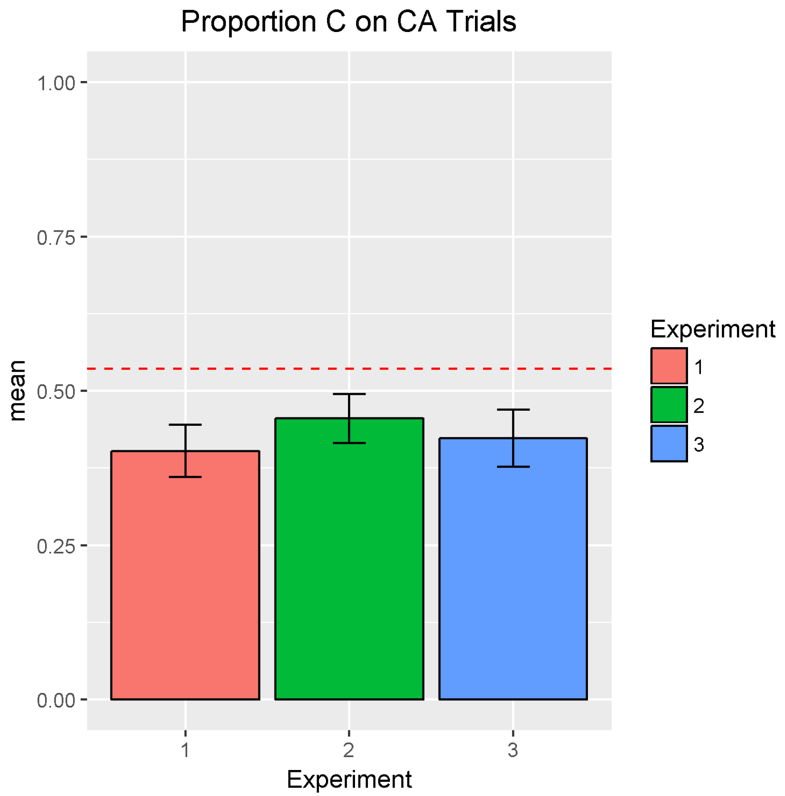
Results – Training Accuracy

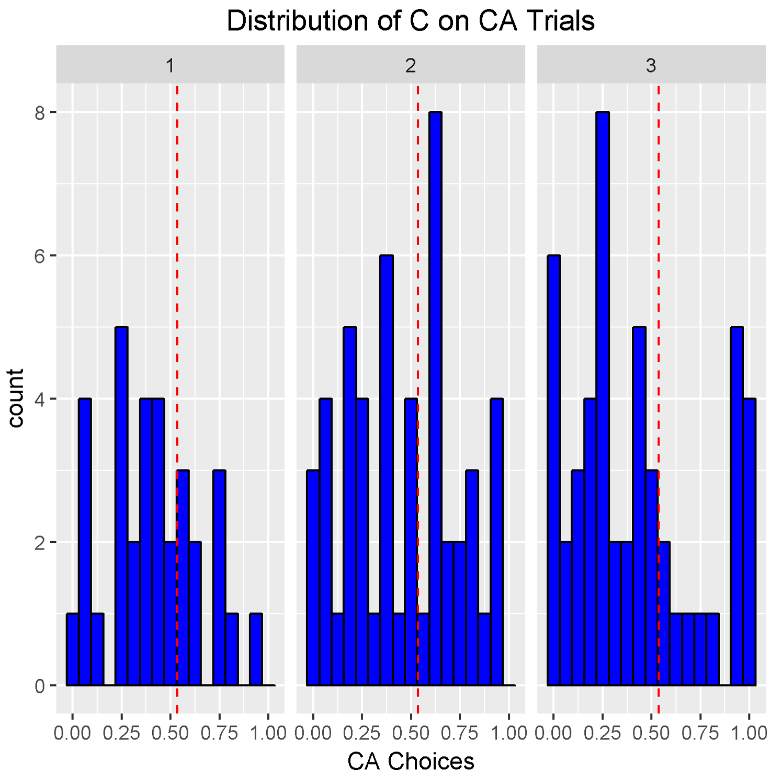




- Training accuracy well above chance for both trials types in each experiment (BF_{10}~10^{13} for AB, BF_{10}~10^{12} for CD).
- BF_{10}=.097 comparing AB versus CD accuracy.

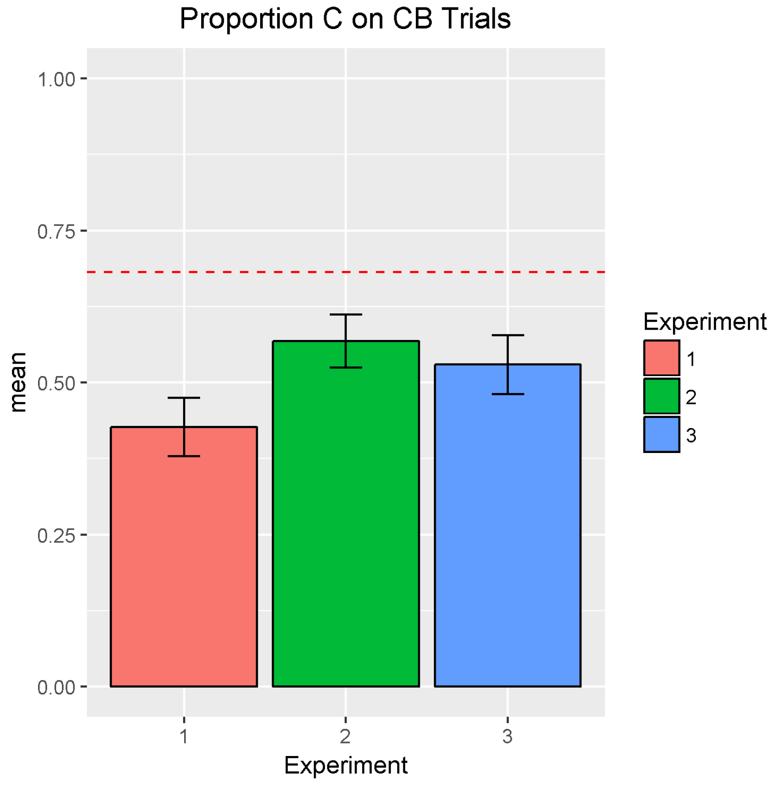
Results - CA Choices

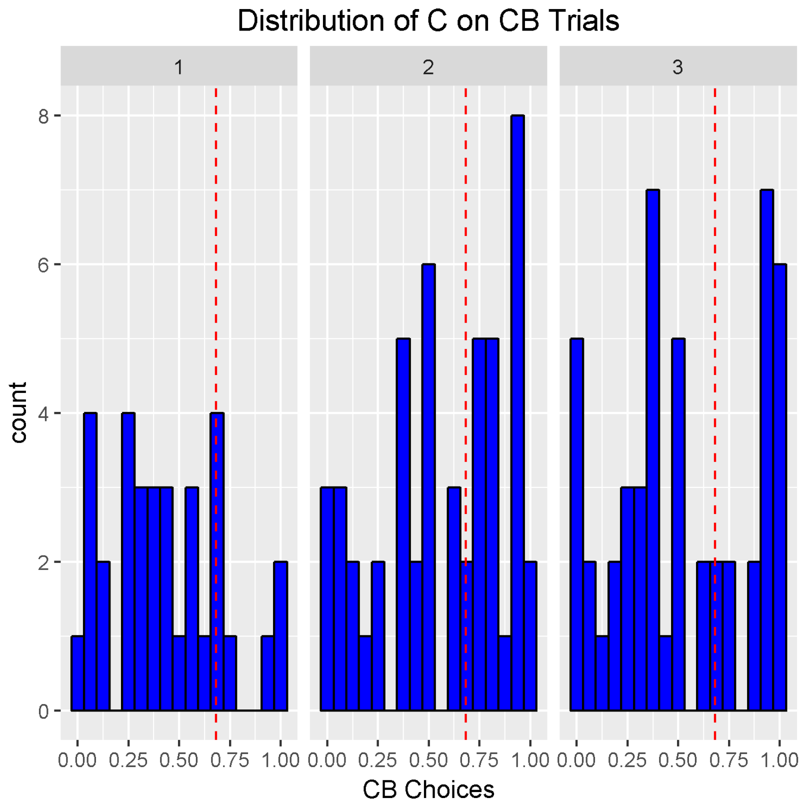




- Participants selected C less often than the reward ratio BF_{10}= 347
- Compared to .5, BF_{10}= 3.96
- Overall Mean = .43, Median = .40

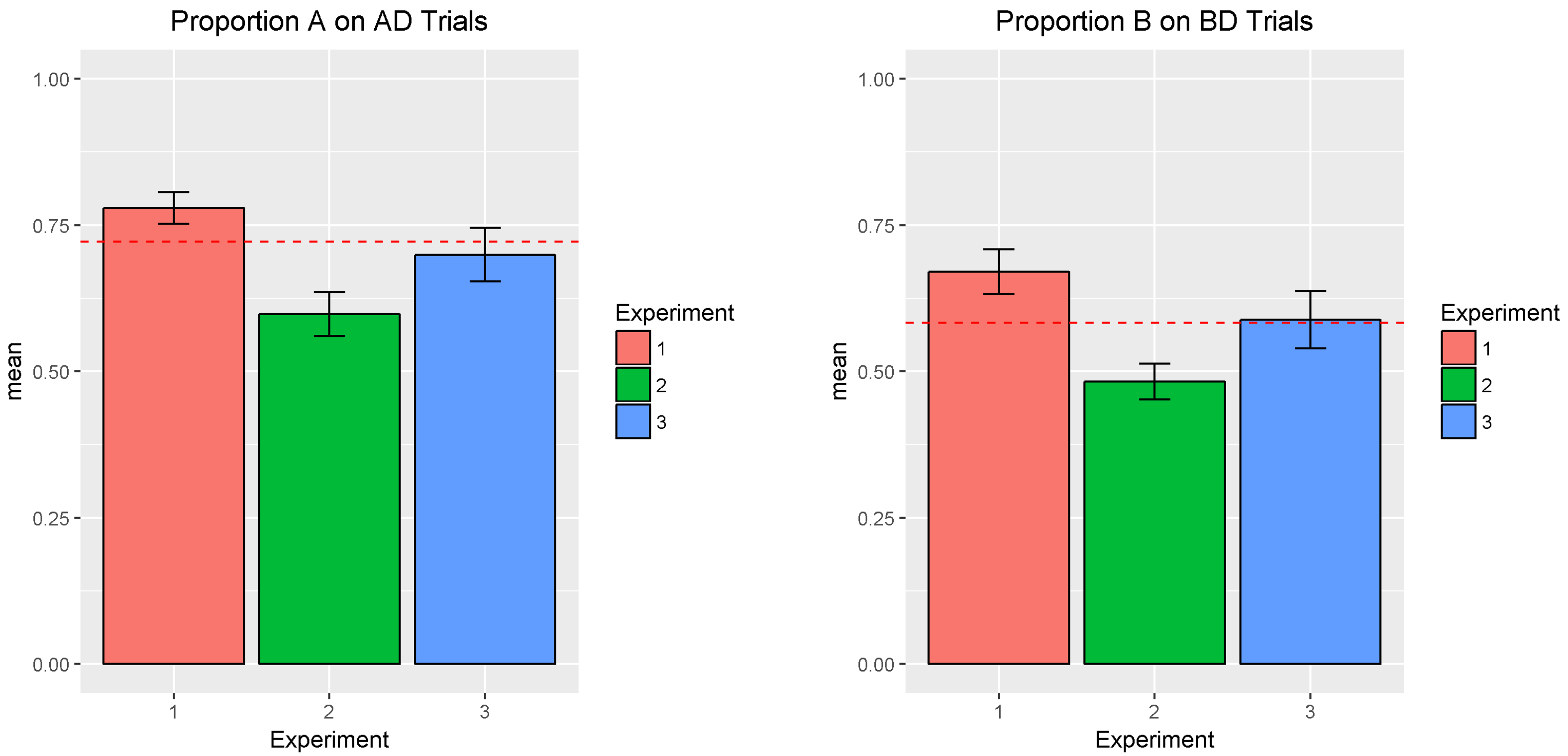
Results – CB choices





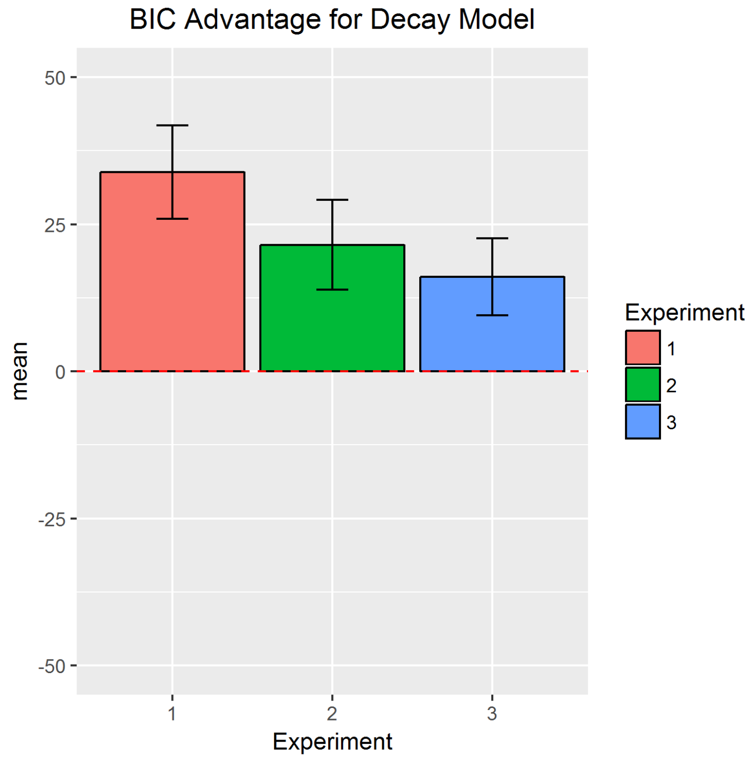
- Participants also selected C less often than the objective reward ratio on CB trials despite a .4 difference in probability of reward, BF_{10}= 426,603
- Decay model predicts this effect better than the Delta model, but may be some link between B and A that the model does not account for.

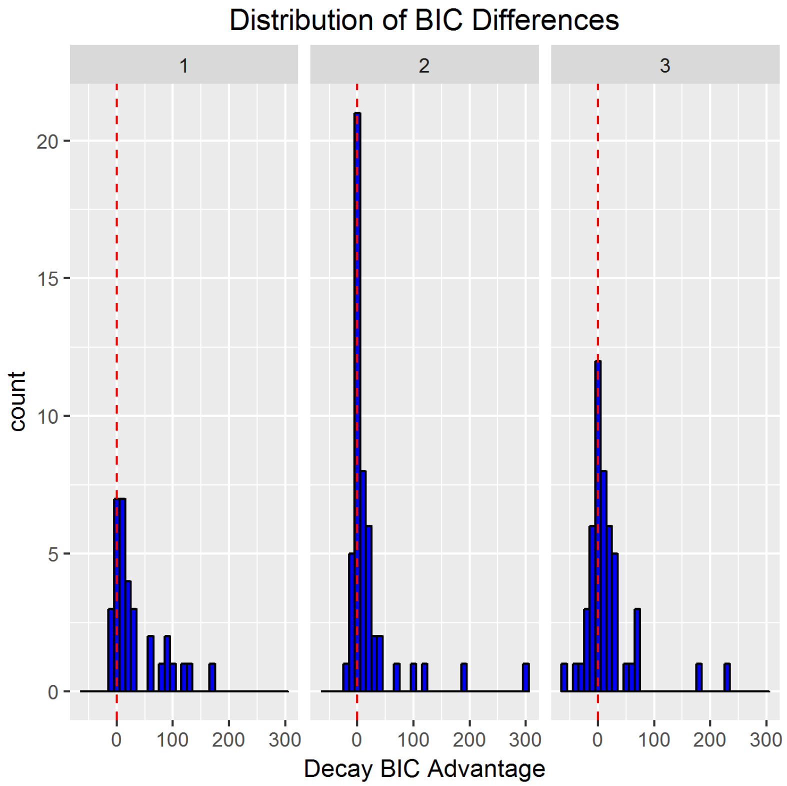
Results – AD and BD Trials



- A and B presented more frequently; choice proportions consistent with objective reward ratios.
- Exp. 2 had the poorest learning, less consistent results overall

BIC Comparison – (BIC_{Delta} – BIC_{Decay})





- Average BIC difference was 22.51, corresponds to a BF of 77K in favor of Decay model (Wagenmakers, 2007).
- 72.9% of subjects best fit by Decay model, BF_{10}= 195,511
- Decay model's advantage negatively correlated with C choices (r=-.28, BF_{10}= 19.2)

Discussion

- Results were more in line with Decay rule predictions.
- Consistent with theories that people do not learn probabilities directly, but translate memory representations into probability judgments (Estes, 1976).
- The Decay model could be a simplification of Instance based models (IBL) where instances of past events are blended into value representations (Gonzalez & Dutt, 2011; Stewart et al. 2006).
  - However, the IBL blends instances within each option so it could be more similar to the Delta model
  - Currently identifying situations where Instance-based models diverge from models where EV is iteratively updated.

Discussion

- Decay model needs more tests and improved generalizability:
  - Probability learning tasks where losses are given.
  - Continuous reward tasks.
  - Dynamic tasks where state information or eligibility traces need to be incorporated to predict long-term reward.
- Developing a categorical decay model that tracks frequency of positive and negative prediction errors.
- Developing a Bayesian version of the Decay rule model where Kalman gain replaces the decay rate (Gershman, 2015).
- Key point here is that EV representations may align more with cumulative reward than average reward.

Acknowledgements

Thanks to our research assistants – Shannon Yap, Tuyet Linh Huynh, Sumedha Rao, Kirsten Downs, Ashton Wilson, Lilian Garza, Josh Philip, Mikayla Herman, Samantha Rumann, Kaila Powell, Kavyapriya Murali, Kinsey Blackburn, Shannon Pavloske, Marena De-Angelis, Catherine Lee, Melissa Hernandez, Tiffany Dobry, Xavier Jefferson,

Lab manager: Kaitlyn McCauley

- Data (Exp. 1) and simulation code: osf.io/t85n6/
- Worthylab.org
- worthyda@tamu.edu
- @worthy_da

Delta and Decay Rules

- The Delta rule model is a simple RL model that has become a default model fit to data from bandit tasks.
  - Tasks where people choose from multiple options and must learn which options are most rewarding.
- Delta-based learning is common in models of **reinforcement learning** (RL; Sutton & Barto 1998; 1981; Jacobs, 1988, Busemeyer & Stout, 2002), **adaptive systems** (Widrow & Huff, 1962), **connectionism** (Rumelhart & McClelland, 1986), and **animal behavior** (Rescorla & Wagner, 1972).
- The Decay rule has been less widely adopted.
- Erev and Roth (1998) developed the current version, and it has been a component of the Prospect Valence Learning model (Ahn et al., 2008); also similar to linear operator models (e.g. Bush & Mosteller 1955)

Delta and Decay Rules

- The expected value (EV) for each *j* option is updated by the Delta rule on each trial (*t)* as:
  - $EV_j(t) = EV_j(t) + \alpha \cdot (r(t) - EV_j(t)) \cdot I_j$
- EV for the chosen option is incremented by the prediction error (PE).
- EVs converge to the *average* reward given by each option
- The Decay rule updates the EV for each *j* option according to:
  - $EV_j(t) = EV_j(t) \cdot A + r(t) \cdot I_j$
- All EVs decay; EV for the chosen option is incremented by the reward.
- EVs are decaying representations of *cumulative* reward for each option.

Delta versus Decay Rule

- The key difference between these two models is that the Delta rule model learns average rewards, while the Decay rule model learns cumulative rewards, both weighted more by recent outcomes.
- The Decay rule model's lack of a prediction error, and its assumption that all EVs decay is responsible for these differences.
- The Decay rule model often leads to better fits, but poorer generalization (Steingroever,et al., 2014; Ahn et al., 2008).
- The models have not been evaluated in tasks where average and cumulative reward differ for each option.
- Here were manipulate the base rate at which options are presented to provide to test whether human behavior is more in line with EVs represented as average (Delta) versus cumulative (Decay) reward.

Manipulating Base Rates

- Inspiration for this study comes from Estes' 1976 paper: On the Cognitive Side of Probability Learning.
- An example of a probability learning task is choosing between an option A that provides a reward with p=.65, versus B with p=.35.
- Estes addressed whether people learn probabilities of reward *per se*, or whether people store memories of instances when they've been rewarded for selecting each option, and translate these memories into probability judgments.
- This distinction between learning reward probabilities versus translating memories of rewarding events into reward probabilities maps onto the mechanics of Delta and Decay models, respectively.

Manipulating Base Rates

- As in Estes' study we manipulate the base rate at which options are presented to decompose average and cumulative reward for each option.
- Four options labeled A-D are learned in pairs - AB on some trials and CD on others, randomly interspersed across training.
- A-D are associated with reward probabilities of [.65; .35; .75; .25].
- C is the most rewarding, and A is dominant over B
- However, over 150 training trials there are 100 AB trials but only 50 CD trials.
    - C gives the highest reward on average while A provides the most total reward.

Manipulating Base Rates

- This manipulation of base rates (twice as many AB as CD trials) should not affect the Delta rule model's predicted EVs
  - They will correspond with the actual reward probabilities [.65; .35; .75; .25]
- However, because the Decay rule model learns cumulative rewards, or a decaying memory of total reward associated with each option, the EV for option A should be the highest, notably higher than option C.
- A 100 trial test phase can be given with novel pairs – CA, CB, AD, BD
- CA trials are of most interest because the Delta model predicts more C choices, but the Decay model predicts more A choices.