Disentangling the influence of positive and negative incentives on cognitive effort

ROBERT J. & NANCY D. CARNEY INSTITUTE FOR BRAIN SCIENCE
BROWN UNIVERSITY

Xiamin (Jason) Leng, Debbie Yee, Amitai Shenhav

Cognitive, Linguistic & Psychological Sciences, Carney Institute for Brain Science
Brown University

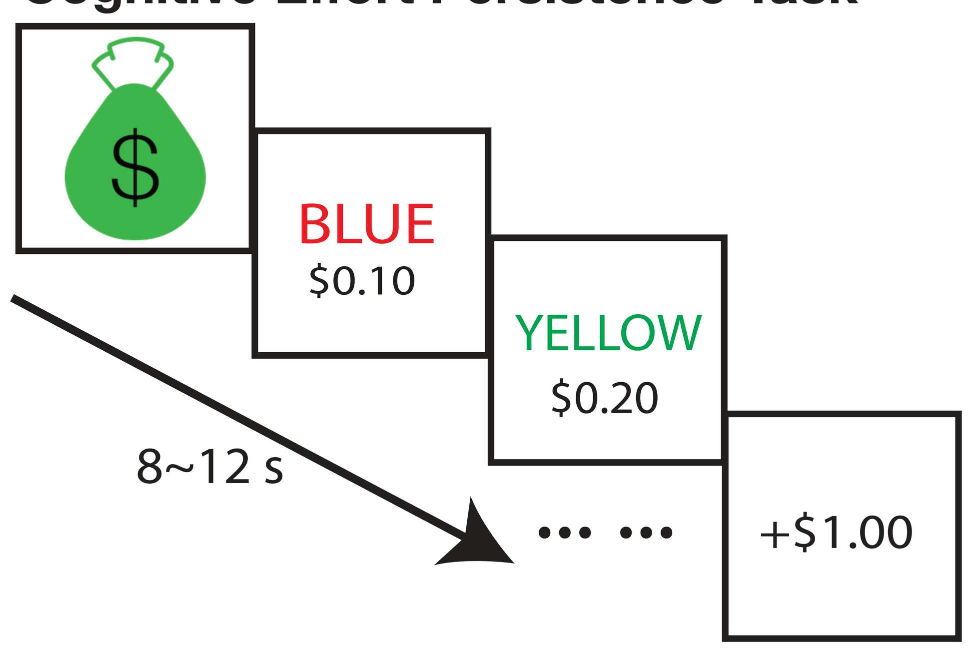
SHENHAV LAB

Introduction

- A large body of research has examined the influence of how potential incentives influence cognitively effortful task performance[1, 2]. These studies have typically been constrained by fixed time and/or number of trials.
- Different incentives motivate behavior differently[3, 4]. Relatively little work has explored how cognitive effort differs based on positive incentives (e.g., potential gains) vs. different kinds of negative incentives (e.g., potential losses avoided vs. punishment).
- We developed a novel **Cognitive Effort Persistence Task** to investigate both sets of questions. In this task, participants are given fixed intervals of time (e.g., 10s) to complete as many trials as they want of a Stroop task.
- We fit drift diffusion models[5] to investigate the influence of positive and negative incentives on cognitive effort strategies.

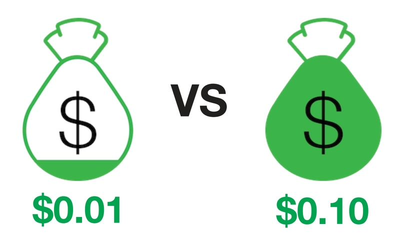
Methods

Cognitive Effort Persistence Task



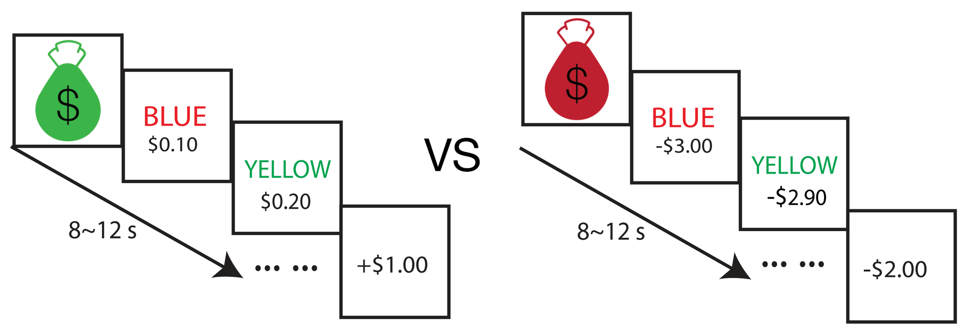
- Within each interval, participants completed as many trials as they wanted of a Stroop task. They were paid based on the number of correct responses.
- Intervals varied in duration and incentives.
- Data were collected on Amazon's Mechanical

Study 1: High vs Low Rewards



- Reward accrued for each correct response

Study 2: Gain vs Loss Avoidance



In loss avoidance condition:

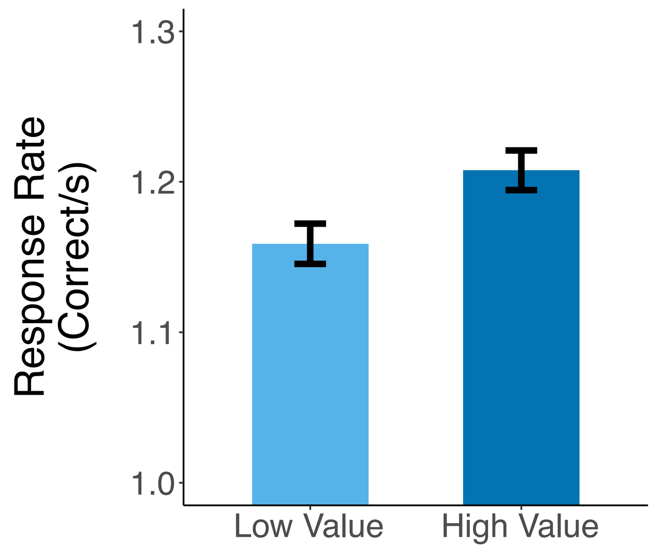
- At the start of each interval, participants face a potential loss from an initial endowment.
- Each correct response reduces the potential loss.

Study 3: Mixed Rewards and Penalties

✓ Reward for correct responses.

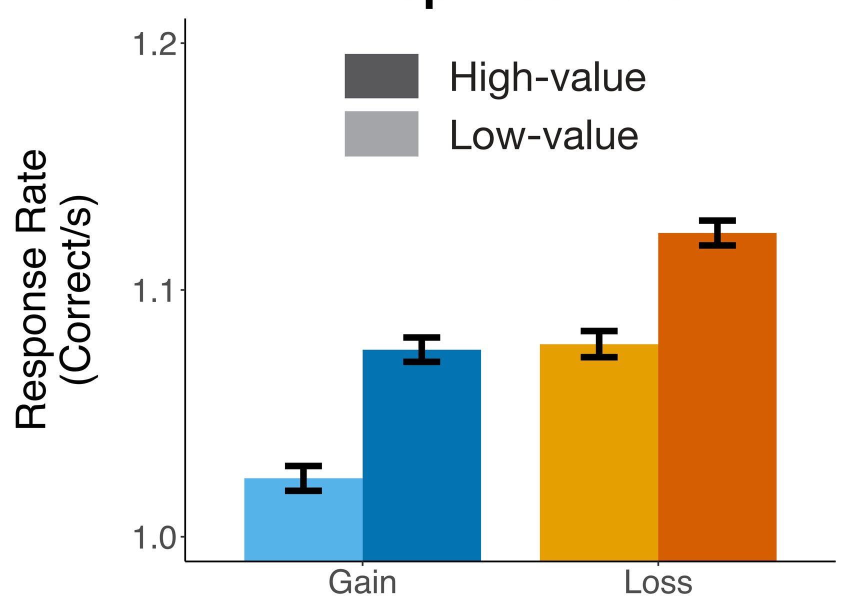
✗ Penalty for incorrect responses.

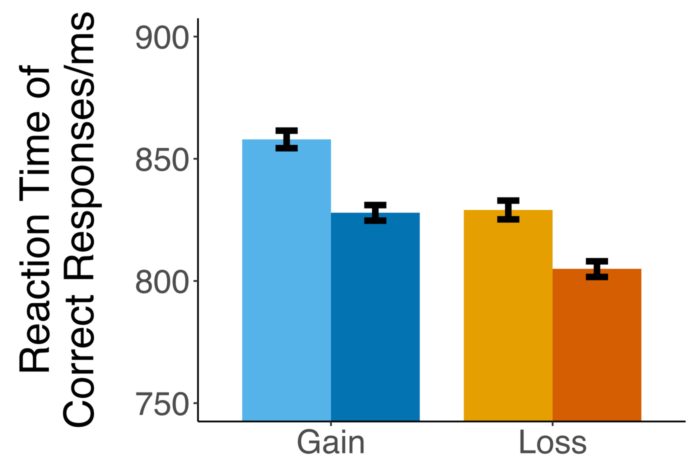
Study 1: High vs Low Rewards (n = 61)

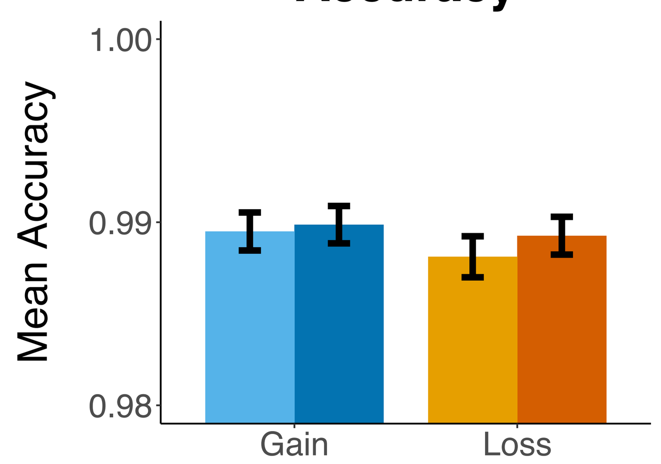


- Ps completed more correct responses per second in high-value intervals ($p < 0.001$).
- These differences were reflected in shorter RTs for high-value intervals ($p < 0.001$), without concomitant changes in accuracy.
- A follow-up study showed that higher potential rewards led to faster trial initiation ($p = 0.012$) and faster responding to the Stroop stimulus ($p < 0.001$).

Study 2: Gain vs Loss Avoidance (n = 91)

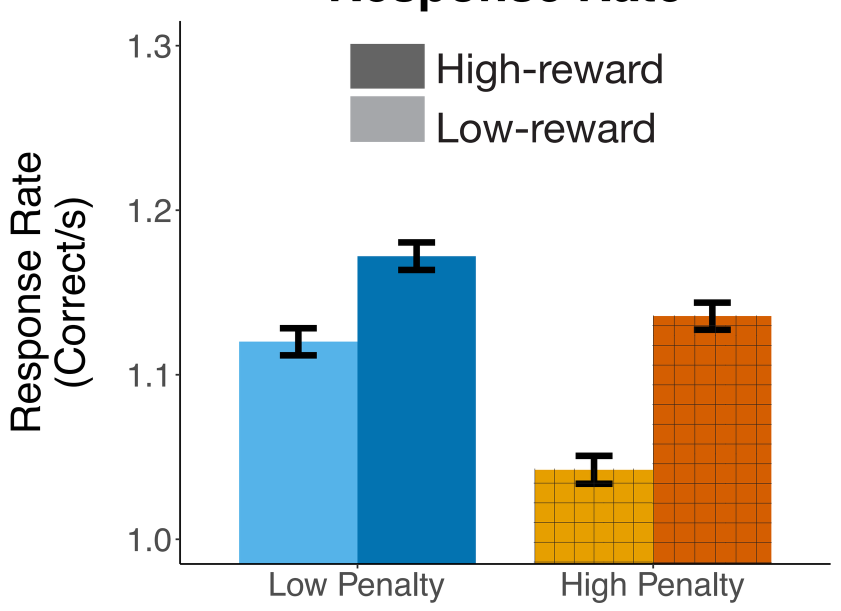


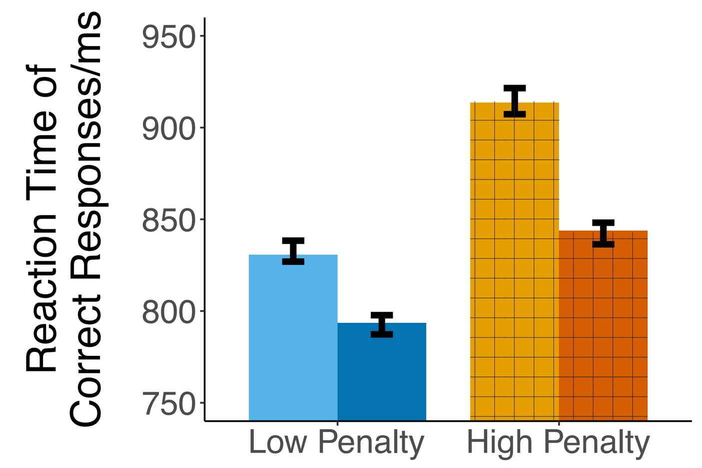


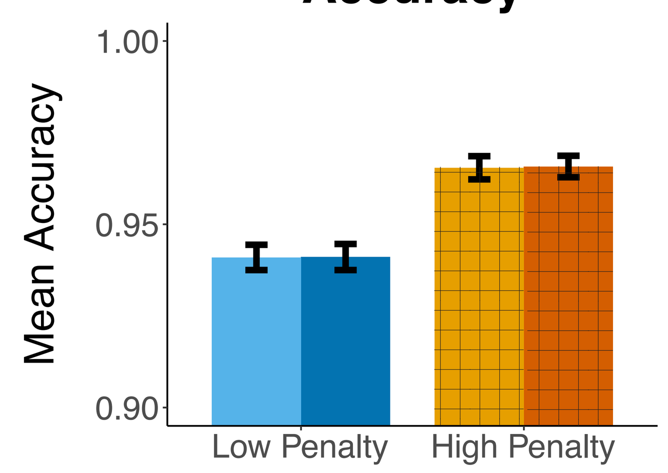


- Ps completed more correct trials when intervals were higher in value ($p < 0.001$) and when they were avoiding losses rather than gaining rewards ($p = 0.003$).
- These were reflected in differences in RT but not accuracy with incentive magnitude (large vs. small) and type (loss vs. gain).
- There is order effect when loss-avoiding conditions are after gain conditions.

Study 3: Mixed Rewards and Penalties (n = 32)







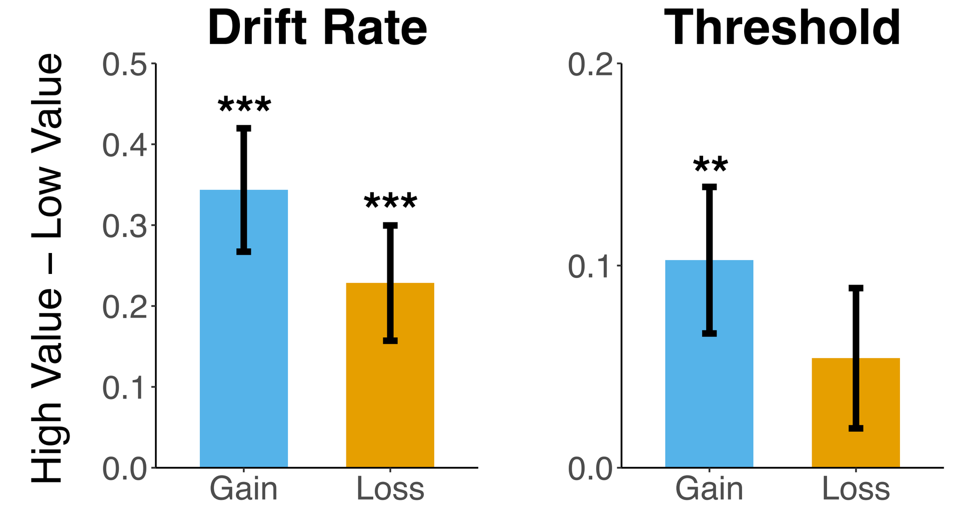
- As the reward for a correct response increased, Ps were again faster but not more accurate.
- As the penalty for an error increased, Ps were slower and more accurate (ps < 0.001).
- Performance was more sensitive to reward level in the high-penalty condition ($p=0.038$).

Drift Diffusion Model

Gain vs Loss Avoidance (Study 2):

In gain-pursuing conditions, both drift rate ($p < 0.001$) and threshold ($p=0.002$) increase signficiantly when higher in value.

In loss-avoiding conditions, only drift rate increases significantly when higher in value ($p < 0.001$).

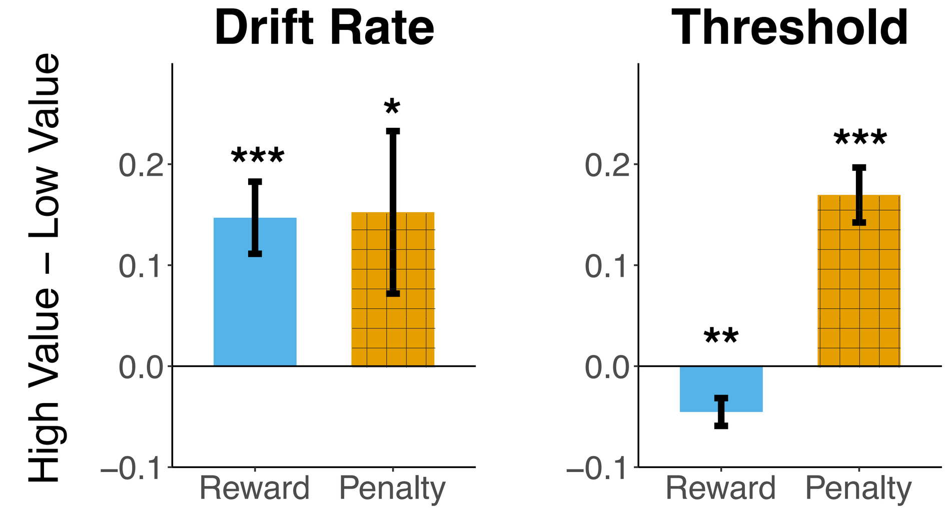


Mixed Rewards and Penalties (Study 3):

Drift rate increases signficantly with higher reward ($p < 0.001$).

Threshold decreases significantly with higher reward ($p = 0.0016$). This is different with Study 2 in which there is no penalty for errors.

Dissociated with reward, higher level of penalty primarily increases the threshold ($p < 0.001$) compared to drift rate ($p = 0.024$).



Conclusion

- The potential for greater reward and greater loss avoidance led to increased effort (i.e., response rate). The potential for greater penalty leads to increased caution.
- The motivating effects of negative incentives depend on their target (e.g., loss avoidance for correct vs. penalty for error).
- Drift diffusion models revealed different strategy in cognitive effort exertion under different types of negative incentives. In loss-avoiding conditions, value level primarily affects drift rate, while in conditions with penalties, the level of penalty primarily affects threshold.

References

1. Chong, T.T.-J., Bonnelle, V., Husain, M. (2016). Quantifying motivation with effort-based decision-making paradignms in health and disease. Progress in Brain Research. 299, 71-100.
2. Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. Annual review of psychology, 66, 83-113.
3. Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. Psychological Review, 64(6p1), 359.
4. Collins, A. G., & Frank, M. J. (2014). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychological review, 121(3), 337.
5. Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. Frontiers in neuroinformatics, 7, 14.