The impact of stimulus value on goal-directed aversive reinforcement learning

Björn R. Lindström[1,2], Armita Golkar[1,2], Sofie Ährlund-Richter[1], Paulina Wiktor[1], Andreas Olsson[1,2]

[1] Department of Clinical Neuroscience, Karolinska Institutet, Sweden [2] Stockholm Brain Institute

www.emotionlab.se

Introduction

- Classical conditioning studies have shown that fear is most easily acquired and difficult to extinguish to certain biologically relevant classes of stimuli (Pavlovian triggers; 1).

  These include dangerous animals (e.g., snakes), threatening con-specifics (2), and racial out-group members (3).
- We asked how different Pavlovian triggers would impact goal-directed aversive reinforcement learning (RL) across three experiments.

  Prediction: Enhanced RL when the Pavlovian trigger stimuli is most predictive of shocks, and impaired RL when the neutral stimuli is most predictive of shocks, reflecting competition between instrumental and Pavlovian systems.

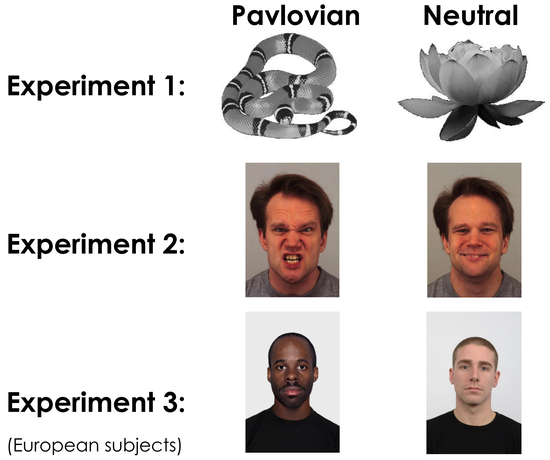
Methods

Probabilistic two-choice shock avoidance task:

Mixed 2 (Group:PavlovianToNeutral/NeutralToPavlovian)

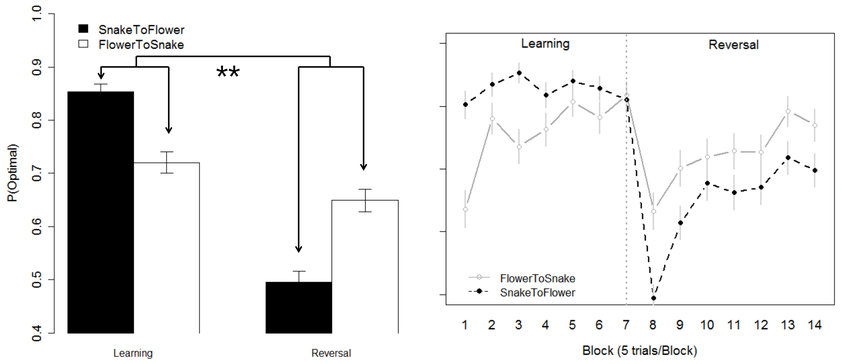
* 2 (Phase: Learning/Reversal) design.

After 35 (of 70) trials, the stimulus – shock contingency switched:
P(Shock | A, Learning) = .25 => P(Shock | A, Reversal)= .75



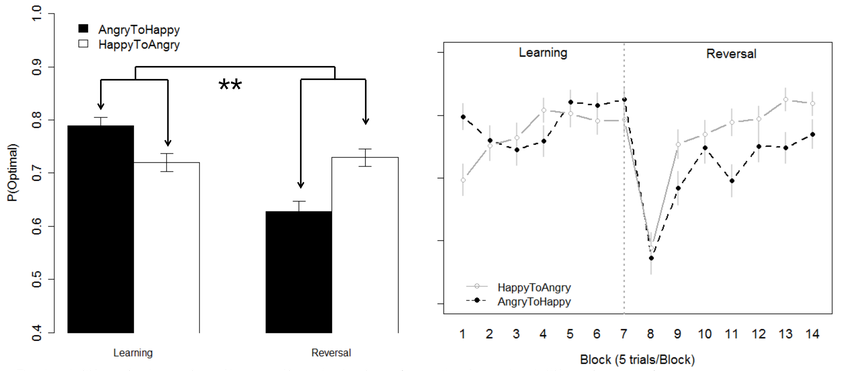
Results

Experiment 1 (n = 32)



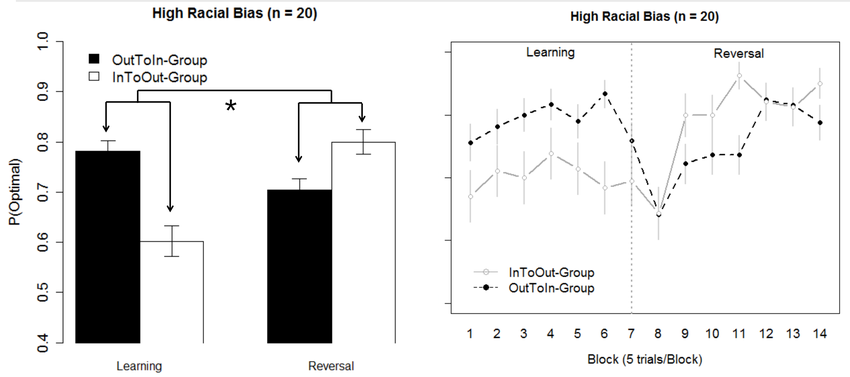
Probability of choosing the optimal choice (smallest probability of shock) by group and condition in Experiment 1. ** p < .01

Experiment 2 (n = 40)



Probability of choosing the optimal choice (smallest probability of shock) by group and condition in Experiment 2. ** p < .01

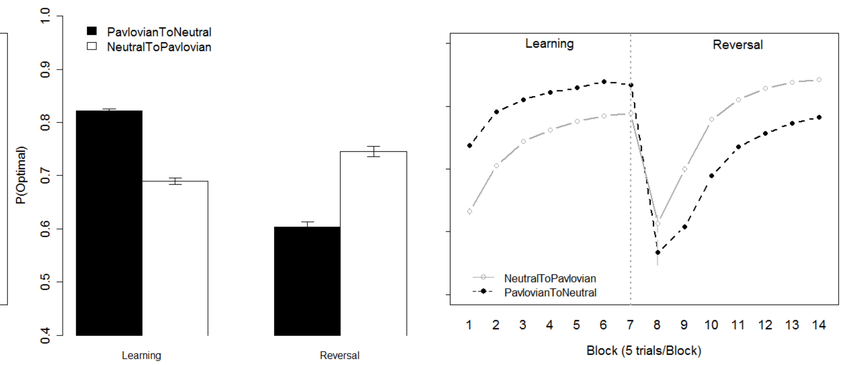
Experiment 3 (n = 48)



Probability of choosing the optimal choice (smallest probability of shock) for high racial bias subjects (mean split on the Modern Racism Scale), group and condition. Low racial bias subjects did not exhibit the interaction.* p < .05

Reinforcement Learning Model

RL Model Predictions



Predictions from the RL model (Pavlovian Value = 0.1, a = 0.25, β= 0.4).

RL Model Implementation

Q-learning model (Exp. 1):

$$Q_{snake}(t+1) = Q_{snake}(t) + \alpha * \delta(t)$$
$$\delta(t) = R(t) - Q_{snake}(t)$$

Softmax decision function:

$$p_{snake}(t) = \frac{\exp((Q_{snake}(t) - Pavlovian_{snake})/\beta)}{\exp((Q_{snake}(t) - Pavlovian_{snake})/\beta) + \exp((Q_{flower}(t))/\beta)}$$

Conclusions

- Confirming the predictions, Pavlovian trigger stimuli had a strong and lasting impact on instrumental behavior. This impact could have important implications for social behavior.
- The results are consistent with a view of aversive Pavlovian triggers as eliciting behavioral inhibition that is transferred to instrumental behavior (Pavlovian to instrumental transfer).
- The RL model provides a parsimonious account of how both enhancement and impairments can result from the congruency or incongruity between an action (i.e., avoiding shocks) and the intrinsic value of Pavlovian triggers.

References

1. LeDoux, J. (2012). Rethinking the Emotional Brain. *Neuron*, *73*(4), 653-676.
2. Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological review*, *108*(3), 483-522.
3. Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science* , *309*(5735), 785-7.