Inferences about Uniqueness in Statistical Learning

Anna Leshinskaya* & Sharon L. Thompson-Schill

University of Pennsylvania

*alesh@sas.upenn.edu

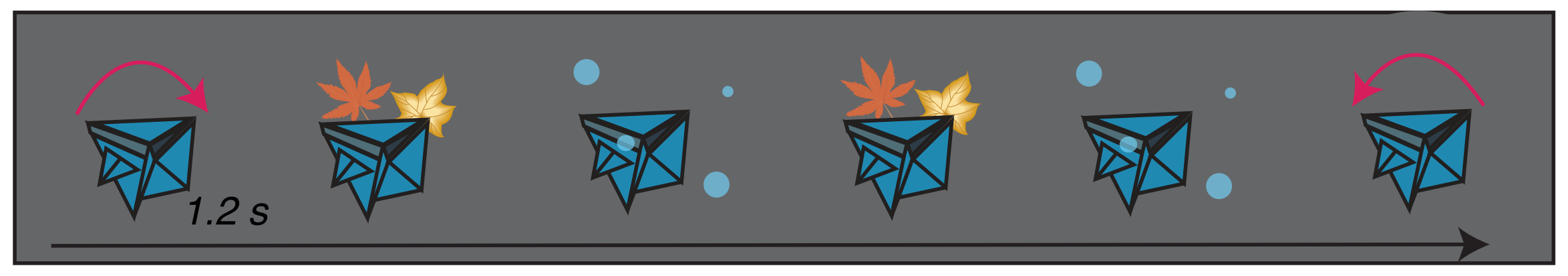
Our minds adeptly register stable patterns in experience. This phenomenon, statistical learning, takes place without conscious effort, feedback, or reward. How inferentially complex is learning under these conditions, and how similar is it to more explicit forms of learning[1]?

Prior work shows that learners do more than register that two stimuli co-occur, but also compute whether they predict each other uniquely and independently, as described by this formula[2,3,4]:

$$\Delta P = P(A|B) - P(A|{\sim}B)$$

Visual statistical learning paradigm

Exposure task: "Decide if the event is common or rare"

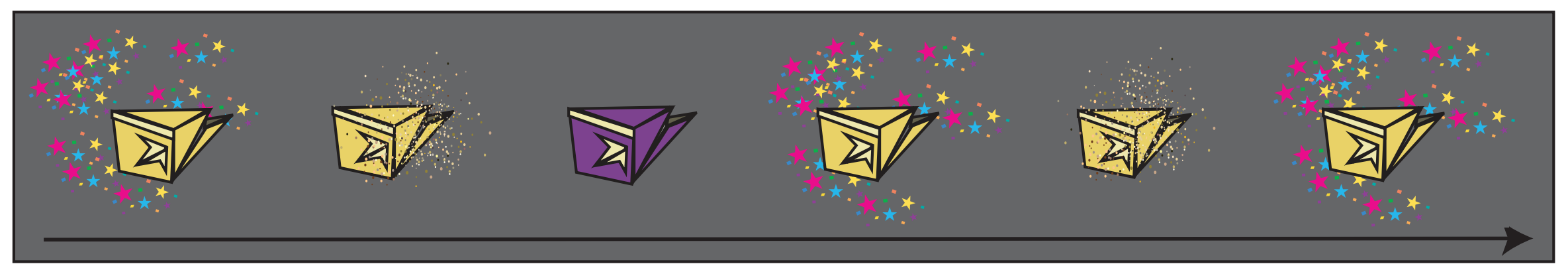


500 events / ~10 minutes over 3 segments

common alternates

rare alternates (replace common events on 10% of occurences, chosen randomly)

Each participant saw 2 sequences, each cued by a distinct object, and showing distinct events

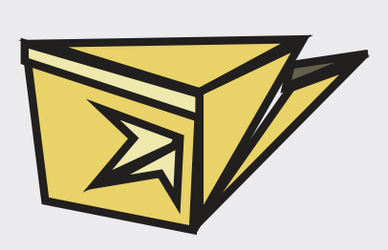


ΔP varied separately from conditional probability

high delta P condition (.97)

trial n+1

		cause	effect	random1	random2	static
trial n+1	**cause**	0.00	**0.98**	0.00	0.00	0.02
	effect	0.07	0.00	0.47	0.45	0.00
	random1	0.15	0.01	0.38	0.41	0.05
	random2	0.16	0.01	0.40	0.38	0.05
	static	0.28	0.01	0.28	0.29	0.14
frequency		0.13	0.14	0.35	0.35	0.04



low delta P condition (.61)

cause	0.00	**0.97**	0.00	0.00	0.03
effect	0.03	0.14	0.41	0.41	0.00
random1	0.27	0.66	0.00	0.00	0.08
random2	0.28	0.65	0.00	0.00	0.08
static	0.27	0.01	0.29	0.28	0.16
frequency	0.13	0.44	0.20	0.20	0.04

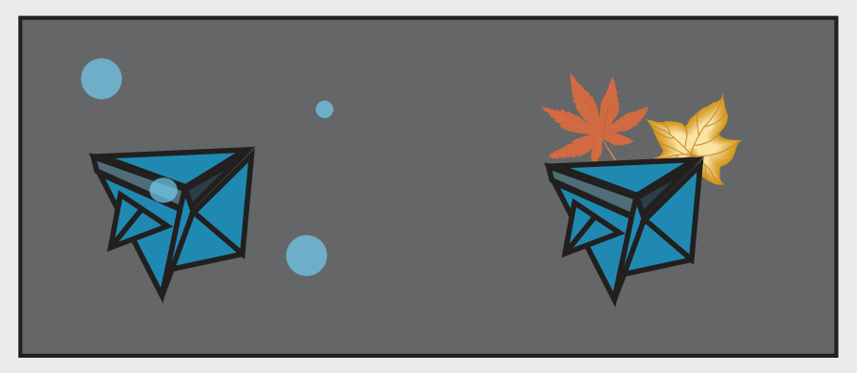
Transition probability structure governing each sequence of events

conditions matched on conditional probability, P(A|B) and joint probability/chunk frequency, P(A&B)

Learning reflects deltaP

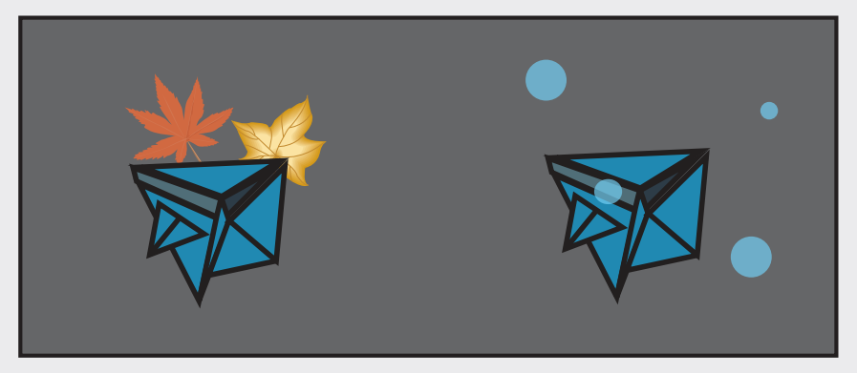
Forced Choice test:

"Which is more typical?"

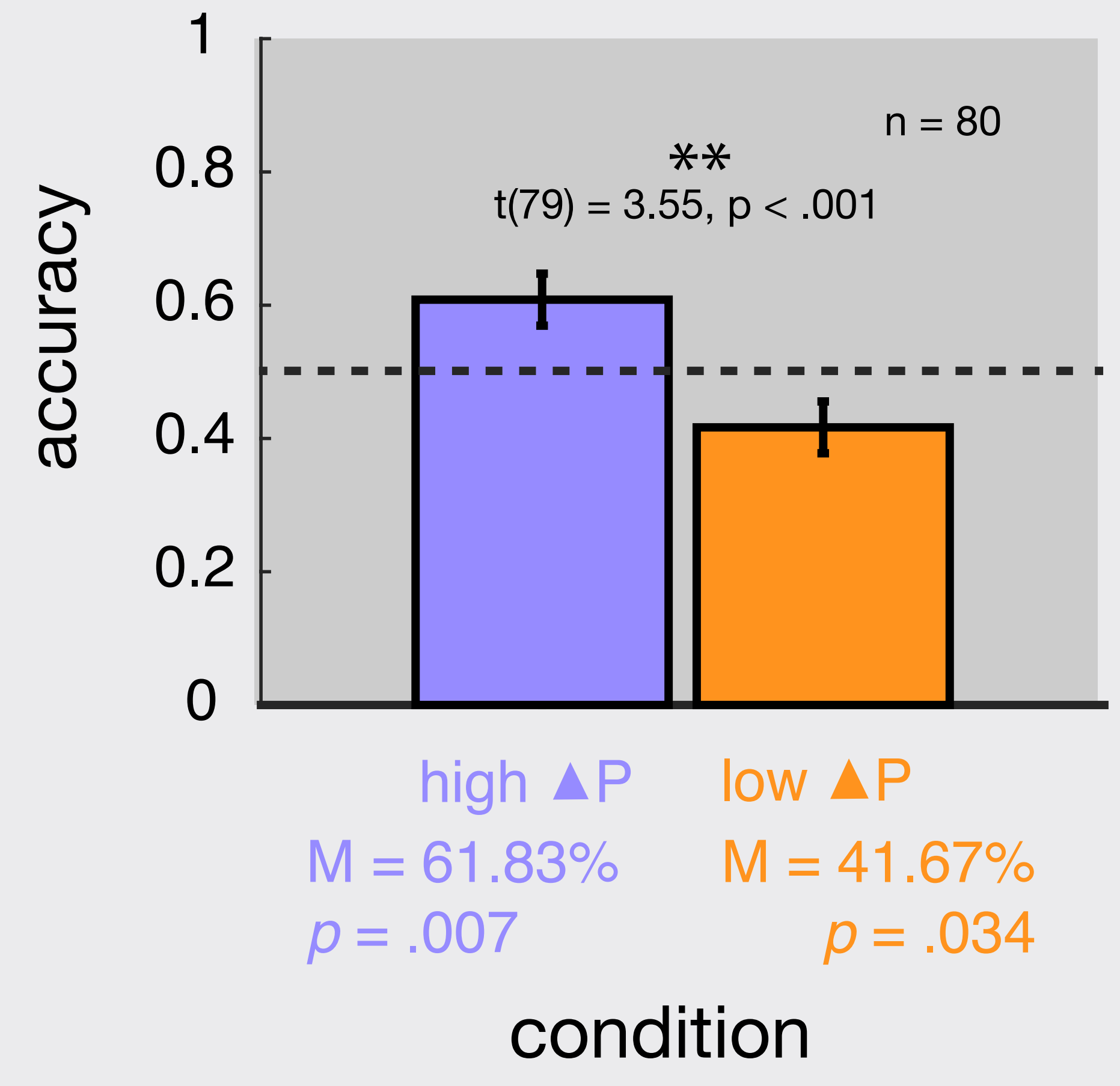


cause-effect

vs.



effect-cause



ΔP predicts noticing

"Did certain events follow each other more often than others? Describe any you noticed for the first set and for the second set of videos."

20/80 participants described a relation between the cause and effect for one sequence, but only 2/80 did so for both.

Participants were more likely to describe it for the high ΔP events (19/80) than the low ΔP events (5/80; Chi² (1) = 9.61, p = .002).

A modified RW learning model explains effects

current prediction strength based on prior n trials

$$a_i = \sum_{k=1}^{n} \boldsymbol{w}_{\mathrm{ki}}$$

error in prediction

$$d_i = 1 - a_i$$

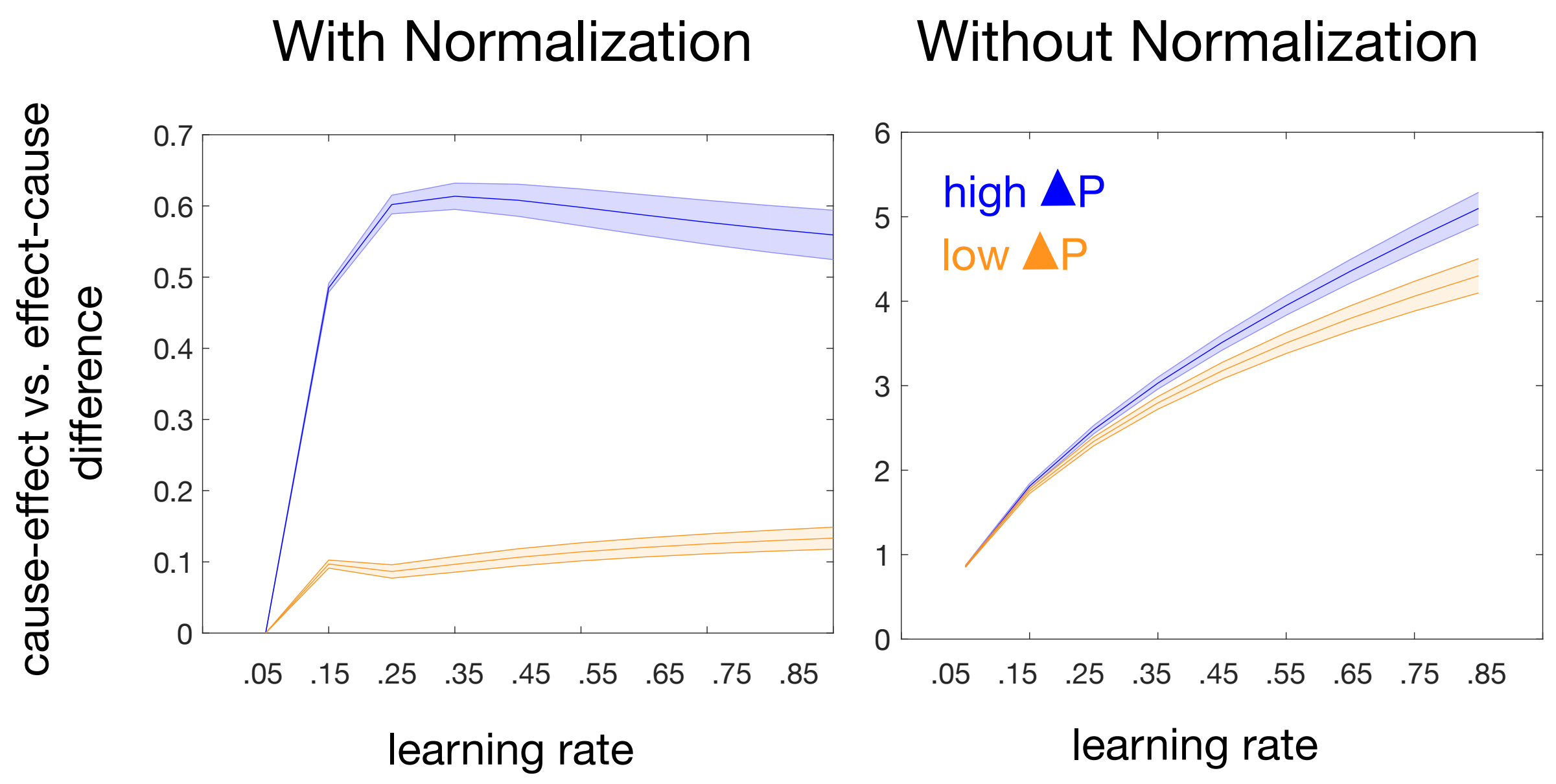
update to weight

$$\Delta\mathbf{W}_{ki} = \alpha\, d_i$$

*Aim is to learn the weights **a** in matrix **W**; weights are updated at each observion **i** with learning rate **alpha.***

0.00	**0.97**	0.00	0.00	0.03
0.03	0.14	0.41	0.41	0.00
0.27	0.66	0.00	0.00	0.08
0.28	0.65	0.00	0.00	0.08
0.27	0.01	0.29	0.28	0.16

With normalization: *After each step, columns are normalized to sum to 1, allowing weights to trade off*



A simple adaptation to a classic model explains the effect of delta P on learning.

We demonstrate that statistical learning is subject to considerations of uniqueness: that learning reflects not just the conditional probability relating two events, but whether that relation is unique. This is despite the incidental, spontaneous and largely implicit nature of such learning.

online materials

https://osf.io/up8qz/.

pdf of this poster

This work was supported by NIH grant R01DC015359 to S.L.T-S.

1. Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. Behavioral and Brain Sciences, 32(02), 183.
2. Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. Bulletin of te Psychonomic Society, 15(3), 147–149.
3. Shanks, D. R. (1995). The Psychology of Associative Learning. Cambridge, UK: Cambridge University Press.
4. Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II Current Research and Theory, 21(6), 64–99.