Inferences about Uniqueness in Statistical Learning

Anna Leshinskaya (alesh@sas.upenn.edu)

Sharon L. Thompson-Schill (sschill@psych.upenn.edu)
Department of Psychology, University of Pennsylvania
425 S. University Ave, Philadelphia PA 19104 USA

Abstract:

The mind adeptly registers statistical regularities in experience, often incidentally. We used a visual statistical learning paradigm to study incidental learning of predictive relations among animated events. We asked what kinds of statistics participants automatically compute, even when tracking such statistics is task-irrelevant and largely implicit. We find that participants are sensitive to a quantity governing associative learning, ΔP, rather than conditional probabilities or chunk frequencies as previously thought. ΔP specifically reflects the uniqueness, as well as strength, of conditional probabilities. This finding opens the possibility of common, sophisticated inferential mechanisms shared between statistical learning, associative learning, and causal inference scenarios.

Keywords: statistical learning; associative learning

Introduction

A core phenomenon in causal reasoning, contingency learning, and classical conditioning (for a review: Mitchell, De Houwer, & Lovibond, 2009) is that learners do more than register that two stimuli co-occur, but also compute whether they predict each other uniquely and independently, as if attempting to determine a causal model. Suppose two events A and B coincide, such that after most occurrences of A, B occurs. However, B also occurs *without* A at a very high rate. One would not represent a strong link between A and B in this case. This consideration is captured by a foundational learning formula, ΔP (Allan, 1980; Rescorla & Wagner, 1972; Shanks, 1985):

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

This equation states that learning is a product of both how often B follows A, as well as how often it appears without it. Surprisingly, it is not known whether this uniqueness principle governs in *statistical learning* tasks: cases where learning takes place incidentally, below awareness, and in absence of feedback or reward, and in which participants passively observe sequential streams of events (Brady & Oliva, 2008; Kim, Seitz, Feenstra, & Shams, 2009), although it has been reported in a paradigm somewhere in between statistical learning and conditioning (Sobel & Kirkham, 2006).

Experiment

We tested whether learners are sensitive to uniqueness in a visual statistical learning task. Participants saw two distinct event sequences, each composed of a unique set of animated events (Figure 1A), while performing a cover task. Each sequence contained one strongly predictive event pair—a *cause* and an *effect*—whose uniqueness we varied. In both

A

B

high △P condition (.97)

trial n+1

trial n	cause	effect	random1	random2	static
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 △P condition (.61)

	cause	effect	random1	random2	static
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

Figure 1. A) Static images depicting several of the stimuli, with common vs. rare alternates depicted in the top vs. bottom row, and the two objects which cued the distinct sequences. B) Mean transition matrices governing the appearance of events in each condition, and event frequencies below.

sequences, the first term in the ΔP formula (above) was matched: the probability that the effect appeared given that the cause appeared on the previous trial was equally high in both. However, in the low ΔP sequence, we increased the value of the second term, P(effect|~cause), by having the effect follow two other events and itself more often than in the high ∆P sequence. Thus, the two conditions were matched in terms of the transition probability from cause to effect, as well as in the number of times a cause-effect pair appeared overall (*chunk frequency*), but differed in terms of how uniquely the cause, rather than other events, predicted the effect. We expected learning to be worse in the low ∆P condition.

Following all videos, a surprise forced-choice test probed participants' knowledge of the cause-effect relation in both sequences, separately. The critical questions showed the cause followed by the effect in one video, and the effect followed by the cause in the other; participants had to choose the video that seemed more typical. These questions were matched across conditions in chunk frequency and transition probability. We found that participants were above chance for the high ∆P sequence (M = 61.83%, SE = 3.90%, t(79) .79, = .007, d = 0.31) but below chance for the low ∆P sequence (M = 41.67%, SE = 3.89%, t(79) = -2.16, p = .034, d = -0.24), which were significantly different from each other (CI [8.43, 29.90], t(79) = 3.55, p < .001, d = 0.55), as shown in Figure 2. These differences were this due specifically to the difference in ∆P. Participants had a weaker representation of the cause-effect relationship when uniqueness was low—despite the fact that in both conditions, cause-effect transitions occurred twelve times as often as effect-cause transitions. On the other hand, participants' confidence that they noticed any systematic order among the events was not reliably above 'unsure' for either condition (*high* M = 3.20, SE = 0.13, t(79) = 1.57, p = .121; *low* M = 3.06, SE = 0.13, t(79) < 1), with no difference between them (t(79) = 1.52, p = .132). Thus, learning was largely implicit, and effects of condition were on the output of this form of learning.

Overall, we conclude that participants' incidental learning is automatically informed by computations of uniqueness, in that neither participants' cover task nor the test questions demanded it or benefitted from it. Answers based on chunk frequency or conditional probability were both valid, and computationally simpler, but could not explain the difference in conditions. Thus, participants' incidental learning process can be described as a computation of ∆P.

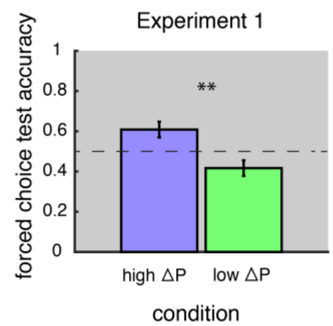

Figure 2. Forced-choice test accuracy ** = p < .001.

Model

We developed a computational account of this finding by adapting the Rescorla-Wagner (R-W) learning rule (Rescorla & Wagner, 1972) to the case of sequentially presented stimuli in which the objective is to learn all pairwise strengths among events. To account for the difference in conditions, we required that the weights from all causes to an effect to sum to 1, similar to Bayesian versions of R-W (Kruschke, 2008). This simple normalization step enabled us to capture the difference between conditions, yielding significantly stronger weights for cause-effect than effect-cause links, in both conditions (*high* ∆P, cause-effect M = 0.70, SE = 0.02; effect-cause M = 0.11, SE = 0.01, t(79) = 21.92, p < .001; *low* ∆P, cause-effect M = 0.19, SE = 0.01, effect-cause M = 0.07, SE = 0.01, t(79) = 9.07, p < .001), with a significantly larger difference in the *high* ∆P condition (t(79) = 17.35, p < .001). Without such normalization, the difference between conditions was not well captured. This is intuitive: the weight from B to X will only be affected by evidence of A to X trials if the weights to X trade off, such that as one gets stronger, the rest weaken (Kruschke, 2008). This can also be seen as representing each link as proportional to the base rate of X, if the base rate of X is, as here, captured in how often it appears following the other events in the state space. Normalization—conversion to relative rather than absolute values—is a cognitively realistic and adaptive mechanism for explaining such effects.

Conclusion

Our key finding was that participants in a statistical learning task were sensitive to not only the conditional probability between two events, but also the uniqueness of that relation. This can be seen as the result of normalization: the assumption that predictors of the same effect trade off, and to be considered effective, must raise the probability of the effect above its rate of occurrence otherwise. Overall, these findings bring statistical learning in closer contact with the rich literature in associative learning and causal reasoning, despite differences in the nature of these learning tasks.

Acknowledgments

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

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147–149.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678–685.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149.

Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning and Behavior*, *36*(3), 210–226.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183.

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.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology*, *37*(B), 1–21.

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: the development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*(6), 1103–15.