One rule to learn them all: The logistic perceptron vs. Widrow-Hoff / Rescorla-Wagner

Zachary N. Houghton[1], Emily Morgan[1] and Vsevolod Kapatsinski[2]

[1]University of California, Davis, and [2]University of Oregon

vkapatsi@uoregon.edu

A fundamental question in learning theory concerns the role of prediction error in learning. Error-driven learning rules predict that how much is learned from encountering an outcome depends on prediction error. Two widely used error-driven learning rules are the Rescorla-Wagner (RW) rule [7] and the logistic perceptron rule [8]. Both propose that the change in weight of a cue-outcome association $\Delta w(cue \rightarrow outcome) = \lambda(\Lambda - A(outcome))$, where λ is the learning rate, A(outcome) is the activation of the outcome given the current cues and association weights from these cues to the outcome, and Λ is 1 for a present outcome or 0 for an absent one. However, A(outcome) is simply the sum of cue-outcome association weights in Rescorla-Wagner but the inverse logit of this sum in the logistic perceptron. The inverse logit transformation ensures that A(outcome) is always between 0 and 1.

Recent work has argued that the logistic perceptron is more appropriate than Rescorla-Wagner for modeling learning whenever the outcomes are categorical (e.g., linguistic forms), having a probability of occurrence but not a magnitude. In particular, [3] showed that RW, but not the logistic perceptron, predicts a 'spurious excitement' effect in which a cue that co-occurs with two strong inhibitors of an outcome will become a very strong excitor of that outcome even if it never co-occurs with it. For example, if A→X, B→Y, C→Y, and BDC→Y, RW predicts D to be a strong cue to X, stronger than A. This is because training on A→X yields w(A→X) = +1, while training on B→Y and C→Y yields w(B→X) = w(C→X) = –1. Since w(B→X) + w(C→X) + w(D→X) must be 0 (as X never occurs after BDC), and w(B→X) + w(C→X) = –2, w(D→X) must be +2. In contrast, the logistic perceptron predicts D to be a cue to Y because it does not allow the sum of cue-outcome associations to overshoot 0 or 1. Human participants in a miniature artificial language learning experiment, where X and Y were distinct meanings and A–D were morphs, showed the behavior predicted by the logistic perceptron, responding to D alone with Y.

However, there is an effect that is captured by Rescorla-Wagner and not the logistic perceptron, and this is overexpectation [2]. Overexpectation is observed if training on AB→X weakens A→X and B→X associations after training A→X and B→X. According to RW, after stage 1, w(A→X) = w(B→X) = +1, but Stage II demands that w(A→X) + w(B→X) = +1, because the sum of cue weights from Stage I would overshoot the limit (1). Overexpectation has never been tested with linguistic stimuli. Linguistic stimuli are of particular interest because the same acoustic dimension can be perceived either categorically or continuously depending on context. It therefore allows us to test the hypothesis, raised in [3] and [6], that RW learning and therefore overexpectation occur with outcomes that are perceived as continuous and therefore have an unlimited magnitude as opposed to a bounded probability.

We implement an overexpectation design with a phonetic outcome (high fundamental frequency; F0) for which both continuous and categorical interpretations are possible. Learners hear a speaker react to some creatures with the ambiguous syllable [k/ga], which can be disambiguated by F0 [5] when F0 is raised at vowel onset. For us, X is F0 raised either where it cues the [k/ga] difference [4] or in the middle of the vowel, where it cues excitement (a continuous dimension). Post-tests determine how learners interpreted F0, and whether they are sensitive to F0 as a cue to voicing (not everyone is; [5]). If overexpectation occurs only with continuous outcomes, we expect overexpectation to occur when high F0 (X) is perceived as degree of excitement, but not when it is perceived as differentiating [ka] from [ga]. Data collection is ongoing.

References

[1] Dawson, M. R. (2008). *Minds and machines: Connectionism and psychological modeling*. Wiley.

[2] Dawson, M. R., & Spetch, M. L. (2005). Traditional perceptrons do not produce the overexpectation effect. Neural Information Processing-Letters and Reviews, 7(1), 11-17.

[3] Kapatsinski, V. (2023). Learning fast while avoiding spurious excitement and overcoming cue competition requires setting unachievable goals: Reasons for using the logistic activation function in learning to predict categorical outcomes. Language, Cognition and Neuroscience, 38(4), 575-596.

[4] Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from “true voicing” languages. The Journal of the Acoustical Society of America, 140(4), 2400-2411.

[5] Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. Journal of Phonetics, 59, 40–57.

[6] Packheiser, J., Pusch, R., Stein, C. C., Güntürkün, O., Lachnit, H., & Uengoer, M. (2020). How competitive is cue competition?. Quarterly Journal of Experimental Psychology, 73, 104-114.

[7] Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory (pp.64–99). Appleton-Century-Crofts.

[8] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors, Nature, *323*, 533-536.