Toward a Unifying Connectionist Model of Attention in Human and Animal Associative Learning

John K. Kruschke
kruschke@indiana.edu
Department of Psychology
Indiana University
Bloomington, IN USA

Recent research in human learning highlights the importance of selective attention, as has past research in animal learning. Mackintosh (1975) presented a highly influential model of animal conditioning which incorporated learned selective attention. In the present work, I derive two variants of Mackintosh's approach from connectionist architectures previously applied as models of attentional learning in humans. These connectionist architectures include the mixture of experts (Erickson & Kruschke, 1998; Jacobs, 1997) and an extension of ADIT (Kruschke, 1996).

This connectionist analysis yields several benefits: Mackintosh's (1975) model is shown to be (virtually) a special case of independently motivated human learning models. Unlike the somewhat arbitrary formulas proposed by Mackintosh (1975), the connectionist formulas for changes in associative strength and in attention are derived consistently from a common objective: error reduction. The connectionist framework also provides a natural generalization of Mackintosh's (1975) ideas, from the presence or absence of a single unconditioned stimulus to multiple categorical outcomes.

The connectionist models are fit to data from human learning experiments, including data showing the inverse base-rate effect, blocking of associative learning, latent inhibition, overshadowing, perseveration of attention through shifts of relevance, etc. New experiments show that lack of learning, à la Kamin (1969) and Rescorla and Wagner (1972), cannot adequately account for blocking in humans. Instead, learned inattention better accounts for the data.

Latent inhibition is treated as a special case of blocking, unlike any previous explanation of which I am presently aware. Latent inhibition is thereby also treated within a unified framework, instead of by alternative heuristics (e.g., Frey & Sears, 1978; Lubow, 1989; Pearce & Hall, 1980).

The analysis and modeling demonstrate that these statistically non-normative behavioral phenomena are "rational", that various animal and human models can be formally unified, that attentional learning in animals and humans might be simultaneously addressed within a common model, and that attentional learning is a crucial component of associative learning in humans.

References

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, –. In press.

Frey, P. W., & Sears, R. J. (1978). Model of conditioning incorporating the Rescorla-Wagner asssociative axiom, a dynamic attention process, and a catastrophe rule. *Psychological Review, 85*, 321–340.

Jacobs, R. A. (1997). Nature, nurture, and the development of functional specializations: A computational approach. *Psychonomic Bulletin & Review, 4*, 299–309.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In Campbell, B. A., & Church, R. M. (Eds.), *Punishment*, pp. 279–296. Appleton-Century-Crofts, New York.

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22*, 3–26.

Lubow, R. E. (1989). *Latent inhibition and conditioned attention theory*. Cambridge University Press, Cambridge UK.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276–298.

Pearce, J. M., & Hall, G. H. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*, 532–552.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In Black, A. H., & Prokasy, W. F. (Eds.), *Classical Conditioning: II. Current Research and Theory*, pp. 64–99. Appleton-Century-Crofts, New York.