Simply Too Many Notes

J. E. R. Staddon¹

Published online: 21 February 2017

© Association for Behavior Analysis International 2017

Keywords Killeen · Internal state · Reversal learning · Skinner

“Simply too many notes...” is the line reportedly suggested to Austrian Emperor Joseph II as a comment on Mozart’s brilliant first opera in Vienna. Well, the opera was brilliant but sometimes there really are too many notes—words—and much of operant conditioning seems to be made up of arguments about them. This is, presumably, a legacy of B. F. Skinner’s devotion to and skill in using language. But science is more than language, and we should surely begin a discussion of consequential learning with some actual examples.

What are we trying to explain? Killeen and Jacobs (2016) do not really make this clear, so an example may help. Here is one that makes more concrete some of the ideas there expressed. Many years ago, Derick Davis (Davis, Staddon, Machado, & Palmer 1993) looked again some data from an earlier experiment on reversal learning (Davis & Staddon, 1990; see also Staddon & Frank, 1974). The problem was simple: How best to explain the fact that pigeons trained on a daily discrimination-reversal task improve across days, but without, apparently, learning to reverse spontaneously? After considering several alternatives, we settled on a very simple, entirely deterministic model. But it did diverge from the usual behaviorist credo in one respect: it required us to assume a *state*, which sounds suspiciously cognitive but really is not. This simply means that the behavior of the organism cannot be accurately predicted from the single variable of response rate. Everybody knows this, of course. Numerous experiments, beginning with the phenomenon of behavioral contrast, show that response rate by itself cannot predict the effect of, for example, a change in relative reinforcement rate.

✉ J. E. R. Staddon
jers@duke.edu

¹ Department of Psychology and Neuroscience, Duke University, 417 Chapel Drive, Campus Box 90086, Durham, NC 27708-0086, USA

Even more alarming to some is that the concept of *state* implies that the future behavior of an organism cannot be deduced just from its present behavior. Again, everyone knows this in a commonsense way.

Killeen and Jacobs (2016) acknowledge the need for some kind of state variable(s): “Reinforcers and other affordances are potentialities rather than intrinsic features. Realizing those potentialities requires motivational operations and stimulus contexts that change the state of the organism...” (p. XXX). In this note, I will try through an example to show parallels between some of their ideas and the Darwinian variation-and-selection view of operant conditioning.

The experimental situation is very simple: two response keys paying off on a variable-ratio (probabilistic) basis (2-armed bandit). In the reversal-learning part of the study, one key each day was paid off on VR 8; the other produced no reinforcement. The “hot” key alternated from day to day. The solid line in Figure 1 shows the percent correct responses for a single pigeon that was well trained to respond equally to both choices at the beginning of the experiment. The dashed line shows learning rate, (the estimated value of the single parameter for a Markovian learning model fitted to the data, which is scaled to the right ordinate). The volatility of the parameter signifies changes in the learning rate, which in turn suggests that this model does not work in this situation (although it fitted data from several other conditions of these experiments).

The main finding is simply that the pigeon reversed faster and faster across daily reversals. How was he doing it? An obvious explanation, implied by earlier results (Staddon & Frank, 1974, Fig. 3; see also Staddon, 2016, Chapter 15), is memory confusion: the bird learns faster and faster each day because he is less and less sure which stimulus is positive, so persists less and less in a wrong choice. The more uncertain, the quicker the switch away from a losing bet. Note, too, that there was no evidence of spontaneous reversal each day. Lack of spontaneous reversal also implies that the bird cannot remember yesterday’s “hot” stimulus.

The Darwinian variation-selection approach to operant learning, which is accepted by Killeen and Jacobs (2016), suggested an easy way to model this process. Wrote Killeen and Jacobs (p. XXX): “Selection by consequences is different than the concept of strengthening a response.” Well, yes and no. It depends on what you mean by

Fig. 1 Reversal-learning performance (top function) of a Pigeon 017 on a variable-ratio schedule that alternates daily between two response keys. The bottom function shows the estimated value of the single parameter for a Markovian learning model fitted to the data. See text for additional explanation. Reproduced, courtesy of the American Psychological Association, from Fig. 6 of Davis et al. (1993)

response, what you mean by *selection*, and what you mean by *strengthen*. Suppose first a pool of responses—a *repertoire*—that compete according to their strengths (repertoire in this sense is similar to what Killeen and Jacobs term *affordances*, response opportunities that are offered by the environment and recognized by the organism). The competition is winner-take-all (WTA), and the strongest response becomes the *active* response; the others are *silent* responses in that they do not occur in that instant of time.

The idea of silent response may seem like heresy to radical behaviorists, but consider the following comment:

Our basic datum ... is the probability that a response will be emitted ... We recognize ... that ... every response may be conceived of as having at any moment an assignable probability of emission ... A *latent* response with a certain probability of emission is not directly observed. It is a scientific construct. But it can be given a respectable status, and it enormously increases our analytical power ... It is assumed that the strength of a response *must reach a certain value* before the response will be emitted. This value is called the *threshold*. [italics added]

If we substitute the word *silent* for *latent*, this proposal is different from what I am proposing only in one respect. I suggest that the *threshold* is simply competition from other latent/silent responses. The quote, interpreted in the Darwinian selection-variation way, addresses the target article's comment: "the probability of a response will increase given that it is followed by a reinforcer, is notated: $p(R_I|R_C) > p(R_I)$. But what gives R_C that power on what occasions has yet to be adequately addressed in our science ...". The quote itself is of course from B. F. Skinner (1948, p. 25). In Killeen and Jacobs' (2016) terms, then, *response* is to be replaced by *response tendency*, which is *strengthened* by reinforcement and reduced by nonreinforcement, and *selection* is accomplished by WTA competition among response tendencies.

This approach is empty without some plausible process or mechanism by which selection can work. A strengthening rule that we found to work quite well is basically Bayesian: the degree of change in response strength caused by reinforcement¹ or nonreinforcement depends on the weight of past data. The assumptions of the *cumulative effects* model are as follows:

1. Each response is has a value, V_A or V_B , which is just *reinforcement probability*:

$$V_A = \frac{\Sigma(\text{reinforcements} + \text{initial conditions})}{\Sigma(\text{responses} + \text{initial conditions})}, \text{ and similarly for } V_B.$$

2. WTA competition: The highest-value response is the one that occurs (becomes the active response).

¹ I assume simple strengthening by contiguity. The limits of this process are a topic for another time.

3. The value of a silent response does not change.

For example, suppose each of two competing responses has been reinforced ten times. Since the schedule is VR 8, the strength of each response (defined in this Bayesian way) is just $10 \text{ (reinforcements)}/80 \text{ (responses)} = 0.125$ (I am ignoring the random nature of the schedule for this illustration, but it is in fact very important for the overall behavior of this otherwise totally deterministic process.). If response *A* now occurs, and is not reinforced, its strength will decrease from 0.125 to 0.123, so the next response will be *B*, and so on for several iterations. The two will alternate. But after very many alternations, this is no longer true. Indeed, given initial conditions considerably greater than zero (many responses and many reinforcements), the process shows pretty good improvement in daily reversals, as Fig. 2 illustrates. This cumulative effects Bayesian model has no free parameters. But its behavior does depend on initial conditions, the initial response and reinforcement totals. If they are large, the model improves across reversals; if they are small, it does not. This is also in agreement with our data: experienced pigeons improved across reversals; naïve ones did not (Davis et al., 1993).

This model is not the only Bayesian possibility. Another and even simpler one does as well, albeit with one free parameter, *a*. For this model, Assumption 1 is changed follows:

1. Each response is has a value, V_A or V_B , depending on whether the response was reinforced.

Reinforced : $V_A = \Sigma(\text{reinforcements} + \text{initial conditions})$ and similarly for V_B .

Fig. 2 Reversal-learning performance of a Bayesian cumulative effects model. See text for further explanation. Reproduced, courtesy of the American Psychological Association, from Fig. 12 of Davis et al. (1993)

Fig. 3 Typical reversal-learning performance of a second Bayesian model. Cumulative responses on the two choices, (A) and (B), are plotted against each other. Daily transitions indicated by $\square$. See text for further explanation

$$\text{Unreinforced : } V_A = \Sigma(\text{reinforcements} + \text{initial conditions} - a), \text{ where } a > 0.$$

The other assumptions are as before. In words, the strength of each response tendency in this model is simply the total number of reinforcements it has received, minus an amount proportional to the number of unreinforced responses that have occurred. The model could hardly be simpler. Yet, it works pretty well. Figure 3 shows a typical simulation: cumulative responses, A vs. B, on nine successive days (day transitions marked by $\square$). The model at first responds indifferently to both choices: The relatively linear leftmost portion of the function indicates that response distribution among the alternatives was the same from day to day. But eventually performance begins to differentiate. The “zig zag” shape of the right side of the function indicates more A responding on A days and more B on B days after day five. Like the other Bayesian model, this one shows reversal improvement only after considerable experience and high initial reinforcement totals for both choices. Behavior is capricious if the initial totals are in single digits.

Finally, a version of the second model which, following an unreinforced response, subtracts from total reinforcements not a constant, but a constant times the accumulated reinforcement total, works better than both the others. This third model learns almost immediately to reverse rapidly as schedules change. But, unlike the others, this model is not sensitive to initial conditions. It learns to reverse as well when all are unity as when all are 100. It is too good, so it is probably not the best model for our actual pigeon.

Although details remain to be worked out, it does seem clear that pigeons choose, in this simple time-free situation, according to a basically Bayesian process that weights past reinforcements heavily. This completely rules out “local” models like simple

contiguity learning, Bush-Mosteller and its variants, or melioration. One small step, perhaps, towards a full understanding of reinforcement learning in a modest vertebrate species.

This theoretical analysis of an experimental paradigm may seem rather tangential to the Killeen and Jacobs (2016) article, whose aim is to draw the reader's attention to language—notation—rather than deal with any particular empirical question. The purpose of that article is therefore more philosophical and organizational than explanatory. But any philosophical inquiry about a scientific topic should also benefit the science. Organizing and clarifying terminology can help. But proliferating terms, and—a radical behaviorist weakness—obsessing about the precise definition of words in advance of full scientific understanding, is probably not helpful. The Killeen and Jacobs article is not guilty of the latter, but does contribute a bit to the former. Readers must judge for themselves whether resurrecting terms like affordance, disposition, and satisfier adds enough to our understanding to balance the added memory load. On the other hand, it should now be obvious—as the article ably points out—that behavior cannot be understood through stimuli and responses alone. Some notion of internal—and no, not necessarily physiological—state is essential. If we all agree on that, the next step is to identify these states and how they work. How well does the behavior of our hypothetical model match the experimental facts? I have tried to illustrate one approach with a simple example.

References

- Davis, D. G. S., & Staddon, J. E. R. (1990). Memory for reward in probabilistic choice: Markovian and non-Markovian properties. *Behaviour*, 114, 37–64.
- Davis, D. G. S., Staddon, J. E. R., Machado, A., & Palmer, R. G. (1993). The process of recurrent choice. *Psychological Review*, 100, 320–341.
- Killeen, P. R., & Jacobs, K. W. (2016). Coal is not black, snow is not white, food is not a reinforcer: the roles of affordances and dispositions in the analysis of behavior. *Behavior Analyst*. doi:10.1007/s40614-016-0080-7. Advance online publication.
- Skinner, B. F. (1948). Verbal behavior. William James Lectures, Harvard University. Download from <http://www.behavior.org/resources/595.pdf>.
- Staddon, J. E. R. (2016). *Adaptive Behavior and Learning*, 2nd edition. Cambridge University Press.
- Staddon, J. E. R., & Frank, J. (1974). Mechanisms of discrimination reversal. *Animal Behaviour*, 22, 802–828.