INTERNATIONAL JOURNAL OF COMPARATIVE PSYCHOLOGY

2014, 27(4), 569-584
Stanley J. Weiss, Editor
Special Issue
Peer-reviewed

On Choice and the Law of Effect

J. E. R. Staddon
Duke University, USA

Cumulative records, which show individual responses in real time, are a natural but neglected starting point for understanding the dynamics of operant behavior. To understand the processes that underlie molar laws like matching, it is also helpful to look at choice behavior in situations such as concurrent random ratio that lack the stabilizing feedback intrinsic to concurrent variable-interval schedules. The paper identifies some basic, non-temporal properties of operant learning: Post-reinforcement pulses at the beginning of FI learning, regression, faster reversal learning after shorter periods, and choice distribution on identical random ratios at different absolute ratio values. These properties suggest that any operant-learning model must include *silent* responses, competing to become the active response, and response *strengths* that reflect more than immediate past history of reinforcement. The cumulative-effects model is one that satisfies these conditions.

Two months before his engagement to his cousin Emma Wedgwood, Charles Darwin wrote copious notes listing the pros and cons of marriage. This is typical human choice behavior, though perhaps more conscious and systematic than most. No one would imagine that animals choose like this. But folk psychology often slips in unannounced. In his famous matching-law experiment, to make his procedure work Richard Herrnstein added what is called a *changeover delay* (COD): neither response could be rewarded for a second or two after each change from one response to the other. The reason was to prevent a third response: *switching*. Matching is got only when switching is suppressed by the COD: "The precise correspondence between relative frequency of responding and relative frequency of reinforcement broke down when the COD was omitted." (Herrnstein, 1961, p. 271).

But a little thought reveals that the response of *switching* is entirely hypothetical. It was derived neither from direct observation nor from proven theory. It came from intuition. If you, like the pigeon, had to choose repeatedly between two options, you might well consider switching as a third option to be assessed for payoff along with the other two.

As long as the pigeons respond on both keys they must switch, of course. But we have no independent justification for treating switching as a third response type. After all, the birds would switch even if the process that generates pecks just allocated them randomly to each key, or if each key generated its own random pecking. The only justification for Herrnstein's argument is – it works. When he added the COD, choice behavior showed a simple invariance.

Please send correspondence to Dr. Staddon, Duke University, USA (Email: jers@duke.edu).

Herrnstein, and many others before and after, have followed the same strategy: tweaking their experimental procedures until they produce orderly results. The strategy is both venerable and respectable. B. F. Skinner was fond of quoting Pavlov: "control your conditions and you will see order." Who could argue? But order is not valuable for its own sake. Order, yes, but what does the order represent? It is order in *what* exactly? Every operant conditioning experiment is a feedback system, a system of which the organism – which is the subject of the inquiry – is only a part. The order in Herrnstein's results is a property of the *system as a whole*, not of the subject organism.

The problem is that negative feedback suppresses variation. A thermostatically controlled heating system reduces the variation in building temperature that would otherwise occur as the outside temperature goes up and down. House temperature does vary a bit of course, even when weather conditions change slowly. But if we are interested not in the weather but in the heating system itself we might disconnect the thermostat, killing the feedback, and control the furnace directly. If that is not possible, we would be more interested in the dynamics (how does the house temperature lag behind the outside temperature? What is the effect of wind? …and so on) than in the steady state, the stable temperature when weather conditions are constant. Similarly, beyond a certain response rate, on a VI schedule response-rate variation makes no difference to the rate of food delivery to the pecking pigeon. If we want to understand how the pigeon works, we will therefore be less interested in the steady state than in the dynamics: when we vary the pattern of reinforcements, how does the pattern of responses vary?

In fact, operant conditioning, as a topic distinct from the group-average experiments of Hull, Tolman and their followers, *began* with dynamics. The novelty and excitement was provided by moment-by-moment changes in the behavior of single animals, not the static average of a group. The cumulative record is a real-time account that shows exactly how a given pattern of reinforcements produces a given pattern of pecks. Schedule pioneers, Ferster and Skinner (1957) even tried to summarize what is going on. Figure 1, is their schematic of the acquisition of fixed-interval (FI) behavior. (Alas, they did not pursue this approach for long.)

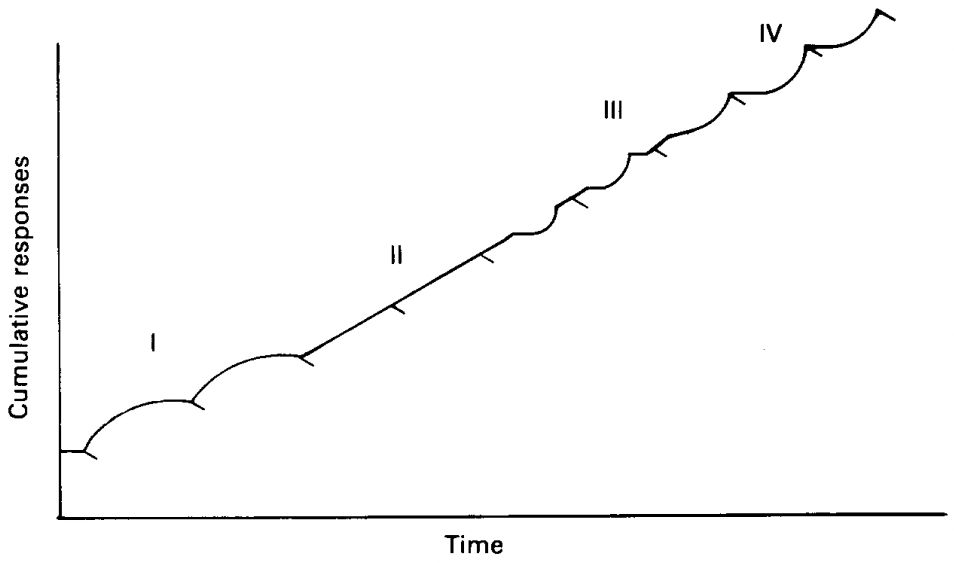

Figure 1. *Acquisition of fixed-interval responding.* Schematic cumulative record of the changing patterns of responding as a pigeon adapts to a fixed-interval schedule. (Adapted from Ferster & Skinner, 1957, Figure 117.)

But as technology advanced, and automatic schedule programming and data recording became easier and easier, the seductive orderliness of the wonderful organism-schedule feedback system took over. Skinner, whose own philosophy was partly responsible, even wrote a paper in 1976 entitled "Farewell, my lovely!" bemoaning the loss of interest in cumulative records (Skinner, 1976). To no avail: operant conditioning had embarked on a course where dynamics was almost ignored and averaging was standard practice. The search for steady-state empirical laws, like the matching law and others, came to dominate. But these laws describe not the behavior of individual organisms, but the behavior of an organism *plus* a schedule system. The nominal subject of the experiment (rat, pigeon, monkey) is just a component. Molar laws are not without interest. But the dynamics that underlie them are still largely unknown.

In this note I suggest how a dynamic analysis of choice might proceed, beginning with a choice situation with no stabilizing feedback at all. I believe that it gives us a clue to what the organism does when freed of constraint. The answers are surprisingly simple.

The Law of Effect

Herrnstein's 1961 *matching law* dominates the study of free-operant choice. But Herrnstein himself hinted at a more profound possibility. A later paper on choice between ratio rather than interval schedules raised the possibility that "matching and maximizing may be dual aspects of a single process, which is the *process of reinforcement itself* [emphasis added]" (Herrnstein & Loveland, 1975, p. 113). Subsequently, there have been a few attempts to identify such a process (Baum & Davidson, 2009; Crowley & Donahoe, 2004; see review in Staddon & Cerutti, 2003).

Surprisingly, perhaps, there seems to have been no real attempt to begin at the beginning, that is, with the individual cumulative records that first defined operant conditioning as a distinct field. Ferster and Skinner made a start: They divided the record of fixed-interval (FI) acquisition into four phases (Figure 1). Phase I is pure law-of-effect (LOE): "Just reinforced? Then respond more. Not reinforced? Respond less" – a burst of responses after each rewarded peck. Phases II and III are transitions to Phase IV, which is the standard fixed-interval pattern: post-reinforcer wait-and-then-respond until the next reinforcer (temporal discrimination). But this attempt to analyze cumulative records into their constituents was never followed up either by Ferster and Skinner or by others.

Ferster and Skinner's summary implies the existence of at least two processes: law-of-effect and timing. Both are presumably involved in matching: LOE because it is involved in all reinforcement; temporal discrimination because the variable-interval schedules used in the standard matching experiment set up a temporal contingency: reinforcement is more likely the longer the time elapsed since the subject's last response. I am concerned here just with the LOE process.

The cats in the original experiments that led Thorndike (1898) to the law of effect learned to repeat the response that got them out of the puzzle box, but the opportunity to repeat could either come immediately, one trial following the next, or after a delay. Although Thorndike provided detailed accounts of his cats' struggles and complaints before they escaped from their puzzle boxes, he was not interested in their behavior for its own sake but as a sign of the formation of *associations* between behavior and situation *stamped in* by reward. Associations and delays between trials require a memory that is both long-term and context-sensitive. But the law of effect in its barest form just refers to the repetition of a reinforced activity: "In operant conditioning we *strengthen* [emphasis added] an operant in the sense of making a response more probable or, in actual fact, more frequent" (Skinner, 1953, p. 65). This is the sense with which I begin.

First, I look at an example of an LOE process, Model 1, and show that it does indeed duplicate Phase 1 of Ferster and Skinner's FI summary. Next, I look at how a simple extension of this model fares in a choice situation lacking any temporal contingency. Model 2 works quite well but has a couple of features that seem not to be reflected in the data. Model 2 also fails to predict regression, a basic phenomenon of extinction. Which leads to the third model, the *cumulative effects* model, which can handle both regression and many of the properties of discrimination-reversal learning. Comparing these three models suggests two key properties for the dynamic process of operant conditioning.

Variable-, or, to be exact, *random*-ratio (RR) schedules, involve no temporal contingencies. The principle is the same as the Las-Vegas-style one-armed bandit. Payoff probability is the same whenever a response occurs; it does not change with passage of time, as on interval schedules, or with number of responses, as on fixed-ratio schedules. One way to get at the LOE, therefore, is through the study of random-ratio schedules, where temporal discrimination should not occur.

There are many ways to implement the law of effect as a real-time process. Perhaps the simplest is the following. For computational convenience I assume that time is discrete. The aim of any LOE model is then to describe how the probability of a response in each time step is increased by reinforcement and decreased by non-reinforcement. A reinforcer may occur in the time step immediately following a response (contiguous) or after a delay. I deal only with the contiguous case here (Catania's 2005 model of Skinner's reflex reserve – a rare effort to model cumulative records – is a version of Model 1 that incorporates delay). The assumptions of LOE Model 1 are as follows (time is discrete):

1. A response occurs in each time step with probability p.
2. p does not change unless a response occurs.
3. If the response is reinforced (in the next time step), p increases according to Equation 1.
4. If the response is not reinforced, p decreases according to Equation 2.

Reinforced:
$$p(t+1) = p(t) + k_R[1 - p(t)] \tag{1}$$

Unreinforced:
$$p(t+1) = k_N p(t) \tag{2}$$

where $0 < k_R, k_N < 1$; k_R and k_N are parameters representing the effects of reward and non-reward, respectively.

These two equations can be combined into Equation 3:

$$p(t+1) = k_N p(t) + k_R[1 - p(t)] \tag{3}$$

Figure 2 shows typical cumulative record generated by this process. The resemblance to the pigeon data (inset) is obvious; both show a burst of responding after each reinforcement. Unlike the pigeon record, however, the model pattern is stable. The pigeon will eventually develop the typical FI *scallop* but Model 1, lacking any process for temporal discrimination, always shows the post-reinforcer-burst pattern.

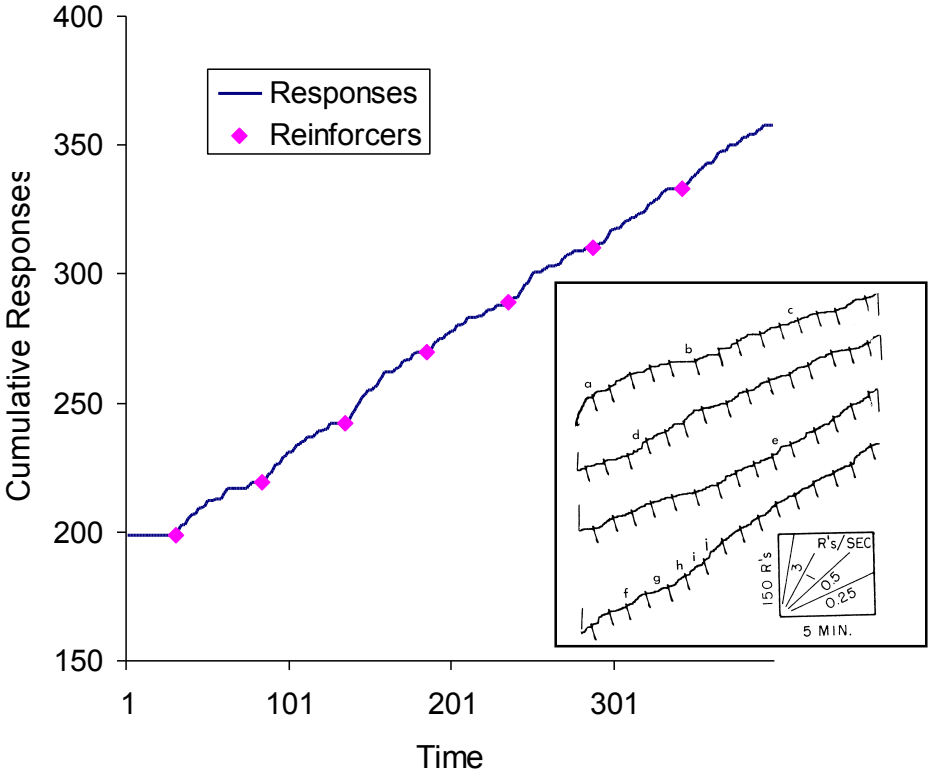

Figure 2. *Law of Effect responding on a fixed-interval schedule.* **Main panel***:* Cumulative record generated by a law-of-effect model (Equation 2 in the text) on a fixed-interval schedule. The size of the post-reward burst depends to some extent on the values of the parameters in Equation 2, but the pattern shown is the commonest. $k_R = k_N = 0.9$. **Inset***:* Cumulative records from a pigeon when first exposed to a fixed-interval 60-s schedule. Hash marks show reinforcements. (From Ferster & Skinner, 1957, Figure 118.)

Random-Ratio Choice

Choice between random-ratio schedules, which are probabilistic in their own right, allows us to eliminate the stochastic (probabilistic) assumption I used for the single-response case. The CE model, which I discuss in a moment, and Model 2 are both entirely deterministic. Model 2, which exchanges *strength,* X_i, for probability, otherwise resembles Model 1 in all but one key respect. Assumptions:

1. X_i does not change unless response i occurs in a given time step.
2. If the i^{th} response is reinforced, its strength increases, thus:

$$X(t+1) = X(t) + k_R[1 - X(t)]$$

which is the same linear operator as Equation 1.

3. If a response is not reinforced, its strength decreases thus:

$$X(t+1) = k_N X(t)$$

which is the same linear operator as Equation 2; $0 < k_R, k_N < 1$.

4. As before, these two equations can be combined into one:

$$X(t+1) = k_N X(t) + k_R[(1 - X(t)] \quad (4)$$

5. The big difference between Models 1 and 2 is this: In Model 2, in each time step, only the response with the highest strength occurs (*winner-take-all*: WTA). The response that does not occur is termed the *silent* response.

Model 2 is about as simple as it can be. It is completely deterministic. The only stochastic element is provided by the random-ratio schedules, which are unavoidable if we are not to introduce a temporal contingency.

The assumption of WTA competition is something of a novelty in learning theory. But it is familiar in other areas of biology. The distinction between expressed (active) and recessive (silent) genes (alleles) has been known since Mendel, for example. WTA competition is critical to the successes of this approach. I believe it will turn out to be an essential ingredient in any model for operant learning.

Assumptions 1 (no change in silent response) and 5 (WTA) have important implications. Consider how Model 2 behaves in the following two choice experiments: The animal can make two responses, A and B. The first experiment has three phases. In Phase 1, A is reinforced on some schedule and B is not reinforced: A gains strength, B remains close to zero. In Phase 2, B is reinforced and A is not: A's strength declines until it falls below the strength of B and ceases to occur (extinction). At that point, since it is not occurring (is silent), its strength remains constant. In Phase 3, response B also is extinguished. Response A still has some strength in Phase 3, because it dropped out in Phase 2 when its strength was less than B, but presumably still greater than zero. Hence, in extinction in Phase 3 both responses have some strength and both will occur as their strengths decline below some *other* activity. The final pattern, late in the extinction phase, is approximate alternation between A and B.

Now we repeat the procedure with a new and presumably identical organism. Experiment 2 has two phases. In Phase 1 only response B is reinforced; in Phase 2, B is again extinguished. To no one's surprise, in extinction either B alone occurs and A does not occur at all, or A occurs at a rate much lower than at the end of the first experiment. Why? Because A has never been reinforced so has no or very low strength in the second experiment. In experiment 1, on the other hand A's strength was reduced by extinction only to a level below B's, not to zero. Hence it could recover in Phase 3 when B was extinguished. The recovery of A in extinction, termed *regression* or *spontaneous recovery*[1], is a reliable finding in learning studies of all kinds. Assumptions 1 and 5 are essential to regression.

To simulate a single-response FI record with Model 2 (or the CE model, discussed in a moment), you need to assume some *other* behavior as well as the reinforced response. This is not implausible. A subject in an operant-conditioning experiment is not restricted just to the operant response. Observation shows that animals engage in many other activities, and that interim activities are important causal contributors to variations in strength of the reinforced response (Hinson & Staddon, 1978; Staddon, 1977). The assumption that they exist and compete with the operant response can explain molar effects such as behavioral contrast that have resisted other accounts.

[1] There is also a purely time-based spontaneous recovery, derivable from Jost's memory law, which I have discussed elsewhere (e.g., Staddon, Machado & Lourenço, 2001; Staddon, *in press*).

To simulate the FI *pulse* in Figure 2 with Model 2, the *other* response can have either a fixed strength or be *reinforced* (in some unspecified way) on a random ratio or interval schedule. I have not done the explorations necessary to see which assumption is better. In the two-choice situation it doesn't matter, because if the *other* response occurs, and the choice response(s) are silent, their strengths do not change. In other words, the relative strengths of the silent choice responses are independent of the strength of the active other response.

In the single-choice case, three states of the system are possible, depending on the value of the RR and the strength of other behavior: exclusive choice of the FI or the RR, or, if the RR value is intermediate, an FI record resembling Figure 2.

When confronted with two different RR schedules, Model 2, like most other choice models – as well as animal (Keasar, Rashkovich, Cohen & Shmida, 2002; Krebs, Kacelnik & Taylor, 1978; Macdonall, 1988) and human subjects – behaves *rationally*. The steady-state, stable pattern is exclusive choice of the smaller-ratio, higher-probability option. But, as one might expect, the smaller the absolute numbers, the fewer responses, and fewer reinforcers even, are needed before exclusive choice is reached. Figure 3 shows typical cumulative-X vs. cumulative-Y trajectories for a low-valued comparison, 5 vs. 10 and a high-valued, 50 vs. 100. The 2:1 payoff ratio is the same for both, but the system settles down to exclusive choice of the smaller 5:10 ratio much faster than the larger 50:100.

Identical Random Ratios

So far so unsurprising. But in some ways the most informative case is choice between two *identical* random-ratio schedules. This is not an experimental setup that would ever occur to someone interested in *choice* as a subject in its own right, because there is nothing to choose between. The two options are identical; it makes no difference which one the subject picks. But of course for this very reason, the procedure tells us more about the effect of each reward on the pattern of responses.

A rule that applies to all operant behavior is that reinforcement reduces variability. For example, Lowe and Harzem (1977) found that pigeons trained on a fixed-interval schedule and then shifted to a fixed-*time* schedule (same interfood interval, but no responding required) continued to respond if the interval was short, but not if it was long. In a standard matching-type choice experiment, pigeons tend to overmatch (choice ratios more extreme than reinforcement ratios) if the reinforcement rates are high, but match if they are lower (Elliffe & Alsop, 1996; Fantino et al., 1972; Logue & Chavarro, 1987). Gharib, Gade and Roberts (2004) found that rats had more variable bar-press hold times when reward was less frequent. Behavior on random-ratio choice schedules also shows less variability with more reinforcement. The two LOE models I will discuss both predict this effect, but differ in the details.

Acquisition data on concurrent random-ratio schedules are sparse, probably because the usual, dull result is exclusive choice of the majority option. Data on identical RR schedules are even rarer. But the equal-values case is interesting just because neither maximizing nor matching makes any prediction. Both are satisfied by any pattern of choices. Matching is forced by the identical ratios: the ratio of responses and reinforcers must be the same no matter what the distribution of choices. Maximizing is not possible since all choice patterns give the same payoff. It follows that a real-time model that can provide more differentiated predictions for identical concurrent random-ratio is worth exploring.

Model 2 does make some predictions for choice between identical ratios. Absolute ratio value is critical. Below a certain ratio value, the system converges rapidly on exclusive choice of one or other option – because the increment in strength to a reinforced majority choice minus the repeated decrements from non-

reinforcement until the next reinforcement is (for a given ratio and pair of parameter values) greater than the increment to the minority of a single reinforcement. Hence, the minority can never catch up to the majority. With the two parameters both equal to 0.9, that value is RR7: at 7 or below the system converges rapidly on exclusive choice – which choice is eventually favored is determined by which is chosen first. At ratios of 8 or larger, the behavior is highly variable.

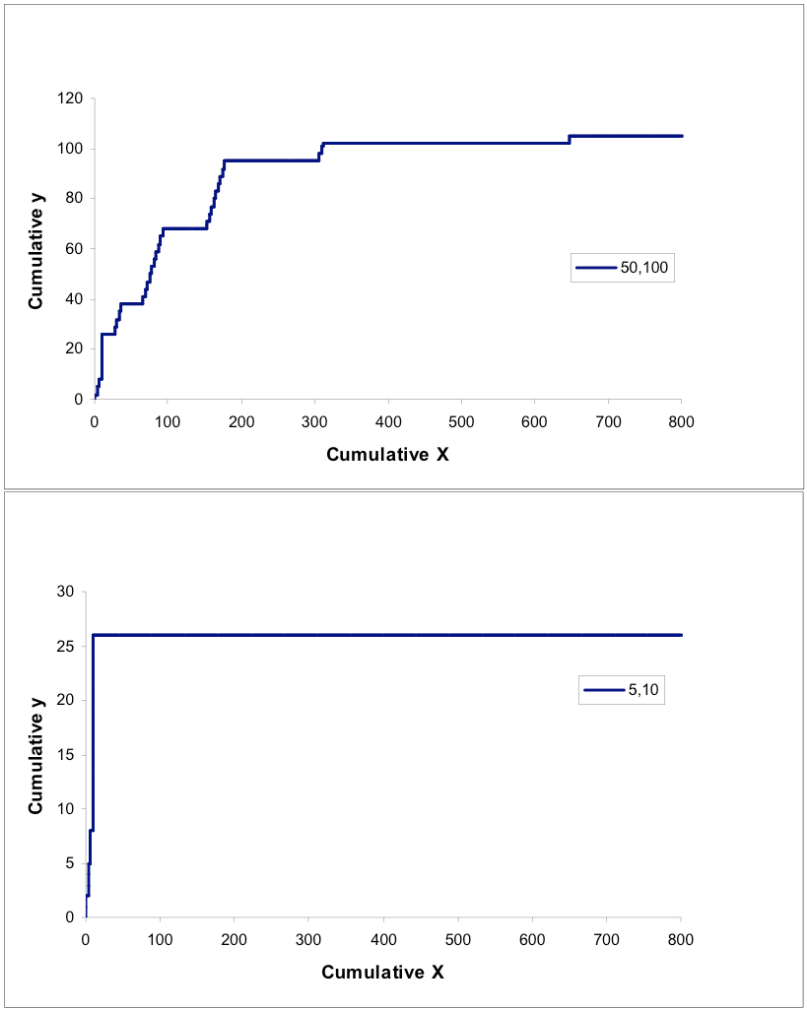

Figure 3. *The development of preference by Model 2.* Cumulative-response trajectories, X vs. Y, when Model 2 is exposed to either RR50 (X), RR100 (Y) or RR5, RR10. In both cases the initial preference was 50:50 Indifference). $k_R = k_N = 0.9$; initial strengths $X = Y = 0.5$.

Figure 4 shows how this looks. The *state space* for Model 2 has two dimensions, not one as in most learning models, including Model 1. For Model 2 and both parameters equal to 0.9, on RR7 (top panel), the two strengths, X and Y, change from initial values of 0.5, 0.5 to a final value which varies as a function of where in the last ratio the simulation ceases, but always falling somewhere on the horizontal line shown, which corresponds to fixation on X. (If Y had been chosen first, the line would be vertical, indicating final fixation on Y.) For ratios of 8 and above the X-Y trajectory is highly variable, including (depending on the total number of responses sampled) partial as well as exclusive preference. Figure 4 (bottom) shows an example of a complex trajectory under RR8.

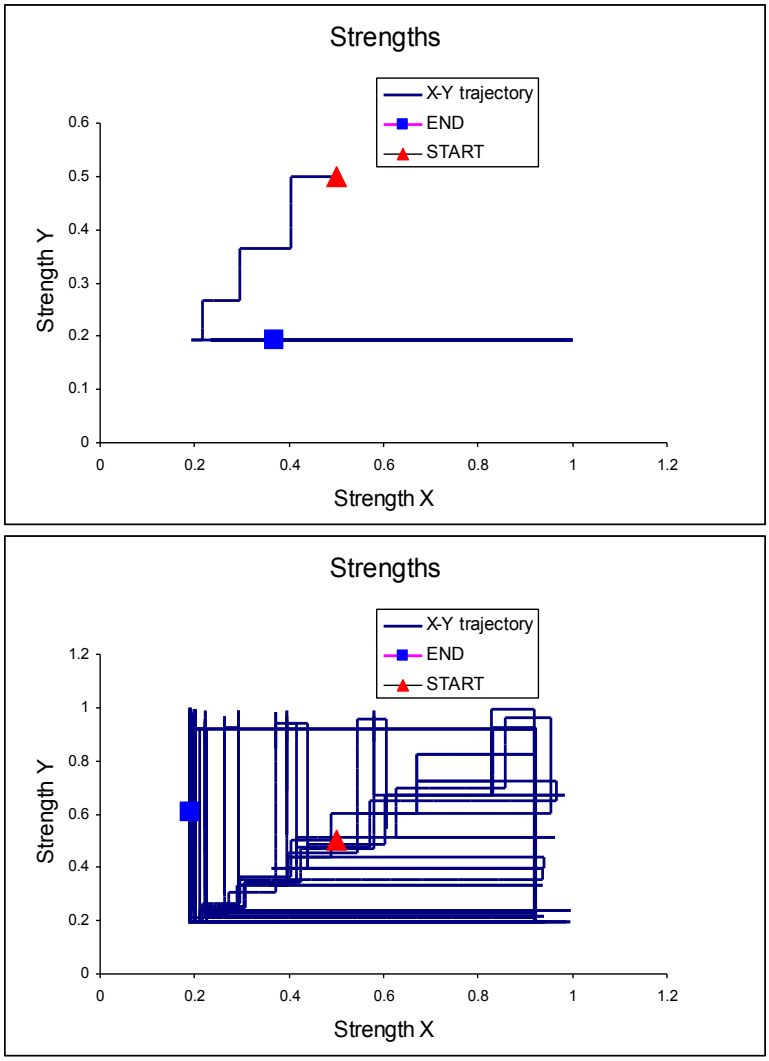

Figure 4. *State spaces, identical ratios: Model 2, Typical Records.* **Top**: Trajectory of strengths X and Y during 4993 time steps on concurrent RR7 RR7. The end state is exclusive choice of the initially chosen option. The model follows this identical pattern for ratios from 2 – 7. **Bottom**: Typical complex and variable trajectory for RR8 RR8 and greater; $k_R = k_U = 0.9$; initial strengths $X = Y = 0.5$.

The top panel of Figure 5 compares the pattern of preferences generated by Model 2 when confronted with two identical RR8 schedules vs. two RR100 – two values (8 and 100) where the model shows non-exclusive choice for parameter values 0.9, 0.9. Each point is a single simulation (experimental session), 20 for each ratio, showing the X and Y totals for a single run of 4993 time steps. So long as both ratios exceed the critical value of 7, there is no obvious difference in the *preference range* between small and large: for both the small and large ratio, choice proportions are spread across the full range from close to zero (almost all 4993 responses are X) to close to one (almost all Y). In other words, for Model 2 there are two choice modes – fixation and what looks like random preference – with a sharp break between the two.

So what are the data? Unfortunately there has been little parametric exploration of random-ratio choice, much less choice between identical ratios. Herrnstein and Loveland (1975) did have two conditions in which the variable-ratios were identical, at 30 and 60. But their plots only show relative response rate as a function of relative reinforcement rate and do not separate out by the absolute values of the ratios. They did not explore really small ratios, such as 5, so we do not know whether pigeons would show perfect exclusive choice under those conditions.

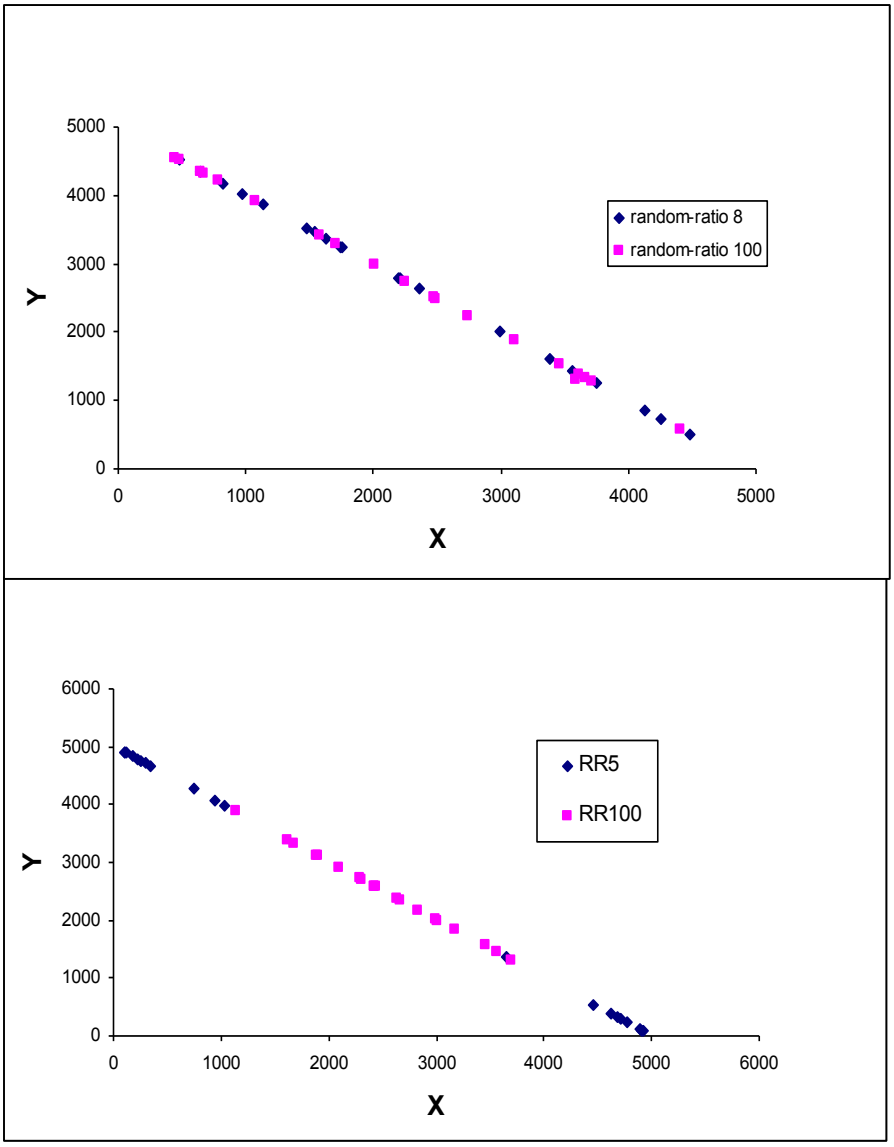

Figure 5. *The pattern of choice between identical schedules as a function of absolute reinforcement rate.* **Top***: Model 2.* Two identical random ratios. Preferences after twenty runs of 4993 time steps. Graph plots total X vs. total Y for two ratios: 8 and 100. Since a response occurs in every time step, the total is always 4993 and thus the points lie on a diagonal. Preference varies across the whole range at both RR values. $kR = kU = 0.9$; initial strengths $X = Y = 0.5$. **Bottom***: CE model*: Two identical random ratios. Preferences after twenty runs of 4993 time steps. Graph plots total X vs. total Y for two ratios: 5 and 100. Exclusive choice is favored at the smaller ratio, indifference at the large value. Initial values: $R_L = R_R = 5$, $x = y = 10$.

There seems to be no evidence for a sharp divide between small and large ratios of the sort predicted by Model 2 in the (admittedly limited) data on equal-random-ratio performance. But there is some evidence for a gradual shift in preference as a function of absolute ratio value. Horner and Staddon (1987) found that pigeons were more or less indifferent between RR75 schedules, but strongly favored one or the other when the ratio was 20 (Figure 6). This pattern was not invariable, however: "In various replications, the exclusive-choice pattern is the commonest, and we have not yet been able to isolate the necessary and sufficient conditions for the simple switch between indifference and exclusivity shown in the figure." (Horner &

Staddon, 1987, p. 62). No one seems to have looked at very small ratios to see if pigeons show the kind of rigid fixation predicted by Model 2. And if they do, is final choice determined by the initial response? Given what we know about operant choice, both seem unlikely. But the general pattern is clear: fixation on one choice or the other is typical when both ratios are small, but indifference, 50:50, is more likely when they are large. Model 2 predicts at large ratios not indifference, but choice scattered across the whole range from 0 to 1.

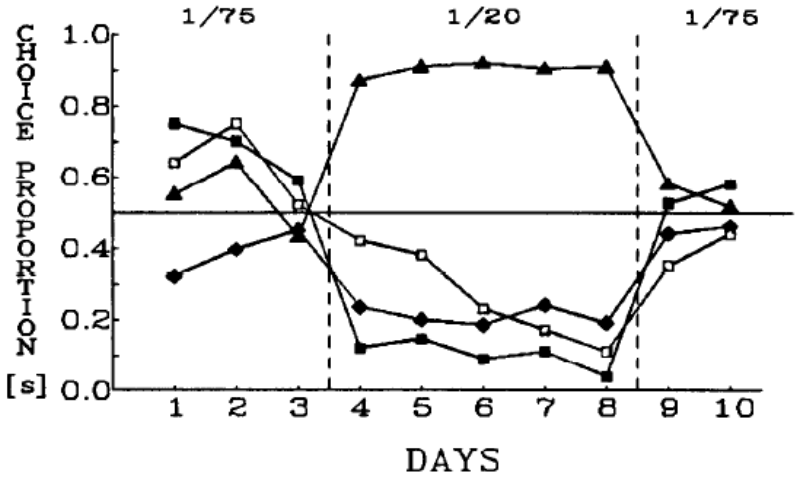

Figure 6. *The effect of absolute reward probability on choice between identical probabilistic alternatives* (concurrent random-ratio schedules). The figure plots the proportion of choices of the right-hand alternative across daily sessions for each of four pigeons for two conditions, p = l/75 and p = l/20, in ABA sequence (Horner & Staddon, 1987, Figure 1).

But Model 2 is not the last word. Another phenomenon points to a third model. If animals in a two-choice situation are paid off probabilistically (RR schedules) for response A on (say) odd-numbered days and for response B on even days, they learn over successive days to switch their preference rapidly every day. But, if payoff is reversed not every day but every fourth day, the pigeons may switch their preference more slowly on each reversal day (Davis, Staddon, Machado, & Palmer, 1993).

A WTA strength model of the right sort can explain these *reversal-learning* effects and regression, as well as the effect of absolute ratio size. Improvement in daily reversal performance just reflects the fact that the competing response tendencies become similar over days. The effect of going from daily reversal to reversal every four days is trickier to model. Each day, the strength of the reinforced response grows larger than that of the unreinforced response. This gain must be reversed when payoff reverses. When reversal occurs every day, the two strengths cannot grow too large, thus the pigeon can change preference swiftly. But when one choice is reinforced for four days before the reversal, the strength of the reinforced choice may, depending on the details of the model, be much larger than the nonreinforced response strength. Hence, preference takes longer to reverse after four days than after one.

The simplicity of this account is muddied by the stochastic nature of the random ratio. For example, a long run of non-reinforcement will tend to equalize response strengths under almost any set of model assumptions. Model 2, although it includes WTA competition, is particularly susceptible to the equalizing effect of a long period of non-reinforcement. Even after a history of equal reinforcement of both, reinforcing only one choice weakens the unreinforced choice only to the point that it loses the competition. Additional

reinforcement of the winner neither strengthens it further nor further weakens the loser. A long period of non-reinforcement usually leads to alternation. Hence, strength disparity between active and silent response may not increase more after four days on one side compared to just one. Consequently, Model 2 does not show slower reversal after 4-day reversal training than after daily reversal.

A better fit to these data is the *cumulative effects* (CE) model (Davis et al., 1993), which was invented explicitly to explain reversal learning. It is in some respects simpler than Model 2. It has no free parameters[2]. It can explain successive discrimination reversal and similar effects of past history on discrimination performance. The response strengths, X_i, are simply equal to *total reinforcement probability*:

$$X_i = \frac{\sum R_i + R_i(0)}{\sum x_i + x_i(0)}, \tag{5}$$

where R_i is the total number of reinforcements received in the experiment for response i and x_i the total number of responses made. In other respects – WTA choice rule and constancy of the *silent* (non-occurring) response – the CE model is the same as Model 2. The two variables: cumulated reinforcements, R_i and cumulated responses, x_i, have *initial values* $R_i(0)$ and $x_i(0)$.

Figure 5, bottom panel, shows CE-model pattern of preferences for two ratio values spanning a similar range as the upper panel, RR5 and RR100. In this case there is a clear difference: preference is usually more extreme at the small value. RR5 data cluster at the extremes; RR100 points tend to indifference. Unlike Model 2, and in agreement with data, at high ratios the pattern tends to indifference, rather than preference evenly scattered from 0 to 1 as in the top panel. And the shift from indifference to exclusive choice is gradual rather than abrupt, which is more in accord with the (admittedly skimpy) data.

The state space for the CE model is more complex than for Model 2. Rather than two, it has four dimensions: response and reinforcement totals for each response. One way to represent it is as two cumulative trajectories: R_i *vs.* x_i for each response; the slope of the resulting curve is just X_i for each response. Because of the WTA competition rule, the two X values usually remain pretty close to one another, but they can diverge under some conditions.

Figure 7 shows three more examples of CE model behavior. The inset shows FI behavior where the other response is reinforced on a random ratio. The top panel shows quite rapid improvement in performance over successive discrimination reversals. The CE model can show faster reversal when reversals are more frequent (the equivalent of daily reversal vs. reversal every 4 days – see Davis et al., 1993 their Figure 7), but its behavior is surprisingly complex. Because of the highly nonlinear WTA rule there is never a direct link between the level of an active response and the state of system. The way the model adapts to different ratio-schedule combinations also depends on the initial conditions in ways that have not been fully mapped.

The model shows regression (Figure 7, bottom panel). There are two phases in extinction after two equal periods where response A and then response B have been reinforced. First, B, the last-reinforced response, persists; then both responses occur *at rates reflecting the number of reinforcers received for each* (arrows). In the absence of a competing (interim) activity, the two responses never cease to occur. In a comparable experiment with Model 2, after the period of persistence, the two responses simply alternate.

[2] This is not quite true. I assume that the model accumulates total reinforcements and responses over the entire history of an experiment, but some broad parameter-defined limits may be necessary. The properties I discuss just require a longish memory.

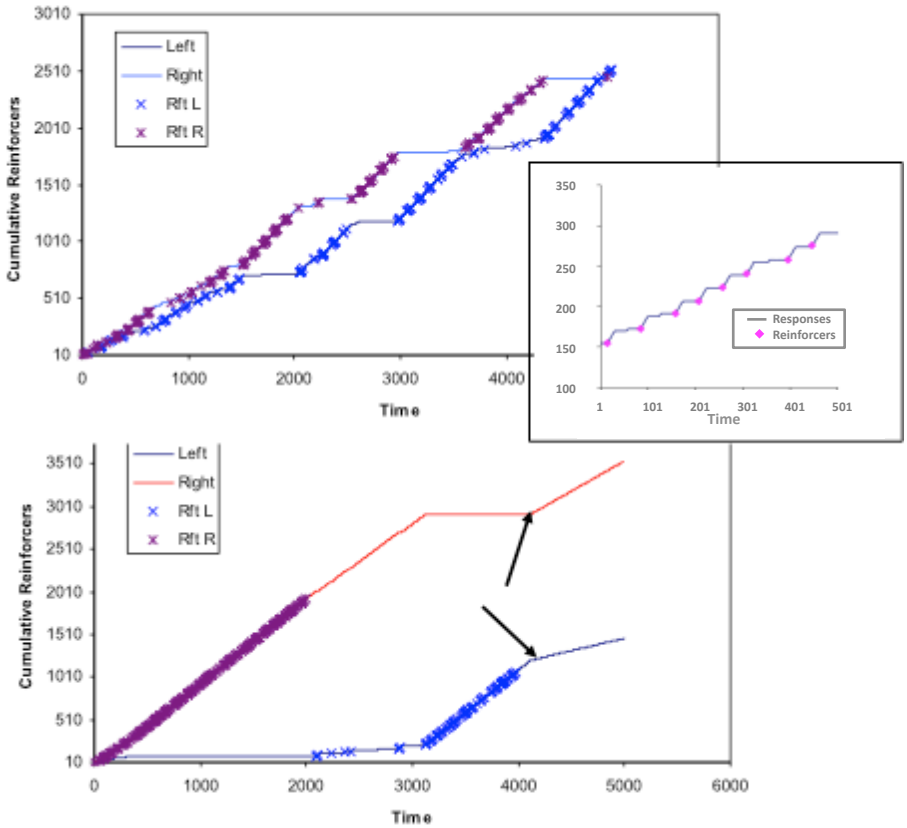

Figure 7. *Cumulative Effects Model, reversal and regression: Typical Cumulative Records*. **Top:** Reversal acquisition: Curves are cumulative left and right responses on RR 8 vs. EXT. Symbols are reinforcers. Components (reversals) alternated every 1000 time steps. After a brief period of indifference, responding alternates between components. **Bottom:** Regression: Right is reinforced on RR8 for 2000 time steps followed by Left reinforced for 2000 steps, followed by extinction for both. Note that both responses continue to occur in extinction. Final relative rates in extinction reflect the number of reinforcements received in training (arrows). Once the random influence of the schedule is eliminated in extinction, the deterministic nature of the model is obvious. Initial conditions: 10 for both responses and reinforcers for both alternatives. **Inset:** FI responding by the CE model.

Concurrent Variable-Interval Schedules

How applicable is the literally time-less CE model to the main arena of free-operant choice: variable-interval schedules? The CE model is consistent with matching because the end-state of the model is equal reinforcement probability for all options. Equation 5 defines this probability over the system's entire experimental history (plus initial conditions). With increasing experimental exposure the contribution of initial conditions becomes relatively smaller and smaller. So the model in effect implies that the system will approach equality of total payoff probability for each choice option. In matching studies using concurrent VI-VI schedules, in the absence of effects from temporal processes and a key procedural feature – the changeover delay (COD) – the CE model will tend to produce under-matching (Todorov, Castro, Hanna, de Sa, & Barreto 1983), since the average of many conditions is likely to be close to 50:50 – equal payoff probability for both choices.

There is some evidence for an effect of absolute reinforcement rate on preference in interval schedules, as the CE model predicts. Fantino et al. (1972) looked at pigeons' preference under concurrent variable-interval schedules. They found, *contra* matching, that response ratios were more extreme than reinforcement

ratios at small VI values, but less extreme at larger values, a result similar to the lower panel of Figure 5. This result is consistent with the general principle that reinforcement acts to select. The lower the reinforcement rate, the weaker the selection and the more variable the behavior.

In the most-studied choice situation – concurrent variable-interval schedules – the role of the law of effect has usually been thought minimal. The emphasis has been on higher-order, time- or rate-based processes like momentary maximizing, melioration or the switching-based kinetic model. But careful analysis of individual, averaged steady-state concurrent-VI data has nevertheless shown what the authors term *post-reinforcer preference pulses* of exactly the sort implied by any LOE model.

Figure 8, from a concurrent VI experiment by Davison and Baum (2003), shows the average ratio of left-right responses for the next 40 responses after each reinforcement for a left or right response across several experimental sessions. The data show a swing to the right after a right reinforcement and a swing to the left after a left reinforcement, and "Larger reinforcers produced larger and longer post-reinforcer preference pulses than did smaller reinforcers" (p. 95) just as the LOE would predict.

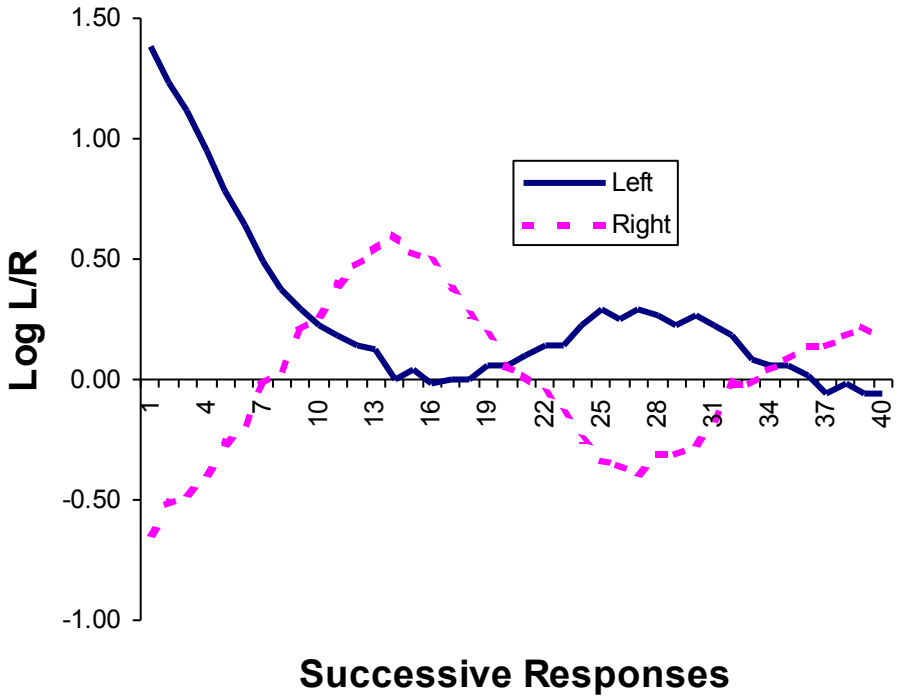

Figure 8. *Preference Pulses. Redrawn* from Davison & Baum (2003, Figure 2). Experiment 1, condition 2 (six reinforcers per minute). Log left/right response ratios at each response up to 40 responses after left- and right-key reinforcers. The data were averaged across 6 pigeons.

These data look like confirmation of the LOE process, even on interval schedules. But a subsequent analysis by McLean, Grace, Pitts, and Hughes (2014) pointed out that much of the effect may be artifactual. The reason is simple. On concurrent VI-VI schedules, especially when there is a changeover delay, animals adapt by spending some time on one choice before switching to the other. They respond in bouts. Because of the temporal contingency built-in to interval schedules – the longer you wait, the higher the probability of payoff – the first post-COD response is much more likely to be reinforced than later responses in the bout. Or, to put the same thing in a different way, a reinforced response is more likely to be early than late in a bout.

The result, of course, is that a reinforced response is likely to be followed by more responses on the same key than a response chosen at random. The result will look like a reinforcer-induced pulse of responding. But the process seems to be the reverse of the obvious one. Not "reinforcement makes more pecks", but rather, the schedule selectively reinforces responses that will anyway be followed by others.

In fact, both processes, LOE as well as the bout pattern, seem to operate. McLean et al. (2014) conclude: "the delivery of reinforcers had modest lengthening effects on the duration of the current visit, a conclusion that is quantitatively consistent with early research on short-term effects of reinforcement." (p. 317). There is a small pulse effect, larger for larger reinforcers.

Conclusion

It is possible to dissect cumulative records into their component processes. This note is a first step, beginning with the simplest version of the law of effect and excluding temporal discrimination. The question is: What properties must a learning model have to be consistent with basic non-temporal learning phenomena: post-reinforcement pulses at the beginning of FI learning, regression, reversal learning after shorter and longer periods, and choice performance on identical random ratios at different absolute ratio values? My conclusion is that to be consistent with all these well-established effects, any operant-learning model must include silent responses, competing in a nonlinear way, such as winner-take-all, to become the active response; and response strengths that reflect more than immediate past history of reinforcement.

References

Baum, W. M., & Davidson, M. (2009). Modeling the dynamics of choice. *Behavioural Processes, 81,* 189-194.

Catania, A. C. (2005). The operant reserve: A computer simulation in (accelerated) real time. *Behavioural Processes, 69*, 257–278.

Crowley, M. A., & Donahoe, J. W. (2004). Matching: its acquisition and generalization. *Journal of the Experimental Analysis of Behavior, 82*, 143-159.

Davis, D. G. S., Staddon, J. E. R., Machado, A., & Palmer, R. G. (1993). The process of recurrent choice. *Psychological Review, 100*, 320-341.

Davison, M., & Baum, W. M. (2003). Every reinforcer counts: reinforcer magnitude and local preference. *Journal of the Experimental Analysis of Behavior, 80,* 95-129.

Elliffe, D., & Alsop, B. (1996). Concurrent choice: effects of overall reinforcer rate and the temporal distribution of reinforcers. *Journal of the Experimental Analysis of Behavior, 65,* 443-463.

Fantino, E., Squires, N., Delbrück, N., & Peterson, C. (1972). Choice behavior and the accessibility of the reinforcer. *Journal of the Experimental Analysis of Behavior, 18,* 35-43.

Ferster. C. B., & Skinner, B. F. (1957). *Schedules of reinforcement.* New York, NY: Appleton-Century-Crofts.

Gharib, A., Gade, C., & Roberts, S. (2004). Control of variation by reward probability. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 271–282.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4,* 267-272.

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior, 24*, 107-116.

Hinson, J. M., & Staddon, J. E. R. (1978). Behavioral competition: a mechanism for schedule interactions. *Science, 202,* 432-434.

Horner, J. M., & Staddon, J. E. R. (1987). Probabilistic choice: A simple invariance. *Behavioural Processes*, *15*, 59-92.

Keasar, T., Rashkovich, E., Cohen, D., & Shmida, A. (2002). Bees in two-armed bandit situations: foraging choices and possible decision mechanisms. *Behavioral Ecology, 13*, 757–765.
Krebs, J. R., Kacelnik, A., & Taylor, P. (1978). Tests of optimal sampling by foraging great tits. *Nature*, *275,* 27-31.
Logue, A. W., & Chavarro, A. (1987). Effect on choice of absolute and relative values of reinforcer delay, amount, and frequency. *Journal of Experimental Psychology: Animal Behavior Processes, 13*, 280-291.
Lowe, C. F., & Harzem, P. (1977). Species differences in temporal control of behavior. *Journal of the Experimental Analysis of Behavior, 28,* 189-201.
Macdonall, J. S. (1988). Concurrent variable-ratio schedules: Implications for the generalized matching law. *Journal of the Experimental Analysis of Behavior, 50*, 55-64.
McLean, A. P., Grace, R. C., Pitts, R. C., & Hughes, C. E. (2014). Preference pulses without reinforcers. *Journal of the Experimental Analysis of Behavior, 101*, 317–336.
Skinner, B. F. (1953). *Science and human behavior.* New York, NY: The Free Press.
Skinner, B. F. (1976). Farewell, my lovely! *Journal of the Experimental Analysis of Behavior*, *25*, 218.
Staddon, J. E. R. (1977). Schedule-induced behavior. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior.* Englewood Cliffs, NJ: Prentice-Hall.
Staddon, J. E. R. (*in press*). *Adaptive behavior and learning* (2nd ed.). New York: Cambridge University Press.
Staddon, J. E. R., & Cerutti, D. T. (2003). Operant behavior. *Annual Review of Psychology, 54*, 115-144.
Staddon, J. E. R., Machado, A., & Lourenço, O. (2001) Plus ça change…: Jost, Piaget and the dynamics of embodiment. *Behavioral and Brain Sciences*, *24*, 63-65.
Thorndike, E. L. (1898). Animal intelligence: an experimental study of the associative processes in animals. *Psychological Monographs*, *2*, 109.
Todorov, J. C., Castro, J. M. O., Hanna, E. S., de Sa, M. C., & Barreto, M. (1983). Choice, experience and the generalized matching law. *Journal of the Experimental Analysis of Behavior, 40*, 99-111.

Financial Support: This work received institutional support.
Conflict of Interest: The author of this paper declares no conflict of interest.

Submitted: *March 8th, 2014*
Resubmitted: *October 9th, 2014*
Accepted: *October 22nd, 2014*