Chapter 8
Is Animal Learning Optimal?

John E. R. Staddon

Optimization theory has always been the first choice for understanding adaptive systems. The behavior of animals, both learned and instinctive, is clearly adaptive. Even before Darwin, biologists noted the close match between the form and behavior of animals and the "conditions of existence," the environment—the *niche* as it would now be called—in which they live. And utility maximization was until quite recently the predominant approach in neo-classical economics. But since Darwin, biologists have known that the form and behavior of organisms is quite often not optimal, in any reasonable sense. The human appendix, the peacock's ungainly tail, and the vestigial legs of snakes are clearly not optimal: humans, peacocks, and snakes would clearly live longer and move more efficiently without these redundant appendages.

But what of *behavior*? Economics is a pretty successful social science and, in the 1970s, psychologists and behavioral ecologists had high hopes that these techniques could shed light on animal learning and behavior (Krebs and Davies 1978; Staddon 1980). But this hope also had to be given up. This article is a brief summary of why, and what the alternatives are.

8.1. Reinforcement Learning

Reinforcement learning is the study of how behavior is guided by its consequences—either *reinforcement*, which increases the probability of behavior it follows, or *punishment*, which decreases the probability. Most learning of "higher" animals—mammals, birds, and humans—falls into this category. It is easy to see how under normal circumstances it will result in organisms getting more of what they like and less of what they dislike, in a sort of hill-climbing optimization process. And indeed, two decades ago, much theoretical research supported the idea that learned behavior maximizes the rate of reinforcement. Apparent failures could usually be attributed to cognitive limitations. Rats and pigeons cannot learn even quite simple sequences, for example. Thus, they cannot master a task that requires double alternation (LLRR) to get food reinforcement. But once *cognitive constraints* of this sort were taken into account, it looked as if all *operant behavior* (as behavior guided by consequences is called) could be accommodated by optimization.

8.1.1. Instinctive Drift: Do Animals "Know" What to Do?

But even in those early days, some striking violations of optimality were known. Keller and Marion Breland were students of B. F. Skinner, coiner of the term "operant" and pioneer in the experimental study of reinforcement learning in animals. They learned to train pigeons to guide missiles on Skinner's "Project Pelican" during World War II, and later went into the business of training animals for commercial purposes such as advertising. They found that animal behavior is not so malleable as Skinner taught. They found that Skinner's memorable phrase "reinforcement shapes behavior as a sculptor shapes a lump of clay" wildly exaggerates the power of reward and punishment to mould behavior. In an article entitled *The Misbehavior of Organisms* (1961), the Brelands described numerous violations of the reinforcement principle. One of the more dramatic involved a raccoon (Fig. 8.1).

> Raccoons condition readily, have good appetites, and this one was quite tame and an eager subject. We anticipated no trouble. Conditioning him to pick up the first coin was simple. We started out by reinforcing him for picking up a single coin. Then the metal container was introduced, with the requirement that he drop the coin into the container. Here we ran into the first bit of difficulty: he seemed to have a great deal of trouble letting go of the coin. He would rub it up against the inside of the container, pull it back out, and clutch it firmly for several seconds. However, he would finally turn it loose and receive his food reinforcement. Then the final contingency: we put him on a ratio of 2, requiring that he pick up both coins and put them in the container.
>
> Now the raccoon really had problems (and so did we). Not only could he not let go of the coins, but he spent seconds, even minutes, rubbing them together (in a most miserly fashion), and dipping them into the container. He carried on this behavior to such an extent that the practical application we had in mind—a display featuring a raccoon putting money in a piggy bank—simply was not feasible. The rubbing behavior became worse and worse as time went on, in spite of nonreinforcement (Breland and Breland 1961).

The raccoon's problem here isn't really a cognitive one. In some sense he "knows" what is needed to get the food since he learned that first. Yet this other "washing" behavior soon takes over, even if it blocks further rewards. The Brelands termed this aberrant behavior and others like it in other species, *instinctive drift*. The fact that animals may learn effective Behavior A successfully, yet revert irreversibly to ineffective Behavior B, suggests that cognition isn't everything. An organism may be cognitively capable of an optimal pattern of behavior, yet fail to persist in it.

8.1.2. Interval Timing: Why Wait?

Here are two other, more technical, examples from the extensive experimental literature on reinforcement learning.

FIGURE 8.1. A raccoon at night, its usual active time

8.1.2.1. Ratio Schedules

Reinforcement schedules are simply rules that relate an organism's behavior to its reinforcing consequences. For example, the behavior may be lever-pressing by a hungry rat, and the rule may be "one lever press gets one food pellet", this is called a *fixed-ratio 1* (FR 1) schedule. Obviously, there is nothing magical about the number one. Animals will respond on ratio values as high as 50 or more. And the ratio need not be fixed; it can vary randomly about a mean from reinforcer to reinforcer: *variable ratio* (VR).

After a little exposure, behavior on most reinforcement schedules appears to be pretty optimal: animals get their reinforcers at close to the maximum possible rate with minimal effort. On ratio schedules, e.g., they respond very rapidly, but expend little energy. But FR schedules do offer a small puzzle. When the ratio value is relatively high, say 50 or so, animals don't start responding at once after each food delivery. Instead, they wait for a while before beginning to respond.

The wait time is approximately proportional to the ratio value: the higher the ratio, the longer the wait. Indeed, if the ratio is high enough, the animal may quit entirely.

The obvious explanation is some kind of fatigue. Perhaps the animal just needs to take a breather after a long ratio? But no, this can't be the explanation because they don't pause on a comparable *variable* ratio. For the right explanation, we need to look at *interval* reinforcement schedules.

8.1.2.2. Interval Schedules

Interval schedules use a more complex rule than ratio schedules—although the principle is still simple enough. An example is "a lever press 60seconds after the last pellet gets another pellet." This is termed a *fixed-interval 60* (FI 60) schedule.

Behavior on FI schedules is also close to optimal. Animals wait before beginning to press the lever and lever-pressing thereafter accelerates up to the time when food is delivered. Figure 8.2 shows a typical "scalloped" record of cumulative responding (in this case, pecking on a disk by a hungry pigeon), rewarded with brief access to grain. The figure also shows the typical *wait time* after food delivery before pecking begins and the accelerated peck rate thereafter. Wait time is proportional to the interfood interval: if the animal waits 15seconds before beginning to respond on an FI 30second schedule, it will wait 30second on an FI 60. Thus, it wastes relatively few responses and gets the food as soon as it is available: altogether a pretty optimal pattern.

Contrast that behavior with what they do on a very similar procedure called a response-initiated delay (RID) schedule (Fig. 8.3). A RID schedule is almost the same as an FI schedule. The only difference is that the time, instead of being measured from the last reinforcer, is measured from the first response after a reinforcer: the organism starts the clock itself, rather than the clock restarting

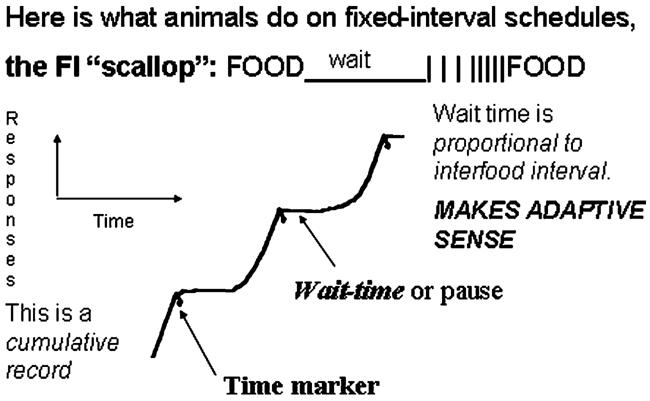
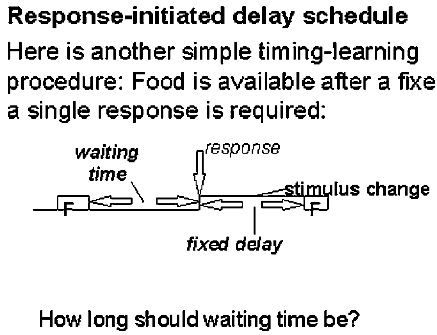

Here is what animals do on fixed-interval schedules,
the FI "scallop": FOOD wait | | | |||||FOOD
Responses
Time
Wait time is proportional to interfood interval.
MAKES ADAPTIVE SENSE
Wait-time or pause
This is a cumulative record
Time marker

FIGURE 8.2. Adaptive behavior—responding at an accelerating rate, adjusted to the usual time of food delivery—on fixed-interval reinforcement schedules



Response-initiated delay schedule

Here is another simple timing-learning procedure: Food is available after a fixed delay, a single response is required:

How long should waiting time be?

FIGURE 8.3. Another time-based reinforcement schedule: response-initiated delay. Food is delivered at a fixed time after the first post-food response. The subject should respond as soon as possible—but animals (and humans, sometimes) do not

after each reinforcement. The optimal behavior here is similar to that on FI: wait after the first response, a time proportional to the delay time. But the first response should not be delayed, since any delay adds to the total time between reinforcers. Thus, the optimal behavior is simple: respond at once after each reinforcer, then quit until food is delivered.[1]

This is not what most animals do. Instead, they ignore the clock-starting response requirement and treat the schedule as if it were fixed interval, waiting for a time proportional to the actual interfood interval before the first response.

The process can be analyzed formally in a very simple way as follows. On fixed interval, the wait time, WT, is approximately proportional to the interfood interval, I (this is termed *linear waiting*). Hence,

$$\mathrm{WT} = \mathrm{kI} \tag{8.1}$$

where constant k is around 0.5 and I is the interval value. On an RID schedule, the interfood interval is necessarily the sum of the wait time plus the imposed delay. So if the same process operates, $\mathrm{WT} = \mathrm{k}(\mathrm{WT} + \mathrm{I})$, which yields

$$\mathrm{WT} = \frac{\mathrm{kI}}{(1 - \mathrm{k})} \tag{8.2}$$

Because the quantity $1 - \mathrm{k}$ is less than one, this yields a wait time that is much too long. Equation (8.2) in fact describes what pigeons do on RID schedules (Wynne and Staddon 1988).

This is the answer to the fixed-ratio schedule puzzle. Evidently, pigeons have a built-in automatic timing mechanism that suppresses responding during

[1] On some RID schedules, a second response is required after the delay to produce the reinforcer. The optimal behavior in this case is to pause after the first response for a time proportional to the delay before responding again.

a predictable delay in between reinforcer deliveries—even when pause is maladaptive, as it is on ratio and RID schedules. It also explains the lack of a post-reinforcement pause on *variable*-ratio schedules, because here the interreinforcement interval is *not* predictable. It is as likely to be short as long.

So what can we conclude from these examples:

1. That animals behave optimally only under restricted conditions.
2. That behavior often follows quite simple, mechanical rules—such as linear waiting.
3. That these rules have evolved to produce optimal behavior only under *natural* conditions, i.e., conditions encountered by the species during its evolutionary history. Fixed-interval schedules are to be found whenever food occurs in a regular temporal pattern, but RID schedules, in which the timing of the food is initiated by the animal's own behavior, never.

8.2. What are the Alternatives to Optimality?

Historically, there are three ways to understand adaptive behavior:

1. *Normative*: Behavior fits the environment: it minimizes energy, maximizes food intake, reproduction, flow access (constructal theory), or, in economics, utility.
2. *Mechanistic*: Behavior is explained by a cause–effect process, which may be either physiological, as in Sherrington's account of the reflex, the Hodgkin-Huxley equations of nerve action, or conceptual/computational, as in Lorenz's hydraulic model of instinct or neural-network models of behavior.
3. *Darwinian*: This is also a causal approach, but more open-ended than option 2. The basic idea is simply that adaptation is best understood as the outcome of two complementary processes, *variation*, which generates behavioral options, and *selection*, which selects from among the resulting repertoire according to some criterion.

Option 1 is still the dominant approach in economics, but recent developments there—experimental economics, game and prospect theory, bounded rationality—increasingly call it into question as a general theory. Constructal theory, the topic of this book, is an optimization approach that has proven very successful in describing many physical and some biological systems. Given the serious problems encountered by other versions of optimality theory in describing learned behavior, it may be less successful there. On the other hand, nerve growth in the brain—a "morphing system" of the sort where constructal theory applies—maybe an area where these ideas will prove useful.

Option 2, a fully mechanistic account, is obviously the most desirable, since it provides a causal account of behavior that in principle may be related directly to brain function. Such accounts are not yet available for most interesting behavior.

Option 3 is a useful compromise. The Darwinian, selection–variation approach can be applied to any adaptive behavior even if details of the process are uncertain. All operant behavior can be usefully regarded as selection from a repertoire of behavioral variants (or, more precisely, a repertoire of generative *programs* or subroutines of which the observed behavior is the output: Staddon 1981). Selection—reinforcement—is then limited by the variants offered up by the processes of behavioral variation. In the raccoon and instinctive drift, e.g., once the animal had learned to manipulate the token, once it became predictive of food, the processes of variation generated the instinctive "washing" behavior, which in fact delayed reinforcement.

The same process has been studied in pigeons in a Pavlovian conditioning procedure called autoshaping. In autoshaping, a disk is briefly illuminated just before food is delivered. The hungry pigeon soon comes to peck the lit disk. The pecking seems to be elicited by the light much as Pavlov's dogs salivated to the bell. Moreover, the pigeon continues to peck, albeit at a reduced rate, even if the experimenter arranges that pecking turns off the disk and prevents the delivery of food. Apparently, a food-predictive stimulus powerfully elicits species-specific behavior which will occur even if it prevents the food.

This variation-selection process can be formalized as a causal model (e.g., the Staddon–Zhang model of assignment of credit in operant conditioning, which also offers an account for instinctive drift: Staddon 2001, Chapter 10). But it is useful even if not enough is known to permit causal modeling.

References

Breland, K. and Breland, M. (1961) *The Misbehavior of Organisms*, http://psychclassics.yorku.ca/Breland/misbehavior.htm.

Krebs, J. R. and Davies, N. B. (eds.) (1978) *Behavioral ecology*. Sunderland, MA.

Staddon, J. E. R. (ed.) (1980) *Limits to Action: The Allocation of Individual Behavior*. Academic Press, New York.

Staddon, J. E. R. (1981) Cognition in animals: Learning as program assembly. *Cognition* **10**, 287–294.

Staddon, J. E. R. (2001) *Adaptive Dynamics: The Theoretical Analysis of Behavior*. MIT/Bradford, Cambridge, MA: Pp. *xiv*, 1–423.

Wynne, C. D. L. and Staddon, J. E. R. (1988) Typical delay determines waiting time on periodic-food schedules: static and dynamic tests. *J. Exp. Anal. Behav.*, **50**, 197–210.