Psychology of Learning

Volume 2

PSYCHOLOGY OF LEARNING

VOLUME II

EDITED BY
DAVID SHANKS

Contents

Part 1:
Connectionism and Learning Theory (*Continued*)

~ 14 ~

On the Law of Effect

R. J. Herrnstein

A review of the evidence for the law of effect would quickly reveal that the simple notion of "stamping-in" (Thorndike, 1911, *e.g.*, p. 283) does not suffice. Animals do not just repeat the first successful act; they are likely to improve upon it until they find something like the optimal performance. In Thorndike's puzzle box, in the maze, or in Skinner's operant conditioning chamber, animals tend toward faster, easier, and more congenial movements, unless the performances are virtually optimal to begin with. Although some theorists find enough stereotypy to suggest a quasi-mechanical process of stamping-in (*e.g.*, Guthrie and Horton, 1946), others have remained unconvinced (*e.g.*, Tolman, 1948). Something more than the static form of the law of effect is needed for a really persuasive theory. The temptation to fall back on common-sense and conclude that animals are adaptive, *i.e.*, doing what profits them most, had best be resisted, for adaptation is at best a question, not an answer. And it is not hard to find evidence that violates both the Thorndikian principle of stamping-in and common-sense motions of adaptation, as the following two examples show.

Ferster and Skinner (1957) reported that an animal, when shifted from an interval to a ratio schedule, typically showed a change in its rate of responding. As regards stamping-in, the rate should remain unchanged for ratio schedules reinforce all rates of responding with equal probability (Morse, 1966). Although the deviation from the theory is large and reproducible, its direction is somewhat unpredictable. For example, in an experiment with pigeons (Ferster and Skinner, 1957, pp. 399–407), one subject's rate of responding increased while the other's virtually ceased, when the schedule changed from a variable interval to a variable ratio matched for numbers of responses per reinforcement. While both of these findings – both the increase and the decrease in the rate of responding in the shift from interval to ratio schedule – violate the Thorndikian

law of effect, only the increase is plausibly seen as adaptive. By responding faster on the ratio schedule, one animal increased its reinforcements per unit time, but, by the same token, the other one reduced its rate of reinforcement by responding more slowly. If the acceleration is adaptive, then the deceleration is not, and both findings are well substantiated.

A related finding, also violating both the Thorndikian law of effect and adaptiveness, has been obtained with the conjunctive schedule, schematically shown in Figure 1. This graph plots on the coordinates of a cumulative record the region over which responses are unreinforced. For the conjunctive schedule, it is the entire plane minus the shaded area within the right angle. In other words, the conjunctive schedule reinforces the first response after the occurrence of a certain number of responses (n on the figure) *and* the passage of a certain period of time (t). The schedule is specified by its component members: fixed interval and fixed ratio in the present instance. The conjunctive schedule in Figure 1 would be called a "conjunctive fixed-interval t, fixed-ratio n + 1." In the simple fixed-interval schedule, rapid responding is implicitly penalized, for faster responding increases the work per reinforcement. In contrast, ratio schedules exact no such penalty for responding quickly, for the amount of work per reinforcement is held constant. In fact, ratio schedules may favor rapid responding by arranging a direct proportionality between the rate of responding and the rate of reinforcement. The conjunctive schedule concatenates these features of ratio and interval schedules, since the rate of reinforcement is directly proportional to the rate of responding only for rates of responding no larger than n/t, beyond which the rate of responding covaries with the responses per reinforcement.
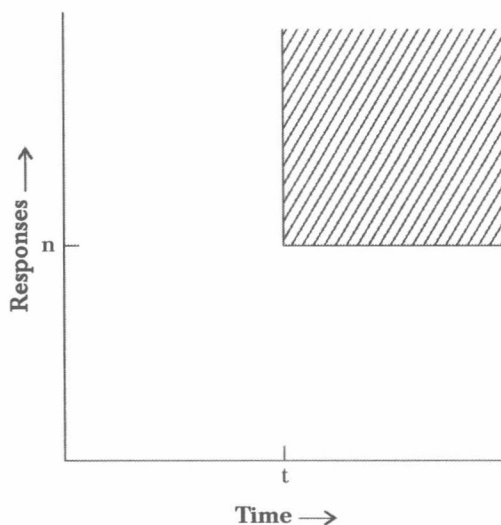
Figure 1. The shaded area shows the region of reinforced responding on a conjunctive schedule of reinforcement. The ordinate is cumulated responding; the abscissa is elapsed time

The relevant finding was obtained in an experiment (Herrnstein and Morse, 1958) that held the interval component constant at 15 min, but varied the ratio component from zero (which is a simple fixed-interval schedule) to 240. Figure 2 shows the relation between rate of responding and the number requirement imposed by the ratio component. Although the pigeons were responding more than an average 300 times per reinforcement on the fixed-interval schedule, a number requirement as small as 10 (for one of the pigeons) or 40 (for either), caused a detectable slowing down of responding. The range of rates of responding within individual fixed intervals is large enough so that number requirements even this small are likely to make contact with the behavior. Larger requirements caused progressively larger decrements in responding. This falling rate of responding reduced the rate of reinforcement, as Figure 3 shows. Here for the two pigeons are the average interreinforcement times as the number requirement was increased. For the fixed-interval schedule, the interreinforcement time was as small as the procedure permits, which is to say, 15 min. Even the smaller requirements produced some reduction in the rate of reinforcement. For one pigeon, the rate of reinforcement fell very sharply as the number requirement was increased; for the other, the decline was more gradual, but in either case, the requirement took its toll throughout.
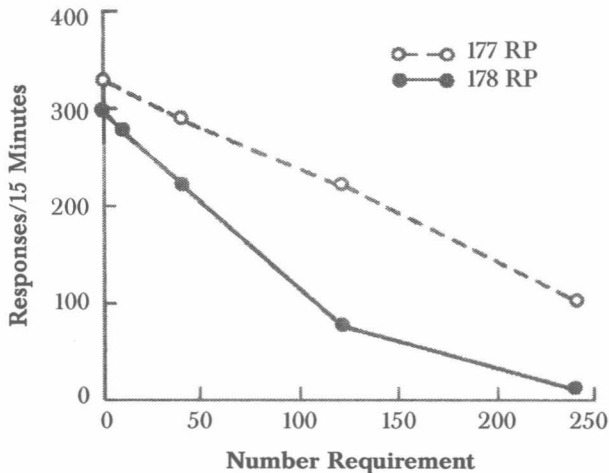
Figure 2. For two subjects, the rate of responding as a function of the size of the number requirement in a conjunctive schedule with a time requirement of 15 min throughout

The conjunctive schedule lends itself poorly to a Thorndikian analysis, for what could be getting stamped-in when responding and response requirement vary inversely? Moreover, the conjunctive schedule makes poor sense as regards the animal's best interests, for the animal may be emitting substantially more (in Figure 2, it was 30-fold more) behavior on the fixed interval than the number requirement demands, and yet the behavior is nevertheless depressed.
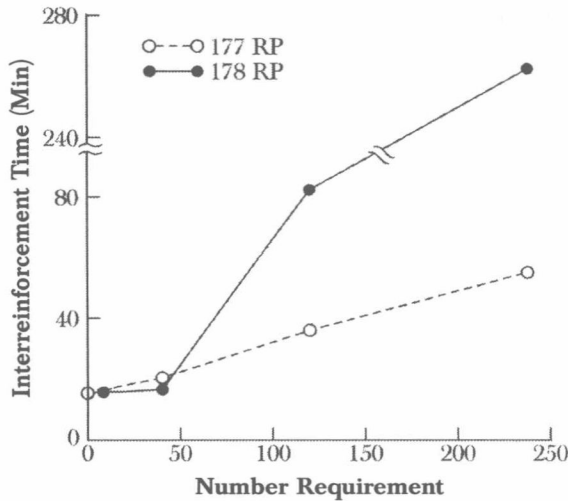
Figure 3. For two subjects, the interreinforcement time as a function of the size of the number requirement in a conjunctive schedule with a time requirement of 15 min throughout

These problem cases are troublesome only within the confines of theory. In the broader sphere of common sense, reinforcement is affecting what might be termed the "strength" of behavior, as reflected in its rate. For example, consider the change from interval to ratio schedules. If the first effect is a higher rate of reinforcement, the rate of responding might increase. But this would further increase the rate of reinforcement, further "strengthening" responding, causing it to rise again, which again pushes the rate of reinforcement upwards, and so on. If, on the other hand, the first effect is a lower rate of reinforcement, then the rate of responding should fall. The rate of reinforcement would then also fall, further weakening the responding, and so on again. This dynamic process occurs with ratio, and not interval, schedules because only the ratio schedule arranges a proportionality between the rate of reinforcement and the rate of responding. The proportionality is the basis for an instability in ratio schedules that should produce either maximal responding or none at all, an implication that is confirmed by the tendency of ratio responding to be "two-valued" (Ferster and Skinner, 1957, chap. 4).

The conjunctive schedule similarly exemplifies the notion of strength. However slightly the number requirement increases the interreinforcement interval, it should reduce the strength, and therefore the output, of responding. If the strength of responding is sufficiently reduced so that the rate of responding is less than n/t. (see Figure 1), then the conjunctive schedule becomes identical to a ratio schedule, and responding may reasonably be expected to vanish altogether, as it does with too-large ratios. One of the pigeons in Figure 2 had, in fact, virtually stopped responding at the largest ratio studied, even though the number requirement was still less than the responses per reinforcement freely emitted on the simple fixed-interval schedule.

Those two examples, and others like them, show that neither stamping-in, nor adaptation, nor the two of them together, can account for what is here being called the strength of behavior. This paper specifies more formally than heretofore the shape of this intuitively obvious concept, while staying within the general outlines of the law of effect.

Reinforcement as Strengthening

Reinforcement as strengthening is not being offered as a new idea, for "to reinforce" means to strengthen, and only by metaphor, to strengthen behavior. The earliest psychological usage concerned Pavlovian conditioning, where "reinforcement" of a reflex in a physiological sense was already familiar in classical work on facilitation and inhibition. The use of "reinforcement" in the vocabulary of instrumental learning was promoted in the mid-1930s, particularly by Skinner and primarily as a substitute for the traditional term "reward", whose very age tainted it with the suspicion of mentalism. Mentalism notwithstanding, "reward" was more neutral than "reinforce", for while reward simply names a class of events that have some effect on the organism, "reinforcement" implies what the effect is, namely a strengthening. The extra connotation was tolerable only so long as it was not contrary to fact, which it was not. The leading advocates of the law of effect – Thorndike, Skinner, and others – had from the beginning spoken in terms of a "strengthening" of behavior.

What, though, does it mean to strengthen behavior? Thorndike's answer was the notion of stamping-in, which may, in fact, be adequate for the acquisition of new behavior. But for behavior already learned, stamping-in seems inappropriate. The response form as such is then no longer changing, and yet, as the examples in the previous section show, reinforcement is still affecting what might be considered the strength of the behavior. The answers of others, like Skinner and Hull, addressed themselves sensibly if not successfully to the underlying problem, which is one of measurement. To say that behavior is strengthened is to imply some dimension of behavior along which it changes when its strength changes.

The measurement problem is empirical, not conceptual, which is not to deny the virtue of clear and original thinking. It is, rather, to point out that the only persuasive argument for any measure of response strength is to show orderly relations between the parameters of reinforcement – its frequency, quantity, quality, and so on – and the designated parameter of behavior. The traditional measures of response – probability, rate, amplitude (*i.e.*, work or effort), latency, resistance to extinction – have all failed to gain unequivocal support simply because orderly data with *quantitative* and *general* significance have not been forthcoming. Although there is no doubt that behavior is affected by its consequences, the law of effect is still expressed qualitatively, rather than as a relation between measurable variables, which it clearly must be at some level of analysis.

The notion of response probability comes closest to being a generally accepted measure of strength, cutting, as it does, across theories as diverse as those of Tolman (1938) and Hull (1943), Brunswik (1955) and Skinner (1953). But the agreement is more apparent than real, for the abstractness of "probability" masks the diversity of methods used for its extraction. For example, in some experiments, particularly those concerned with acquisition, the changing probabilities of response are estimated by the proportion of subjects doing something at successive points in training. In other experiments, single subjects are the basis for estimation of the probability by integrating over successive trials. In still others, the probability is estimated by the proportion of trials, or proportion of subjects, showing the choice of one response alternative out of a known set of alternatives. Not even the use of relative frequencies – the measure in modern probability theory – is common to all theorists, for according to Skinner, the rate of responding is the proper estimate of the organism's probability of responding. This is not an estimator in the formal sense – a mathematical probability is a dimensionless quantity between 0 and 1.0, and response rate is neither dimensionless nor bounded in principle – but rather an index of the animal's disposition to respond over some interval of time. Given the present state of knowledge, this abundance of measures is more likely to confuse than to enrich.

To reduce the confusion, and hopefully to advance the state of knowledge, the present approach focuses initially on a single relative-frequency measure as its index of strength. No "probability" will be inferred simply because to do so might suggest an equivalence with other empirical measures for which there is no evidence. The measure is exemplified by an experiment in which pigeons had two keys to peck (Herrnstein, 1961). The keys were available continuously during experimental sessions and pecking was reinforced with two variable-interval schedules, mutually independent and running simultaneously. The relative frequency is obtained by dividing the number of pecks on one key by the sum to both. In the context of operant conditioning this is a concurrent schedule, but it is clearly a version of the familiar "choice" experiment. It is, however, different in two significant respects. First, it uses continuous exposure to the alternatives instead of discrete trials. Second, reinforcements come on interval, instead of ratio, schedules. In the typical choice experiment, as well as in gambling casinos, the over-all probability of winning is constant for a given play (or response alternative), so that the number of wins is proportional to the number of plays. With interval schedules, there is no such proportionality, as noted earlier. Instead, the probability of winning on any given play is inversely related to the rate of play (response), and the number of wins is virtually independent of the number of plays, given a high enough rate of play.

The pigeons, then, had a pair of keys to peck, and the experiment rewarded their efforts with brief access to food at irregular intervals. The schedules set a maximum rate of reinforcement throughout the experiment at 40 per hour, but the number allocated to one key or the other was systematically varied, to see how the distribution of responses was affected. The question was whether,

to use the vocabulary of this paper, response strength as relative frequency was some plausible function of reinforcement frequency. The answer was both plausible and attractively simple, as shown in Figure 4. The ordinate is the proportion of responses on the left key; the abscissa is the proportion of reinforcements delivered thereon. The points fall near the diagonal, which is the locus of perfect matching between the distribution of responses and of reinforcements. A simple equation summarizes the finding (P is number of pecks, R is number of reinforcements, and the subscripts denote the two alternatives).

$$\frac{P_L}{P_L + P_R} = \frac{R_L}{R_L + R_R} \tag{1}$$

Unless the number of pecks far exceeds the number of reinforcements, the matching function (equation 1) is trivially true. For example, if the variable-interval schedules were assigning reinforcements to the two keys more quickly than the pigeons pecked, then every peck would be reinforced. The data point would necessarily fall on the diagonal, but the result would have little empirical content. If reinforcements were being assigned at the rate of one for every other response, or one for every third response, the possible range of variation for the points would still be narrowly constrained around the diagonal, and the findings would still be essentially vacuous. In fact, the possible range of variation is exactly fixed by the actual numbers of reinforcements and responses.
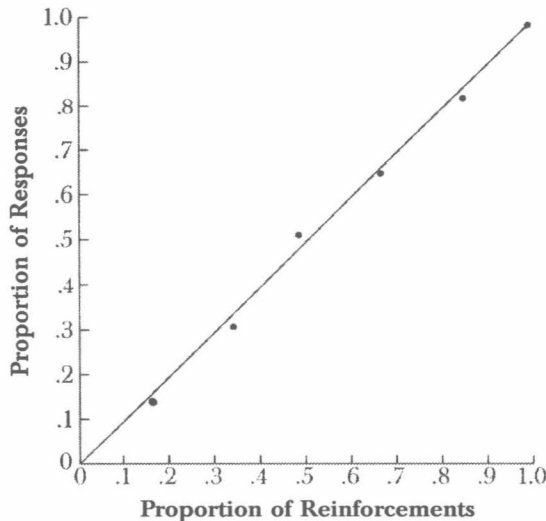

Figure 4. The relative frequency of responding to one alternative in a two-choice procedure as a function of the relative frequency of reinforcement thereon. Variable-interval schedules governed reinforcements for both alternatives. The diagonal line shows matching between the relative frequencies. From Herrnstein (1961)

Since there are at least as many responses on each key as there are reinforcements, the smallest relative frequency of response for a given relative frequency of reinforcement $R_L/(R_L + R_R)$ is

$$\frac{R_L}{P_L + P_R} \tag{2}$$

This fraction approaches $R_L/(R_L + R_R)$ as the total number of responses approaches the total number of reinforcements. The largest relative frequency of response for a given relative frequency of reinforcement is also dependent upon the fact that there can be no fewer responses than reinforcements. Thus, the ratio of responses for a given $R_L/(R_L + R_R)$ can go no higher than 1.0 minus the minimum value for the other key – $R_R/P_L + P_R)$ – which may be written as

$$\frac{R_L + P_R - R_R}{P_L + P_R} \tag{3}$$

This fraction, too, will approach $R_L/(R_L + R_R)$ as the total number of responses approaches the total number of reinforcements, which is to say when the responses on the separate keys each equal the reinforcements thereon. The result, then, is that as the number of responses approaches the number of reinforcements, the possible range of variation for the ratio of responses converges on to the matching relation in Figure 4. In the case of the experiment summarized here, however, the ratio of responses to reinforcements was approximately 100. At 0.5 on the abscissa, therefore, the possible range of variation of responding was from 0.005 to 0.995. At other values of the abscissa, the possible range was comparably broad. The close agreement between the matching relation and the distribution of responding says something, therefore, about the animal, not just about the procedure itself. What this is, and how it operates in a variety of situations, will occupy the remainder of the paper.

Response Strength and Choice

The experiment summarized in Figure 4 had an additional procedural wrinkle. Each time the pigeon shifted from one key to the other, there was a brief period of time during which any reinforcement called for by the variable-interval schedule was delayed. This "change-over delay" (COD) lasted for 1.5 sec, and was imposed in order to prevent the pigeons from switching after virtually every response. It was found, without the COD, that the distribution of responses tended to stay around 50-50, without regard to the distribution of reinforcements. If the matching relation were an accident of the duration of the COD, it would hardly be a principle of either response strength or choice. The most direct test of the COD is contained in an experiment in which it was varied systematically to see whether the matching relation was a peculiarity of only a certain range of durations.

The experiment, by Shull and Pliskoff (1967), varied a number of conditions besides the duration of the COD. Instead of pigeons, their experiment used albino rats. The reinforcer, instead of brief access to food for a hungry animal, was the opportunity for the rat to get electric current passed through its brain in the region of the posterior hypothalamus. A reinforcement consisted of the lighting of a small light in the presence of which the rat could press the lever 20 times, each time obtaining 125 milliseconds of a 100-Hz train of sine waves of 150 to 300 microamps across the electrodes implanted in its head. The variable-interval schedule resumed at the end of the last burst of current. The schedule itself was a variation of the simple concurrent procedure, one that had been originally described by Findley (1958). Instead of a pair of response alternatives associated with a pair of variable-interval schedules, the Findley procedure had the two variable intervals associated with a pair of stimuli, but responses on only one of the two levers could produce reinforcement. At any one time, only one of the two stimuli was present, and while it was present, reinforcements from only its associated variable-interval were forthcoming. The second lever in the chamber switched from one stimulus to the other, along with its associated variable-interval schedule. Actually, the two variable-interval programmers ran concurrently and continuously, just as they do in a conventional concurrent procedure. Shull and Pliskoff varied the COD, which is here the minimum possible interval between a response to the switching lever and the first reinforced response. Their finding was that matching occurred as long as the COD was greater than a certain minimal duration, as was found in the earlier study, but that beyond that value, matching was maintained whatever the duration of the COD in the range examined (0 to 20 sec). As the COD is made larger, however, it begins to affect measurably the obtained rates of reinforcement by interacting with the schedules themselves, as might be expected. Matching is always with respect to obtained, rather than pre-specified, reinforcement rates.

The experiment by Shull and Pliskoff extended the generality of the matching relation more than merely by showing that the COD is not the controlling variable. It extended the finding to rats, to Findley's procedure, and to intracranial stimulation as the reinforcer. Other studies have extended it further. Reynolds (1963a) showed matching with three response alternatives instead of two. Holz (1968) found matching even when each of the responses was punished with electric shock, in addition to being reinforced by the usual variable-interval schedules. Holz varied the intensity of the punishment until it was so severe that the pigeons stopped responding altogether. However, as long as they were responding, and as long as the punishment for the two responses was equally intense, the distribution of responses matched the distribution of reinforcements. Catania (1963a) and Neuringer (1967b) found matching with respect to total amount of food when the two reinforcers differed not in their rate of occurrence, but in the grams of food per reinforcement. In another study, Catania (1963b), using the Findley procedure, found matching both for the proportion of responses and for proportion of time spent with each of the two stimuli. Baum and Rachlin (1969) showed matching (correcting for a position bias) when the "responses" consisted of standing on one side

of a chamber or the other. The proportion of time spent in each location was found to be distributed as the associated proportion of reinforcements. Along these lines, Brownstein and Pliskoff (1968) found that the Findley procedure can be further modified so that when the animal selects one stimulus condition or the other, reinforcement comes along independently of any response. The pigeons here are simply choosing between one rate of reinforcement or the other. Their finding, too, is described by the matching relation, suitably adapted. The proportion of time spent in each stimulus condition is equal to the proportion of reinforcement received therefrom. Nevin (1969) noted that matching is found in human psychophysical studies when the proportion of "yes" responses is plotted against either the proportion of trials containing a signal or the relative size of payoff (*i.e.*, the frequency or magnitude of reinforcement). Shimp (1966), and the present author in unpublished work, have found matching in discrete-trial procedures of various sorts.

The list of confirmatory studies could be extended, but without profit at this point. It has been consistently found that responding is distributed in proportion to the distribution of reinforcement, as long as the responding and the reinforcements across the alternatives are not unequal qualitatively. Thus, the matching relation would not be expected if the reinforcer for one response were a preferred food and that for the other were a nonpreferred food, unless the scale values for the reinforcers expressed the difference quantitatively. Nor would it be expected if the two responses differed in some important way, *e.g.*, that one involved considerably more work than the other. In fact, the matching relation may be used to construct equivalences between qualitatively different responses or reinforcers, although no such undertaking has come to the author's attention. It should, however, be possible to scale reinforcers against each other or responses against each other by assuming that the subject must be conforming to the matching relation whenever it is in a choice situation of the general type employed in these experiments, and by adjusting the measures of response or reinforcement accordingly.

The main opposition to the matching relation is found in the literature on so-called "probability learning". If an experiment arranges a certain probability (excluding 1.0, 0.5, and 0) of reinforcement for each of a pair of response alternatives, and if the subject distributes its responses in proportion to these pre-assigned probabilities, then the matching relation, as defined here, is violated. Imagine that the two probabilities are 0.4 and 0.1. In a sequence of 100 responses, probability learning requires 80 responses to the better alternative and 20 responses to the poorer one. The number of reinforcements would be $80 \times 0.4 = 32$ for the one, and $20 \times 0.1 = 2$ for the other. With respect to the matching formula, this is a violation, for

$$\frac{80}{80+20} \neq \frac{32}{32+2} \tag{4}$$

The literature does not, however, claim strict conformity to probability learning. Instead, responding is often confined exclusively to one or the other of the two alternatives, typically toward the alternative with the higher probability

of reinforcement. But even when the two reinforcement probabilities are equal, responding tends to become exclusive for one of the choices. These deviations from probability learning are said to be instances of "optimizing" or "maximizing", since exclusive preference is the optimal strategy in the sense that the subject will, on the average, maximize its winning if it stays with the better bet. Even when the two probabilities are equal, nothing is lost by exclusive preference, and perhaps something is gained, for the subject is thereby spared the effort of switching from one alternative to the other.

Maximization in the probability type experiment actually conforms to the present matching function. Equation (1) is satisfied when all or none of the responses and all or none of the reinforcements occur for one of the alternatives and is therefore consistent with all of the experiments that deviate from probability learning and find maximization instead. Not all the experiments, however, find exclusive preference, and it is not yet clear whether apparently minor procedural factors are inadvertently affecting outcomes, as some have suggested (Bush and Mosteller, 1955), or whether there is, in addition, a phyletic factor. Bitterman (1965) argued that the higher organisms tend to maximize, while the lower ones, like fish and pigeons, tend to "probability learn". However, the experiments by Bitterman and his associates often use procedural features, such as forced trials to the alternatives, that complicate the calculation of the obtained frequencies of reinforcement, if not the psychological processes at work. In any event, pigeons do not invariably show probability learning; Herrnstein (1958) showed maximizing in pigeons given a choice between differing probabilities of reinforcement.

It is, in other words, not clear how much of the literature of probability learning actually violates equation 1, since it is not clear how much of this literature can be taken as firm evidence for probability learning. Nevertheless, suppose for the sake of argument that there is some valid evidence for probability learning, which is to say, that the responses are in the same ratio as the *probabilities* of reinforcement. How does this differ from the findings with rate of reinforcement, according to which the responses are in the same ratio as the *numbers* of reinforcement? The two findings turn out to be closely related mathematically, as the following set of equations shows, starting with the equation for probability learning:

$$\frac{P_L}{P_R} = \frac{\dfrac{R_L}{P_L}}{\dfrac{R_R}{P_R}} \tag{5a}$$

$$P_L^2 \, R_R = P_R^2 \, R_L \tag{5b}$$

$$P_L \sqrt{R_R} = P_R \sqrt{R_L} \tag{5c}$$

$$P_L \sqrt{R_R} + P_L \sqrt{R_L} = P_R \sqrt{R_L} + P_L \sqrt{R_L} \tag{5d}$$

$$\frac{P_L}{P_L + P_R} = \frac{\sqrt{R_L}}{\sqrt{R_L} + \sqrt{R_R}} \tag{6}$$