5 Instrumental learning

The things which move the animal are intellect, imagination, purpose, wish and appetite."

Aristotle, *De Motu Animalium*, 700b.

Although any mechanism which reduces responsiveness to a repetitious and insignificant series of events may have biologically useful functions, if only in energy-saving, and any mechanism which produces anticipatory reactions to the precursors of pain and pleasure has rather more obvious utilitarian possibilities, in neither of these cases is any account taken of on-the-spot cost-benefit analyses, which might lead to either active seeking of goals or more passive abstention from possible courses of action, according to evaluations of the pay-offs, positive or negative, thereby obtainable. The basics of habituation and classical conditioning do not therefore entail a truly purposive form of goal-oriented behaviour, even though Pavlov suggested that the arousal of food-seeking instincts by remote signs of the presence of food was part of the utility of the Pavlovian conditioning process. Thus, after considering theories which confine themselves to the effects of repeated stimuli, or of repeated stimulus pairings, we are entitled to ask whether a new theory is needed to encompass the learning of motor activities, especially those which are associated with wish and appetite. Various detailed ways in which the phenomena of classical conditioning may be contrasted with those of active and purposive, or, more conventionally, 'instrumental' learning, are discussed in chapter 6, but clearly the question of motivation, in the form of drives, goals and incentives, is an initial reason for wanting to add something to theories of classical conditioning pure and simple.

A second obvious reason why one would want to add something to theories of learning concerned only with habituation and classical conditioning is that they say almost nothing about the learning of new response strategies or skills. Typically, in habituation an instinctive response disappears, and in classical conditioning an already present instinctive response is seen on new occasions, and thus there is no direct allowance for large increases in the repertoire of available behaviours. It is possible, though far-fetched, to propose that anticipatory changes due to classical conditioning underlie new response patterns and skills (Watson, 1914/1967; Spence, 1956; Bindra, 1968, 1976). In a sense, all any of us can ever do is contract the muscles we were born with in different sequences and degrees of intensity, but the setting up of new sequences, whether in the human skills of carpentry and carpet-weaving or in the more limited performances of circus and laboratory animals, surely requires principles of response organization which could never be made obvious from the study of salivation or withdrawal reflexes.
Thorndike’s stimulus-response connections and the Law of Effect

A very ancient and very general principle of behavioural organization is that of reward and punishment responses are directed so that the unpleasant is avoided and the pleasant pursued. However, there are many and various possible internal mechanisms which could produce behaviour which corresponds to this vague descriptive generalization. Thorndike (1898, 1901) provided, at the same time as Pavlov’s discoveries in stimulus-pairing experiments, both a more specific descriptive law and a theory of the mechanism of response selection, both of which were extremely influential. At the descriptive level, filtering out as much as possible of Thorndike’s hypothetical mechanisms, his ‘Law of Effect’ says that individual responses are initially made at random, but are selected according to their effects, that is according to their consequences for the responding animal. This is ‘the method of trial and error with accumulated success’ (Thorndike, 1898, p.105), which has something in common with Darwinian natural selection, since particular behaviours which turn out to be advantageous to the animal are retained from a range of unorganized variation.

Figure 5.1 Thorndike’s problem box.

See text. After Thorndike (1898).
In the experiments which Thorndike performed, the change from variable reactions in the ‘trial and error’ phase to a single advantageous response was easily observed, and in some cases very gradual. The main experiments were performed by putting hungry cats inside small crates or ‘problem boxes’ (see Figure 5.1) from which they could escape: there were 15 boxes, each with a different mechanical system — a loop of string to be pulled, or a lever to be pressed, for instance — which allowed a cat to open a door if an appropriate response was made. A piece of fish was available outside the boxes, which a cat could eat if it got out, but the initial reactions of most of the cats he tested appeared to Thorndike to be directed at getting out of the box rather than reaching the food. In all the boxes these initial reactions were very similar in 11 of the 13 cats Thorndike tested (the other two being more sluggish). ‘When put into the box the cat would show evident signs of discomfort and of an impulse to escape from confinement. It tries to squeeze through any opening; it claws and bites at the bars or wire; it thrusts its paws through any opening and claws at everything it reaches; it continues its efforts when it strikes anything loose and shaky; it may claw at things within the box’ (Thorndike, 1898, p. 13). These various behaviours were counted by Thorndike as relatively undirected trial and error. But, in the course of these ‘impulsive struggles’ it was likely that the cat would, eventually and by accident, make the response which opened the door to the box it was confined in. For
example, in the simplest box, Box A, there was a large wire loop hung 6 inches above the floor in the centre, in front of the door, which opened the box if pulled, and on the first occasion on which the cats were put in this, most of them had accidentally opened the door within three minutes. But of course the important experimental result is that all the animals learned to do better than this, and escaped in less than a minute on the third and subsequent tests, on average, and in less than 10 seconds by the twentieth trial (see Figure 5.2). This is the experimental demonstration of the Law of Effect — the effects or consequences of a response change the probability that it will be performed in the future. Learning curves such as that in Figure 5.2 assess this in terms of the time taken to perform the response, or its latency, but the topography of the movements involved change as well, so that a well-trained cat,

instead of struggling wildly, will, as soon as it is confined ‘immediately claw the button or loop in a definite way’ (Thorndike, 1898, p. 13).

However, it was the theory which Thorndike used to explain the result, as much as the result itself, which attracted comment. Applying the rule of parsimony in explanation advocated in Lloyd Morgan’s Canon (Morgan, 1894), Thorndike attacked Lloyd Morgan’s own view that animal learning of this kind was an example of the association of ideas, that is, that the cats should be assumed to have associated the response of pulling a loop of wire and the outcome of getting out of the box and eating fish (Thorndike, 1898, p. 66). As an alternative, Thorndike proposed that learned associations took a much simpler form, being always between the sense-impression of an external stimulus situation and a direct impulse to act. The consequences of actions, in terms of perceived pleasure or pain, or in terms of ‘satisfying’ or ‘annoying’ states of affairs, were important in Thorndike’s theory only because pleasurable results ‘stamped-in’ or burned in the situation-response connection which led to them, while, at least in the early versions of his theory (see chapter 7), annoying results of an action stamped such connections out again. ‘The impulse is the sine qua non of the association’, and thus animals must have the impulse to act, and perform an act, before any associations are formed, and animals must always learn by doing and by actual experience of actions followed by reward. Thus Thorndike did not expect, nor did he find, learning taking place by one animal imitating what another animal did, even in his experiments on cebus monkeys (1901). For monkeys and other primates, and even eventually for people (Thorndike, 1931), Thorndike proposed that situation-response, or stimulus-response, connections were the basis of all learning. This proposal clearly bears some resemblance to Pavlov’s idea of the formation of temporary connections in the nervous system (pp. 60—2), but the crucial addition made by Thorndike was immediate effect of reward and punishment in stamping-in or stamping-out connections between quite arbitrary preceding responses and their eliciting stimuli (see chapter 6).

As we shall see, the basic results of Thorndike’s exper-
reward and punishment take place via the formation of stimulus-response connections. It is therefore worth noting briefly some further details reported by Thorndike which, he admitted, did not seem consistent with the most extreme position which he considered — the possibility ‘that animals may have no images or memories at all, no ideas to associate’ (1898, p. 73, original italics). One such detail is the fact that cats were very slow to learn to lick or scratch themselves if these responses were rewarded by the opening of the door (by Thorndike). This would not be expected if the only principle at work was a mechanical tendency to repeated any rewarded act, but various additional explanations can be offered, such as ‘an absence of preparation in the nervous system for connections between these particular acts and sense-impressions’ (Thorndike, 1898, p. 76; see Seligman, 1970). The more elaborate alternative is to suppose that the idea of clawing a loop to get out is more readily formed than the idea of licking oneself to get out. Indeed, a general feature of Thorndike’s discussion of his results is an appeal to the process of attention. A particular response such as pushing a lever was more likely to be learned if the animal ‘pays attention to what it is doing’ (1898, p. 27). This factor is also appealed to to explain why it is that individual cats learned more and more quickly as they had experience of more and more difficult problem boxes. Part of Thorndike’s explanation for this was that ‘the cat’s general tendency to claw at loose objects within the box is strengthened, and its tendency to squeeze through holes and bite bars is weakened’, but in addition he noted that ‘its tendency to pay attention to what it is doing gets strengthened, and this is something which may properly be called a change in degree of intelligence’ (1898, p. 28).

Finally, in ignorance of Pavlov’s work, Thorndike performed a stimulus-pairing experiment, which led him to the view that his cats formed representations of sense-impressions (1898, pp. 75—6). A hungry cat was kept in a large enclosure with a side of wire netting. Outside this Thorndike practised a routine of clapping his hands (and saying out loud, ‘I must feed those cats’), ten seconds before going over and holding a piece of fish through the netting, 3 feet up from the floor. It will surprise few readers to learn that, after this had taken place 30 times, the cat climbed 3 feet up the wire netting immediately Thorndike clapped his hands, without waiting for the actual sight of the fish. Thorndike tentatively took the view that this was because the cat had some anticipatory idea or representation of the fish being presented in this case, without, however, acknowledging any anticipation of getting out and eating fish as a necessary factor in the string-pulling and lever Pressing experiments.

Hull’s stimulus-response principles and reward by drive reduction

It can be seen from the example above that Thorndike was less than completely rigorous in his application of the principle that learning takes place when motivationally significant events stamp in connections between sense-impressions and arbitrary impulses to action. His account of learning relies to a large extent on subjective factors, even though he
minimized the importance of inference and reasoning in animal experiments, since the learned connections were supposedly between the internal perception of external stimuli, which could be significantly modified by attentional processes, and a subjective compulsion to perform a given response. Equally obvious is the fact that the ‘satisfaction’ or ‘pleasure’, which is responsible for burning in preceding stimulus-response connections, is a hypothetical mental state of the subject, albeit one that can be inferred on the basis of behavioural evidence. Hull (1937, 1943, 1952) attempted to provide a far more rigorous and systematic theory of learning based on the principle of automatic stimulus-response connections, which had a far-reaching influence on many areas of both academic and applied psychology, but which stands today mainly as a monument to the limitations of this approach.

Hull attempted to strip away from Thorndike’s Law of Effect all the more mentalistic or psychological extra processes involved in the perception of the environment, the organization of response output, and the emotional assessment of reward and punishment. Learned connections were no longer between sense-impressions of what the animal was attending to and its impulse to perform a particular act, but directly between the neural impulses from activated sensory receptors and certain muscular reactions — the term Hull used was ‘receptor-effector connections’. The complexities of attention and perception were thus (temporally) side-stepped. Although this approach was initially plausible in terms of biological function, Hull also attempted what turned out to be an even more dangerous simplification by tying all motivation to states of physiological need, avoiding any reference to subjective states of pleasure or satisfaction. Animals have a biological need for food; lack of food was thus supposed to produce internal physiological stimuli described as the Qdrive’ of hunger. This would serve two purposes: the drive would in the first place be expected to activate and energize behaviours, but it also had an essential place in Hull’s theory of reward, since eating food was supposed to affect the animal only in so far as it reduced the drive of hunger or the physiological deficit in nutrition.

Hull adopted Pavlov’s (and Skinner’s) term of ‘reinforce-merit’, for the event which strengthened stimulus-response connections, and thus Thorndike’s Law of Effect became the Law of primary reinforcement, stated as follows:

> Whenever an effector activity occurs in temporal contiguity with the afferent impulse, or the perseverative trace of such an impulse, resulting from the impact of stimulus energy upon a receptor, and this conjunction is closely associated in time with the diminution in the receptor discharge characteristic of a need, there will result an increment to the tendency for that stimulus on subsequent occasions to evoke that reaction. (Hull, 1943, p. 80)

It was clear even to Hull that there might be a digestive delay before
eating food actually changed physiological need states, and therefore he proposed that, normally, eating food reinforced preceding stimulus-response connections not by the ‘primary reinforcement’ of reducing the state of need, but by ‘secondary reinforcement’ caused because the stimuli involved in eating will have been ‘closely and consistently

associated’ with eventual need satisfaction (Hull, 1943, p. 98). There is no doubt that particular experiences of eating do indeed acquire motivational significance partly because of subsequent metabolic changes (see for instance pp. 232ff on taste-aversion learning, and Rozin and Kalat, 1971), and the appeal to drive reduction has the advantage of being directly applicable to the learning of responses which allow escape from external imposed aversive states (see chapter 7). However, the notion that the reduction of needs or drives is in any sense a necessary condition for learning has been almost universally abandoned (although see Hall, 1983, pp. 206—13). The most important reason for this is that the learning which occurs in the course of habituation, more active exploration of the environment (see discussion of latent learning, pp. 150—2) and in several forms of Pavlovian conditioning, all lack the support of any physiological imbalance as obvious as those involved in hunger or lack of food. Further, learning under the influence of general social motivation (Thorndike rewarded chicks by allowing them access to the company of their fellows: see discussion of imprinting, pp. 23—4), or that rewarded by specific sexual stimuli (Sheffield, et al., 1951 showed that proximity to a female rat, even without completed sexual activities, was rewarding to males of the species) appears to be governed more by the properties of external events than by hypothetical internal needs (Hinde, 1960, 1970). But, even in the prototypical case of the reduction of hunger by eating, motivational effects are considerably more complex than originally suggested by Hull. The tendency of human beings to eat things which are not medically good for them (for instance, large amounts of sweetened saturated fat in the form of chocolate or ice cream) itself suggests that an abstract conception of biological need is not an adequate guide to motivational significance. A general alternative to drive reduction, indicated for instance by the apparent equivalence of low-calorie and high-calorie sweeteners as reinforcers for rats (Sheffield and Roby, 1950), is that the stimulus properties of food substance when ingested, as well as their post-ingestional consequences, may have built-in rewarding effects.
Figure 5.3 The emotional effects of rapid incentive learning.

Results of an experiment in which the amount of food obtained as reward by rats for running to the end of a straight alley was varied, using units of a 50th of a gram. Rats trained with larger rewards ran faster than a trained with a small reward, but when all then received that small reward, a rapid reduction in speed of running took place in those rats which had formerly been given large rewards, so that they ran more slowly than rats already accustomed to receiving small rewards. After Crespi (1942).

However, the experimental results which led Hull to modify his theory (though not to abandon drive reduction altogether) suggested that even the physical properties of reinforcing agents themselves do not adequately predict their behavioural effects, but rather, an internal evaluation of a given reward determines its effects on the learned performance of hungry rats. Crespi (1942) compared the performance of rats trained to run down an alley for food rewards of just less than a third of a gram with that of groups trained first with larger or smaller rewards, and then shifted to this reward amount. Figure 5.3 illustrates the case of rats which first experienced
5-gram rewards, and were then given a fraction of that amount for the same response, showing that their speed of running dropped sharply, and to a level well below that of animals’ which had received the smaller reward consistently. Crespi’s own interpretation of this was that the reduction in reward size engendered an emotional depression in the animals. Hull (1952, p. 142) retained the notion of a temporary emotional effect caused by reward reduction to account for the drop in response speed below the normal level, but also added a special process of ‘incentive learning’, detached from both drive reduction and the formation of stimulus-response connections, which he used to explain both the rapid change in performance in this experiment, and the immediate improvement in accuracy of maze-running when rewards are introduced following a period of unrewarded exploration (‘latent learning’ — see pp. 150—2). There are a number of other results from standard laboratory experiments, under the heading of ‘contrast effects’ (Dunham, 1968; Mackintosh, 1974, pp. 213—6) or ‘post-conditioning changes in the value of the instrumental reinforcer’ (Mackintosh, 1983, pp. 80—6) which seem to indicate that a re-assessment of the emotional value to be put on a given reward can take place in separation from the response-learning process, and some of these will be discussed in a later section (pp. 137—7). It is usual now to conclude that ‘instrumental learning cannot be reduced to an S-R association’ (Mackintosh, 1983, p. 86), and since the drive reduction factor in the reward process tends also to be discounted, Hull’s reformulation of the Law of Effect has lost both its mainstays.

**Skinnerian pragmatics**

The best-known modern protagonist of the Law of Effect is B. F. Skinner, who has gradually abandoned almost all theoretical speculation as to the role of drive, stimulus-response connections, or any other internal mechanism or process which might explain how or why learning takes place, in favour of empirical demonstration of the practical power of reward procedures in the animal laboratory, and rhetorical assertions about the ubiquity of strictly analogous effects in human life. I shall for the present confine myself to a discussion of the former, and ignore the latter (see Walker, 1984, chapter 10).

Thorndike’s technique for studying learning in cats and dogs required him to replace the animal inside a box each time it made the successful response of getting out. In his experiments on cebus monkeys however, Thorndike found it expedient to change this procedure so that the animal remained inside an experimental enclosure, a piece of banana being dropped to it down a chute whenever it depressed a lever, Skinner (1938) used a similar set-up for laboratory rats, with automated devices for recording lever presses and delivering food pellets, and later developed a version which was used in very extensive experiments on pigeons (Ferster and Skinner, 1957) thus inventing what is now a very widely used piece of apparatus in animal laboratories, known as the Skinner box. The isolation of the animal subject, and the automatization of the experimental procedures, may be partly responsible for the fact that a wide variety of
extremely reliable (and replicable) behavioural results can be obtained by this method, and it is of course extensively used today by researchers with no commitment to Skinner’s theoretical views. In fact, Skinnerian apparatus and behavioural techniques are used in so many theoretical contexts that it is quite inappropriate to gather together all the obtainable results here, but the basic phenomena associated with ‘schedules of reinforcement’ can usefully be reviewed at this point.

Schedules and contingencies of reinforcement

There is a good deal in common between certain kinds of Skinnerian reinforcement theory and general principles of economics (see, e.g., Lea, 1978, 1981; Rachlin et al., 1981). It is assumed in economics that certain global aspects of human business — savings, investment, employment and so on — may be ultimately determined by certain quantifiable variables — money supply or interest rates for instance, without too much detailed attention being given to the actual psychology of individual decision-making or the gradual change or growth of business enterprises. The success (or failure) of such assumptions rests partly on the degree to which people or companies do indeed act predictably in always acting to maximize their own economic advantage. In the same way it is a central assumption of reinforcement theory that animals will learn anything and everything which maximizes their receipt of food rewards, with many subtle and global aspects of learned behaviour thus being ultimately determined by quantifiable variables concerned with the conditions of availability of food rewards (or ‘contingencies of reinforcement’). The end relationship between the scheduling of food rewards and the exact patterns of behaviour which are eventually determined by various conditions of reward can be and often is discussed without too much attention being given to gradual learned changes in behaviour (and the mechanisms which might be responsible for these). However, there can only be fixed relationships between contingencies of reinforcement and patterns of behaviour engendered by them if animals act predictably in accordance with the Law of Effect, and at the descriptive level always select behaviour which has beneficial consequences. This is not always the case — individual species are influenced by particular reinforcers in ways which reflect instinctive processes that are relatively immune from experienced payoffs (Roper et al., 1983; Breland and Breland, 1961; see chapter 6) — but with the now standard laboratory procedures first used by Skinner, the predictability of relationships between contingencies of reinforcement and behaviours learned thereby is surprisingly high.

Continuous and ratio reinforcement

The simplest possible Skinner-box experiment, analogous to the Pavlovian case where a buzzer always signals food, is performed when an animal is able to deliver to itself a small amount of food each time it performs a single response such as pressing a lever or pushing a button. Left to itself, an individual of any common laboratory species will eventually discover this by trial and error, if the response is such as is
made occasionally by accident; but it is usual to give prior adaptation to
the working of the food delivery mechanism, and

there are other methods of training, such as autoshaping (see pp. 84—5
and chapter 6), or ‘shaping’ the response by care-fully watching the animal
and delivering reward for successive approximations to the required
behaviour (e.g. delivering rewards initially when a rat sniffs at a lever,
then only when it touches it, then only when it presses the lever down).
Once such a response has been learned, then if the animal is already
hungry when it is placed in its experimental box, it will press the lever at
regular intervals, gradually slowing down if it is allowed to become
satiated. After such training, an additional learning process can be
observed by disconnecting the food mechanism, since a hungry animal
will at first continue vainly to make the response but will gradually cease
to do so — the process of response extinction. Little can be concluded
from this simple confirmation of trial and error learning, but the nature of
the procedure allows for many extensions of the basic task. For instance,
the animal may be required to make a fixed number of responses for each
reward — a fixed-ratio schedule. This involves gradual learning,
especially if the fixed number is large. Clearly the animal is not able to
obtain rewards as frequently if it has to make 100 responses before each
one, but additional standard results are that there is a pause in responding
after each reward, roughly proportional to the size of the ratio, and that a
run of responses is then made at a fast and gradually increasing rate
(Ferster and Skinner, 1957). Obviously, the higher the ratio the higher in
some sense is the cost in time and effort of obtaining rewards, and this is
reflected in the variation both in vigour of performance and the size of the
highest ratios that animals will continue to perform, which are
systematically related to motivational factors such as degree of prior
depprivation and the size and palatability of the rewards used (Hodos, 1961;
Hodos and Kalman, 1963; Powell, 1979). Ecologically speaking, the
relationships between effort and food reward would be expected to be
vastly different between species, depending on niche. For instance,
herbivorous browsers, need to feed for long periods but carnivores
(especially cold-blooded ones), eat relatively infrequently and difficult-to-
obtain but richer meals. However, it appears that rats, being omnivorous,
are extremely flexible in this respect, when they

are left permanently in a Skinner box with various ratios of lever-pressing
necessary to obtain food rewards (Collier et al., 1972). If allowed one
small pellet for every press, then about once every hour they would go and
press 20 times in succession, eating 20 pellets in less than 10 minutes.
When required to make 160 presses for each pellet, they were able to
obtain almost as much food as before, by spending 14 hours a day making
the necessary 70,000 responses. Even more flexibility was shown if the
unusual procedure was used of rewarding rats by allowing them to enter a
tunnel and eat for as long as they liked — but with the constraint that if
they left the tunnel for more than 10 minutes then a fixed ratio of lever
presses was necessary to make food available again. In this case rats
required to make only one response to produce free access to food ate 10
separate meals a day, but if the ratio was increased they chose, reasonably enough, to eat fewer but longer meals, eating only five times a day when 40 responses were necessary to gain access to food, and eating only one large meal per day when just over 5,000 lever presses were required to gain access to food. In this experiment (Collier et al., 1972) water was always available, but exactly analogous results were obtained when fixed ratios of responses were necessary to obtain water, with food always available, except that with indefinitely long access to water as the reward, a ratio of only 300 bar presses was required to persuade the rats to take just one (long) drink of water per day (Marwine and Collier, 1979). This strongly suggests that frequent access to food is rather more rewarding for individual rats than frequent access to water. Lore and Flannelly (1978) deduced that access to water is ecologically important from their naturalistic study of the location of the burrows of wild rats on a large landfill, since these tended to be located within 50 metres of a stream but only within 100 metres of the food source, (a mound of refuse). Clearly many factors, notably in this case an obvious preference for sloping ground, may determine burrow location.

The general descriptive conclusion from experiments on fixed ratio schedules is simply that the size of the ratio significantly affects the pattern of responding for a given reward.

The theoretical implications of this are not always clear, but several experiments suggest that laboratory animals are capable of learning some representation of the size of the fixed ratio (Adams and Walker, 1972; Mechner, 1958; Rilling and McDiarmid, 1965). That this is not strictly necessary for performance on ratio schedules is shown by the possibility of variable ratio schedules, in which a probabilistic device such as random number generator ensures that the exact point at which reward is given is always unpredictable, even though the average value of the ratio can be specified. Only a limited amount of evidence about the effects of this is available, but some rather complicated comparisons suggest that a variable ratio schedule produces somewhat faster responding than a fixed ratio of the same average value, with shorter pauses after reward is received, due to the effects of the occasional reinforcement of a single response or a very short run of responses. Given the choice, pigeons appear to prefer to work on the uncertain variable ratio than on a fixed ratio of equivalent average value (Ferster and Skinner, 1957; Fantino, 1967; Sherman and Thomas, 1968).

Interval schedules of reinforcement

It also appears to be the case that variable ratio scheduling of reinforcements produces a higher rate of response than the same number of reinforcements (that is the equivalent frequency over time, and thus the same rate of gain of food) which are made available according to a different rule, based on the minimum interval of time between one reward and the next. For a variable interval schedule of reinforcement, a certain period of time, random about some mean, must elapse after one reward has been given before a single response produces the next. Since only one response per reward is necessary, it is not surprising that fewer are made
than when multiple responses are mandatory on variable ratio schedules (Ferster and Skinner, 1957; Thomas and Świtalski, 1966; Peele et al., 1984; Zurif, 1970). Variable interval schedules are very widely used because they produce a very consistent and steady rate of responding, which accelerates only slightly as time without a reward passes (Ferster and Skinner, 1957; Catania and Reynolds, 1970). For instance, if hungry pigeons can obtain reward for a single key peck on a variable interval of 1 minute, they typically begin pecking the key again within a few seconds of receiving food. Very occasionally they are rewarded at this point, but at other times they may continue to peck regularly, without anything else happening, for 4 or 5 minutes, until the next reward finally becomes available.

The theoretical analysis of performance on interval schedules is easier in the alternative case of a fixed interval, where the minimum period between successive rewards is always the same. Here the typical performance in a rat or pigeon, depending on the length of the interval, is that there is a distinct pause after a reinforcement has been received (as with fixed ratios) followed by a gradual increase in the rate of responding, such that the highest rates of response occur just before the next reinforcement becomes due. This increase in response rate with time, along with other measurements, has been used to deduce that laboratory animals make use of an ‘internal clock’, which can be accurate to within a few seconds, to govern responses (Dews, 1962, 1970; Roberts and Church, 1978; Roberts, 1981; Meck and Church, 1984; Gibbon, 1977). According to Roberts (1981), this clock is a linear measure of time, which can be reset by food rewards, stopped at a given point by appropriate external signals, and used to time intervals of several different lengths concurrently. This is rather hypothetical, but it is clear that one of the factors that may be involved in performance on interval schedules is the engagement of some sort of representation of the passage of time. Whether the internal clock is associated with expectancies of reward, or serves rather as a stimulus for more mechanical habits of response, is not usually obvious.

**Choice in reinforcement schedules**

The reinforcement schedules described so far have all concerned the relationship between a single response and the availability of reward. A strikingly more complex state of affairs arises if an animal is confronted with two available responses, each assigned its own schedule of reinforcement. A degree of orderliness prevails, however, if both available responses are rewarded on variable interval schedules. Suppose a pigeon is allowed for one hour every day to obtain grain in a box where pecking the left key is rewarded on average every 3 minutes (VI3min) and pecking the right key is rewarded on average every 6 minutes (VI6min). What is the optimal way for it behave? The answer, of course, depends on what one means by ‘optimal’. To gain the most food with the least amount of effort, physical or intellectual, it would be sufficient for the bird to peck the keys alternately, about once every 10 seconds. In fact, with the schedule as
described, and a not-very-hungry pigeon, this is exactly what happens — the bird pecks at roughly the same rate on both keys, even though it gets 20 rewards per hour from the left key but only 10 from the right. The nature of the variable interval schedules means that the distribution and rate of responding has very little effect on the distribution of rewards, provided that both responses are made occasionally. But, by increasing the psychological separation between the two responses, the effects of rewarding one twice as often as the other can soon be observed: the standard procedure is to ensure that a left response is not rewarded if it is made just after a right response, or vice versa (a ‘change-over delay’, Herrnstein, 1961; Silberberg and Fantino, 1970).

Once this is done, data of impressive regularity may be obtained, since the procedure of giving twice as many rewards for left pecks as for right pecks then means that pigeons peck left twice as often as they peck right (Herrnstein, 1961, 1970). The interpretation of this unsurprising result has, however, proved to be difficult, and theories of increasing mathematical complexity are now proposed to account for it (Herrnstein, 1970; Baum, 1974; Rachlin et al., 1981; Prelec, 1982, 1984; Shimp, 1969; de Villiers and Herrnstein, 1976). For present purposes it is sufficient merely to stress that the regularity does not arise only because a response is strengthened in proportion to the exact number of times it is rewarded.

There is a descriptive generality, called the ‘matching law’, which can be given either as

\[
\frac{B_1}{B_1 + B_2} = \frac{R_1}{R_1 + R_2}
\]

(1)

or as

\[
\frac{B_1}{B_2} = \frac{R_1}{R_2}
\]

(2)

where \(B_1\) and \(B_2\) are the absolute frequencies of two behaviours, and \(R_1\) and \(R_2\) are the absolute numbers of rewards each receives per hour. There are two general kinds of explanation for why this equation satisfactorily predicts behaviour, and therefore predicts in some sense the choice that a pigeon makes between two responses. The first kind says that the matching law arises because in some way or other animals are sensitive to the ratio of reinforcements for two responses and adjust their response choices accordingly. The second supposes that the matching relationship arises because the animals are attempting to respond optimally and that the observed allocation of responses results from their ‘maximizing’ their momentary chances of reward. Although these theories sound very different, as it happens close examination of what the two theories predict about experimental results suggests that they are ‘empirically indistinguishable’ (Ziriax and Silberberg, 1984; see Mackintosh, 1983, p. 258).

Relative reward value

There is, however, an extremely unambiguous empirical distinction between two possible rules which would both produce the matching law in
practice. One rule implies that any response is performed according to the reinforcements it itself receives; the other says that a response is performed according to the *relative* value of the reinforcements it receives — most directly according to the proportion of the total rewards that are assigned to it in circumstances of choice. Thus

\[
B_1 = kR_1
\]  

(3)

is the first case and

\[
B_1 = \frac{kR_1}{R_1 + R_2}
\]  

(4)

is the second. Either equation (3) or equation (4) would produce equations (1) and (2). But, ‘the data unequivocally support’ equation (4): (Herrnstein, 1970). This is because there is a very simple experimental test: a pigeon is rewarded at a standard level — say 20 times per hour, for pecks on the left key; while over a period of days or weeks, the frequency of rewards for right key pecks is varied between zero and 40 per hour. Even though the left key pecks are being rewarded at the same level, the rate of left-key pecking varies systematically, roughly in line with equation (4), although the fit of the data is even better when the total frequency of rewards in the denominator is raised to a fractional power (Catania, 1963). As with the matching law itself, the reason why equation (4) works is more obscure than might be expected. We can rule out the argument that left-key pecks are made less often if right-key pecks are rewarded a lot, because the animal is then too busy making right-key pecks, since it can be arranged, by giving a visual signal for exactly when they will be rewarded, that it gets many rewards for right-key pecks without having to spend much time on that behaviour (Catania, 1963).

The descriptive alternative is that the effectiveness of a certain level of reward on the left key declines when an alternative source of reward is available. It is likely that this sometimes involves perceptual or emotional contrast effects between two levels of incentive experienced by the same animal, but a simpler possibility is that the satisfying or even satiating effects of one reward alternative detract from overall drive in general, or from the incentive properties of all alternatives. (Crespi, 1942; Walker et al., 1970; de Villiers and Herrnstein 1976). Hull (1952) found it necessary to distinguish between the drive-reducing and the incentive effects of rewards, and (at least) these two factors probably have to be taken into account. The satiating effect of rewards seems obvious, for instance, in the experiment of Rachlin and Baum (1969), in which pigeons received 20 rewards per hour (VI 13 min) and 4 seconds access to grain, for pecks on their right key, while the same number of signalled rewards on the left key varied in duration from 1 to 16 seconds. When this alternative was 16 seconds, the birds gained 10 grams more weight in the experimental hour than when it was 1 second, and this would be sufficient to account for the fact that the
rate of response for the standard reward was slower in the first case.

However, a fairly general finding is that the duration or magnitude of rewards given in concurrent schedules has much less effect on behaviour, and produces much less close matching of behaviour to relative reward properties, than does the variable of frequency of reinforcement (Walker et al., 1970, Walker and Hurwitz 1970; Schneider, 1973; Todorov et al., 1984). In other words, animals are much more likely to notice that they get an extra small reward from an alternative source, than that they get a reward which is twice as big. Frequent access to small amounts of food is generally much preferred to infrequent access to large amounts (see Collier et al., 1972, p. 131), and thus it is the details of the distribution of rewards over time, rather than the gross total of consumption, that has the strongest psychological effect in the experiments.

**Skilled performance and the form of a response**

In the use of Skinnerian techniques for detailed laboratory work on ‘the experimental analysis of behaviour’, the emphasis is very often on the rate at which a given response is performed, with little attention given to the initial learning which determined its qualitative form. But reward procedures have just as much of an effect on ‘how’ as they do on ‘how often’ a response is performed. Skinner (1938) described the training of rats to press a lever with a certain force, or to press it down for a certain length of time. As a general rule, he found that ‘Rats tend to adjust to a force which secures only slightly above the reinforcement of every other response’ (1938, p. 317). Thus, if a lever needs to be pressed with a force equivalent to a dead weight of 20 grams, even experienced animals will make up to 40 per cent of their attempts at pushing it down with insufficient effort. They are strong enough to do better than this, since if the criterion is shifted, and a mechanical adjustment made so that a press of 60 grams or more is needed, then the animals quickly learn to make harder presses, but still make a third or more partial presses of less than 60 grams in force. Skinner found that one animal which weighed less than 200 grams itself could sustain presses averaging over 100 grams each, but even at much lower force requirements there seemed to be a principle of least effort bringing responses down below the minimum force necessary.

Normally for food rewards rats press a lever for only a fraction of a second. However, they may be gradually trained to hold it pressed down for 20 or 30 seconds at a time (Skinner, 1938), and various more complex techniques can be used to assess the accuracy of such timing, when there is no external feedback to signal when the lever has been pressed long enough. A further possible variation in the topography of a rat’s lever-pressing is the amount of angular movement:

Herrick (1964) succeeding in training his animals to depress a lever through an angular displacement of at least 20.35 degrees, but not more than 25.50 degrees. Rewards were delivered only when the lever returned to its home position after excursions which met this criterion, and therefore the judgment of a correct lever press required some muscular
finesse on the part of the rats (and some technical ingenuity in the design of the apparatus). However Herrick’s rats satisfied Skinner’s expectation that they should adjust their behaviour so that at least every other response gained reward.

Although the acquisition of motor skills in animal learning is sometimes overlooked, it is clear that laboratory animals, starting with Thorndike’s cats, are able to learn to operate with considerable accuracy mechanical apparatus that is quite different from anything members of the same species would ever encounter in a natural environment. Often natural and instinctive behaviour patterns are also elicited (see chapter 6), but it is undeniable that new muscular patterns of some skill emerge under the influence of artificial training procedures.

**Creativity in response selection**

It is a Skinnerian dogma that changes in the form of a response always reflect the shaping effect of contingencies of reinforcement in the environment on a passive subject. This probably overstates the case even for laboratory rats and pigeons, but an explicit counter-example is available from

systematic experiment on two rough-toothed porpoises, members of the only mammalian group, the whales, better endowed with brain tissue than the primates. Pryor et al. (1969) were initially attempting to demonstrate only the power of the Law of Effect in the shaping of new behaviours by reward. For five days a porpoise was, in the course of public performance at a sea life park in Hawaii, rewarded for a different particular behaviour each day, but the procedure had to be aborted because the animal began to perform a large number of response sequences which not only had not been specifically trained, but which had not previously been observed by the experimenters. These included various aerial flips and swimming with the tail out of the water. Pryor et al. assumed that ‘novelty was an intrinsic factor’ in the origin of the new responses, and that what had happened was that the procedure of giving rewards for something different each day had inadvertently rewarded novelty as a category of response. A second and more docile animal was therefore tested with a similar procedure with two observers and systematic recording of data. For the first seven days only breaching, beaching, porpoising and swimming upside down were observed and the animal tended to adopt a rigid pattern of these previously reinforced tricks. After a further seven sessions of training with particular new tricks, this second animal began to display the same varied behaviour as the first, emitting eight different kinds of behaviour in the sixteenth session, including a spin, a flip, an aerial spin, an upside-down tail slap, and a tail side-swipe, which were all seen for the first time. These were then rewarded one by one until, in five final sessions (28—33), the animal had to come up with a new response each time to obtain its fish rewards. On session 30, it performed 60 different patterns of movement, but all had been seen before, and thus it got no fish. On the next three sessions however, it managed to come up with a backwards flip, an upside-down porpoise, and finally the response of standing on its tail and spitting water at the trainer. It may be stretching a point to refer to this as ‘creativity’, but
clearly, the cumulative effects of the training procedure resulted in the animal learning something other than merely the mechanical repetition of stereotyped and unchanging patterns of response.

Figure 5.4 The Hampton Court Maze for rats. See text. After Small (1901).

Spatial learning and ‘cognitive maps’

The conclusion that instrumental conditioning typically consists of something more than the repetition of particular stamped-in stimulus-response connections was of course reached earlier by Tolman (1932, 1948) on the basis of rather different evidence, namely, the behaviour of rats in mazes. Hardly anything about the behaviour of rats in mazes supports the existence of specific stimulus-response connections, since, in the first place, it is impossible to identify a specific stimulus, or a specific response. Small (1901) trained rats to follow the correct path to food in a modified form (suitably scaled down) of the maze built for human amusement at Hampton Court Palace (see Figure 5.4). Yew hedges are replaced by high sided boards, which the animals cannot see over, and they must thus move in enclosed alleys. Clearly in this type of maze, detailed visual cues will be likely to be less important than they are in the other main type, the ‘elevated maze’, where the rats run on planks without sides fixed several feet above the ground. Both Small (1901) and later Watson (1907) who used the same maze with rats variously handicapped by blindness, deafness or lack of the sense of smell, speculated that the
learning of the maze was accomplished as a motor pattern — by associations of ‘motor images of turning’, for Small, and ‘a serially chained kinaesthetic reflex-arc system’ for Watson. An enormous amount of evidence (summarized by Munn, 1950) suggests that associations between successive movements are rarely important in maze-learning. MacFarlane (1930) and Evans (1936) by flooding enclosed alley mazes with water, showed that rats which had learned a maze by swimming, running or wading could follow the connect path by an alternative form of locomotion, that is using very different kinaesthetic cues. More drastically, Lashley and Ball (1929) and Ingebnietsen (1932) lesioned ascending proprioceptive nerve tracts in the spinal cord, producing in many cases aberrant posture or patterns of movement, but with relatively little effect on the rats’ ability to learn or retain the correct pattern of turns in a maze. Since direct associations of series of movements does not thus seem important, and since also no individual sense modality appears crucial to maze-learning (Watson, 1907; Honzik, 1933, 1936), the alternative theory is that maze performance has to be under the control of multiple stimuli, and feasible with multiple responses (Hunter, 1930) or, in what may be another way of saying the same thing, controlled by ‘central’ or ‘symbolic’ factors (Lashley, 1929).

Tolman’s contribution can be summarized as the designing of elegant behavioural experiments to demonstrate that rats use various cues to achieve a sense of geographic place, and as the providing of the long-lived term ‘cognitive map’ as part of his theory of ‘purposive behaviourism’. The best examples of experiments used by Tolman to support his theory are thus those on the phenomenon of ‘place learning’. The normal laboratory rat will give every appearance of expecting food at a particular familiar place in a maze it has learned: it will sniff and search for the food if it is absent, and attempt to burrow under or climb over obstacles placed in its normal path. Alternative routes to the same food may be used if several of equivalent length are available (Dashiell, 1930).

There are many anecdotal reports of rats finding short cuts (e.g. Helson, 1927). Tolman et al., (1946) attempted a systematic study of this, by first training animals to take a right-angled dog’s leg path to a goal box which was situated in a direction of 60 degrees from the starting point, and then providing them with 18 paths at 10—degree intervals. More than a third of Tolman’s rats did indeed take the sixth path, going directly towards the site of the goal box, but another 17 per cent took the first path on the right, and other experimenters trying to replicate the result found rats apparently trying to retrace the dog’s leg. However, a simpler experiment, which opposes turns in a particular direction against movement towards a particular place, produces easily replicable results. The procedure is to start rats from different sides of a cross-maze (see Figure 5.5), and to compare animals always rewarded for turning in the same direction (and therefore rewarded in different places) with others always rewarded in the same place (and therefore rewarded after both left—and right-turns). Tolman et al. (1946, 1947) found that their animals were much better at always going to the same place, using an elevated maze with asymmetrical lighting and other obvious possible landmarks. It is not now doubted that
this sort of place-learning comes relatively easily to rats, but, if a similar experiment is run in the dark, or under a homogeneous white dome, then, not surprisingly, it is easier for the animals to learn a consistent turning response (e.g. Blodgett and McCutchan, 1948).

Figure 5.5 Place versus response learning.
A maze in which rats may be trained to always make the same turn, to different places, or to turn either left or right, in order to get to the same place. See text.

Several more kinds of experiment have more recently confirmed the complicated nature of place learning. O’Keefe has identified not only particular external landmarks used by his rats (coatracks, etc., which when moved produced errors in a predictable direction), but also particular brain cells (in the hippocampus) which, among a large population, fire whenever the rat goes to a certain place in its apparatus (O’Keefe and Nadel, 1978; O’Keefe, 1979). The prize for the most Tolmanian experiment of the 1980s will probably go to R. G. M. Morris (1981), since he combined the swimming which was used in Tolman’s laboratory by MacFarlane (1930) with an appeal to remote landmarks. Rats were put in a large circular tank of water made opaque by the addition of milk (mud would have been more naturalistic). At a certain point in the tank was a cylinder with its top 1 centimetre below the surface — as soon as they found this rats would climb up on to it, and the test of learning was that, with experience, they would swim eventually to the invisible submerged platform from any point in the tank, thus demonstrating their degree of knowledge of its location.
Figure 5.6 The radial maze.
If rats are placed on the central platform of a maze with a plan such as this, food rewards being available but out of sight at the end of each arm, experienced animals will visit each arm only once, thus demonstrating some form of memory for their activities in the recent past.

However, the most widely used spatial learning technique in modern animal laboratories is undoubtedly the radial maze, introduced by Olton and Samuelson (1976). In this several arms (initially eight elevated planks of different widths) radiate out from a central point, where a hungry rat is placed to start with, and a single pellet of food is put, out of sight, at the end of some or all of these arms (see Figure 5.6). If, as in the original experiment, all the arms are baited with food, then the rats’ best strategy is to run down each arm only once, on a given daily test. After several days’ training, rats on average choose about 7.5 novel arms in their first eight choices. Control experiments show that this statistically significant result is still obtained if odour cues are eliminated either by the crude method of drenching the apparatus with Old Spice aftershave, or by the more revealing technique of confining the rat in the central area between choices, and surreptitiously exchanging an arm which had already been run down with another, or by rotating all the arms to different positions (Olton et al., 1977; Olton et al, 1979). Thus on any given day, the rats’ cognitive map and associated memory is sufficiently detailed to record which seven places the animal has already been to; and even with a 17—arm maze, an average of 14 different arms were chosen in the first 17 choices, (Olton et al, 1977). This is a matter of recent memory, since the procedure works without the animals having to
visit arms in any fixed order, and therefore it works when the memory is of exactly what has happened on a particular day (see chapter 10). A more traditional form of spatial learning is obtained if the radial maze is used with only the same one or two (or four) arms baited each day, so that the animal has to remember that food is always in some places but not in others (Olton, 1979; Olton and Papas, 1979).

**Multiple levels of representation in instrumental conditioning**

Tradition in the theoretical analysis of instrumental conditioning requires that we ask the questions of ‘what is learned?’ and ‘what is the necessary and sufficient condition for learning?’. Modern research suggests the answers that there is no single kind of thing invariably learned in all cases of instrumental (or operant) conditioning; that there is no obvious essential ingredient; and that there are several limiting cases which are just sufficient to demonstrate one of the many phenomena which instrumental conditioning involves. However, in general the answers previously given to the questions of what is learned, and in what way, are not irrelevant to modern concerns, since it may be argued that almost every proposed theoretical form of association should be retained as one of several possibilities, and almost every supposedly critical condition is a variable which may affect the course of learning, even if it is not critical.

Although the views of earlier theorists such as Thorndike and Hull should thus not be dismissed out of hand, it is undeniable that there has been a general shift of opinion away from the theory of learning-as-doing towards more cognitive theories of learning-as-knowing, even for laboratory animals. Part of this distinction corresponds in fact to a difference between two kinds of knowing, knowing how’ and ‘knowing that’, which Dickinson (1980) has imported into learning theory from cognitive science as the contrast between ‘procedural’ and ‘declarative’ representations. It is important that neither kind of representation necessarily has to be in verbal form; but the distinction can be illustrated by the example that an instruction in the form ‘pull the gear lever back when the engine makes a loud high-pitched noise’ is procedural; while instructions which are more explanatory, such as ‘moving the lever changes the gear’; ‘higher gears are used when the car goes faster’, or ‘the positions of the gears are as follows’, are initially only declarative, but are well suited to the deriving of a wide range of procedural instructions, by processes of inference and integration (see Dickinson, 1980, p. 23). In human practice, the distinction between procedural and declarative representations usually maps onto the difference between automatic habits and more considered and controlled reflections (cf Shiffrin and Schneider, 1984). This is the case in most people’s experience of learning to change gear, along with other aspects of learning to drive. First attempts, and the early stages of learning, require a surprising amount of concentration and effort for activities which later on become second nature. When this has happened, the declarative aspects of early learning may become irrelevant, with eventual skills including some accomplished but fairly isolated
habits. I hope I am not the only driver to have had the experience of hiring a car on the Continent, and on the first occasions of reaching a suitable speed in first gear, putting in the clutch and making sweeping movements with my left hand, where the gear lever normally is (and, even more embarrassing, doing the same thing the opposite way around when first driving my own car on returning home). There is a wealth of anecdotal evidence, first emphasized by James (1890/1983) that well-learned human skills produce automatic habits, which are independent of rational purpose; Reason and Mycielska (1982) have examined the degree to which this contributes to ‘action slips’ of mixing up habits (pouring the water from the kettle into the tea caddy and so on).

In the context of learning theory, the contrast is between automatic habits and planned actions, but also between the level of representation, whether procedural or declarative. By and large, more habitual processes apply to muscle or limb movements, while purposive actions are coded in terms of goals, or sub-goals, but it is possible to speak of mental habits, and emotional reflexes, and relatively goal-directed muscle-twitches. As previously discussed (pp. 118—22), Thorndike’s (1898) answer to the question of what is learned was very procedural, since it corresponded to the instruction ‘when in the box, press the lever’, but it was not at the lowest level of representation of the stimuli and responses involved, since it concerned learned connections between the sense-impression of being in the box, and an impulse to make a response, directed at a controlling feature of the environment, such as a lever or a loop of wire. By comparison, Hull’s refinement of Thorndike’s Law of Effect was equally procedural, but at a lower level of representation, since the learned connections were proposed to be directly between stimulus receptors and muscle movements. Schematic versions of both these theories are given in Figure 5.7.
Opinion has decisively swung against them, and there is now much support for the Tolmanian alternative answer to the question of what is learned in instrumental conditioning, sketched at the bottom of Figure 5.7. It should be noted that the Tolmanian version differs from the others in at
least two respects. First, it is at a higher level of stimulus and response representation than Hullian theory; it assumes internal ideas or representations of wanted or expected rewards, and coherent impulses for actions rather than direct conditioning of muscle movements.

Second, it does not include a direct link between environment input and response output at any level of representation: in Tolman's terms, what is learned is a 'means-end readiness', or a 'sign-Gestalt-expectation'. In the sketch in Figure 5.7, the sequence is that an animal first wants food, then searches for an action which, if performed, will produce the wanted food. The main learning in instrumental conditioning is thus the learning of what response is necessary to gain desired consequences, or, in its simplest form, an association between a response and the reward which follows it (Mackintosh, 1983; Adams and Dickinson, 1981b; Dickinson, 1980). In freer language, it is now said that in instrumental conditioning an animal is required to 'reach certain conclusions' about what its behaviour should be, or to 'track causal relationships' between its behaviour and reward (Mackintosh, 1983, p. 112; Dickinson, 1980, p. 143). This kind of theory, that is (c) in Figure 5.7, is certainly considerably more complicated than the other two which it has supplanted, and we should thus briefly review the evidence in its favour.

The evidence in favour of the Tolmanian two-stage theory of instrumental learning, in which responses are selected according to whether they are associated with a desired goal,

is in part indirect, in the form of evidence against stimulus-response principles, and in part more positive, in the form of evidence that an evaluation of goals determines instrumental performance and that responses become associated with consequent rewards.

Spatial learning as opposed to stimulus-response connections

The difficulty of accounting for an animal’s learned movements in space in terms of specific stimulus-response connections was a niggling problem for S-R theory of the Watsonian type from its inception, and Tolman’s criticisms of S-R theory, with recent developments, are discussed above, (pp. 141—2). Only two further points need be added. First, the recent use of the radial maze, and similar techniques in which animals learn not to return to a depleted food source, emphasizes the failings of the Thorndike principle that rewards stamp in only immediately preceding behaviours. Theories about cognitive maps, or working memory, are vague, but the vagueness is necessary to account for the behavioural results. As a second point, we may resurrect an old anecdotal result which may have considerable ecological validity. Helson (1927) argued against stimulus-response theories on the grounds that reward procedures could induce relatively novel behaviours. He had attempted to train rats to distinguish between two shades of grey card that indicated which of two parallel routes they should take over a raised barrier which was placed across the end of a long deep box, with a partition down the middle dividing these routes and two food compartments on the other side of the barrier. The barrier was covered with wires, the side with the
incorrect card being electrified. The animals showed no sign of detecting
the difference between shades of grey. However they quickly acquired a
distaste for wires which delivered shock. One small rat of the four tested
refused to climb over either barrier, while two others circumvented the
barriers by climbing up the wall of the experimental enclosure, and
walking round on top of it to the food boxes. Helson argued that this was
not an accidental response which was stamped in. There is very little doubt
that laboratory rats, and most wild animals, will attempt alternative
methods of locomotion, via alternative routes, to get to a desired location,
if this becomes necessary. A strict interpretation of the Law of Effect
would require that animals be permanently halted if a tree falls across their
accustomed route to a water hole. Stimulus-response theorists have always
made attempts
to account for the emergence of relatively novel goal- directed behaviour,
in terms of generalization of previously acquired responses, ‘habit-family
hierarchies (Hull, 1934, 1952) and the like. These attempts have never
been particularly convincing, but are now less often attempted because of
an accumulation of other evidence against stimulus-response theory, and
against the explanatory version of the Law of Effect.

Re-evaluation of goals and response consequences

When ‘stimulus-response connections’ was the answer to the question
of ‘what is learned?’, ‘reward’ or ‘drive reduction’ was the force which
made the connections, and hence was once regarded as a necessary
condition for all learning. This view can also now be speedily dismissed,
and has already been bypassed by the previous chapters in which learning
is presumed to take place in the course of habituation and classical
conditioning, without the operation of a reward mechanism being
necessary. Recent evidence has however been directed at the proposition
that expectations of a goal or specific reward is one of the mechanisms
responsible for the results of instrumental conditioning procedures.

Latent learning

A phenomenon which weakened the assumption that drive reduction
was necessary for learning, and which suggests that performance in mazes
is influenced by expectancies of rewards, was investigated in Tolman’s
laboratory by Blodgett (1929) and Tolman and Honzik (1930b). Rats
allowed to find food in the goal box of fairly complicated mazes (e.g. 14 T
sections in succession) gradually improve their speed of getting through,
and gradually make fewer errors, over a matter of weeks (that is 10 or
more daily trials). This could be interpreted, albeit implausibly in view of
the number of possible wrong alternatives, as the gradual strengthening of
accidental correct responses, by the experience of reward. However, if
other animals are allowed an equivalent amount of experience of
wandering about the maze without getting food at the end, and therefore
without very much change in

behaviour, it can be shown that a form of ‘behaviourally silent’ learning
has taken place, since the sudden introduction of reward in the goal box
leads to a sudden change in performance, animals with unrewarded experience immediately catching up with those always rewarded after their first reward in the experiment of Tolman and Honzik (1930b). Rats allowed to explore a maze thoroughly will return efficiently on the next opportunity to any point in the maze where food is first given (Herb, 1940) and thus the evidence is very strong that geographical information is learned for its own sake.

The independence of expectations with regard to goals, and methods of response which achieve goals, is best illustrated by experiments in which these two factors are separately manipulated, with monkeys or chimpanzees, showing an animal where food is being hidden will induce strategies of response designed to retrieve it, and there is little argument about the role of internal representations of the hidden food in these cases (see chapter 10). With laboratory rats, it appears to be possible to change expectations of what will be found in the goal box of a maze simply by placing the rat in the goal box, and thus ‘placement studies’ supply evidence that expectations can in some circumstances dictate choice and intensity of response. For instance Seward (1949) allowed 64 rats to explore a simple T-maze with goal boxes of differing texture and colour, these differences not being visible from the choice point. He then placed the rats in one of the goal boxes by hand, where they were allowed to eat. When they were next put in the start box of the T-maze, 54 of them turned towards the goal box in which they had been previously fed. Since they had not made the turning response before being fed, the assumption must be that they had previously learned that turning in a given direction brought them to a given goal box, and that when a particular goal box was made attractive by an association with feeding, this new evaluation of the goal could be applied to the old spatial knowledge. Similarly, if rats have already learned to go in a particular direction to food in a T-maze, then if they are placed in the usually correct goal box with no food present, this is sufficient to increase subsequent turns in the opposite direction (‘latent extinction’: Deese, 1951). A slightly different form of latent learning occurs if changes in motivational states occur after an explanatory phase. Rats may remember where food was in a maze when they are made hungry, even if they were not hungry at the time when they saw it (Bendig, 1952). Rats allowed to drink salty water for lever pressing when just thirsty, will press that lever, but not one which has been previously rewarded with plain water, if they become seriously salt-deficient (Kriekhaus and Wolf, 1968). ‘Latent extinction’ of a food-rewarded lever-pressing response is obtained if the animals are pre-exposed to the functioning of the reward-delivery mechanism without the usual rewards (Hurwitz, 1955).

De-valueation of reward

In all the above cases, learning about the consequences of a response took place without performance of the response itself, and subsequent effects on behaviour are therefore probably attributable to the prior association between the response and its consequences. In latent learning, it can be said that expectations as to the value of responding in a certain
way are increased, while in latent extinction, expectations with regard to the value of responding are lowered. A further example of changes in the value of a goal is provided by the taste-aversion technique (Holman, 1975; Chen and Amsel, 1980; Dickinson, 1986; Adams, 1982; Dickinson et al., 1983). Animals poisoned after eating food of a certain flavour or smell eat very much less of that particular food subsequently (Garcia et al., 1977a, see pages 232—42). Under certain circumstances, rats poisoned after eating the particular flavour of food which they have previously received as rewards for bar-pressing in a Skinner box will thereafter show a greatly reduced tendency to perform this response (Adams and Dickinson, 1981a, 1981b). In these circumstances, it is arguable that the lever-pressing should have the status of ‘a true action that is under the control of the current value of the reinforcer or goal’ (Adams and Dickinson, 1981b, p. 163).

Specific associations between responses and rewards

In the ideal form of the expectancy theory of instrumental learning (Figure 5.7(c)), all responses are made for specific consequences. If a rat knows that food is available at one end of its cage but water at the other, it will go to the appropriate place when hungry or thirsty, and sometimes similar behaviour can be demonstrated in less familiar T-mazes, with water on one side and food on the other (Thistlethwaite, 1951). It is unlikely, however, that an exact representation of reward is necessary, in all cases of instrumental conditioning. For instance, rats rewarded unpredictably with 10 or zero units of food, or with 8 or 2, learn to run up an alley more vigorously, if anything, than rats always rewarded with 5 units (e.g. Yamaguchi, 1961). Though not strictly necessary, it has always been suspected that very specific expectancies of reward are sometimes formed in experiments with animal subjects. The best-known study was performed with rhesus monkeys by Tinkelpaugh (1928), under Tolman’s supervision. The main purpose was to demonstrate what would now be termed ‘representational’ factors (Tinkelpaugh called them ‘representative’) in monkeys by allowing them to watch items of food being put in particular places or containers at some time before they were permitted to retrieve the food for themselves. Overnight delays of up to 20 hours proved to be possible (see chapter 10). However, the study is now remembered for the checks made on memory for the identity as well as the location of hidden food. Films were taken of the animals’ facial expressions and searching activities aroused when a piece of banana had been hidden, but a piece of lettuce (which the animals liked much less than banana) substituted during the delay interval. Reactions to lettuce were sufficiently pronounced to convince Tinkelpaugh that the monkey had specific expectations of finding ‘banana’, rather than ‘some food object’, in the container, although an expectation of ‘highly desired food object’ would account for the results. It is difficult to obtain any more direct evidence of the specificity of representations of the reinforcer in instrumental conditioning.

Lorge and Sells (1936) claimed to have repeated Tinkelpaugh’s finding when rats were trained in Thorndikean problem boxes.
Nine rats had already been trained with two trials per day in each of three different boxes. The responses required for escape were designated as ‘face-washing’, ‘stand-up’ and ‘begging’ which after training were essentially three different stereotyped postural responses. Throughout training adoption of the correct posture for a particular box resulted in a door being opened so that the animals could exit from the box to a small cup containing bread and milk. To examine ‘representative factors’ the nine well-trained rats were simply run as usual, but with sunflower seeds in the reward cup instead of the bread and milk. Although rats’ facial expressions are not as revealing as those of monkeys, their general behaviour was changed considerably by the substitution of sunflower seeds. For the two days that they were observed, all rats on all occasions, on making the appropriate response and getting to the food cup, picked up one of the seeds, threw it aside, then ran back into the box and repeated the response. Although the animals had previously eaten sunflower seeds, this experiment would have been better controlled if some of the animals had received sunflower seeds throughout the experiment. However, precisely this comparison was possible in the earlier experiment of Elliot (1928), who trained rats on a multi-unit maze for either sunflower seed or branmash as reward, and found that animals switched from mash to seeds began to make many errors, whereas animals always rewarded with seeds performed accurately.

Although there is very little doubt that chimpanzees may form specific expectancies about particular food stuffs, which then determine both direct goal-seeking and other forms of behaviour which make the expectancies obvious to an observer (Savage-Rumbaugh et al., 1983; see chapter 10), evidence supporting a role for specific expectancies of reward in instrumental learning by laboratory rats and pigeons is understandably less strong. The experiments in which rewards are changed, such as those of Elliot (1928), just mentioned, and Crespi (1942, see pp. 126—7) can be interpreted as indicating an overall motivational value of reward associated with instrumental behaviour, with reductions in motivational value having various emotional effects (including ‘frustration’ in the theory of Amsel, 1962; see Capaldi et al., 1984).

However, a different kind of experiment, first suggested by Trapold (1970) provides an additional source of evidence for the theory that the qualitative nature of rewarding events, other than just their degree of goodness or badness, is usually noticed by rats and pigeons, and occasionally assists in the selection of appropriate responses. In Trapold’s experiment the task for rats was to press one lever, on the left, when a clicker was sounding, but another lever, on the right, when they heard a tone. Providing the two stimuli, and the two separate responses to be made to them, are easily distinguishable, rats are perfectly capable of learning such tasks, by trial and error, although they may need thousands of trials to get things completely right (that is, to be right 9 times out of 10). Trapold’s finding was that the task was made much easier if two different rewards were used, food pellets for the left lever and sugar solution for the right, even if both rewards were presented at the same place. Furthermore, in a separate experiment, he showed that associating particular rewards
with the clicker and tone signals, before these were used as cues for the bar-pressing task, also made the eventual learning of the task faster if the rewards were kept the same way around, but slower if they were reversed. This supports the expectancy theory of instrumental learning illustrated in Figure 5.7(c), in which the animal is supposed to think of the reward first, and then decide which response must be made to obtain the reward. Clearly this decision should be easier if separate rewards need separate responses.

Various elaborations of this type of experiment have confirmed that what is known as the ‘differential outcome procedure’ (Peterson, 1984) usually enhances learning or has other effects predictable in terms of differential expectancies of specific rewards, with both rats (Carlson and Wielkiewicz, 1976; Kruse et al., 1983) and pigeons (Peterson, 1984; Delong and Wasserman, 1981). For pigeons a rather more difficult task has been used, in which one signal, which may be either green or red, say, determines which of two stimuli, say black or white lines, should be chosen next. If the first green signal means that the bird should now choose black to get a reward of food, but a red signal means it should choose white, to get a reward of water, then learning is quicker, and performance more accurate, than when all correct choices are reinforced by either food or water at random (Peterson et al., 1978). With this procedure, if the task is made even more difficult by a delay between the red/green signal and the opportunity to make the next choice, between horizontal or vertical lines, then differential reward outcomes of food and water for the two correct choices allowed 95 per cent correct performance with a 10—second delay, while the use of only a single reward reduced accuracy to chance levels. It was thus argued that it was the expectancy of a particular reward, retained over the delay interval, which was responsible for directing the eventual choice (Peterson et al., 1978; Peterson, 1984). If green and vertical are two visual stimuli both associated with food, and red and horizontal both associated with water, then representations of food or water (or peripheral responses specific to these) will be sufficient to bridge the delay, and that should certainly make the task easier by comparison with that of remembering whether the first signal has just been green or red, and knowing that green means horizontal and red means vertical.

Various alternative sorts of instrumental learning

The evidence in the section above might tempt one to conclude that the Tolmanian explanation for what is learned is right, whereas Hull’s and Thorndike’s ideas were wrong (see Figure 5.7). But not only are there some reasons for hesitation in interpreting this favourable evidence, there are also many results which argue quite unequivocally against specific expectations of rewards being always a necessary factor in instrumental learning. For instance, some attempts to demonstrate latent learning fail (Thistlethwaite, 1951; e.g. rats which find food in a T-maze when thirsty but not hungry do not go to it when they are hungry, Spence and Lippit, 1946). More important, there are many cases where laboratory animals appear to learn rigid and mechanical automatic habits which become more
or less completely independent of

the desirability of the goals used to establish the habits. For instance Holman (1975) trained rats to press a lever for a reward of saccharin solution, and then made the rats averse to saccharin by injections of lithium chloride. This stopped the rats drinking saccharin solution, but had no effect on their habit of pressing the lever. Adams (1982) and Dickinson et al. (1983), having previously shown that in some circumstances a conditioned aversion to previous rewards will immediately reduce the strength of habits they reinforced, agree that with long-established habits, especially those conditioned on fixed interval schedules of reinforcement, the learned performance is very little affected, if at all, by manifest devaluation of the ostensible goals.

The relative independence of habits from goals was stressed theoretically by Thorndike (1898) but with the empirical support of the fact that cats which had learned to press a latch to get out of a box often (but not always) continued to do so when he cut a hole in the roof. More dramatic apparent obliviousness to the final purpose of their activities is easy to demonstrate in rats, which, after being thoroughly trained to run down an alley for food pellets, may then run quickly over identical food pellets strewn in their path. Similarly, it is a reliable result that rats trained in Skinner boxes to press a lever for food pellets may ignore similar rewards presented in a bowl, preferring to obtain the same objects in the more usual way (Neuringer, 1969; see Morgan, 1974, 1979 for reviews). This is partly a trade-off between accustomed habits and degree of effort, since, understandably, rats required to press many times for a single food pellet are relatively eager to accept free pellets from a new place. However, there is sufficient rigidity in rats’ learned behaviour to support a good measure of Thorndikean or Hullian habits — it must be remembered that human behaviour, and not only that classed as neurotic or abnormal, is composed of established routines performed unthinkingly, as much as of rational purposes.

There is thus good reason to believe that rats and people, though occasionally and fleetingly directing their behaviour towards expected goals, make extensive use of the less taxing mechanism of automatic habit for behaviours which are often repeated. Morgan (1894), just before Thorndike, had clearly stated the proposition that goal-seeking is always responsible for the early stages of learning, and for dealing with new problems, but that there is rapid drift to the habit mechanism whenever possible. There is no reason to assume that a goal-seeking phase is always necessary. There is something to be said for the position that even human learning may on occasion occur in a Thorndikean or Skinnerian fashion. This certainly goes for muscle co-ordination in complex skills, but there is experimental evidence that mannerisms, habits of speech, and social skills are sometimes unknowingly stamped in (Rosenfeld and Baer, 1969). However, for the general theory of learning, the more important cases are those at the opposite end of the scale of intellectual capacity (or, if you prefer, at some other corner of a diffuse array of psychological abilities).
Although claims for instrumental learning in *Aplysia* are at present muted (Hawkins and Kandel, 1984; see chapter 4), the basic feature of a tendency to repeat rewarded acts is claimed for creatures as diverse as earthworms and flatworms (Fantino and Logan, 1979) and not only for advanced arthropods like honey-bees (Couvillon and Bitterman, 1982, 1984) but also for less advanced insects like cockroaches, and even for the decapitated cockroach, and legs of cockroaches detached from the rest of the body (Horridge, 1962; Distenhoff *et al.*, 1971). There are ample physiological reasons for doubting whether these neural systems are capable of sustaining, for instance, specific expectations of reinforcers which are as rich in information as those claimed for rats and pigeons (Peterson, 1984). Further, for rats and pigeons, and other laboratory animals; it is possible to demonstrate forms of instrumental conditioning after radical interference with normal brain functions such as complete removal of the neocortex of the cerebral hemispheres (decortication) or removal of most of the hemispheres themselves (decerebration; e.g. Heaton *et al.*, 1981). Oakley (1979b; 1983) has reviewed such results, obtained by himself and others. Decorticate rabbits will eventually, after slightly slower learning, press levers for food in Skinner boxes, on basic fixed interval and fixed ratio schedules (Oakley, 1979b). There must therefore exist subcortical mechanisms which are sufficient for the demonstration of instrumental learning. This raises the possibility, at least, that there are subcortical methods of stimulus-response association, which differ in some way from the full range of psychological processes available to the normal animal.

![Diagram](image)

**Figure 5.8** Alternative types of association in instrumental learning.

Evidence from common laboratory animals, from cross-species
comparisons and from physiological investigations, all suggests that there may be more than one answer to the question of what is learned. As a general rule of thumb, any mechanism proposed in the last 100 years as the universal principle of learning may be accepted as a form of learning in some or other species or preparation. Instead of choosing between the alternatives presented in Figure 5.7, a more accurate and comprehensive answer to 'what is learned?' is obtained by combining all three processes illustrated in Figure 5.7, with some additions suggested by the results discussed above, to give a diagram such as that shown in Figure 5.8. The purpose of this is to show that more than one process may be involved in instrumental learning, and there are many more possibilities than can conveniently be drawn in. For instance, the example chosen is the pressing of a lever for food rewards by a hungry animal in a Skinner box. Instrumental learning to escape from dangerous or unpleasant external stimuli has not yet been considered, and would require some changes, though not wholesale ones (see chapter 7). Also not immediately obvious is the fact that learning of some kinds is possible without any strong motivational events occurring; latent learning of the kind which commonly takes place when rats explore mazes is subsumed along with much else in the box which refers to responses which are 'appropriately associated with the wanted event'. Although this covers a multitude of possible forms of association, including any method of getting to the known location of a wanted event, it does not seem unreasonable to assume that getting a food reward after a particular response may contribute to the association between the reward and that response. This may work in both directions, so that wanting a reward will in future produce an impulse for that response (1), and when the response is made, a firmer expectation of specific impending reward might very well be justified (2). In the situation in which rewards are regularly obtained, it is assumed that some representation of these goals is conditioned to the context (3); but this needs to be modifiable by the effects of specific signals, which often have stronger conditioned effect than background or contextual stimuli (Wagner, 1976; Lovibond et al., 1984). Clearly, ordinary classical conditioning should provide representation of the reinforcer associated with particular signals, as at (3), and similar effects are frequently now supposed to play a role in instrumental conditioning (Rescorla and Solomon, 1967; Kruse et al., 1983). But as Jenkins (1977) pointed out, and as Mackintosh (1983) has cogently emphasized, this cannot be the only role for what is known as a discriminative stimulus in operant conditioning for, to put it loosely, this does not merely inform the animal that rewards may be forthcoming (as in classical conditioning); it also instructs the animal that if it wants a reward, then a particular response is necessary. Thus, laboratory animals learn easily, if they have to, that rewards are sometimes delivered free but that when a signal is on they must work for them (Weiss, 1971; Jenkins, 1977). The signal in this case does not indicate any increase in expectations of rewards, but rather that a given response is now appropriately associated with reward, and thus there is an input from the discriminative stimulus at (1).

Now, in order to accommodate the evidence for Thorndikean habits,
we can further assume that the pure frequency of occurrence (not illustrated), or the effects of reward (4), will increase an automatic impulse to make the behavioural act normally rewarded, either in response to the background cues, as in Thorndike’s original experiment, or connected with specific experimental signals (or discriminative stimuli). This is a fairly high level of representation of the response, since a variety of muscular coordinations, some undoubtedly unlearned, will be necessary to actually do something. However, both in animals lacking that level of representation of the response and for detailed stimulus-response co-ordination, such as is involved in gear-changing in higher animals, then the muscular coordinations themselves will be one of the things learned in instrumental conditioning procedures (5). We might expect that these will be mainly internal (some kinesthetic) associations (Watson, 1914/1967) and less tied either to the external cues or to the motivating effects of reward, but we have Locke’s example of dance steps learned in a particular room to remind us that skills may sometimes need familiar external conditions, and the rewards are likely to be involved both as feedback to confirm correctness and in various kinds of motivation (Barber and Legge, 1986).

Finally, although considerable attention is currently given to expectations of reward, representations of the reinforcer and such like, concern with the ‘wanted’ character of reward representation (Tolman, 1959) and the role of specific drive states (Hull, 1952) is probably due for a comeback. Drive states such as hunger, thirst or salt deficiency presumably have an innate multiplying effect on internal representations of food, water and salt: there may be many other in-built drive states, and there can certainly be alterations of these by which cravings for cigarettes, chocolates, morphine and numerous other ingestible substances are controlled metabolically (see pp. 73—7, under classical conditioning). There-fore the level of specific drive is shown as modulating the degree to which external cues arouse ideas of reward and the intensity with which any such ideas are followed by the various response-producing processes encompassed by the ‘perform response’ box (top of Figure 5.8). In the variation of Hullian theory adopted by Spence (1956) and perpetuated by Bindra (1968, 1976) and others, the effect of rewards was interpreted as ‘incentive motivation’, which appeared to consist in part of general activation and energy available, with which to perform conditioned responses. Whether or not such a process is of any great importance in conditioning experiments has been disputed (Mackintosh 1974) but, depending on the exact reward (Roper et al., 1983; Peterson, 1984), it is possible that some or other form of arousal or emotion can become conditioned to external cues (6). The restlessness exhibited by caged animals in anticipation of regular feeding times would seem to be partly attributable to general arousal, and since a similar effect is observed in general movements of decorticated rabbits, and in measures of heartbeat in goldfish (Oakley 1979a; Savage, 1980), it seems wisest to allow associations of arousal with external cues which are separable from those of more concrete and informative expectations. However, the distinction is not intended as a
hard and fast one: some kinds of conditioned arousal may be reward-specific — possibly including salivation, licking of lips and so on in the case of food, and hormonal and autonomic changes in anticipation of sexual contacts — while what are regarded as more cognitive expectancies, as of a goal in a lucky dip, or a prize behind the curtain, may have general motivating effects, while being altogether vague, nebulous, or mistaken.

**Instrumental conditioning — conclusion**

The extent to which instrumental conditioning differs from classical conditioning and habituation, the degree to which it reflects the biological adaptations of particular species, and varied requirements it may place on brain capacities — these are all topics left to the next chapter. The present conclusions might be given by saying that everyone is right about instrumental learning. The achievement of goals, the fulfilment of needs and the enactment of inner purposes are of fundamental biological importance, as both Aristotle and Hull suggested. Both in descriptive terms for most animals, and in mechanical process for some of them, these biological functions are served by the establishment of connections between external circumstances and response output, reinforced by practice and by successful outcomes in a relatively direct fashion, as proposed at the end of the last century by Thorndike. Descriptively also, the experimental control of the delivery of food rewards to hungry animals has the power and reliability in its effects on behaviour that was claimed for it by Skinner. However, the tide of theory at present flows in favour of more cognitive interpretations of the descriptive Law of Effect, which is the principle of Skinner’s operant conditioning techniques, and it is now commonly supposed that responses learned by reward are guided by specific expectancies of achievable goals, thus supporting the role of purpose in instrumental behaviour, more or less according to the views of Tolman. But instrumental learning is a chapter heading, not a psychological unity with a single explanation. The internal psychological effects of reward procedures encompass, at the very least the mechanisms of association both between stimulus and response, and between response and expected reward, which have been previously presented as mutually exclusive possibilities.