

3 Thorndike (1874—1949), Pavlov (1849—1936) and twentieth-century theories of animal learning

Thorndike's stimulus-response connections

It was really Thorndike and Pavlov, rather than Lloyd Morgan, who established the trend towards minimalist explanations of animal learning within comparative psychology. Lloyd Morgan not only assumed that 'the higher grades of animal life' consciously adjusted their actions to serve their needs, but intended to provide an account of mental evolution of essentially the same kind as Darwin's and Romanes's. Thorndike was keen to sweep away all such theoretical frills, and took up a fiercely combative position against the 'anecdote school' of Romanes and his North American counterparts such as Wesley Mills (there is a remarkably vituperative exchange between Mills and Thorndike in the *Psychological Review* of 1899).

Thorndike's PhD thesis was published in 1898, under the title of *Animal Intelligence*, and is notable for its innovations in experimental technique. In a sense, Thorndike introduced the laboratory experiment into animal psychology. Lloyd Morgan's and Wesley Mills's experiments were naturalistic in so far as birds or other animals were observed in the open air under free-range conditions. Thorndike did not have a laboratory as such, since he kept chicks, cats and dogs in his own lodgings, and the restrictions on space enforced in themselves a different sort of data gathering from that which Mills, for instance, was able to make with semi-wild foxes. Just as important as the physical restrictions on space was Thorndike's theoretical attitude. Although he adopted the conventional title for his thesis, he pointed out that previous authors had paid much more attention to animal intelligence than to animal stupidity. Thorndike's tone throughout is that of a man who is going to remedy this omission.

The experimental technique with which Thorndike collected behavioural evidence to support his belief in animal stupidity was the

use of 'puzzle boxes'. Both dogs and cats were tested in these, and there were maze experiments with chicks, but the data which had pride of place came from the performance of twelve cats, each of

which was tested in several different puzzle boxes.

The puzzle-box experiments

The boxes were small crates, hammered together from wooden slats, and only 20 inches by 15 in plan, and 12 inches high (see Bitterman, 1969, for an illustration). Although the cats were young—mostly between 5 and 8 months old during the experiments—all except two struggled violently when confined in such a small space. They were tested in what is described as a state of ‘utter hunger’, with a piece of fish visible outside the box, but Thorndike was in no doubt that the struggles were directed at getting ‘out of’ the box rather than to ‘the food’. One of the criticisms put forward by Mills (1899), which seems reasonable under the circumstances, was that the state of panic which Thorndike succeeded in inducing in the cats was not likely to be conducive to considered and insightful solutions to problems.

The problem which the cats were required to solve was the detection of the mechanism by which the door to the box they were in could be released. Each box had a different system of pulleys, strings, and catches, arranged so that pulling a certain loop of string, or pushing a button, or pressing a combination of levers, would allow a door to fall open. Romanes’s emphasis on the understanding of such mechanical devices as a distinguishing capacity of mammals was thus put to the test. Thorndike was able to say with some confidence that his cats did not work out or understand the sequence of strings and catches, and indeed the solutions are not always obvious to the human observer (see the illustrations in Bitterman, 1969). However, if the cats had simply failed to get out of the box, the experiments would not have been worth reporting. What happened was that the cats did get out, and eventually became very good at getting out, but Thorndike claimed that automatic and crude psychological processes were responsible. No inference, insight, thought, or even any ‘looking over the situation’ was, in Thorndike’s opinion, used by the animals. Rather, simple associations between the stimulus of being in the box, and the response necessary to get out, were ‘stamped in’ in the course of trial-and-error learning.

This argument is strongest when the animal’s tactics change very slowly, and Thorndike liked to quote his most gradual ‘learning

63

curve’. Cat 12, during the testing sequence in Box A, showed few signs of intelligence. Box A was one of the easier puzzles, since clawing at a loop of string near the front released the door. The first time cat 12 was confined in this box it, like the others, scratched, clawed and thrust its paw through any openings it could find. After almost three minutes of this (1 60 seconds) it succeeded in pulling at the loop of string, and opened the door. Now, according to Lloyd Morgan, the discovery that the pulling of the loop was followed by the opening of the door ought to have impressed itself on the consciousness of the animal, so that the next time it wanted to get out, it should pull the loop immediately. What actually happened

was that the second trial was a quick one—the cat got out in 30 seconds— but the third and fourth tests did not suggest that it had fully grasped the solution to the puzzle, since the third and fourth scores were 60 and 60 seconds. The full list of scores, over 22 successive trials in this box, was 160, 30, 90, 60, 15, 28, 20, 30, 22, 11, 15, 20, 12, 10, 14, 8, 8, 5, 10, 8, 6, 6, 7. What Thorndike deduced from this is that cat 12 never really understood what was going on, but that there was a gradual, unthinking, and automatic improvement in the efficiency of its escape response. The theory is that a mental connection is strengthened, but that the connection is a direct one between the stimulus of being in the box, and the response of pulling the string, rather than an association between pulling the string and getting out. The cat is not assumed to think ahead about getting out, but the pleasure and relief it feels when it does get out act backward to make an automatic link between preceding stimulus and response. This is a parsimonious theory, which explains the success of the animal in eventually ‘solving’ the puzzle, while leaving out any process of inference, reasoning, or mental anticipation.

It is interesting that Thorndike did not go all the way with this stimulus and response theory, since he quite clearly left in mental states which appear to be largely superfluous. Although he firmly rejected all ‘reason, comparison or inference, perception of similarity, and imitation’, Thorndike believed that animals had subjective sense impressions, evaluative feelings, and impulses to respond connected to prior sense impressions. In his theory, animals have no mental life that is not concerned with the mediation between stimulus and response, and in this they differ fundamentally from man, who can have purely mental connections and free ideas and impulses which need not be directly tied to reflex-like responses to stimuli. Thus any progression in mental evolution ‘is not from little and simple to big and

64

complicated but from direct connections to indirect connections’. Only in man, and possibly in other primates, should we expect discrimination, generalisations and abstractions (Thorndike, 1898, pp. 108—9). This is in line with the speculations of Wundt, and Thorndike’s mentor William James (*Principles of Psychology*, 1891, vol. II, pp. 348—60), but is the theory supported by Thorndike’s new evidence?

First, it should be noted that the performance of cat 12 in Box A is not altogether representative of that of the other eleven animals. The average escape times of all the cats tested in Box A shows that about 5 minutes were taken to escape on the first and second trial, but less than 30 seconds on the third trial, with slow improvements thereafter. There are plenty of individual cases when cats appear to get the idea after the first or second trial, and escape immediately on all subsequent tests. Thus the gradual modification of response is not universal, or even typical. In some cases, Thorndike admits, especially when a cat had been paying attention to what it was doing when the release device worked, ‘a single experience stamps the

association in so completely that ever after the act is done at once' (Thorndike, 1898, p. 27). But in these cases, is it more sensible to say that act is performed as a reflex response to stimuli, rather than saying that the act is performed because the animal remembers that this is the way to get out? This is not an easy question to answer, but it is obvious that the mere fact of the performance of the response does not exclude one or the other explanation. Additional information would help. For instance, do the animals appear to anticipate getting out once they have made the response? Would they refrain from making the releasing response if for some reason they did not wish to get out? The only variables available in Thorndike's results are how difficult the puzzle seemed to be, and how gradually the improvements in skill took place, and even these allow an interpretation in terms of a gradual realisation of the relation between a particular manipulation and the opening of the escape door. It is worth noting that the most gradual improvements in performance took place when, in Thorndike's words, there was least congruity between the act and the result, such as when cats were released if they licked or scratched themselves, or when chicks were freed for pecking their feathers.

Immediately after his postgraduate work on animal learning, Thorndike took a job at the Teachers College of Columbia University, and devoted himself to the psychology of education. Although he had

65

stressed the differences between human and animal mental processes, this did not prevent him from recommending 'animal-like methods of learning' in education. As he was sceptical about the influence of imitation and understanding on his animal subjects, so he tended to emphasise the importance of practice and activity in human education. Although the routine and mechanical formation of stimulus- response connections was a rather inferior 'form of intellection', 'this hitherto unsuspected law of animal mind may prevail in human mind to an extent hitherto unknown' (Thorndike, 1898, p. 105). *Elements of Psychology* (1905) and the three-volume *Educational Psychology* (1913) were extremely influential in their day (see Joncich, 1968 and Hilgard and Bower, 1975). The essence of Thorndike's recommendations are contained in his 'Law of Practice' and 'Law of Effect'. The Law of Practice was simply the increase in habit formation due to repetition, which in Thorndike's hands meant that 'the only way to teach fractions in algebra, for example, is to get the pupil to do, do, do' (Thorndike, 1898, p. 105). But practice by itself is not enough, since it is the *effects* of actions, in terms of state of satisfaction or dissatisfaction which result from them, rather than merely the repetition of actions, which is crucial in determining whether the actions become stamped in or not.

Pavlov and conditioned reflexes

In spite of Thorndike's great influence in ensconcing practice and motivation, rather than understanding and insight, as key

concepts in learning theory, Thorndike's thunder as the arch-mechanist in early twentieth-century psychology has been stolen by the more extended and systematic investigations of Pavlov (1849—1936). It is inevitable that Pavlov's experiments raise more questions about the necessity of conscious mental experience in the alteration of response, since the responses measured were such as to be almost by definition involuntary, unintentional, and not necessarily known to the responding animal.

The behaviour measured in Pavlov's early work was the behaviour of the stomach, or rather of certain glands in the stomach. Pavlov's hypothesis was that the secretion of gastric and pancreatic juices was under the control of the brain, and not simply a matter of reaction to local stimulation. In fact, local mechanical and chemical stimulation

66

and reactions to internal hormones also influence gastric secretions, but what took Pavlov's interest was the effect of remote and psychological factors on these internal activities.

The experiments on digestion, for which Pavlov received the Nobel prize of 1904, involved a number of surgical operations, performed on dogs. A first step, developed by others in the 1840s, was the chronic implantation of a metal tube in the dog's stomach, through which gastric juices (consisting mainly of hydrochloric acid) could be collected. Reasonable amounts of gastric juice are normally only obtained when the animal eats, and this means that the fluids collected are contaminated by food. The refinement introduced by Pavlov was the anatomical separation of the mouth from the stomach: the oesophagus was cut and the two ends independently brought out to the throat. This meant food could be eaten by the dog without reaching the stomach, or food could be dropped into the stomach without the dog's having eaten it. Although this sounds horrendous, animals survived these procedures for some years.

The physical separation of eating and digestion allowed the demonstration of the psychological dependence of the latter on the former. Bread dropped into the stomach without the animal noticing it is not digested, but eating bread with relish induces gastric activity, even if nothing is present in the stomach. This is not a simple matter of taste stimuli driving internal secretions, since the mere sight of food produces gastric juices, if the dog takes an interest in it, and the effectiveness of food in the mouth depends on how far it 'suits the dog's taste' (Pavlov, 1955, p. 142).

These results in themselves suggest that the previous experience of animals may determine the reactions of their internal organs to external stimuli. This conclusion was established beyond doubt in the further experiments by Pavlov on the secretion of saliva rather than the secretion of gastric juices. Practically all of the experiments reported by Pavlov as examples of 'conditioned reflexes' (see Pavlov, 1927 and 1955) employ precise measurements of the amounts of saliva secreted by dogs in response to various external stimuli, surgical externalisation of the salivary glands being used to allow saliva to run off and be measured, usually in terms of number

of drops. The standard Pavlovian experiment presents a restrained and hungry dog with an electric buzzer, which if sounded is followed 10 seconds later by a piece of meat dispensed in front of the dog by a remotely controlled delivery apparatus. Dogs do not normally salivate when they hear a buzzer, but

67

even after one experience of the conjunction of the buzzer with food, a dog may salivate when the buzzer is sounded again a few minutes later (Pavlov, 1927, p. 27). To establish a reliable response, a dog might be given half a dozen pieces of meat, each preceded by the buzzer, at 5-minute intervals, every day for a week or so. After this a demonstration of the 'conditioned reflex' could be given by simply sounding the buzzer, for 30 seconds, without giving any meat. This should now produce the same effect as showing the dog real food—copious amounts of saliva will be secreted, and the dog will lick its lips, look at the food dispenser, and perhaps wag its tail.

-

Much has been deduced from this sort of simple experiment. The first point is that salivation is no more an intentional action than gastric secretion, and so the idea of the dog salivating as a reasoned and deliberate action does not really arise. There are plenty of instances when 'conditioned reflexes' seem extremely simple, rigid, automatic and unconscious. If both the stimulus and the response are confined to the digestive tract for instance, as in the experiments of Bykov (1957) one would be reluctant to appeal to higher mental activities as the intervening mechanisms. Similarly investigators subsequent to Pavlov have claimed to find conditioned reflexes mediated by the spinal cord in mammals (see Ince *et al.*, 1978 for recent evidence of a conditioned reflex in lower spinal cord of a human paraplegic) and it is possible to get fairly conventional Pavlovian conditioning in animals whose cerebral cortex has been removed (see Russell, 1971). The beginnings of this trend lie in the experiments by Twitmeyer (1902) which seemed to show classical conditioning of the human 'knee-jerk' reflex. This result was undoubtedly influential in establishing the conventional view of Pavlovian conditioning as a process requiring only very rudimentary 'stimulus-response' associations. In any case the interpretation of Pavlov's results given by learning theorists such as Guthrie (1930) and Hull (e.g. 1929, 1937, 1943) has meant that 'conditioning' has become a watchword for explanations of animal (and human) adaptations to experience which exclude thought, mental ideas or inference, or anything of the kind.

It is important to point out that these minimalist interpretations are in violation of much of Pavlov's evidence, and that Pavlov himself, although he attacked the vague mentalisms of contemporary introspectionism and 'Gestalt' psychology, held relatively subtle views on the relation between conditioning, brain function, and cognition. His intention, already evident in his Nobel speech of 1904, was to provide a

68

scientific account of 'higher nervous activity'—the workings of the

cortex of the mammalian cerebral hemispheres. 'The cerebral hemispheres stand out as the crowning achievement in the nervous development of the animal kingdom' is the first sentence of *Conditioned Reflexes* (1927) which is subtitled *An Investigation of the Physiological Activity of the Cerebral Cortex*. Pavlov was aiming at a theory of the crowning achievement, not a reduction to the lowest common denominator of neural adaptability. He hoped for a fusion of the psychological with the physiological, of the subjective with the objective' (1955, p. 455). In the case of man, evidence concerning the fusion can be obtained because 'distortion of the human subjective world is linked, obviously, with anatomical and physiological disturbances of the higher part of the brain' (ibid). In the case of other mammals, it is impossible to know how to define subjective understanding, but, by the use of tools such as the method of conditioning, the animal's adaptation to experience can be discussed in terms of the 'analysing and synthesising' activities of the cerebral cortex in their brains. Thus Pavlov's treatment of 'higher nervous activity' was somewhat more liberal and flexible than many of his followers. And, moreover, it is arguable that the evidence from Pavlovian conditioning is consistent with even more liberal - and flexible theoretical interpretations than Pavlov ever put on it.

Pavlov's theorising became detailed and complex, but it is possible to give a summary in terms of two general mechanisms, as he did himself succinctly in 1909 in 'Natural Science and the Brain' (see *Selected Works*, 1955) and more discursively in 1927. The first mechanism is the formation of temporary connections; this explains how arbitrary stimuli are able to function as signals, as in the case of the buzzer signalling food in the standard experiment discussed above. Note that these connections are not quite so restricted as those proposed by Thorndike, since they are not necessarily just connections by which a stimulus forces a particular response out of the organism. Signals for food may direct animals to acquire it, rather than only eliciting salivation (Pavlov, 1955, p. 210). Exactly how this happens is never very clear, but it would be Pavlov's line to say that various parts of the brain are used to react to food itself, and signals for food are stimuli which, because of previous associations, get these parts of the -brain going even when food isn't there.

The second mechanism, treated separately from the first sort of simple association, is related to traditional concerns about the methods

69

by which the external world is perceived and classified. Pavlov discussed these methods in terms of 'analysers'. 'An analyser is a complex nervous mechanism which begins with an external receiving apparatus and ends in the brain' (1955, p. 215). This sounds like what might nowadays be called a sensory pathway, but naturally the emphasis is on the way in which sensory information is received in the brain so that 'the given analyser can decompose the external world' (Pavlov, 1955, p. 216).

The salivary conditioning technique was used (a) to establish how far the sensory analysers could become sensitive to subtle distinctions in the external world, and therefore make fine discriminations, and (b) to discover the extent to which the analysing apparatus can synthesise experiences, this mainly in the sense of responding selectively to combinations of individual stimuli, or to members of sets of stimuli. Some of the results are well known, but others have tended to be disregarded by subsequent commentators on Pavlov's work. The simplest result is that a large amount of association by similarity is the usual concomitant of conditioning with a single stimulus. Thus if a note of middle -C is made a signal for food, almost as much salivation will also be obtained if B flat or D is sounded, with progressively diminished responses to notes further away from the original. This is called by Pavlov 'generalisation', which has now become a common term; but he also saw it as a form of synthesis. It is perhaps not quite the same thing as what Aristotle meant by association through similarity, but clearly similarity is involved. Sound frequencies are not perceived in isolation from one another, but are organised on an internal scale of pitch.

Fine discriminations

However, generalisation along the scale is not a fixed and invariant process, but depends on experience. In particular, there appears to be a process of comparison, or contrast, within particular dimensions. If a dog continues to be given food whenever middle C is sounded, but gets no food for interspersed B flats, it will very quickly cease to give the generalised response to these non-rewarded tones. This Pavlov called the method of contrasts, and it was used because it led so rapidly to dogs salivating only to very precisely defined sensations. Protracted experience of just middle C may mean conditioned responses for any

70

variation in pitch are much reduced, but this is an unreliable and extremely lengthy procedure for obtaining fine discriminations.

Pitch was one of the sensory dimensions examined extensively in Pavlov's laboratory. Usually wind instruments were used to produce the various tones, but Galton's whistle was needed for high frequencies, inaudible to the human ear, and the results were replicated with pure tones from automatic resonators. Dogs became 100 per cent accurate even with notes only one-eighth of a tone apart, which was the limit of the available apparatus. They were sensitive to the timbre of the wind instruments (although the limits of this were not established) and could detect differences in the rhythm of repetitions of the same note (e.g. between 100 and 96 beats a minute) which were not apparent to the human observer. So much for Thorndike's contention that animals are not capable of fine discriminations.

Synthesis

The 'synthesising function of the nervous system' was seen by Pavlov as a rather general characteristic. But as usual its nature was illustrated, rather surprisingly, by particular techniques of salivary conditioning. The most obvious type of result, which was said by Pavlov to be easy to obtain, without much -further discussion, is selective response to a simultaneous combination of stimuli. For instance, if the dog gets food when a whistle and a tuning fork sound together, but not when these events occur singly, it only salivates to the combination. Similarly, if a lamp and a tone turned on together precede food, but are experienced individually without the food, only the double stimulus is perceived as a food signal. This procedure works equally well in the other direction; that is, if either of the two stimuli individually signal food, and elicit salivation, but the two presented together are always unrewarded, and are therefore *not* responded to in combination. The effects of combined sensations on salivation is not perhaps what Locke had in mind in discussing the compounding of ideas, but the effectiveness of arbitrary combinations of sounds in functioning as a distinct signal should throw doubt on whether an inability to perceive compounds is a defining limitation on animal thought.

Much more detailed work was done on a different sort of compound stimulus, where individual components were presented successively, in various orders or rhythms. In this case, says Pavlov, 'the synthesizing

71

function of the nervous system is still more obvious' (1927, p. 145). Using a single brief tone, a signal could be composed of three soundings of it, with a 2-second interval before the first repetition, and then a 1-second break. This particular rhythm could be made the signal for food, by comparison with the alternative pattern of a 1-second break between the first two tones, followed by the 2-second interval. This alternative would certainly be responded to the first time it was given, but if the original sequence continued to signal food whenever it was given, but the reversed sequence was never rewarded, then dogs came to salivate when they heard one rhythm, but not the other.

Another sort of compound discrimination tested was that of an ordered sequence of different events. For instance, a rising arpeggio of the notes C-D-E-F within one octave was used as a signal for food, the dogs differentiating this from the other 23 sequences of the same notes such as C-E-D-F, C-D-F-E, and so on to F-E-D-C. This indicates a considerable degree of integration over time, or sensitivity to stimulus order. The effect of order was apparent also over longer intervals, even when it was less obvious that a particular order was a necessary condition for getting food. When a dog was usually fed once every 10 minutes over a period of 80 minutes each day, but with a fixed order of 4 individual signals, repeated, which was a metronome the first time, an electric lamp the second, a whistle the third, and a touch on the body for the fourth, the dog was measurably disturbed when this accustomed sequence was changed. It was this result which led Pavlov to the view that what he was investigating was a 'complex dynamic system' rather than a

collection of isolated reflexes (1927, p. 232).

Pavlov's evidence can therefore be used just as well to support a theory of animal psychology which supposes that the inner causes of behaviour arise from 'complex dynamics' as to deduce that the inner mechanisms are like simple clockwork. Indeed, the results just quoted can be read as a demonstration that what Locke regarded as the underlying elements of human thought—discernment, comparison, and composition—are general aspects of the operation of mammalian cerebral cortex.

It should be admitted that although Pavlov occasionally took an imaginative view of the psychological implications of his results, the main thrust of his theorizing was rather cumbersome and mechanistic. The essential axiom of his theories was the existence of physiological processes of excitation and inhibition in the brain, whose ebbing and flowing determined behavioural responses, and it is discussed in terms

72

of simple mechanisms of excitation and inhibition that has been the major legacy of Pavlovian theory in English-speaking psychology (see Gray, 1979). Thus although Pavlov expressed the hope, as a final conclusion to *Conditioned Reflexes*, that his experiments would throw light on the relation between conscious and unconscious thought, his influence has not been in this direction. As he himself tended to discuss individual results as evidence for simple physiological mechanisms, he encouraged the tendency to view animal behaviour as collections of knee-jerk-like responses to external stimuli. However, it has recently become more common to give explanations for classical conditioning results in terms of the existence of more complex perceptions and cognitions. Thus it may be said that the dog in the standard experiment 'expects food' or has an 'internal representation' of food, when it hears the buzzer (cf. Mackintosh, 1974; Rescorla, 1979). The usefulness of this type of explanation can be illustrated by a minor experiment mentioned by Pavlov in passing. A dog was initially touched on the side before being given a mixture of breadcrumbs and dried meat powder, and developed the conditioned tendency to salivate when touched. But, when an additional association was introduced to the same animal, of a visual signal followed by Dutch cheese, salivating to the touch was much reduced, even if just the signal for cheese came at the beginning of an hour in which the touch/dry mixture sequence was tested. There are many ways of interpreting this phenomenon. Pavlov said it showed 'analysing activity of the cerebral part of the chemical analyser for taste' (1927, p. 139). It would be just as parsimonious to suggest that if the dog thinks it might get cheese, it is not going to be very interested in the possibility of breadcrumbs.

However, Pavlov was the archetypal brain-state theorist, and consistently advocated the view that the subjective world of the animal should be disregarded, in favour of descriptions of hypothetical brain activities. Even though the nature of his experimental techniques means that the Pavlovian view of animal behaviour is taken to be that their activities are usually reflexive and determined by external stimuli in a mechanical and unthinking

manner, it should not be forgotten that Pavlov firmly asserted that the brains -of man and dog work in fundamentally similar ways. This can be interpreted, and usually is, as a slight on human capacities, but can also be used to emphasise the potential of the animal brain. Pavlov himself oscillated between unifying human and animal mental phenomena under the banner of conditioning ('habits based on training, education and discipline of

73

any sort are nothing but a long chain of conditioned reflexes') and setting them apart at the familiar dividing line of speech. Speech 'provides conditioned stimuli' (1927, p. 407) and so is an expression of the fundamental signalling characteristic of the cerebral cortex, but at the same time the 'second signalling system', which is language, is self-contained and is independent of the 'first signalling system' of nonverbal perceptions, which is shared by man and other mammals (1955, p. 537). Thus although Pavlov's techniques established that animals can make extremely fine sensory discriminations, Of both individual and compound sensations, the net result of his work has been to reinforce the Cartesian opinion, that animals may be rather efficient machines, but lack anything corresponding to human awareness or imagination.

Watson (1878—1958) and Hull (1884—1952)

The mechanistic and minimalist treatment of animal abilities which began with Thorndike and Pavlov was firmly implanted into the mainstream of academic psychology by the behaviourist tradition begun by Watson and continued by Hull and Skinner. 'Behaviourism' as a term was coined by Watson in a paper in 1913, which was followed up by textbooks in 1914 and 1919. Oddly, the three Americans, Watson, Hull and Skinner, were influenced just as much by Pavlov as by Thorndike—it is 'conditioning' rather than 'trial and error learning' which is the dominant concept. Watson's research had been in maze learning and colour vision in animals, but the appearance of the conditioned reflex method, rather than his own work, seems to have been the inspiration for the idea of behaviourism. There has always been a number of separate themes within behaviourism: first, a denial that subjective or mental events should be appealed to as causes of overt acts; second, the substitution of simplified explanations, derived from Pavlovian conditioning; and third, a reluctance to admit into the explanatory system not only innate ideas, but innate influences of any kind, except in the form of innate reflexes.

Watson is important as the originator of the behaviourist movement, but his academic career was cut short when Johns Hopkins University fired him in 1920 because of an affair with his research assistant. After this he worked his way up to a prominent position on Madison Avenue as a partner in J. Walter Thompson. He is best

74

known for his application of behaviourist principles to people rather than animals—his belief was that human behaviour is a result of

‘millions of conditionings’ during life, and not affected by inherited predispositions.

If Watson’s presidential address to the American Psychological Association in 1915 sounded a behaviourist call to arms. Hull’s, in 1936, announced a battle won: in the 1940s 70 per cent of the papers about animal learning or related subjects in the main journals referred to Hull’s theories (Spence, 1952). After putting forward ‘A functional interpretation of the conditioned reflex’ in 1929, Hull went to Yale, and was a dominating figure there until he died in 1952. His domination was due almost entirely to theoretical systematisation, rather than the discovery of new facts—although enormous amounts of experimental work were done in the name of the theory, it is the theory itself, rather than major additions to the experimental findings of Thorndike and Pavlov, which one looks back to.

Hull’s theory could be described as the apotheosis of the simplified explanation. The aim was to reduce all behaviour, and therefore all psychology, to a few underlying principles, common to rats and human subjects. Postulates of the theory (16 of them in Hull, 1943) were written using an elaborate system of mathematical symbols, in imitation of Newton’s *Principia*, the intention being to construct a rigid account of psychology, in which the behaviour of rats in mazes could be related to underlying principles in the same way that the movements of the heavenly bodies can be predicted from Newton’s laws of motion. Although it would probably be now agreed that this attempt was always doomed to fail, during its heyday it sustained the most mechanistic aspects of Pavlov’s and Thorndike’s beliefs.

The underlying assumption in Hullian theorising is that *all* behaviour can be construed as responses to stimuli. After Watson and Thorndike, this was not unusual—‘Stimulus- response’ or ‘S-R’ psychology became the conventional wisdom. In Hull’s theory (although there are various minor qualifications) whether or not an animal makes a certain response, when for instance it is put in a puzzle box, or hears a buzzer in a salivary conditioning experiment, depends on two aspects of its current state: drive and habit. Drive is a function of need, but only in so far as needs create inner tensions, not because needs make animals desire future goals. Habit is simply an accumulation of previous experience, expressed as the tendency for a stimulus to evoke a response. The essence of this theory is that there is a deliberate

75

exclusion of anything like a conscious awareness of perceived stimuli, or an expectation of future events. Hull had found he could make very rudimentary electrical circuits mimic a response to a conditioned stimulus (Hull and Kruger, 1931) and directed much of his theoretical efforts to interpreting more complex phenomena as outcomes of the simple mechanisms of habit and drive (e.g. Hull, 1937). As with Pavlov, it was not Hull’s intention to inculcate the opinion that animal intelligence worked in a fundamentally different way from human thought; indeed, he wanted to apply his general theories to child psychology, and psychoanalysis (Hull, 1938, 1939).

But it has been one of the legacies of his theory that many concerned with explaining human perception and action, although they may regard stimulus-response theories as self-evidently inadequate when it comes to human behaviour, suppose that stimulus-response theories work reasonably well for the simpler task of explaining perception and action in rats, cats and dogs.

In fact, the plain postulates of drive and habit work for only the crudest sorts of Pavlovian conditioning experiments, and do not get very far in explaining how or why a dog retrieves thrown sticks. Cumbersome elaborations are needed to account for orientation in space and for rapid reactions to changed rewards. The emphasis on these features in animal experiments, which began to undermine stimulus-response theory almost from its inception, was largely due to Tolman.

Tolman (1886—1959)

Unlike the stimulus-response theorists, Tolman assumed that even rats possess some generally accessible knowledge about where they are, and what is likely to happen next, their actions being based on this knowledge rather than driven by automatic habits. He called himself a behaviourist, and relied on experimental evidence, rarely worrying about the question of subjective awareness in animals in so many terms. However, he took the behavioural evidence to require the supposition of some inner purposes and cognitions, which ‘intervened’ between perception and action or, more strictly, between the objective environment and the observed behaviour (see Tolman, 1932, 1948, 1949). An important form of cognition is knowledge of local geography—the term adopted by Tolman, which has now passed into

76

general use, was the ‘cognitive map’ which, he said, was used by both rats and men to guide their movements from place to place. The behaviour of most animals in their natural environment suggests some fairly sophisticated methods of navigation, the migrations of birds and the homing of pigeons in particular still posing puzzles as to exactly what the methods are. A mammal that has a nest and uses a water hole does not usually appear to learn every possible route between them by Thorndikean trial and error, but often gives every sign of ‘knowing’ its territory well enough to make the term ‘map’ seem plausible.

The demonstration of spatial knowledge in laboratory tests is not quite so straightforward, but Tolman conducted a number of experiments on the ability of rats to find their way about artificial mazes that were sufficiently convincing to require major additions to simple stimulus-response theory.

Place learning

An experimental test devised relatively late in the Hull—Tolman controversy used a cross-maze, with four end-points that can be

labelled as North, South, East and West for convenience, though compass orientation as such has not been shown to influence rats. Rats were started at North and South on alternate trials, some of them getting food at one place—East, say—every time, so that they had to turn left or right depending on where they started. Food reward for the others was shifted from side to side so that to get it they had to turn left all the time, whether they started from North or South. When this experiment was done by Tolman (Tolman, Ritchie and Kalish, 1946), the rats that had to go to the same *place* every time learned to do this correctly very quickly, whereas most of the rats supposed to make the same *turning response*, to end up in different places, never managed to do this consistently. Tolman's conclusion was that the exact nature of the response was not particularly important and that it was much easier for the animals to learn that food was always in the same place, even if they had to do different things to get there, compared to having the food in different places, but always available after a left turn. Although clearly the ability to learn about places must depend on the availability of some landmarks or spatial cues, there is really no doubt at all that animals in general and laboratory rats in particular learn something about geographical locations which is not coded in terms of

77

particular sequences of movements, and this rather ordinary fact has never been incorporated satisfactorily into strict stimulus-response theories.

Hull's attempt to deal with evidence such as this, which suggests the use of a 'cognitive map', was the elaborate idea of 'Habit Family Hierarchies'. A successful response, such as turning left coming from North, would not only increase the strength of that particular habit, but would also increase other habits, perhaps as yet unperformed, such as turning right from South. This is what is known as an *ad hoc* modification of a theory, since it is not clear why turning right from South should be such a close relation of turning left from North in the rat's family of habits. Here another elaborate idea of Hull's comes in—the 'fractional antedating goal response' (see Hull, 1934, 1937, 1952). The way this should work can be summarised as follows. After turning left from North, the rat runs along the 'going East' bit of the maze just before it gets its food. 'Going East' comes before food, and is thus like a conditioned stimulus in a salivary conditioning experiment. Something about going East might thus make the animal salivate, or produce some other internal response characteristic of the goal—hence the 'fractional goal response'. But the 'going East' part of the turning left from North is shared by the response of turning right from South, and thus this second turning response will share in the anticipatory goal response—and that is why rats turn right from South after being rewarded for turning left from North.

This makes it pretty obvious why Tolman's cognitive maps have proved to be more popular than Hull's little goal responses and families of habits—Hull's system seems to require more cognitive effort by both the human theorist and the rats. The important part of

both theories is that they assume a large degree of *organisation* of knowledge (or of habits or behaviour determining processes of whatever kind) inside the animal's head. A labelling which corresponds to 'food at point X' is available to influence many kinds of movement towards X by a hungry animal, which may start from various other points of space, and which may make use of various muscular movements. A rat possessing the information 'food at X' will jump, climb, dig, gnaw and swim in the course of getting there (e.g. MacFarlane, 1930).

More complicated experiments (Tolman and Honzik, 1930a) involved the choice of different routes to get to 'food at X' when the most direct path from a starting place to 'X' was blocked. If there were three paths which all led to X, of different lengths, then rats very

78

sensibly learned to choose the shortest one. But if the shortest route was blocked, they switched to the next longest one. However, if the shortest path was blocked at a point where the middle length route had already rejoined it, then, according to Tolman and Honzik (1930a) most rats did not bother to try out the middle-length path, but went the long way round, demonstrating, according to Tolman, 'insight'. Not everyone has agreed that this is a reliable result, but Deutsch and Clarkson (1959) were able to obtain something similar by having two short routes to 'food at X' and one very long path, beginning at the same point as the two short ones, which eventually led to 'food at Y'. If one of the short paths was blocked, then a significant majority of rats (after experiencing the block) took the alternative short route. But if the animals found 'X' empty of food, or were prevented from eating by wire netting, they went back and tried the long path to 'Y', the alternative goal. More recently, with a technique which has been easily replicated, it has been shown that, given several bits of food in the same maze, for instance at the end-points of radiating spokes, rats are very good at not retracing their steps, that is, not visiting again a place where they have already eaten the food (Olton, 1978, 1979). All these results mean that some kind of memory for places is being utilised. It may not involve anything like human conscious memory, but the evidence as it stands suggests that what rats do in mazes is determined more by internal organisation of spatial information than by the minimal Hullian habit.

Approach to objects in open space, and around barriers

The advantage of searching for the simplest possible mechanisms to explain orientation in space is that it demonstrates the complexity of some very ordinary achievements. If a bowl of food is put down in the middle of a room, we are not surprised if a cat turns towards it from various distances and directions or even if the Cat detours around or over other objects placed in the way. It is convenient to appeal to an internal 'cognitive map' to account for this, but in some ways Hull's more tortuous explanations are more revealing (Hull, 1952, ch. 8). Even to deal with simple approach to objects, special assumptions about the 'generalisation' of locomotor

responses have to be made, and similarly, perception of objects has to be assumed to generalise over a 'distance-reception continuum'. To account for detouring round

79

barriers (the 'Umweg' problem) Hull has to bring in whole systems of habit-family hierarchies, in which the shortest routes to goals have highest priority, but if these are not available (strictly speaking, after these have been knocked out by trial and error experience), more circuitous movement habits are substituted. It is evident that a considerable amount of inner shifting of gears, selection among possibilities, and application of principles, would be needed to make such a system work. Certainly, it is more straightforward to say that a cognitive map is being used, but it should not be forgotten that utilisation of a cognitive map may involve internal computations of considerable complexity. Equally, cognitive maps are not always perfect. Animals can get lost, and there are many occasions when they do not do the intelligent thing, and detour around barriers, but persist with rigid attempts to push their way through. The lesson to be drawn from Tolman's emphasis on orientation in space, and Hull's attempts to deal with it, is that to explain overt moving about, it is necessary in some cases to assume that there are means of inner mental organisation of spatial knowledge.

Cognitive maps and 'latent learning'

The emphasis on purely spatial aspects of an animal's internal knowledge was a comparatively late development in Tolman's theorising. His early concern was to provide a behaviourist account of the facts of consciousness and imagery ('A new formula for behaviourism', 1922) and of emotional experience, and his main theme remained close to that of 'Purpose and cognition: the determiners of animal learning' (1925) throughout his career. As a general strategy, he recommended that ideas, intentions, belief and expectancies could best be studied by linking them with observable acts, rather than by examining them directly in subjective experience, and that the laboratory rat should serve as a 'furry white test tube' in which these sorts of cognition may be seen to take place. Tolman's explicit conclusion in 'A behaviouristic theory of ideas' (1926) is that the best way to find out about the private mental contents of thought is to look at 'a really good rat in a really good maze'.

Few would now wish to follow Tolman in attempting to derive a universal theory of cognition and thought from the behaviour of rats in mazes, but we are interested here in whether rats demonstrate any

80

thought at all. (Tolman wrote occasionally about wider issues, as in *Drives towards war*, 1942; and he spent the first year of the war advising the OSS, which afterwards became the CIA; but one would not want to advance this as support of the general applicability of his theories.)

The crucial aspect of cognition that Tolman felt could be

demonstrated in rats is some sort of awareness of the likely consequences of actions. There is an absolute distinction between Tolman's view that mammals perform responses because they expect them to bring about desirable ends and the more conventional behaviourist assumption that movements only occur as automatic reactions to stimuli. Evidence in favour of Tolman's view comes partly from experiments in which variation in the availability of goals seems to have a direct effect on the animal's actions. One type of experiment of this kind is said to demonstrate 'latent learning', but is now more important in indicating relationships between behaviour and purpose.

The original experiments in Tolman's laboratory at Berkeley (Blodgett, 1929; Tolman and Honzik, 1930b) employed mazes in which rats had to make a sequence of 6 or 14 choices (to turn left or right) in correct order. Hungry rats given food at the end of the maze each day seem gradually to learn the correct sequence, performing randomly on the first day, but almost perfectly after one or two weeks (depending on the length of the maze). This could be interpreted as a gradual 'stamping in' of correct habits. Also animals allowed to run through the maze for an equivalent number of days without food rewards apparently learn very little, since they continue to turn into many blind alleys. At first sight this supports the idea that no reward means no 'stamping in' of correct responses. But if, after this preliminary experience, food was put at the end of the maze on one day, the rats ran the mazes almost perfectly on the next day.

The first deduction is that by the time the food was introduced, the animals already knew what the quickest way to run the maze was, even though they did not bother to take it—hence 'latent learning'. Just as important, however, is the fact that having found food at the end of the maze, after an incorrect sequence of responses, the rats ran through by the quickest route the next time they had the chance. A possible interpretation is that (a) the rats knew the quickest way to get through the maze, but also, (b) they now expected that food might be at the end of the maze and (c) they wanted to get to the food as quickly as

81

possible, and (d) putting together all these resulted in the action of running the maze correctly.

There have been various modifications of these original latent learning experiments. If rats are allowed to explore a simple 'T-maze' incorporating a single left/right choice, and then fed once in either the left or the right arm of the 'T', they turn towards the place where they were fed when next allowed to run the maze from the beginning (Seward, 1949). If on the other hand they are put in one side of a T maze and given an electric shock, they avoid going to that side in a subsequent run (Tolman and Gleitman, 1949). If rats experience food on the right and water on the left when running through a one-choice maze when neither hungry or thirsty so that they do not actually eat or drink, then if half of them are tested when hungry, they will run to the right, and others tested when thirsty will go to the left (Spence and Lippitt, 1946). There are some variations

in the degree of success of this kind of experiment: it is necessary, for instance, to make sure that the animals can tell one part of the maze from another by differences in visual or textural cues, but it is now accepted by authorities such as Mackintosh (1974) that the evidence for latent learning in rats is unassailable.

Expectancies, beliefs and means-ends readiness

Precisely what theoretical conclusions should be drawn from the empirical phenomena of latent learning, and other sorts of intelligence exhibited in maze running, is not so clear. Tolman himself used a bewildering variety of terms to describe the internal mental processes he felt were involved, from 'initial cognitive hunch' (1925) and 'Sign-Gestalt' (1927) to 'expectancies', 'representations' and 'means-end readiness' (1959). There is currently a consensus of using 'expectancy' rather loosely as an explanation, together with the assumption that associations between responses and their consequences is a typical form taken by learning in the rat (Bolles, 1972; Mackintosh, 1974; Roitblat *et al.* 1984). In all these cases it is obvious that some kind of mental activity is being attributed to the animals; that is, there is considered to be some internal sifting and selection of information rather than simply the release of responses by a certain set of environmental conditions. Knowledge of goals, knowledge of space and knowledge of actions that

82

may lead to goals seem to be independent, but can be fitted together by animals when the need arises.

Tolman's theory of consciousness and habit

It may be necessary to dispel the impression that rats are all-knowing, and never perform automatic actions. It is not difficult to perform experiments in which rats are apparently oblivious to everything except laboriously acquired mechanical habits. Tolman described an occasion on which he trained rats to run from one end to the other of a long narrow box, pushing through a white Curtain halfway down, to get into a food compartment at the end. He then put them directly into the food compartment but gave them an electric shock instead of food. On the assumption that the animals knew that they would end up in the same food compartment if they ran through the box, he was confident that they would now refuse to run. But when he carried the first rat back to the starting point it 'immediately dashed off gaily and just as usual through the whole discrimination apparatus and bang whack into the very food compartment in which he had just been shocked' (1933, p. 250), and the others did the same. Only when they had been shocked after getting themselves into the food compartment did they refuse to run again. This is fair enough in any case but Tolman supposed that long practice had made running automatic and 'fixated', and more like a traditional conditioned response. Any theory of cognition must include the possibility that much practised actions, which may have initially required active thought, can eventually be made

habitually and automatically.

The opposite of a habitual and automatic response should perhaps be an action performed with a conscious intention. Since Tolman supposed that the purposes of humans and rats were of the same type (e.g. Tolman, 1932), and that human thinking 'is in essence no more than an activated interplay among expectancies', it is reasonable to align him with theorists such as Lloyd Morgan who assumed that animal behaviour may be guided by conscious memories and expectations. However, he was enough of a behaviourist to neglect the question of whether mental activities deduced from the responses of rats ought to be equated with what is available to human introspection. His early attempt to pin down the behavioural criteria for a moment of animal consciousness was not a success. It went like this: 'whenever an

83

organism at a given moment of stimulation shifts then and there from being ready to respond in some relatively less differentiated way to being ready to respond in some relatively more differentiated way, there is consciousness' (Tolman, 1927, p. 435). This seems to say that any novel response is a product of conscious mental organisation (cf. Shallice, 1972), and such a view is implicit in much of Tolman's writing, but on the whole he preferred to talk of 'Sign-Gestalts' and 'beliefs', tying these down to observed behaviours rather than putative subjective experience.

Skinner's operant conditioning: doctrines and techniques

Since Hull's death in 1952, Skinner has been the leading light of behaviourism. He solved the theoretical disputes between Hull and Tolman by ignoring them entirely, and thus under his influence most of the problems of animal cognition were left in abeyance. Skinner's behaviourism is of the most radical kind, since all questions concerning inner mechanisms, whether these are in terms of neural machinery, or of learned habits, or of mental states, are held to be improper. Naturally such an extreme position has led to a degree of academic isolation for Skinnerians —investigations of neurophysiology, and of human cognition, have flourished quite independently, but the existence of a strong Skinnerian school built up during Skinner's time at Indiana from 1945—8 and at Harvard since 1948 has until fairly recently acted as a heavy counter weight against speculations along the lines pioneered by Tolman.

To a great extent, Skinner's influence rests on his development of new and powerful experimental techniques and the addition of new experimental findings to the corpus of data concerning the behaviour of laboratory animals. It is important to distinguish, therefore, between the reliability of these findings and the interpretation (or rather, lack of interpretation) which Skinner chose to put on them. Both the findings, and the theoretical positions, were published in *The Behaviour of Organisms* in 1938 and have changed very little—since 1948 Skinner has become known for his promulgation of contentious behaviourist recipes for human society, rather than for further analysis of the causes of animal behaviour.

A large part of Skinner's technical achievement was in

haviour in animals. The 'Skinner box' is a version of Thorndike's puzzle box. When an animal presses a lever or a button in the box, it does not get out to receive food, but a small amount of food is delivered to it while it remains inside. It is then free to press the lever again, and get another pellet of food, and to go on like this for as long as it, or the experimenter, chooses. In essence, this is simply another way of doing Thorndike's experiment (see pp. 62—5). The experimenter is freed from the task of replacing the animal in the box, and if the response of the animal is recorded electrically (in Skinner's experiments, it was made to pull a pen a small distance over a steadily unrolling sheet of paper, thus plotting an automatic 'cumulative record' of the behaviour), the human experimenter does not need to stand by with a stop-watch. But the results of the experiment in the basic form do not differ in the slightest from those of Thorndike. The finding is that when a hungry rat is left in the box, it will eventually put its paws on the lever and press it down, in the course of natural exploration. If this press on the lever results in the delivery of a food pellet, and it has previously learned to recognise when this happens, it will go and eat the food, and come back and press the lever again almost immediately. I have pointed out that most of Thorndike's cats needed only one experience to exhibit virtually complete learning of a simple successful response, and Skinner (1938) made a great deal of the fact that his rats usually switched to a regular pattern of pressing the lever and eating at a steady rate, as soon as they had tried it once. This counted as a case of 'one-trial learning', and Skinner wrote to a friend at the time that 'it leaves the insight boys with their mouths open' (see Skinner, 1979, p. 90). It may be said that Skinner tended to exaggerate the instantaneity of the results, much as Thorndike exaggerated the gradualness of his—the rats often took several food pellets before they got into their stride (see Skinner, 1938, pp. 66—9).

The fact that rats will perform the simple yet apparently arbitrary movements required to manipulate an object, because this results in their getting food, is the most basic of Skinner's findings. It was not particularly novel, except that pushing down a horizontal metal rod about a centimetre had not attracted much attention previously. It is the assertions made by Skinner about why this happened that are novel. Few before (or since) Skinner have been quite so systematic in denying the necessity for appealing to events inside the animal as explanations for what it does. Skinner contended that following the response with food increased its *strength* as measured by its future

performance. As a descriptive shorthand this is difficult to argue with. But Skinner's argument was that anything else except a descriptive shorthand is necessarily invalid. This is a defensible, if vacuous, methodological point. One may be pessimistic, and say that the evidence that the behaviour occurs will always be

insufficient to decide why it occurs. In other words one might decide to leave the question of explanation open. But this is not the outcome of Skinner's deliberations. Instead, he takes an overtly minimalist view—the rat pressing the lever and then rushing over to retrieve its food pellet is said to be exhibiting a chain of reflexes which are like the reflexes used in swallowing, or in postural adjustment (Skinner, 1938, pp. 54—5) which are quoted precisely because they are distinguished as being 'purely mechanical'. Skinner makes the case that the rat performs the response without knowing why it does so, in any sense of the term 'knowing'. He has recently reiterated his position with the example of a cat chasing a mouse. The movements of the cat, according to Skinner, are exactly like inherited reflexes—we should not assume that there is anything in the cat's brain which corresponds to trying to catch the mouse, or even that the cat likes the chase (Skinner, 1977).

Despite Skinner's protestation to the contrary (Skinner, 1950) this *is* a theory, and not just a methodological strategy. It is a strong statement that rats and cats do not have any cognitions. It is instructive to contrast this theory, applied to very simple behaviours, with that of Tolman (1937). Tolman wrote about the manipulations performed by rats to get food, but the case in point was slightly more elaborate than the lever press in a Skinner box; it concerned the ability of rats to draw a metal tray of food towards them by pulling on a string. Although monkeys and chimpanzees need very little persuasion to pull towards them delicacies tied on the end of string, similar behaviour requires a certain amount of training in rats. The method used by Tolman was first to accustom a rat to eat from a metal pan, which could be slid through a gap under the front of the animal's testing cage. On the day of string-pull training, a string attached to the pan ran through the cage, but was not needed at the start, since the pan was still inside. The first step was to try and slide the pan out of the cage while the rat was eating. The usual reaction of the rats was to immediately clutch the pan and pull it back. Once this began, the superior strength of the experimenter was exerted to drag the pan first to the limit of the animal's reach, and then just beyond it. When the pan itself was beyond reach, some rats managed to get it back by vigorous

86

scrabbling, which was sufficient to move the string, others would immediately pull the string with their forepaws, and occasionally one would seize the string in his teeth. After they had succeeded in pulling in the food several times, they were capable of pulling in the pan from about 50cm away. But, in these early stages, Tolman found it necessary to induce string-pulling as such, instead of general scrabbling, by either jiggling the string, 'pulling the string away and evoking a contesting tug', or, as a last resort, smearing the string with something edible.

Tolman (1937), saw these tactics as aids to the formation of new ideas by the rat, rather than stimuli which elicited only overt responses. His analysis is in terms of 'Sign-Gestalt expectations' (used interchangeably with 'hypotheses'). In paraphrase, he assumes

that before the rat learns to pull the string, the pan as an object is recognised as something from which food may be obtained, and that when the pan is pulled away the animals already have an expectation that if they clutch it or scrabble for it, they will get it back. The crucial learning consists in forming a new expectation that pulling the string, as a separate act, will also result in access to the pan.

Applying this sort of theory to rats in Skinner boxes is relatively straightforward (although Tolman was strangely reluctant to deal with other people's data). A rat which discovers, on one occasion, that the depression of the lever is followed by the delivery of a food pellet is induced to form the mental hypothesis (or expectancy, or belief) that doing things with the lever will produce food, and will thereafter manipulate the lever whenever it wants food. If it is extremely hungry to start with, and only small amounts of food are obtained each time, the lever will be operated with some regularity. Something along these lines seems to be implied by Mackintosh's recent conclusion that Skinner's basic experiment, as an example of instrumental learning, should be explained by 'the statement that an association was formed between the response and the reinforcement and that this association underlies the increase in the probability of that response' (Mackintosh, 1974, p. 222 — 'reinforcement' was Pavlov's and Skinner's term for the food).

However, any explanation of this kind would be anathema to Skinner. It is true that the bare fact that the animal makes a response does not provide grounds for deciding whether it is making it as a reflex with a certain strength, or whether it is making it because it has the idea that the response will get it something. The Skinnerian position,

87

which has been maintained with some success, is simply to refuse to discuss anything except strengths of various responses under various stimulus conditions. Whereas Hull attempted cumbersome modifications of his system, such as the habit family hierarchy, to account for extra inconvenient information like latent learning and orientation in space, Skinner resolutely ignores inconvenient facts or includes them only as examples of changes in behavioural strengths due to reinforcement by food or something acting in the same way as food.

The intriguing empirical aspect of Skinner's work is the degree to which the activities of rats, pigeons, and most other experimental animals can be 'shaped' by the careful provision of external rewards. Skinner's own *tour de force* was to train a rat not only to pull a string, but to pull a string which released a marble from a rack, which the rat then picked up in its forepaws and carried to a tube projecting two inches above the floor of its cage. Dropping the marble down the tube was the response which finally led to the delivery of food (Skinner, 1938). This performance was achieved by first rewarding the animal for pushing a marble down a hole in the floor, then for manipulating it into the tube rising a fraction of an inch from the floor, and so on. Every step was trained by a series of approximations. Our deduction is supposed to be that gradual

changes in response strength, caused by food rewards, can eventually be accumulated into a complex sequence of behaviour, and hence complex sequences of behaviour are 'nothing but' automatically produced bits of responding. It certainly seems unlikely that the rat begins with an internal belief that it must pull the string, carry the marble, and so on, if it is to get its food, although the demonstration does not prove that this sort of mental process in rats is impossible. However, it would remain open to Tolman to see the sequence as a succession of small beliefs about what to do next, held together, surely, by an overall expectation that food would be forthcoming in the end.

One factor in the success of Skinner's tactics is that such phenomena as latent learning and spatial intelligence are less apparent within the limited confines of the Skinner box than in other forms of experiment, or in naturalistic observation. However, there are a number of aspects of behaviour in Skinnerian experiments that, although they are susceptible to analyses in terms of the strengths of unseen categories of response, make the universal application of explanations based only on the strengths of reflexive responses somewhat less plausible.

These have been extensively reviewed by Mackintosh (1974), who came to the conclusion, mentioned above, that, at the very least, it must

88

be assumed that the rat learns something about the relation between the act it performs, and the reward which comes afterwards. I will mention here some of the features of Skinner's results which, it seems to me, are at least as compatible with the view that the animal's behaviour results from mental states as with the Skinnerian opinion that only various strengths of peripheral reflex-like movements need to be considered

First, in the case of the few rats which Skinner actually watched, he noticed that the early stages of lever pressing were characterised by a tendency for the animals to fiddle about with the metal rod, without actually operating it, and then to rush to the food dispenser. Since they in theory should only have had successful pushing movements strengthened, it is not obvious why any other forms of behaviour should appear, although of course Skinner interprets the fiddling about as a partial or initial stage of the 'proper' response. However, it is conceivable that the rat is working on the hypothesis that doing something rather generally with the lever will be followed by reward, and consequently searches for food after inadequate manipulations. The fact that a rat goes and sniffs and scrabbles around in the tray where a food pellet is usually delivered might well be taken as an indication that the animal expects (or perhaps just hopes) to find a food pellet there, instead of being downgraded as a manifestation of reflex-like conditioned approach movements (as in Skinner, 1938, pp. 53—4). After repeated experience a standard method of depressing the lever may indeed become habitual—this is certainly the case with human actions, such as assembly-line work, or changing the gears of a car, but it does not mean then that associated mental states were never present.

In the human case conscious mental states reappear if something goes wrong—if the gears stick, or the assembly line breaks down, Similarly if the rat's food magazine breaks down, or is disengaged by the experimenter in the procedure known as 'extinction', the behaviour of the animal indicates some change of inner mood. The lever is rattled, banged, pressed harder than usual (e. g. Notterman and Mintz, 1965) and rats under these conditions employ numerous shifts in posture, changes of paw, and so on. Hull would appeal to a related set of habits, but it makes a certain amount of sense to assume that the rat retains the belief that doing something with the lever ought to produce a food pellet, and tries everything that comes to mind. If nothing works, the rat will gradually give up trying, although

89

the original belief is not forgotten, since it can be quickly re-instated by the resumption of food delivery.

Intermittent reward

Every undergraduate text (e.g. Walker, 1975) recounts the tale of how Skinner discovered that very interesting behaviours arise if the rewards are not given for every response, but according to various 'schedules of reinforcement'. The simplest of these were discussed by Skinner in 1938 and the standard work on this subject is still Ferster and Skinner (1957). The easiest schedule of reinforcement to understand, from the point of view of procedures if not from that of theoretical explanations, is referred to as a 'fixed ratio', since the rule is that there is a certain number of responses before each reward. If the fixed ratio is ten, for instance, the rats must make nine presses on the lever without anything happening, before the tenth press delivers a food pellet, then there will be a further nine uneventful presses, and so on. If an animal has just got used to the basic procedure of being rewarded for every push (Continuous reinforcement) and the rule is suddenly Changed to fixed ratio ten, it will spend most of its time vainly searching for accustomed rewards, and may very well cease pushing the lever altogether. However, with a sufficiently persistent animal, or, as in Skinner's 1938 report, animals with prior experience of some other kind (on fixed interval reinforcement: see below), there is a gradual adjustment to the new rule. The eventual outcome is that the rat rests briefly after receiving a food pellet (but rests much longer than when it could get another pellet by the next press) and then goes over to the lever and makes ten rapid presses in succession, causing the release of another bit of food. Skinner's explanation of this is that the animal 'discriminates' the period after a reward as being a bad time as far as getting another reward goes, and 'as a result of this discrimination the rat stops responding for a short period' (1938, p. 288), and that rapid responding is 'preferentially reinforced'. This latter is certainly descriptively true: if the rat responds slowly, on fixed ratio ten, it will get its rewards slowly, and the faster it responds, the more quickly its rewards will come.

But the reader may note that a capacity in the rat for making comparisons between these cases has been surreptitiously inserted in

Skinner's account—a stricter application of the rule of blind reinforcement would say that if the rat responds slowly (as it does to begin with)

90

it will get reinforced for responding slowly, and therefore it should continue to respond slowly. Achievement of the 'typical' fixed ratio performance demands at least that the rat takes notice of the fact that if it happens to respond more quickly than usual it gets its reward sooner than usual. A more Tolmanian version of this would have the rat noticing that a certain amount of responding is necessary for each reward, and deducing that it might as well get the responding over with at a reasonable speed. There is in fact a considerable amount of evidence that the rat is indeed sensitive to the rudimentary demands of the schedule. If for instance the food mechanism is suddenly disengaged the rat will stop after ten or twelve responses and look for the food (or, if you like, make a conditioned approach to the food pan). Something like this was systematically investigated with two pigeons by Ferster and Skinner (1957, pp. 616—19). The birds were trained to make exactly fifty pecks at a disc for each food reward, but then there were arbitrary periods of 20 minutes or more when no rewards were given. During these periods the pigeons showed a clear tendency toward rapid runs of not much more than fifty responses at a time. Mechner (1958) showed that the judgment of the number of responses required in a fixed ratio is rather approximate, that is if rats are supposed to make eight responses before going somewhere else to get their reward, they often make seven or nine instead, but if it is not judgment at all, then response-strength explanations must rely on extremely *ad hoc* inner discriminations. The organisation and patterning of responding produced by fixed ratio schedules of reinforcement are thus quite compatible with the existence of inner purposes and beliefs in the behaving animals, although the usual deduction is that the performances observed in Skinner boxes require, at most, inner organisation of responses.

According to the behaviourist tradition, one might suppose that a belief *is* an 'inner organisation of responses', and nothing else. But the argument is whether the organisation of responses observed when rats (or chimpanzees) react to schedules of reinforcement is, as Skinner claimed, just like the sequences of muscular reflexes in swallowing, or leg co-ordination, which in man we know to be unconscious and automatic; or, alternatively, whether the actions of the animals are more like walking to a bus stop or ringing a door bell, which in ourselves we interpret as the result of mental intentions and expectations. Skinner, like Pavlov before him, sometimes appears to claim that both the learned activities of animals and the intentional actions

91

of men are determined, like spinal reflexes, without the intervention of thought or deliberation, and this may be the origin of some of the confusion and obfuscation which has surrounded the problem of

animal cognition in the behaviourist era.

Time-based schedules of reinforcement

Apart from ratios requiring a certain number of responses, Skinner (1938) also investigated the case where rewards are only available to the rat with a certain minimum interval between them. This is known as a 'fixed-interval schedule'. If, for instance, the fixed interval is one minute, the rat is allowed to get a reward by just one press on the lever, but must then wait for a minute or more before another press on the lever will result in the delivery of food. The most economical thing for the rat to do under these conditions would be to get a food pellet, and rest for at least a minute before going to manipulate the protruding response-device again. This sort of thing will happen, with a not very enthusiastic animal, or very small (or satisfyingly large) rewards, but is usually avoided by experimenters concerned to get decent amounts of behaviour to measure. The textbook example of fixed-interval behaviour is an animal that usually gives no attention at all to the lever for the first fifteen seconds after a reward, begins to press it now and then about half-way through the minute, and gradually accelerates to a more vigorous and rapid working of the instrument just before the first recording of a response after the interval has passed delivers a reward.

The most obvious way in which this may be explained, in terms of response-strengths mechanically stamped in by food rewards, is to say that there is a 'biological clock' which provides an internal stimulus corresponding to the passage of time, and that responses made at about the right time are thus strongly conditioned while responses made too early become relatively weak (Roberts and Church, 1978). This interpretation of a 'temporal discrimination' underlying the distribution of behaviour is the one put forward by Skinner (1938, pp. 270—7), although there is also the additional claim that slow, or spaced-out, responses are 'preferentially reinforced' on interval schedules, in contrast to the automatic selection of rapid responding in the ratio procedure.

However, in an acknowledged lapse into the vernacular, Skinner expands this account by referring to what a person reacting to the conditions might be aware of, for instance the question of 'whether it

92

was time for a reinforcement to occur' and the knowledge that 'he only has to wait long enough and a response will be effective' and the subjective factor that 'having tested the lever once unsuccessfully, he soon feels that it is time to try again' (Skinner, 1938, p. 277). It is arguable that the same sorts of mental processes have some influence on the development of the temporal discrimination in animals. Certainly the theory that rats develop expectancies as to the likelihood that a lever response will be effective can be easily mapped on to the way that responses are actually timed. After many hours of practice on a fixed-interval, or any other, schedule of reinforcement, it would not be surprising if some sequences of responding became fixed and habitual, but the

many fluctuations in the pattern of responding on fixed- interval schedules, which were remarked on by Skinner, and may go on for weeks in rats, could as well be due to interactions between hope and experience, and the development of expectations of reward, as to oscillations in response strength. The main sort of fluctuation in behaviour, apart from the timing of responses between rewards, is that animals will make rather a lot of responses between the rewards, for two or three reinforcements in a row, and then go on to make only a few lever presses in the periods between rewards for some time, before returning to the pattern of making numerous ineffective lever presses. According to Skinner's doctrine of reinforcement, a consistent and rigid pattern of behaving should emerge; he therefore went to some lengths to explain these fluctuations as a strain on response strength during the phases with lots of responses, followed by a compensatory decline during which response strength recovers (1938, p. 123 and pp. 302—3). As he says (p. 303), this is not very convincing, and it seems just as plausible that making many ineffective responses leads to a degree of despondency, whereas a sequence of 'cheap' reinforcements engenders a sense of optimism about the utility of responding, which in turn proves false.

Variable schedules of reinforcement

The single most popular Skinnerian technique is one which, it may be claimed, inculcates the most consistent and sustained expectancies of reward in the animals exposed to it, and thus gives the most reliable and persistent behaviours for the experimenter to record. This is the 'variable-interval schedule' introduced, among many other schedules, by Ferster and Skinner (1957). The procedure is the same as the fixed interval, described above, in that only one response, made at the right

93

time, is necessary for the production of a reward, but differs in that the time when a successful response can be made is unpredictable, and must remain unknown to the animal. A random sequence of intervals between rewards is provided by mechanical or electrical machinery, and thus a reward may sometimes be obtained by responding immediately after another one has just been given, but at other times rewards remain unavailable for long periods. As rewards *may* be obtainable at any time, it is reasonable to suppose that if the animal forms expectancies at all, it expects reward at any time, and is thus led to respond persistently until a reward is given, although there is some evidence that the longer the animal goes without reward, the greater is its tendency to make the required response (Catania and Reynolds, 1968). The alternative formulation, of course, is that the strength of the response is conditioned equally to all values of the internal sense of time passed since the last experience of food.

One of the most notable results of this variable-interval schedule of reinforcement is that, to an even greater extent than with the other Skinnerian schedules, exposure to the procedure means that if the

experimenter disengages the food-reward mechanism at some point, the animal continues to respond in the normal way for an extremely long time. Although it can be said that this is due to the nature of the response, it having been conditioned very strongly to occur during long periods when reinforcement is absent, the data supports equally the contention that this is due to the nature of the expectancies held by the animal.

The final permutation of the rudimentary possibilities presented by the basic Skinnerian experiment is the 'variable ratio' schedule of reinforcements (see Ferster and Skinner, 1957). This is like the 'fixed ratio' discussed above, in that a certain number of responses (rather than just one response as in the interval schedules) is necessary to produce a reward, but the exact number of responses necessary for each successive reward is made to vary at random. This is a more difficult procedure for the animal, and for the experimenter, since if, at random, a number comes that is too large, the animal will give up. (On a variable interval, if the subject of the procedure stops bothering to respond for a while, the relevant time period will continue to elapse, and a subsequent single response will be rewarded, giving further encouragement; but with the ratio version, discontinuation of the animal's attempts will leave it as far from its eventual goal as ever.) But provided the animal is persuaded gradually, by small increases in the

94

average requirement, to make more and more responses per reward, the effect is to generate the most intense devotion to the business at hand that is ever observed in the Skinnerian experimental apparatus. What is made clear in Ferster and Skinner's reports, however, is that the pattern of this responding is not sustained, but is often composed of bursts, or runs of responses, with pauses in between. The general Skinnerian tack is that this is due to a 'correlation between behaviour and reinforcement', and in particular to a preferential reinforcement of very rapid responding. But the brief bursts of responding observed during variable-ratio schedules, or after such schedules in 'extinction' when actual rewards are no longer given, would support the view that the animal is operating according to hypotheses that certain amounts of responding are necessary to produce the chance of reward (cf. Ferster and Skinner, 1957, pp. 397—411).

Skinner's operant conditioning versus theories of animal cognition

Although there has been much opposition to the Skinnerian thesis—that the experimental success of schedules of reinforcement in generating predictable and reliable behaviour in laboratory animals means that all explanatory theories of animal and human behaviour should be abandoned (Skinner, 1950) — the thesis has always carried a certain amount of weight. I have argued here that, on the contrary, the phenomena produced by Skinnerian techniques, rather than serving as a substitute for theories of animal cognition, themselves stand in need of a satisfactory explanation, and that, in some cases, such an explanation should be supplied in terms of the mental states of animals undergoing Skinnerian conditioning.

Since the heyday of Hull and Skinner, there has been a halting and tentative return to questions about animal cognition that were posed by Romanes and Lloyd Morgan, and others before them, and kept alive during the behaviourist interregnum by Tolman. I shall review briefly some of the recent developments in theory, but must begin by mentioning a writer who is not strictly speaking recent, nor strictly speaking concerned with animal behaviour. This is Piaget (1896—1980), a Swiss psychologist whose account of the development of knowledge in children has had a profound influence on educational theories in the English-speaking world over the last two decades.

95

Cognition without language in Piagetian theory

There are two reasons why Piaget's theories can be applied to the study of cognition in animals. First, he himself began academic life as a zoologist, and has provided a unified, if somewhat impenetrable, account of how human thought should be integrated with natural science, in his book *Biology and Knowledge* (1971). Second, and perhaps because of this, he is unusual among major thinkers in human psychology in denying that language is prior to cognition.

The interpretation of Piaget's writings is notoriously difficult, and I shall not attempt to discuss in detail what he says about children. The reader is referred to the assessment of this by Boden (1979). The most basic concept seems to be that of 'schemata'. These are mental entities, cognitive structures, things in the mind—the units of knowledge. According to Piaget animals have schemata, and therefore the question of animal cognition can be put in terms of the qualities of their schemata. Instincts are fixed and inherited schemata, so that a bird building a nest is using inherited *knowledge* about nest building—as opposed to only inherited *reflexes*. This seems reasonable enough: the difficulty with Piaget's style is that he says things like 'the assimilation of the pieces of straw into the forms of nest making activity is at that stage only functional, and so we speak of it as assimilation into a schema' (1971, p. 178). Instinctive schemata are not, however, so fixed as to prevent internal interplay between them, which would amount to a simple form of inference. As an extreme case, Piaget entertains the notion that inference occurs in certain species of land snail, which (a) lay eggs and (b) bury themselves in the ground to avoid predators and extremes of climate. Some species bury their eggs, and Piaget says that 'The tendency to lay eggs below the ground could be seen as the result of coordination and assimilation of the laying schema into the schema of self-protection or sheltering in the ground' (1971, p. 240). This perhaps should not be taken too seriously—Piaget has a soft spot for snails, since they were the subject of his Ph.D. thesis, and laying eggs in the ground is undoubtedly a separate and inherited behaviour pattern. But the variations observed in the location, shape, and composition of nests within the same species of bird, remarked on by Darwin, might well be best explained by the inheritance of mental concepts or schemata, which are susceptible to a certain degree of creative interpretation by individual animals.

Instinctive schemata are certainly subject to alteration and elaboration by accumulated experiences in the life of individuals: in Piaget's system Pavlovian conditioning is the passive assimilation of a new stimulus into the schemata of the original reflex, whereas in the sort of learning demonstrated by Skinner (pulling strings or pressing levers) new schemata are formed. The latter 'is no longer a matter of assimilation into previous schemata, but of construction in the sense of utilization of new relationships' (Piaget, 1971, p. 255 and p. 180).

Cognition in child and man is discussed in terms of progressive elaboration and construction of schemata (and new and sudden combinations among them), which proceeds in certain inevitable sequences, but depends on the activities and experience of the individual, and tends to occur before, rather than after, linguistic expression, especially in children. There are some stages of intelligence which only humans reach, but 'the line of demarcation between "intelligent" and "non-intelligent" is entirely a matter of convention':

there is 'continuity of the learning processes from the most elementary upward' and this 'demonstrates the permanent part played by these assimilatory activities which seem to be common not only to animals and men, but also to all levels of cognitive function' (Piaget, 1971, p. 260).

Piaget thus clearly states a strong theoretical conviction that animal thought is continuous with human thought, and shares many of its properties. However, this should not be taken merely as a dogma; it is supported to some extent by the methodology by which Piaget investigates thought in the child. As an illustration, consider the perception of objects. This may seem to be a given, in so far as we tend to assume that everyone perceives objects in the same way, but it will be remembered that there have always been philosophical disputes about this, and there are obvious questions to be asked about whether a 6-month-old infant perceives objects differently from a 3-year-old, and so on. Piaget uses several practical behavioural tests. When does a baby first reach out to grab something? How successful is it at this? When does the baby begin to follow a falling object with its eyes? If a toy or rattle is hidden under a cloth while the infant watches, does it pull the cloth away to retrieve the toy? To the extent that a child passes tests such as these, it can be said to be developing 'sensory motor intelligence', and a particular form of this is 'object constancy' exhibited by remembering that things are under cloths, anticipating the reappearance of toys passed behind screens, and so on. Clearly,

97

similar tests can be performed on kittens (Gruber *et al.*, 1971) and infant monkeys (Vaughter *et al.*, 1972) and gorillas (Redshaw, 1978). Equally clearly, infant mammals will be able to pass at least the simpler versions of the tests, but the Piagetian hypothesis is that the sequence of successive tests passed will be roughly similar to

the sequence observed in human infants. This appears to be the case:

kittens obviously become capable of following the trajectory of a ball of paper, and retrieving it from behind a chair, after fewer days of existence than a human baby, but the time course of perceptual development in young gorillas is roughly similar to that observed in humans over the first 12 months (Redshaw, 1978; see Chapter 9, this volume).

This may seem unremarkable, but the advantage of Piaget's discussions is that they emphasise that distinguishing between one's own body and the outside world, memory for the disposition of objects in the world, and elementary forms of interaction with objects, are cognitive achievements of some complexity, even though they may pale into insignificance by comparison with the eventual manifestations of fully human intelligence. Very roughly, the latter are divided by Piaget into 'concrete' and 'formal' cognitive 'operations'.

Concrete operations have to do mainly with logical relationships concerning objects, developing over the first 8 or 10 years of age, while formal operations are what most people would call verbal and symbolic, and seem to correspond to what a bright and educated 11 to 14-year-old knows about physics and mathematics. Formal operations are therefore specifically human. In so far as concrete operations concern knowledge of quantitative and qualitative aspects of objects, and the understanding of number, volume and mass, they may also be uniquely human. But elements of these achievements, such as understanding that a large predator (or conspecific) does not become smaller when it is seen further away, or at a less flattering angle (size and shape constancy) may be built in to animals' perceptual systems, or may be attainable with experience.

Further comparisons between the performance of monkeys and apes and human children on some of the more sophisticated Piagetian tests of cognitive understanding will be discussed in a Later chapter. It is sufficient here to stress that many of the less spectacular abilities of children (and in some cases of adults as well), which are routinely taken to indicate thought and mental processing when they are studied in man, are abilities which are shared with animals.

98

The Question of Animal Awareness (Griffin, 1976)

Subjective awareness is not necessarily the same thing as cognitive complexity. Many who have had low opinions of the capability of animals for inference or abstract perception (for example, Locke) never doubted that their subjective experience of sufficiently simple things (pain, heat and cold, colour, sound) was as intense as our own. Since no one has yet discovered a completely watertight way to test whether this is so or not, the behaviourist tactic of ignoring the question still has something to be said for it. Griffin (1976) has however made a plea for more serious consideration of the possibility that the subjective feelings of animals should be granted a status close to that which we afford to

other people's. He does not put forward a theory of cognition as such, but rather argues that the behaviour of animals should be discussed in terms of their awareness of mental events, irrespective of degrees of abstraction or cognitive complexity that may be inferred in various animal species. Consciousness is equated with the presence of mental images, and the existence of mental images in animals is indicated by orientation and navigation in space (the 'cognitive map' argument), and apparent seeking after goals. Griffin's short essay is thus a convenient and lucid restatement of what he acknowledges as Tolmanian theses, directed at the recent biological tradition of accounting for observed natural sequences of animal behaviour in terms of 'fixed action patterns' and 'innate releasing mechanisms' without discussion of possible intervening mental events (e.g. Hinde, 1970).

Griffin may have been rather over-optimistic, however, in supposing that 'a possible window on the minds of animals' is available in 'participatory investigation of animal communication'. In view of the difficulties in interpreting attempts to communicate directly with chimpanzees by means of imposed systems such as gestural sign language (see Chapter 9) one may doubt if there is very much to be gained by dialogues between a participating scientist and an electric fish (Griffin, 1976, p. 91). It is a weakness in Griffin's case that he ignores differences between species, and the factor of biological closeness to man, giving equal weight to the gestures of chimpanzees and the signals exchanged by honey bees, as prototypes for human language (e.g. pp. 59—60). A related gap is the connection between mental states and brain-states. If one ignores the dependence of subjective awareness on neuronal mechanisms, one is more likely to

99

assume that the feelings of bees, or even plants, ought not to be distinguished from those of species more closely related to ourselves. This is the opposite tack from the one I am taking here.

The brain-state theories of Hebb (1949), Konorski (1967) and Bindra (1976)

Although, oddly, there is a biological tradition, followed by Griffin, of discussing animal behaviour in isolation, without any consideration of the role of the brain, which is exemplified by 'ethologists' such as Lorenz (1966) and by the recent vogue for 'sociobiology' (Dawkins, 1976a; Wilson, 1975), there has of course been no shortage, since Descartes, of theories as to how animal behaviour is determined by animal brains. What is relatively rare, however, is a theory which discusses animal *cognition* as a function of brain activity. It might be supposed that, if we adopt the materialist assumption that cognition *is* brain activity, then there is little need to go any further. But the assumption can only be justified if we know how it is that brains enable animals to do apparently intelligent things. Bindra's *A Theory of Intelligent Behaviour* (1976) attempts a comprehensive theoretical explanation of how goal-directed, voluntary, foresightful and conscious actions get produced by mammalian brains, with human cognitive faculties

interpreted as special cases of more general aspects of brain function. In part, this serves as a compilation of support from the 'neural and behavioural sciences' for the contention that the mental processes we call cognition can be interpreted as certain forms of brain activity, but Bindra also puts forward specific suggestions as to what these fundamental processes are (or, at least, what they should be called).

Very broadly, Bindra belongs to the category of brain- state theorists who say that whenever we need to assume that a thought, as an internal act of knowing, takes place, there is a specific set of physical things in the brain which corresponds to this mental event. His particular assumptions and terminology are derived from the earlier explanations of Hebb (1949) and Konorski (1967). Hebb's idea was an elaboration of the widely accepted hypothesis that the brain works by developing links between individual neurons, or nerve- cells. The crucial concept was the 'cell-assembly'—groups of nerve-cells linked together (mainly, according to Hebb, because of a history of

100

synchronous firing in the experience of the individual), are able to act as independent units. Thus inner sensations such as visual images or vivid hallucinations are explained as the activation of cell-assemblies derived from more real visual experience (Hebb, 1968). The organisation of mental states is thus reduced to the organisation of neuronal circuits. There are many rather similar versions of' the hypothesis that connections between individual neurons underlie brain function (e.g. Barlow, 1972). There are also alternatives which follow Lashley (1929, 1950) in supposing there is something more vague and ephemeral about the location of information in the nervous system, and that something analogous to electromagnetic fields, or the coding of information in radio waves, means that specific nerve-cells, or groups of cells, are not so important.

However the other theory, apart from Hebb's, that Bindra draws on, is that of Konorski (1967), which goes to the extreme of assigning extensive cognitive functions to *particular* neurons. The simplest version of this sort of hypothesis suggests that there could be a single neuron corresponding to a single mental concept. A very large variety of visual patterns, verbal associations and emotional contexts may be related to particular mental ideas such as 'grandmother', but, as a limiting case, one may suppose that there is just one brain cell that fires when and only when 'grandmother' is being thought of. This is not to be taken too seriously, except as an illustration that a step-by- step sequence of links between neurons may allow some cells to map more closely on to mental categories than others. In Konorski's terms, there are 'gnostic neurons' whose knowledge is derived from their position in brain circuits. Konorski was the head of a Polish research laboratory working very much on Pavlovian Lines, using dogs in conditioning experiments, and this can be seen as a development of Pavlov's theory of brain function, with 'gnostic neurons' as individual pieces in Pavlov's cortical mosaic.

Bindra retains the notion of the mosaic by concentrating on assemblies of brain cells, rather than individual neurons—his unit of brain function is the ‘gnostic-assembly’. As he points out (p. 91), this is more or less the same thing as Hebb’s ‘cell-assembly’, the main change of emphasis being that there is *not* supposed to be an exact correspondence between particular mental states and particular gnostic-assemblies. The gnostic-assembly for ‘apple’ represents combinations of features or dimensions of real apples, not all of which may be needed for any individual apple (Bindra, 1976, p. 360). The momentary

101

experience of eating or seeing a particular apple is influenced by the existence of a gnostic-assembly about apples, but each bite of each apple is a unique mental state, and a unique brain-state. Bindra coins a new word for a particular mental experience in a particular brain: the ‘pexgo’. While this neologism may not be very attractive, it is useful to make a distinction between relatively permanent aspects of neural organisation, and the precise form of moment- to-moment events in the nervous system. ‘Pexgo’ is a contraction of ‘presently excited, distinctive neural (gnostic) organisation’ (Bindra, 1976, p. 87), which serves as the unit of immediate perceptual knowledge. In a similar way, knowledge about actions is supposed to be contained in neural ‘act-assemblies’ so that the mental intention to act can be described as the activation of this sort of brain-state, which may then engage assemblies lower down in a neural hierarchy, which will deal with the sequencing and timing of movements. The meaning of words is defined in terms of gnostic organisations activated by words, remembering words depends on activating similar assemblies of neurons, and solving problems depends on exciting new gnostic organisations by ‘progressive narrowing of the determining set’ (Bindra, 1976, p. 344).

Whether this strategy of translating psychological terms into their hypothetical neural equivalents has any merit is of course open to dispute. I cite it here since in my view tying down the processes of human cognition to brain processes establishes a clear relationship between human and animal thought. If human thought is to be conceived of as patterns of neural excitation in the cerebral cortex, then it becomes reasonable to wonder about the cognitive functions of the cerebral cortex in other species. As it happens, Bindra himself does not seem to share this view, since his accounts of animal behaviour include very little reference to neural processes which would amount to memories or expectancies. Pavlov’s dogs, fed after the sound of a bell, are assumed to acquire gnostic organisations which lead them to go and lick the food receptacle when the bell is sounded, but are not allowed an internal expectancy or image of the food (Bindra, 1976, p. 215). Skinner’s rats, pressing a lever which gains them their food reward, are specifically denied an idea (or excited gnostic-assembly) of their own response (Bindra, 1976, p. 242), let alone the idea that the response causes the reward (pp. 278—9). Activating brain- states in a way that corresponds to the process of remembering is discussed only in the context of human memory.

Skinner's experiments learn the things that they do is rather peculiar. Responses which appear to be voluntary are said to be made because there is conditioned excitation of inner stimuli which elicit the response (Bindra, 1976, p. 242). This is called an 'incentive-motivational view', harking back to a term of Hull's, and animals appear to be left with a compulsion to act, without memories of the results of previous actions, or anticipations of what might happen next.

Goals and expectancies in Mackintosh's *The Psychology of Animal Learning*

In 1974, Mackintosh published a comprehensive and authoritative review of 'the main areas of research that have developed from the pioneering work of Pavlov and Thorndike'. The question of how learning is produced by brain processes is excluded, but theoretical issues such as the degree to which animals are capable of anticipating future events are assessed on the basis of the enormous body of evidence about how laboratory animals behave in standard experimental tests. The arguments, and the experiments, are sometimes extremely detailed, and are not always conclusive. In general, however, it seems fair to say that the conclusions which Mackintosh derives from research initiated by Pavlov and Thorndike support not the 'conditioned response' theories which these pioneers began with, but rather the 'expectancy' theories of Tolman (see above).

As an example, theories like Bindra's, just mentioned, that animals act because compelled to do so by inner motivational states, irrespective of retained information concerning the consequences of actions, are weakened by several experiments which suggest that, even for laboratory rats, external stimuli, such as buzzers and lights, which were used as signals for 'food, did not promote diffuse states of enthusiasm, but rather, 'specific expectations of the reinforcer with which they had been paired' (Mackintosh, 1974, p. 227). It may be wondered what sort of experiment could demonstrate the existence of such- inner mental events in rats. Suppose that hungry monkeys are tested in an elaborate version of a Skinner box, in which there is a raisin dispenser, and a peanut dispenser. When a buzzer is sounded, they can pull a chain which will deliver raisins, and, if they hear a whistle, pushing a button will operate the peanut dispenser. If they get this

right, they can supply themselves with both peanuts and raisins, but if they get things the wrong way round, and go for the chain when they hear the whistle, or the button when they hear the buzzer, neither peanuts nor raisins are forthcoming. Monkeys should usually get it right, and a very simple incentive conditioning theory would say that they feel enthusiastic about chain-pulling as a conditioned reaction to the buzzer, and are motivated to push the button by the sound of the whistle, without knowing why they are drawn to these

actions. The twist to the experiment is in giving the animals preliminary free trials, *before* they are required to learn to operate the dispensers for themselves. Suppose, in those preliminary experiences, raisins come preceded by whistles, and peanuts preceded by buzzers, and that is the opposite way round to what is going to happen in training, should this make any difference? If what is required is just a motivation to respond, without any expectations as to whether a peanut or a raisin is coming next, it should not. On the other hand, if we were exposed to such preliminaries ourselves, we should surely acquire mental associations that whistles go with raisins and buzzers go with peanuts, which would prove very confusing when the signals were reversed. It is because such reversals of the 'meaning' of experimental signals appear to be confusing for laboratory rats, in versions of the experiment just sketched out (e.g. Trapold, 1970), that Mackintosh suggests that specific information about what should happen next is somehow available to them, at least in the case of something as interesting as food.

An alternative phrase is that, in the Pavlovian kind of experiment, where one event follows another, the first can call up a 'rich encoding' of the second (Rescorla, 1980). Although sometimes disguised as 'the formation of an association between stimuli', Pavlov's result is increasingly being interpreted, because of more elaborate experiments, as due to 'central representations' (Mackintosh, 1974, p. 90; Rescorla, 1979). In other words Pavlov's result has to be explained by a brain process that is to all intents and purposes a mental association, so that the sound of the bell calls up an image, a representation, or a thought, of the previously experienced meat powder.

The utility of being able to think ahead in this way must lie in being able to take action on the basis of anticipated events. This may not always be obvious within the confines of the usual sort of Pavlovian experiment, but in the other kind of laboratory test, where animals are

104

allowed to find their way through mazes, to discover methods of escaping from confinement, or to manipulate moveable objects which work food dispensers, the advantages of internalising experienced relationships between actions and pay-offs are made more transparent. Starting with Thorndike, the preferred assumption was that such an internal storing of prior experience could only take the form of rigid tendencies to make particular bodily movements. Mackintosh's conclusion, however, is that this assumption is inadequate: the experimental evidence is such as to 'suggest very strongly that animals are capable of learning that their behaviour has certain consequences' (Mackintosh, 1974, p. 211). The reason why animals make apparently intentional actions 'is precisely the expectation of an increase in incentive' (p. 269). The expectations are typically established on the basis of experience, so that 'disconfirmation of previously established expectancies' may in turn lead to changed actions (p. 417 and p. 346).

Thus, purely as a means of accounting for the accumulation of evidence from laboratory experiments in animal training, Mackintosh is led to adopt a significant fraction of Tolman's cognitive theory:

animals have expectancies in so far as they are able to anticipate, on the basis of prior experience, specific and general properties of incipient yet absent events, and also have 'if-then' expectancies in so far as they anticipate the results of their own actions. This counts as a recent theory only because Mackintosh's text carries rather more weight than the protestations of Tolman a few decades ago; and it is two hundred years since we were told by Hume that animals 'learn many things from experience, and infer that the same events will always follow from the same causes'.

Conclusions: consciousness, abstraction, memory and brain processes in animals

What may be concluded from the astonishing variety of opinions concerning animal thought which have been reviewed in the last three chapters? It is clear that no single view is likely to command universal assent, but it ought to be possible to map out areas of dispute, to suggest borders that remain to be fought over, and ideally to indicate decisive sources of evidence. I must confess, however, that no such mapping out has suggested itself to me, and I propose instead to list cases where I am now going to take sides.

105

Human thought and consciousness

Evaluating the status of mental events in animals, except by comparison with what is assumed to exist in people, seems to be rarely attempted, and perhaps it is impossible. It may therefore be profitable to rephrase all questions about thought in animals in the form 'To what extent are animals like us?' and this is what I intend to do here. We can thus ask a number of separate questions, such as: 'Do animals have visual systems which work like ours?' 'Do animals have mental imagery like ours?' 'Do animals have hopes and fears like ours?' Such questions may be useful, even if we do not know very much about what is being used for comparison in ourselves. For instance, the function of sleep, and in particular of dreaming, remains shrouded in mystery, but the question 'Do animals dream like us?' is one to which, though there may be no definitive answers, plenty of relevant evidence can be addressed (see Chapter 6).

Although it is perhaps the most difficult of all, consider first the question 'Do animals have consciousness like ours?' Clearly we first have to say something about what features of human consciousness we are referring to. If we mean consciousness as in being awake, as opposed to being asleep or under anaesthetic, then the answer is a qualified yes. Sleep and anaesthetics seem to have all the effects in other mammals, and probably birds, that they do in us. Here we can point out what is probably a general rule, that the animals which are biologically more like us, that is mammals and especially monkeys and apes, are also more like us psychologically.

However, suppose we mean by 'conscious' being able to *say* that we are aware of something in particular, in ordinary human

conversation. Then the answer is obviously no, since there are no animals that can tell us that they are aware of something in the same way that another person can. The uniqueness of human speech is the most commonly agreed boundary line between us and all other species, and the best strategy seems to be simply to conclude that anything that is verbal, and dependent on speech or language in some significant way, is not available to other species. This excludes a lot, it is true. Thought in the sense of 'eighteenth-century thought', 'Greek thought', or 'what Fleet Street is thinking today' just does not arise in any other species we know about. But this does not exclude everything, and it does not necessarily exclude everything that we happen to apply verbal labels to. If we say 'I feel hungry' or 'I think I am going to be sick' (and are telling the truth), then although verbal expression may add a

106

considerable extra gloss to what we are conscious of, there is no reason to suppose that some of the feelings we are experiencing at the time might not be shared by individuals from other species, to the extent that they share the physical states that are responsible for our own hunger, or nausea.

The same argument applies to members of our own species, to the extent that someone who is retching, pale, and clearly suffering from seasickness, is probably feeling sick, even if he assures us that he is not. In particular we would suspect that he *was* feeling sick if he finally threw up.

Here we come to an important general point, and that is the relation of subjective awareness to physical states, and in particular to physical states of the brain. If it is accepted that all human subjective awareness arises as a result of brain processes (cf. Armstrong, 1968 and Chapter 1, p. 17), decisions as to the consciousness of other persons, and of members of other species, are considerably simplified, in theory if not in practice. The simplification is limited, since it could be claimed that the most important brain-states are those which mediate speech, in which case we would not be very much better off. But suppose that we want to decide whether someone who has been born deaf and dumb is enjoying his soup as much as we are. If we could know that (a) that soup is the same, (b) his taste-buds, olfactory apparatus and so on are the same, and (c) all his brain-states except those involved in speech are the same, we ought to be confident that his conscious appreciation of the soup has a good deal in common with our own. Knowing the name of the soup may make a difference, and someone who is convinced that he likes chicken soups may be more inclined to relish one made, with rabbit if he accepts verbal assurances that it is chicken, but such niceties do not require us to abandon a general rule of identifying mental experiences with their physical causes.

If we grant that there is a physical basis to sensations, and that our own subjective awareness of perceiving things is due to particular sorts of reactions in our brains, we would be led to believe that our consciousness of certain well-defined sensations such as stomach-ache, pressure on the fingers, or bright lights flashing in our eyes, is roughly similar to the subjective sensations of animals

which possess roughly similar bodies and brains to ours, and are exposed to roughly similar sets of events. There is some empirical support for this idea, in that, for instance, electrical stimulation of certain parts of the brain of patients who are awake, but have had their brain exposed for surgery, leads

107

them to report feelings of touch or movement in particular places on the body surface; actual touching of the skin at these places causes measurable electrical changes in the appropriate locations in the brain; and similar relationships, apart from the absence of verbal reports of feeling, are observed in other mammals (e.g. Penfield and Rasmussen, 1950; Penfield and Roberts, 1959; Woolsey, 1965).

The proposition that states of the human brain have something critical to do with the conscious aspects of human perception has the clear implication that one of the criteria for whether the same kind of consciousness should be expected in animals is whether they have the same kind of brains. This means that the comparative anatomy and physiology of human and animal brains assumes the role of a decisive source of evidence concerning animal consciousness, and as such will be examined more closely in the next chapter. There is nothing very new in assigning importance to comparative brain anatomy—it is a view that was already being satirised by Charles Kingsley in 1886, in the course of digs at a thinly disguised Professor Huxley, who declares that apes have ‘hippopotamus majors’ in their brains, just as men have:

You may think that there are other more important differences between you and an ape, such as being able to speak, and make machines, and know right from wrong, and say your prayers, and other little matters of that kind; but that is a child’s fancy, my dear. Nothing is to be depended on but the great hippopotamus test. (*The Water Babies*, p. 172)

What is new, since 1886, is a good deal more information about how brains work, and a rather greater readiness to accept that this is pertinent to consciousness and thought.

It may be objected that an ape being consciously aware of a stomach-ache, because it has a similar brain, and a similar stomach, to ours, does not take us very far towards an assessment of animal thought. This is so: apart from Descartes and the behaviourists, most people, from Aristotle and Aquinas to Thorndike and Wundt, have in any case taken it for granted that mammals at least have an awareness of immediately perceived sensations. But it is important to confirm this to get over the barrier represented by the view that animals cannot possibly think because they are incapable of being aware of anything. The real concern is of course not whether an ape thinks that it has a

108

stomach-ache, but whether it is capable of entertaining the possibility that another ape might have a stomach-ache, or whether it can remember that it itself had a stomach-ache the last time it ate unripe fruit.

If it were possible to make the large jump, and say that, because the ape's brain closely resembles ours, it must be capable of the complete range of cognitions available to a human being isolated from normal culture and education, in other words to assume that the imagination and inference available to apes is not violently different from those faculties that would be observed in the human species at an equivalently brutish stage of existence, this might save a considerable amount of further speculation. But even the most convinced materialist cannot make such a large jump, not only because what he knows as imagination and inference may be inseparable from culture, education and social traditions, but also because the advent of language, or some subtle change in brain utilisation prior to language, and conceivably prior to the development of social traditions and so on in *Homo sapiens*, may mean that human brains, while undoubtedly sharing a great many properties with other mammalian brains, are able to do some new and different things.

Descartes's contention that human brains have access to a superior controlling agency has been caricatured as a claim that our brains are only the instruments of thought, as a piano is an instrument for a piano-player. The separate existence of the piano-player, or the 'ghost in the machine', presents problems, especially if one accepts Darwinian theories of evolution. However, if we view brains as pianolas, rather than pianos, operated by sets of instructions within themselves, it is easy to see that as physically similar pianolas may be capable of playing different kinds of tunes, so physically similar brains may have different capacities for the sequences of operations we call thought. We would have to test pianolas by listening to them play, and we must also examine animals' brains by the tunes they generate, that is by assessing the kinds of behaviours they are capable of sustaining.

In other words, a form of behaviourism remains essential for the evaluation of cognition. What animals do will have to be the final test for what they think.

Abstraction and perception

We may try out these criteria of behavioural capacity and available neural mechanisms in relation to the question of whether animals can

be said to perform abstractions. After language, this is the most popular dividing line between animals and men, Locke and Romanes for instance giving it pride of place as a peculiarly human faculty. Kenneth Craik (1943, p. 72) pointed out that, unless one adopts the practice of putting the question in the form 'do animals abstract like men?' the distinction loses much of its force. Abstraction means neglecting particulars, and this is easy, not difficult. Craik's example is a photoelectric cell counting objects passing by on a conveyor belt. Every particular characteristic of the objects is neglected in favour of the non-specific factor of whether the photo-cell beam is broken. Similarly, very rudimentary mechanical devices can respond to weight irrespective of shape, height irrespective of width, or speed irrespective of what it is that is

moving. Is this what is meant by abstraction? It seems to be partly what Locke meant (cf. pp. 23), since his example is the abstract perception of 'whiteness' irrespective of whether it occurs in snow or milk. This degree of abstraction could be achieved by a sophisticated version of Craik's photoelectric cell, and a Pavlovian experiment could easily show monkeys salivating to anything white. It is reasonable to conclude that when Locke said 'brutes abstract not' he was simply wrong.

However, although Locke was quite specific about the abstraction of general ideas such as whiteness as an underlying ability, the reason why this ability has been erroneously raised to such a high position in the panoply of intellectual capacities is probably because of the existence of abstract *words* for such things as whiteness. The special thing about human abstraction probably boils down to the use of verbal labels.

I have already mentioned, in discussing Pavlov's experiments, that Locke's assertions about 'compounding' not being present in animals can be shown to be false, by behavioural testing. It is still common to discuss animal perception in terms of isolated particulars—a robin being said to recognise another robin only in terms of redness, for instance. It may be the case that isolated stimulus qualities can exert a powerful influence on animal behaviour, especially for young animals or creatures such as the frog, which has highly specialised responses to anything that moves like an insect. Similarly, it is sometimes claimed that 'cross-modal' perception—recognising an object by touch, sight or sound as the same thing—is a specifically human capacity (e.g. Geschwind, 1965). Sophisticated behavioural tests have confirmed that monkeys and chimpanzees are perfectly capable of recognising objects

110

in alternative modalities such as touch or sight (Weiskrantz, 1977) and these add to the evidence that this form of human abstraction is shared at least by other primates. This is a case where further behavioural tests can be expected to provide stronger evidence. But it ought to be stressed that object recognition is a general function of the animal sensory systems rather than something confined to primates—a cat is superbly equipped to recognise a mouse by sight, sound, smell and touch, and it is more likely that the cat's perception of 'mouse' relates to all four sources of information than that it responds quite independently to each modality. A visual sense which was not correlated with touch and sound would be much less useful than one that is, and the vertebrate brain is often described as a set of correlation centres which exist precisely to accomplish the integration of information from the various modalities.

Thus, while it is perfectly legitimate to emphasise that animals do not think abstract thoughts in the sense of using abstract words rather than concrete words, and in the sense that human abstract thought requires the grammar of language or a symbolic system such as mathematics, it is quite wrong to push limitations on abstraction back to elementary forms of object perception. In direct opposition to Locke, I would maintain that the dog's perception of

his master occurs as a compound of voice, shape and smell, rather than 'so many distinct marks'. What the dog cannot do is say to himself that he is one of the interacting elements in a master-dog situation. But this is a limitation in language, not in perception.

Memory and learning

Lloyd Morgan maintained that any animal capable of learning from experience could only do so by calling up conscious memories of previous occasions in order to guide present acts. He did not say that much-practised skills and drills could not become routine and automatic, but rather that the early stages of practice that are necessary to establish unconscious habits must require conscious effort, and the active recall of mental images. The thrust of behaviourist scepticism, beginning with Thorndike, and exemplified by Hull and Skinner, has been to suggest that learning can take place without any inner memories or mental processes of any kind. It has to be acknowledged that there are many forms of learning for which this is undoubtedly true.

First, adaptive changes in behaviour resulting from experience can

111

be observed in animals or parts of animals with access only to neural mechanisms so rudimentary that the existence of mental processes which normally require an intact human brain seems extremely unlikely. Learning of a sort can be demonstrated in isolated ganglia of cockroaches and slugs, decorticated rats, the human foetus, humans in a vegetative coma, and isolated parts of the spinal cord in humans and other mammals.

Second, there are many forms of human learning which can take place without any indications of subjective awareness. We are not necessarily aware of the movements of individual muscles which must be co-ordinated by experience in learning to play tennis (an example Thorndike was fond of) or of the movements our eyes must make when scanning a printed page—still less of the learned neural processes which enable us to recognise letters and words. It has been repeatedly pointed out that children learn a good deal more grammar while in blissful ignorance of its existence than when they are required to develop conscious memories about its describable forms.

But because *some* kinds of learning are apparently unconscious and automatic, and do not involve the subjective inspection of recalled memories of past events, we should not make the behaviourist mistake of concluding that *all* forms of learning have a similar character. Only the doctrinaire ever gave credence to the proposition that we could account for human psychology without including unexpressed mental events; but demonstrating similar processes in animals is a much more difficult matter. The traditional touch-stone, still in use, is the goal directedness of animal behaviour (e.g. Tolman, 1932; Bennett, 1976; Dickinson, 1985). The animal runs through a maze, or climbs a tree, because it remembers there is something it wants that can be obtained by these activities. This, in itself is not good enough: Thorndike was able to claim that we may

occasionally make a tennis stroke which has a purpose, without thinking about its purpose until afterwards, if at all, and that animals might do this sort of thing all the time. Similarly, it can be claimed that a cruise missile behaves in a goal-directed and purposeful way, but we would not believe that the cruise missile therefore has intentions and memories like our own, even if, like other manufactured robots, its actions were determined by prior experience.

The only way out is to make detailed examination of the extent to which the behaviours and neural mechanisms observed in animals are like those of our own which we identify with memory. I will return to this topic in a later chapter, but it seems to be becoming clear that

112

detailed evidence suggests that laboratory animals have memories, or expectancies about the future based on experience. Lloyd Morgan may have been right to say that conscious memories guide learned actions in so far as the typical laboratory tests of animal learning have to be interpreted as expectations concerning future events, as well as mechanical elicitations of a response (Mackintosh, 1983). Put in another way, this means that it makes sense to say at least some of the time that Pavlov's dogs salivated because they believed that a buzzer should be followed by food, Thorndike's cats released latches because they remembered that this would allow them to get out of confinement, and Skinner's rats pressed levers because they expected, on the basis of previous experience, that this would enable them to eat food pellets.

Language and thought

As I pointed out at the beginning of Chapter 1, for some purposes it is perfectly reasonable to define 'thought' as something that only man is capable of, and to then say that everything that might occur in the brains of animals should be called something else. The disadvantage of this strategy is that it encourages the belief that every mental process which in ourselves becomes tied up with verbal expression is necessarily absent in all other species. Because we have a word for whiteness, and our own perceptual activities tend to become locked on to our words, many have followed Locke in asserting that species which lack a word for whiteness are therefore not able to isolate this quality perceptually. Because we practically always can put words to our memories and anticipations, we assume that without the words, memories and anticipations cannot exist. When we use a name for individual objects, such as 'the FA cup' or 'the White House' it is clear that using the name embodies a great deal more than sensory information about the objects. Indeed we can use the names perfectly well without ever having seen or touched the FA cup, or the White House. However vague our notion of the meaning of names may be, it is clear that there is much that is independent of the sorts of immediate physical knowledge that can be assumed to be available through animal senses. When we talk of 'the author of the anonymous letter', or 'inaudible sound frequencies', we make use of linguistic abilities

that seem to be peculiar to our own species.

It is clear, therefore, that there must be mental processes made possible by language, which cannot take place without linguistic

113

competence and experience, and which must be denied to animals. The argument is whether language-dependent mental processes are grafted on to, or perhaps develop from, more elementary inner representations and ideas which arise in the brains of animals, or whether, if linguistic devices are excluded, nothing is left which bears any resemblance to human perception, memory, and subjective awareness. I hope to support a strong form of the first alternative: that language-dependent mental processes can be accomplished in the human brain only because animal brains in general work in terms of perceptions, memories, concepts and ideas. Without such internal referents, there would be little point in evolving whatever it is that enables human brains to exchange information via speech. Rigid and restricted methods of communication might be useful if, for instance, instinctive cries of alarm elicited flight in other members of a species, without any intervening inner mental pictures of predators. It is conceivable that fairly elaborate exchange of information between ants or between bees takes place like this, by way of elicitation of preprogrammed response plans. But it is patently obvious that the way human speech works is that someone having the idea of bananas can produce a corresponding idea in someone else simply by saying 'bananas'.

Now, one might suppose that the ability to have the idea 'bananas' depends on the possession of language, and thus that the ability to have ideas evolved after, or in step with, the evolution of language. This is the familiar theory put forward by Locke, that the ability to have abstract ideas is a uniquely human achievement, and a prerequisite for speech, since words are only useful if they are connected to ideas. My proposal is that rather good 'abstract ideas' for such things as bananas were available *before* the evolution of human language. Thus chimpanzees may have the idea of a banana (and related concepts such as 'lots of bananas'), bears may have a pretty clear idea of 'honey', and cats of 'mouse', even if they do not have words for them. The general theory behind this is that a function of brains is to provide an internal model of reality, or something like it (e.g. Craik, 1943; Bindra, 1976) and that ideas are necessary because they correspond to external realities. Thus the original purpose of brains is to internalise the objective external world, and the external world is usefully described in terms of complex relationships and associations, amounting to relatively abstract ideas. Once this inner network of ideas is re-externalised by mapping them on to verbal utterances, new possibilities for the mental manipulation

114

of ideas emerge. Spoken words are themselves both external objects and voluntary actions. Internalisation of words, and relationships between words, clearly allows for a new realm of inner ideas, not necessarily identical with the internal representation of more

tangible physical realities. In the human species, at present, and certainly in its more academically inclined members, mental processes driven by words tend to dominate.

To test the hypothesis that there is such a thing as pre-linguistic thought, one might first wish to examine the psychological processes of young children, in whom ideas might occasionally run ahead of speech, or in adults deprived of language by deafness, or damage to the brain. However, my interest is clearly in the demonstration of non-linguistic mental abilities in other species. This may be done through the investigation of perception and action, if this provides evidence of classification, categorisation and compounding of ideas, and the activation of inner memories and expectancies. A more immediately appealing tactic is to make non-linguistic mental events manifest by forcing animals to reveal their inner cognitive organisation in a way analogous to human speech. The last ten years have seen an enormous amount of effort devoted to the inculcation of human-like modes of communication of ideas in chimpanzees, the species where this tactic should have the greatest chance of success. If a chimpanzee could be persuaded to speak English, or properly utilise the gestures of a sign language like those used by the human deaf, this would certainly provide a short cut to the discovering of what normally goes on in chimpanzee minds. Unfortunately, no chimpanzee has yet been found that can reproduce the remotest approximation to human speech, and it is still unclear whether the numerous apes that have been trained to mimic sign-language-like gestures are doing very much more than what a parrot can do with speech-like vocalisations. However, the evidence produced in these exercises must be important, whichever side it supports, and I shall review it in Chapter 9.

For the present I simply assume that mental states of all kinds are properties of brains, and that it is therefore possible to consider mental evolution as an aspect of brain evolution.