

appropriate signals for this function, including not only information from limbic structures and the orbitofrontal cortex about rewards available; and cognitive information from the association areas of the cerebral cortex: but also information from the motor, premotor, somatosensory, and parietal cortical areas which provides information about action and movements. The hypothesis is that the striatum, followed by the globus pallidus and substantia nigra, provide a two-stage system for bringing these signals together. Such convergence is implied and provided for by the dendritic organization of the basal ganglia (Percheron *et al.* 1984a,b; Rolls 1984b; Rolls and Williams 1987a; Rolls and Johnstone 1992; see Section 6.4). By virtue of such anatomy, the basal ganglia would allow in its output stages signals indicating that food was available, was currently rewarding, and that the costs of obtaining it were not too high as indicated perhaps by frontal cortical inputs, to be combined with, for example, signals from parietal and premotor areas reflecting what position in space is being visually fixated, and actions or movements being made to approach that position in space. These signals might well lead to the activation of individual neurons as a result of previous repeated co-occurrence, and associated synaptic modification in the basal ganglia. The output of such neurons would then indicate that actions should be made to approach and obtain rewarding targets in the environment. If a signal arrived in the basal ganglia from another cortical or limbic area (indicating for example that a novel visual stimulus had appeared), then lateral inhibitory connections and the competitive interactions it makes possible within the striatum would lead to interruption of feeding, and would implement an arbitration system (see further Rolls 1984b; Rolls and Williams 1987a; Rolls and Johnstone 1992). It is suggested that part of what is implemented in this way is a relatively simple interface between sensory and action systems, in that the signals reaching the striatum about reward need only to signify that reward is available, with the details about where it is being implied, for example, by which position in space is being visually fixated, and is the current subject of attention. Limiting sensory analysis in this way allows the reward signal, reflecting the output of the 'what' system in perception, to be linked unambiguously with the output of the 'where' system in the interface between systems that specify what targets in the environment should be the targets for action, and the systems which lead to action. The action system in this case can be general-purpose, in that it can be organized to perform general-purpose operations (which are not different for each type of goal object or reward), to approach and acquire the objects in space which are currently the focus of attention. (Of course, if the object that is the object of attention is associated with punishment, as shown by inputs carried by other neurons from limbic structures, then the basal ganglia should access or 'look up' operations learned previously that will lead to withdrawal from the current focus of attention.) The ways in which reward systems provide for an interface to action systems, and in which the basal ganglia may operate to map signals into actions, are discussed more fully elsewhere (Rolls 1984b, 1990b, 1994a; Rolls and Williams 1987a; Rolls and Johnstone 1992; Rolls and Treves 1998).

These ideas address some of the crucial current questions in understanding how brain systems involved in decoding rewards, including food reward, are interfaced to action systems to produce, for example, feeding. It is likely that understanding the functions of the basal ganglia will be crucial in understanding these issues.

## 3 The nature of emotion

In this chapter the nature of emotion, and its functions, are considered, to provide a foundation for exploring the brain mechanisms involved in emotion in Chapter 4. It is important when considering the neural bases of emotion to know what types of stimuli elicit emotions. This is considered here, and helps to focus analysis on certain types of brain processing in Chapter 4. It is also important to know what the functions of emotion are, for they provides a guide to the neural output systems which may be expected to receive inputs from the brain regions involved in emotions.

Certain more general issues about emotion are left until later in the book, after the brain mechanisms involved in emotion and in related behaviours have been described. One such issue is that of emotional feelings, which is part of the much larger issue of consciousness. This issue is left until Chapter 9. Another issue is the interesting one of emotion and brain design. Why do we have emotions? Are emotions crucial to brain design and evolution? This issue, which deals with the adaptive value of emotion in relation to evolution by natural selection, is left until Chapter 10, where it is suggested that emotion is fundamental to the design of any adaptive brain with flexible behaviour. That will conclude our exploration of reward, punishment, emotion, motivation, their brain mechanisms, and why we have these processes.

### 3.1 A theory of emotion

#### 3.1.1 Introduction

What are emotions? This is a question that almost everyone is interested in. There have been many answers, many of them surprisingly unclear and ill-defined (see Strongman 1996; Oatley and Jenkins 1996). William James (1884) was at least clear about what he thought. He believed that emotional experiences were produced by sensing bodily changes, such as changes in heart rate or in skeletal muscles. His view was that "We feel frightened because we are running away". But he left unanswered the crucial question even for his theory, which is: Why do some events make us run away (and then feel emotional), whereas others do not?

A more modern theory is that of Frigida (1986), who argues that a change in action readiness is the central core of an emotion. Oatley and Jenkins (1996, p. 96) make this part of their definition too, stating that the core of an emotion is readiness to act and the prompting of plans. But surely subjects in reaction time experiments in psychology who are continually for thousands of trials altering their action readiness are very far indeed from having normal or strong emotional experiences? Similarly, we can perform an action in response to a verbal request (e.g. open a door), yet may not experience great emotion

when performing this action. Another example might be the actions that are performed in driving a car on a routine trip—we get ready, and many actions are performed, often quite automatically, and little emotion occurs. So it appears that there is no necessary link between performing actions and emotion. This may not be a clear way to define emotion.

Because it is useful to be able to specify what emotions are, in this chapter we consider a systematic approach to this question. Part of the route is to ask what causes emotions. Can clear conditions be specified for the circumstances in which emotions occur? This is considered in Section 3.1.2. Continuing with this theme, when we have come to understand the conditions under which emotions occur, does this help us to classify and describe different emotions systematically, in terms of differences between the different conditions that cause emotions to occur? A way in which a systematic account of different emotions can be provided is described in Section 3.1.3. A major help in understanding emotions would be provided by understanding what the functions of emotion are. It turns out that emotions have quite a number of different functions, each of which helps us to understand emotions a little more clearly. These different functions of emotion are described in Section 3.1.5. Understanding the different functions of emotion helps us to understand also the brain mechanisms of emotion, for it helps us to see that emotion can operate to affect several different output systems of the brain. With this background of the factors that cause emotion, and what emotion in turn does, we are in a good position to understand the brain systems that determine whether emotion is produced and that implement behaviour, autonomic, and endocrine responses to emotion-provoking stimuli. These analyses leave open though a major related question, which is why emotional states feel like something to us. This it transpires is part of the much larger, though more speculative, issue of consciousness, and why anything should feel like something to us. This aspect of emotional feelings, because it is part of the much larger issue of consciousness, is deferred until Chapter 9. In this chapter, in considering the function of emotions, the idea is presented that emotions are part of a system which helps to map certain classes of stimuli, broadly identified as rewarding and punishing stimuli, to action systems. Part of the idea is that this enables a simple interface between such stimuli and actions. This is an important area in its own right, which goes to the heart of why animals are built to respond to rewards and punishments. This issue is taken up in a comprehensive way in Chapter 10.

The suggestion is that we now have a way of systematically approaching the nature of emotion, their functions, and their brain mechanisms. Doubtless in time there will be changes and additions to the overall picture. But the suggestion is that the ideas presented here do provide a firm and systematic foundation for understanding emotions, their functions, and their brain mechanisms.

### 3.1.2 Definitions

I will first introduce the definition of emotion that I propose (Rolls 1986a,b, 1990b, 1995c). The essence of the proposal is that emotions are states elicited by rewards and punishers including changes in rewards and punishments. A reward is anything for which an animal

will work. A punisher is anything that an animal will work to escape or avoid. An example of an emotion might thus be happiness produced by being given a reward, such as a hug, a pleasant touch, praise, winning a large sum of money, or being with someone whom one loves. All these things are rewards, in that we will work to obtain them. Another example of an emotion might be fear produced by the sound of a rapidly approaching bus when we are cycling, or the sight of an angry expression on someone's face. We will work to avoid such stimuli, which are punishers. Another example might be frustration, anger, or sadness produced by the omission of an expected reward such as a prize, or the termination of a reward such as the death of a loved one. Another example might be relief, produced by the omission or termination of a punishing stimulus, for example the removal of a painful stimulus, or sailing out of danger. These examples indicate how emotions can be produced by the delivery, omission, or termination of rewarding or punishing stimuli, and go some way to indicate how different emotions could be produced and classified in terms of the rewards and punishers received, omitted, or terminated.

Before accepting this proposal, we should consider whether there are any exceptions to the proposed rule. Are any emotions caused by stimuli, events, or remembered events that are not rewarding or punishing? Do any rewarding or punishing stimuli not cause emotions? We will consider these questions in more detail in the next few pages. The point is that if there are no major exceptions, or if any exceptions can be clearly encapsulated, then we may have a good working definition at least of what causes emotions. Moreover, it is worth pointing out that many approaches to or theories of emotion have in common that part of the process involves 'appraisal' (e.g. Frida 1986; Oatley and Johnson-Laird 1987; Lazarus 1991; Izard 1993; Stein, Trabasso and Livrago 1994). This is part, for example, of the suggestion made by Oatley and Jenkins (1996), who on p. 96 write that "an emotion is usually caused by a person consciously or unconsciously evaluating an event as relevant to a concern (a goal) that is important; the emotion is felt as positive when a concern is advanced and negative when a concern is impeded". The concept of appraisal presumably involves in all these theories assessment of whether something is rewarding or punishing, that is whether it will be worked for or avoided. The description in terms of reward or punishment adopted here simply seems much more precisely and operationally specified. In the remainder of this section, we will also consider a slightly more formal definition than rewards or punishers, in which the concept of reinforcers is used, and show how there has been a considerable history in the development of ideas along this line.

Emotions can usefully be defined as states produced by instrumental reinforcing stimuli (Millenson 1967; Weiskrantz 1968; J. Gray 1975, Chapter 7; J. Gray 1981). (Earlier views, to which these more recent theories are related, include those of Watson 1929, 1930; Harlow and Stagner 1933; and Amsel 1958, 1962.) This definition is extended below. (Instrumental reinforcers are stimuli which if their occurrence, termination, or omission is made contingent upon the making of a response, alter the probability of the future emission of that response. Part of the point of introducing reinforcers into the definition of emotion-provoking stimuli, events, or remembered events, is that this provides an operational definition of what causes an emotion. Another reason for introducing rein-

forcers into the definition is that different emotions can then be partly classified in terms of different reinforcement contingencies. I note that the definition provided above should be taken to include the formulation 'emotions can be defined as states produced by stimuli which can be shown to be instrumental reinforcers', for the formal conditions for demonstrating that a stimulus is a reinforcer may not always be present when such a stimulus is delivered, omitted, or terminated.) Some stimuli are unlearned (or 'primary') reinforcers (e.g. the taste of food if the animal is hungry, or pain); while others may become reinforcing by learning, because of their association with such primary reinforcers, thereby becoming 'secondary reinforcers'. This type of learning may thus be called 'stimulus-reinforcement association', and probably occurs via a process like that of classical conditioning. It is a form of pattern association between two stimuli, one of which is a primary reinforcer, and the other of which becomes a secondary reinforcer (see Rolls and Treves 1998). (It is not stimulus-response or habit learning, in which the association is to a response, and in which other brain systems are implicated—see Rolls and Treves 1998, Chapter 9.) If a reinforcer increases the probability of emission of a response on which it is contingent, it is said to be a 'positive reinforcer' or 'reward'; if it reduces the probability of such a response it is a 'negative reinforcer' or 'punisher'. For example, fear is an emotional state which might be produced by a sound that has previously been associated with an electrical shock. Shock in this example is the primary negative reinforcer, and fear is the emotional state which occurs to the tone stimulus as a result of the learning of the stimulus (i.e. tone)-reinforcement (i.e. shock) association. The tone in this example is a conditioned stimulus because of classical conditioning, and has secondary reinforcing properties in that responses will be made to escape from it and thus avoid the primary reinforcer, shock.

The converse reinforcement contingencies produce the opposite effects on behaviour. The omission or termination of a positive reinforcer ('extinction' and 'time out' respectively, sometimes described as 'punishing'), reduce the probability of responses. Responses followed by the omission or termination of a negative reinforcer increase in probability, this pair of reinforcement operations being termed 'active avoidance' and 'escape' respectively (see J. Gray 1975, Chapter 4; Mackintosh 1983, pp. 19–21; Dickinson 1980; and Pearce 1996, for further discussions of this terminology).

A useful convention to distinguish between emotion and a mood state is as follows. An emotion consists of cognitive processing which results in a decoded signal that an environmental event (or remembered event) is reinforcing, together with the mood state produced as a result. If the mood state is produced in the absence of the external sensory input and the cognitive decoding (for example by direct electrical stimulation of the amygdala, see Section 5.2), then this is described only as a mood state, and is different from an emotion in that there is no object in the environment towards which the mood state is directed. (In that emotions are produced by stimuli or objects, and thus emotions 'take or have an object', emotional states are examples of what philosophers call 'intentional states'.) It is useful to point out that there is great opportunity for cognitive processing (whether conscious or not) in emotions, for cognitive processes will very often be required to determine whether an environmental stimulus or event is reinforcing (see further Section 3.1.4).

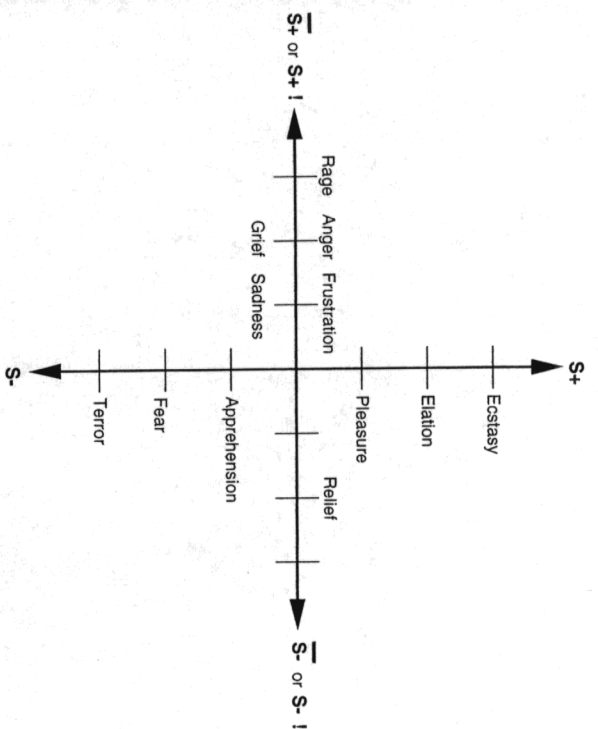


Fig. 3.1 Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the centre of the diagram, on a continuous scale. The classification scheme created by the different reinforcement contingencies consists of: (1) the presentation of a positive reinforcer (S+); (2) the presentation of a negative reinforcer (S-); (3) the omission of a positive reinforcer (S+ I) or the termination of a positive reinforcer (S+ I); and (4) the omission of a negative reinforcer (S-) or the termination of a negative reinforcer (S- I).

### 3.1.3 Different emotions

The different emotions can be described and classified according to whether the reinforcer is positive or negative, and by the reinforcement contingency. An outline of the classification scheme thus created is shown in Fig. 3.1. Movement away from the centre of the diagram represents increasing intensity of emotion, on a continuous scale. The diagram shows that emotions associated with the presentation of a positive reinforcer (S+) include pleasure, elation and ecstasy. Of course, other emotional labels can be included along the same axis. Emotions associated with the presentation of a negative reinforcer (S-) include apprehension, fear, and terror (see Fig. 3.1). Emotions associated with the omission of a positive reinforcer (S+ I) or the termination of a positive reinforcer (S+ I) include frustration, anger and rage. Emotions associated with the omission of a negative reinforcer (S-) or the termination of a negative reinforcer (S- I) include relief. Although the classification of emotion presented here differs from earlier theories, the approach adopted here of defining and classifying emotions by reinforcing effects is one that has been developed in a number of earlier analyses (Mifflenson 1967; J. Gray 1975, 1981; Rolls 1986a,b, 1990b; see Strongman 1996).

The mechanisms described here would not be limited in the range of emotions for



which they could account. First, different classes of emotion could arise because of different reinforcement contingencies, as described above and indicated in Fig. 3.1.

Second, different intensities within these classes can produce different degrees of emotion (see above and Millenson 1967). For example, as the strength of a positive reinforcer being presented increases, emotions might be labelled as pleasure, elation, and ecstasy. Similarly, as the strength of a negative reinforcer being presented increases, emotions might be labelled as apprehension, fear, and terror (see Fig. 3.1). It may be noted here that anxiety can refer to the state produced by stimuli associated with non-reward or punishment (J. Gray 1987).

Third, any environmental stimulus might have a number of different reinforcement associations. For example, a stimulus might be associated both with the presentation of a reward and of a punishment, allowing states such as conflict and guilt to arise. The different possible combinations greatly increase the number of possible emotions.

Fourth, emotions elicited by stimuli associated with different primary reinforcers will be different even within a reinforcement category (i.e. with the same reinforcement contingency) because the original reinforcers are different. Thus, for example, the state elicited by a stimulus associated with a positive reinforcer such as the taste of food will be different from that elicited by a positive reinforcer such as grooming. Indeed, it is an important feature of the association memory mechanisms described here that when a stimulus is applied, it acts as a key which 'looks up' or recalls the original primary reinforcer with which it was associated. Thus emotional stimuli will differ from each other in terms of the original primary reinforcers with which they were associated. A summary of many different primary reinforcers is provided in Table 10.1, and inspection of this will show how quite a number of emotions are produced typically by certain primary reinforcers. For example, from Table 10.1 it might be surmised that one of the biological origins of the emotion of jealousy might be the state elicited in a male when his partner is courted by another male, because this threatens his parental investment.

A fifth way in which emotions can be different from each other is in terms of the particular (conditioned) stimulus which elicits the emotion. Thus, even though the reinforcement contingency and even the unconditioned reinforcer may be identical, emotions will still be different cognitively, if the stimuli which give rise to the emotions are different (that is, if the objects of the emotion are different). For example, the emotional state elicited by the sight of one person may be different from that elicited by the sight of another person because the people, and thus the cognitive evaluation associated with the perception of the stimuli, are different.

A sixth possible way in which emotions can vary arises when the environment constrains the types of behavioural response which can be made. For example, if an active behavioural response can occur to the omission of an S+, then anger might be produced but if only passive behaviour is possible, then sadness, depression or grief might occur.

By realizing that these six possibilities can occur in different combinations, it can be seen that it is possible to account for a very wide range of emotions, and this is believed to be one of the strengths of the approach described here. It is also the case that the extent to which a stimulus is reinforcing on a particular occasion (and thus the emotion produced) depends on the prior history of reinforcements (both recently and in the longer

term), and that mood state can affect the degree to which a stimulus is reinforcing (see Section 4.8).

### 3.1.4 Refinements of the theory of emotion

The definition of emotions given above, that they are states produced by reinforcing stimuli, is refined now (see further Rolls 1990b).

First, when positively reinforcing stimuli (such as the taste of food or water) are relevant to a drive state produced by a change in the internal milieu (such as hunger and thirst), then we do not normally classify these stimuli as emotional, though they do produce pleasure (see Chapters 2 and 7). In contrast, emotional states are normally initiated by reinforcing stimuli that have their origin in the external environment, such as an (external) noise associated with pain (delivered by an external stimulus). We may then have identified a class of reinforcers (in our example, food) which we do not want to say cause emotions. This then is a refinement of the definition of emotions given above. Fortunately, we can encapsulate the set of reinforcing stimuli that we wish to exclude from our definition of stimuli which produce emotion. They are the set of external reinforcers (such as the sight of food) which are relevant to internal homeostatic drives such as hunger and thirst, which are controlled by internal need-related signals such as the concentration of glucose in the plasma (see Chapters 2 and 7). However, there is room for plenty of further discussion and refinement here. Perhaps some people (especially French people?) might say that they do experience emotion when they savour a wonderful food. There may well be cultural differences here in the semantics of whether such reinforcing stimuli should be included within the category that produce emotions. Another area for discussion is how we wish to categorize the reinforcers associated with sexual behaviour. Such stimuli may be made to be rewarding, and to feel pleasurable, partly because of the internal hormonal state. Does this mean that we wish to exclude such stimuli from the class that we call emotion-provoking, in the same way that we might exclude food reward from the class of stimuli that are said to cause emotion, because the reward value of food depends on an internal controlling signal? I am not sure that there is a perfectly clear answer to this. But this may not matter, as long as we understand that there are some rewarding stimuli that some may wish to exclude from those that cause emotional states.

Second, emotional states can be produced by remembered external reinforcing stimuli. Indeed, the remembered neuronal states are, it is thought, very similar to those produced by a real sensory input, in all but the early stages of sensory processing: see Rolls 1989a; Rolls and Treves 1998).

Third, the stimulus which produces the emotional state does not have to be shown to be a reinforcer when producing the emotional state—it simply has to be capable of being shown to have reinforcing properties. The emotion-provoking stimulus has rewarding or punishing properties, and is a goal for action.

Fourth, the definition given provides great opportunity for cognitive processing (whether conscious or not) in emotions, for cognitive processes will very often be required to determine whether an environmental stimulus or event is reinforcing. Normally an emotion consists of this cognitive processing which results in a decoded signal that the



environmental event is reinforcing, together with the mood state produced as a result. If the mood state is produced in the absence of the external sensory input and the cognitive decoding (for example by direct electrical stimulation of the amygdala, see Rolls 1975), then this is described only as a mood state, and is different from an emotion in that there is no object in the environment towards which the mood state is directed. It is suggested that, in order to produce some stability of mood, the firing rates of the neurons that are activated by external reinforcing stimuli to produce mood states must therefore have their spontaneous firing rates carefully controlled by the brain. (Many brain systems use lateral inhibition in order to maintain sensitivity to contrast constant, but this is not possible in the emotion system in which absolute levels of reinforcer must be represented over moderately long time spans.) The difficulty of maintaining a constant absolute level of firing in neurons such as these may contribute to 'spontaneous' mood swings, depression which occurs without a clear external cause, and the multiplicity of hormonal and transmitter systems which seem to be involved in the control of mood (see Chapter 6).

Having said this, it also seems to be the case that there is some 'regression to a constant value' for emotional stimuli. What I mean by this is that we are sensitive to some extent not just to the absolute level of reinforcement being received, but also to the change in the rate or magnitude of reinforcers being received. This is well shown by the phenomena of positive and negative contrast effects with rewards. Positive contrast occurs when the magnitude of a reward is increased. An animal will work very much harder for a period (perhaps lasting for minutes or longer) in this situation, before gradually reverting to a rate close to that at which the animal was working for the small reinforcement. A comparable contrast effect is seen when the reward magnitude (or rate at which rewards are obtained) is reduced—there is a negative overshoot in the rate of working for a time. This phenomenon is adaptive. It is evidence that animals are in part sensitive to a change in reinforcement, and this helps them to climb gradients to obtain better rewards. In effect, regardless of the absolute level of reinforcement being achieved, it is adaptive to be sensitive to a change in reinforcement. If this were not true, an animal receiving very little reinforcement but then obtaining a small increase in positive reinforcement might still be working very little for the reward. But it is much more adaptive to work hard in this situation, as the extra little bit of reward might make the difference between survival or not. A similar phenomenon may be evident in humans. People who have very little in the way of rewards, who may be poor, have a poor diet, and may suffer from disease, may nevertheless not have a baseline level of happiness that is necessarily very different from that of a person in an affluent society who in absolute terms apparently has many more rewards. This may be due in part to resetting of the baseline of expected rewards to a constant value, so that we are especially sensitive to changes in rewards (or punishers) (cf. Solomon and Corbit 1974; Solomon 1980).

The approach described above shows that the learning of stimulus–reinforcer associations is the learning involved when emotional responses are learned. In so far as the majority of stimuli which produce our emotional responses do so as a result of learning, this type of learning, and the brain mechanisms which underlie it, are crucial to the majority of our emotions. This, then, provides a theoretical basis for understanding the

functions of some brain systems such as the amygdala in emotion. Ways in which this approach can be applied are given below.

It also follows from this approach towards a theory of emotion that brain systems involved in disconnecting stimulus–reinforcer associations which are no longer appropriate will also be very important in emotion. Failure of this function would be expected to lead, for example, in frustrating situations to inappropriate perseveration of behaviour to stimuli no longer associated with positive reinforcement. The inability to correct behaviour when reinforcement contingencies change would be evident in a number of emotion-provoking situations, such as frustration (i.e. non-reward), and the punishment of previously rewarded behaviour. It will be shown in Chapter 4 that this approach, which emphasizes the necessity, in, for example, social situations, to update and correct the decoded reinforcement value of stimuli continually, provides a basis for understanding the functions of some other brain regions such as the orbitofrontal cortex in emotion.

### 3.1.5 The functions of emotion

Understanding the functions of emotion is also important for understanding the nature of emotions, and for understanding the brain systems involved in the different types of response that are produced by emotional states. Emotion appears to have many functions, which are not necessarily mutually exclusive. Some of these functions are described next, and in more detail by Rolls (1990b).

The first function proposed for emotion is the elicitation of autonomic responses (e.g. a change in heart rate) and endocrine responses (e.g. the release of adrenaline). It is of clear survival value to prepare the body, for example by increasing the heart rate, so that actions which may be performed as a consequence of the reinforcing stimulus, such as running, can be performed more efficiently. The projections from the amygdala and orbitofrontal cortex via the hypothalamus as well as directly towards the brainstem autonomic motor nuclei may be particularly involved in this function. The James–Lange theory (see Schachter and Singer 1962; Grossman 1967; Reisenzein 1983), and theories which are closely related to it in supposing that feedback from parts of the periphery, such as the face (see Adelman and Zajonc 1989) or body (Damasio 1994), leads to emotional feelings, have the major weakness that they do not give an adequate account of how the peripheral change is produced only by stimuli which happen to be emotion-provoking. Perhaps the most important issue in emotion is why only some stimuli give rise to emotions. We have prepared the way for answering this by identifying the stimuli which produce emotions as reinforcing stimuli. It is then possible to answer the question of *how* emotions are produced by investigating which parts of the brain decode whether stimuli are reinforcing. Such investigations are described in Chapter 4.

The second function proposed is flexibility of behavioural responses to reinforcing stimuli. The thesis here is that when a stimulus in the environment elicits an emotional state, we can perform any appropriate response to obtain the reward, or avoid the punishment. This is more flexible than simply learning a fixed behavioural response to a stimulus. More formally, the first stage in learning, for example, avoidance would be

classical conditioning of an emotional response such as fear to a tone associated with shock. The amygdala and orbitofrontal cortex appear to be especially involved in this first stage of this type of learning, by virtue of their function in forming and correcting stimulus-reinforcement associations (see below). The second stage would be instrumental learning of an operant response, motivated by and performed in order to terminate the fear-inducing stimulus (see J. Gray 1975; Rolls 1990b). This two-stage learning process was suggested as being important for avoidance learning by N. E. Miller and O. H. Mowrer (see J. Gray 1975). The suggestion made here is that this general type of two-stage learning process is closely related to the design of animals for many types of behaviour, including emotional behaviour. It simplifies the interface of sensory systems to motor systems. Instead of having to learn a particular response to a particular stimulus by slow, habit, trial and error, learning, two-stage learning allows very fast (often one trial) learning of an emotional state to a rewarding or punishing stimulus. Then the motor system can operate in quite a general way, using many previously learned strategies, to approach the reward or avoid the punisher, which act as goals. This not only gives great flexibility to the interface, but also makes it relatively simple. It means that the reward value of a number of different stimuli can be decoded at roughly the same time. A behavioural decision system can then compare the different rewards available, in that they have a form of 'common currency'. (The value of each type of reward in this 'common currency' will be affected by many different factors, such as need state, e.g. hunger; how recently that reward has been obtained (see Chapter 10); the necessity in evolution to set each type of reward so that it sometimes is chosen if it is important for survival; etc.) The decision system can then choose between the rewards, based on their value, but also on the cost of obtaining each reward (see Chapter 10). After the choice has been made, the action or motor system can then switch on any behavioural responses possible, whether learned or not, in order to maximize the reward signal being obtained. The magnitude of the reward signal being obtained would be indicated just by the firing of the neurons which reflect the value of reward being obtained (e.g. the taste of a food if hungry, the pleasantness of touch, etc.), as described in Chapters 2 and 4. The actual way in which the appropriate response or action is learned may depend on response-reinforcer association learning, or on some more general type of purposive behaviour that can be learned to obtain goals (see Mackintosh and Dickinson 1979; Pearce 1996 Chapter 6).

A third function of emotion is that it is motivating. For example, fear learned by stimulus-reinforcer association formation provides the motivation for actions performed to avoid noxious stimuli. Similarly, positive reinforcers elicit motivation, so that we will work to obtain the rewards. Another example where emotion affects motivation is when a reward becomes no longer available, that is frustrative non-reward (see Fig. 3.1). If an action is possible, then increased motivation facilitates behaviour to work harder to obtain that reinforcer again or another reinforcer. If no action is possible to obtain again that positive reinforcer (e.g. after a death in the family), then as described in Section 3.1.3, grief or sadness may result. This may be adaptive, by preventing continuing attempts to regain the positive reinforcer which is no longer available, and helping the animal in due course to therefore be sensitive to other potential reinforcers to which it might be adaptive to switch. As described in Chapter 9, if such frustrative non-reward occurs in

humans when no action is possible, depression may occur. A depressed state which lasts for a short time may be seen as being adaptive for the reason just given. However, the depression may last for a very long time perhaps because long-term explicit (conscious) knowledge in humans enables the long-term consequences of loss of the positive reinforcer to be evaluated and repeatedly brought to mind as described in Chapter 9 and in Section 10.6, and this may make long-term (psychological) depression maladaptive.

A fourth function of emotion is in communication. For example, monkeys may communicate their emotional state to others, by making an open-mouth threat to indicate the extent to which they are willing to compete for resources, and this may influence the behaviour of other animals. Communicating emotional states may have survival value, for example by reducing fighting. There are neural systems in the amygdala and overlying temporal cortical visual areas which are specialized for the face-related aspects of this processing, that is for decoding for example facial expression or gesture (see Chapter 4).

A fifth function of emotion is in social bonding. Examples of this are the emotions associated with the attachment of the parents to their young, with the attachment of the young to their parents, and with the attachment of the parents to each other. In the theory of the ways in which the genes affect behaviour ('selfish gene' theory, see R. Dawkins 1989), it is held that (because, e.g., of the advantages of parental care) all these forms of emotional attachment have the effect that genes for such attachment are more likely to survive into the next generation. Kin-altruism can also be considered in these terms (see e.g. R. Dawkins 1989; Chapter 10, Footnote 1).

A sixth function of emotion may be generalized from the above. It may be suggested that anything that feels pleasant to the organism, and is positively reinforcing, so that actions made to obtain it are performed, has survival value. (Stimuli which act as implicit or unconscious rewards should also act to produce pleasant feelings in the explicit or conscious processing system, so that both the implicit and explicit routes to action operate largely consistently—see Chapter 9.) One example of this is slight novelty, which may feel good and be positively reinforcing because it may lead to the discovery of better opportunities for survival in the environment (e.g. a new food). It is crucial that animals that succeed in the genetic competition which drives evolution have genes which encourage them to explore new environments, for then it is possible for the genes which happen to be present in an individual to explore the large multidimensional space of the environment in which they might succeed. Another example is gregariousness, which may assist the identification of new social partners, which could provide advantage. Probably related to the effects of novelty is sensory-specific satiety, the phenomenon whereby pleasant tastes during a meal gradually become less pleasant as satiety approaches (see Rolls 1986c, 1989b, 1993a). This may be an aspect of a more general adaptation to ensure that behaviour does eventually switch from one reinforcer to another. Comparably, it is likely that natural selection acting on genes will lead to unpleasant feelings, and negative reinforcement, being associated with behaviour which does not have survival value, at least in cases where genes can influence matters. (Of course the genes may be misled sometimes and lead to behaviour which does not have survival value, as when for example the non-nutritive sweetener saccharin is eaten by animals. This does not disprove the theory, but only points out that the genes cannot specify correctly for every possible

stimulus or event in the environment, but must only on average lead to behaviour feeling pleasant that increases fitness, i.e. is appropriate for gene survival.)

A seventh effect of emotion is that the current mood state can affect the cognitive evaluation of events or memories (see Blaney 1986), and this may have the function of facilitating continuity in the interpretation of the reinforcing value of events in the environment. A theory of how this occurs is presented in Section 4.8 'Effects of emotions on cognitive processing'.

An eighth function of emotion is that it may facilitate the storage of memories. One way in which this occurs is that episodic memory (i.e. one's memory of particular episodes) is facilitated by emotional states. This may be advantageous in that storage of as many details as possible of the prevailing situation when a strong reinforcer is delivered may be useful in generating appropriate behaviour in situations with some similarities in the future. This function may be implemented by the relatively non-specific projecting systems to the cerebral cortex and hippocampus, including the cholinergic pathways in the basal forebrain and medial septum, and the ascending noradrenergic pathways (see Section 4.7; Rolls and Treves 1998; Wilson and Rolls 1990a,b). A second way in which emotion may affect the storage of memories is that the current emotional state may be stored with episodic memories, providing a mechanism for the current emotional state to affect which memories are recalled. In this sense, emotion acts as a contextual retrieval cue, that as with other contextual effects influences the retrieval of episodic memories (see Rolls and Treves 1998). A third way in which emotion may affect the storage of memories is by guiding the cerebral cortex in the representations of the world which are set up. For example, in the visual system, it may be useful to build perceptual representations or analysers which are different from each other if they are associated with different reinforcers, and to be less likely to build them if they have no association with reinforcers. Ways in which backprojections from parts of the brain important in emotion (such as the amygdala) to parts of the cerebral cortex could perform this function are discussed in Section 4.8, 'Effects of emotions on cognitive processing'; by Rolls (1989a, 1990b, 1992b); and by Rolls and Treves (1998).

A ninth function of emotion is that by enduring for minutes or longer after a reinforcing stimulus has occurred, it may help to produce persistent motivation and direction of behaviour.

A tenth function of emotion is that it may trigger recall of memories stored in neocortical representations. Amygdala backprojections to the cortex could perform this for emotion in a way analogous to that in which the hippocampus could implement the retrieval in the neocortex of recent (episodic) memories (see Rolls and Treves 1998).

It is useful to have these functions of emotion in mind when considering the neural basis of emotion, for each function is likely to have particular output pathways from emotional systems associated with it.

### 3.1.6 The James-Lange and other bodily theories of emotion

James (1884) believed that emotional experiences were produced by sensing bodily changes, such as changes in heart rate or in skeletal muscles. Lange (1885) had a similar

#### James-Lange theory of emotion

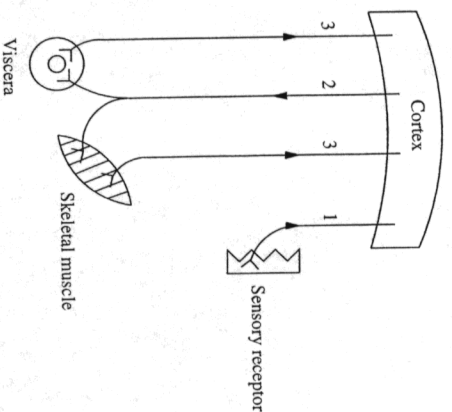


Fig. 3.2 The James-Lange theory of emotion proposes that there are three steps in producing emotional feelings. The first step is elicitation by the emotion-provoking stimulus (received by the cortex via pathway 1 in the Figure) of peripheral changes, such as skeletal-muscular activity to run away, and autonomic changes, such as alteration of heart rate (via pathways labelled 2 in the Figure). The second step is the sensing of the peripheral responses (e.g. altered heart rate, and somatosensory effects produced by running away) via pathways labelled 3 in the Figure). The third step is elicitation of the emotional feeling in response to the sensed feedback from the periphery.

view, although he emphasized the role of autonomic feedback (for example from the heart) in producing the experience of emotion. The theory, which became known as the James-Lange theory, suggested that there are three steps in producing emotional feelings (see Fig. 3.2). The first step is elicitation by the emotion-provoking stimulus of peripheral changes, such as skeletal-muscular activity to produce running away, and autonomic changes, such as alteration of heart rate. But, as pointed out above, the theory leaves unanswered perhaps the most important issue in any theory of emotion: Why do some events make us run away (and then feel emotional), whereas others do not? This is a major weakness of this type of theory. The second step is the sensing of the peripheral responses (e.g. running away, and altered heart rate). The third step is elicitation of the emotional feeling in response to the sensed feedback from the periphery.

The history of research into peripheral theories of emotion starts with the fatal flaw that leaves unanswered this most important question. The history continues with the accumulation of empirical evidence which has gradually weakened more and more the hypothesis that peripheral responses made during emotional behaviour have anything to do with producing the emotional behaviour (which has largely already been produced anyway according to the James-Lange theory), or the emotional feeling. Some of the landmarks in this history are as follows. First, the peripheral changes produced during emotion are not sufficiently distinct to be able to carry the information which would enable one to have subtly different emotional feelings to the vast range of different stimuli that can produce



different emotions. The evidence suggests that by measuring many peripheral changes in emotion, such as heart rate, skin conductance, breathing rate, and hormones such as adrenaline and noradrenaline (known in the United States by their Greek names epinephrine and norepinephrine), it may be possible to make coarse distinctions between, for example, anger and fear, but not much finer distinctions (Wagner 1989, Cacioppo *et al.* 1993; Oatley and Jenkins 1996). Second, when emotions are evoked by imagery, then the peripheral responses are much less marked and distinctive than during emotions produced by external stimuli (Ekman *et al.* 1983; Stemler 1989; Levenson *et al.* 1990). This makes sense in that although an emotion evoked by imagery may be strong, there is no need to produce strong peripheral responses, because no behavioural responses are required. Third, disruption of peripheral responses and feedback from them surgically (for example in dogs, Cannon 1927, 1929, 1931), or as a result of spinal cord injury in humans (Hohmann 1966; Bermond *et al.* 1991) does not abolish emotional responses. What was found was that in some patients there was apparently some reduction in emotions in some situations (Hohmann 1966), but this could be related to the fact that some of the patients were severely disabled which could have produced its own consequences for emotionality, and that in many cases the patients were considerably older than before the spinal cord damage, which could itself have been a factor. What was common to both studies was that emotions could be felt by all the patients; and that in some cases, emotions resulting from mental events were even reported as being stronger (Hohmann 1966; Bermond *et al.* 1991). Fourth, when autonomic changes are elicited by injections of, for example, adrenaline or noradrenaline, particular emotions are not produced. Instead, the emotion that is produced depends on the cognitive decoding of the reinforcers present in the situation, for example an actor who insults your parents to make you angry, or an actor who plays a game of hula hoop to make you feel happy (Schachter and Singer 1962). In this situation, the hormone adrenaline or noradrenaline can alter the magnitude of the emotion, but not which emotion is felt. This is further evidence that it is the decoded reinforcement value of the input stimulus or events which determines which emotion is felt. The fact that the hormone injections produced some change in the magnitude of an emotion is not very surprising. If you felt your heart pounding for no explicable reason, you might wonder what was happening, and therefore react more or abnormally. Fifth, if the peripheral changes associated with emotion are blocked with drugs, then this does not block the perception of emotion (Reisenzein 1983). Sixth, it is found that in normal life, behavioural expressions of emotion (for example smiling at a bowling alley) do not usually occur when one might be expected to feel happy because of a success, but instead occur when one is looking at one's friends (Kraut and Johnson 1979). These body responses which can be very brief, thus often serve the needs of communication, or of action, not of producing emotional feelings.

### 3.1.6.1 The somatic marker hypothesis

Despite this rather overwhelming evidence against an important role for body responses in producing emotions or emotional feelings, Damasio (1994) has effectively tried to resurrect a weakened version of the James-Lange theory of emotion from the last century, by

arguing with his somatic marker hypothesis that after reinforcers have been evaluated, a bodily response ('somatic marker') normally occurs, then this leads to a bodily feeling, which in turn is appreciated by the organism to then make a contribution to the decision-making process. (In the James-Lange theory, it was emotional feelings that depend on peripheral feedback; for Damasio, it is the decision of which behavioural response to make that is normally influenced by the peripheral feedback. A quotation from Damasio (1994, p190) follows: 'The squirrel did not really think about his various options and calculate the costs and benefits of each. He saw the cat, was jolted by the body state, and ran.' Here it is clear that the pathway to action uses the body state as part of the route. Damasio would also like decisions to be implemented using the peripheral changes elicited by emotional stimuli. Given all the different reinforcers which may influence behaviour, Damasio (1994) even suggests that the net result of them all is reflected in the net peripheral outcome, and then the brain can sense this net peripheral result, and thus know what decision to take.) The James-Lange theory has a number of major weaknesses just outlined which apply also to the somatic marker hypothesis. Another major weakness, which applies to both the James-Lange and to Damasio's somatic marker hypothesis, is that they do not take account of the fact that once an information processor has determined that a response should be made or inhibited based on reinforcement association, a function attributed in the theory proposed in this chapter and by Rolls (1986a,b, 1990b) to the orbitofrontal cortex, it would be very inefficient and noisy to place in the execution route a peripheral response, and transducers to attempt to measure that peripheral response, itself a notoriously difficult procedure (see, e.g., Grossman 1967). Even for the cases when Damasio (1994) might argue that the peripheral somatic marker and its feedback can be by-passed using conditioning of a representation in, e.g., the somatosensory cortex to a command signal (which might originate in the orbitofrontal cortex), he apparently would still wish to argue that the activity in the somatosensory cortex is important for the emotion to be appreciated or to influence behaviour. (Without this, the somatic marker hypothesis would vanish.) The prediction would apparently be that if an emotional response were produced to a visual stimulus, then this would necessarily involve activity in the somatosensory cortex or other brain region in which the 'somatic marker' would be represented. This prediction could be tested (for example in patients with somatosensory cortex damage), but it seems most unlikely that an emotion produced by a visual reinforcer would require activity in the somatosensory cortex to feel emotional or to elicit emotional decisions. The alternative view proposed here (and by Rolls 1986a,b, 1990b) is that where the reinforcement value of the visual stimulus is decoded, namely in the orbitofrontal cortex and the amygdala, is the appropriate part of the brain for outputs to influence behaviour (via, e.g., the orbitofrontal-to-striatal connections), and that the orbitofrontal cortex and amygdala, and brain structures that receive connections from them, are the likely places where neuronal activity is directly related to the felt emotion (see further Rolls 1997a,d and Chapter 9).

### 3.1.7 Individual differences in emotion, and personality

H. J. Eysenck developed the theory that personality might be related to different aspects

of conditioning. He analysed the factors that accounted for the variance in the differences between the personality of different humans (using, for example, questionnaires), and suggested that the first two factors in personality (those which accounted for most of the variance) were introversion vs extraversion, and neuroticism (related to a tendency to be anxious). He performed studies of classical conditioning on groups of subjects, and also obtained measures of what he termed arousal. Based on the correlations of these measures with the dimensions identified in the factor analysis, he suggested that introverts showed greater conditionability than extraverts; and that neuroticism raises the general intensity of emotional reactions (see Eysenck and Eysenck 1968).

J. Gray (1970) reinterpreted the findings, suggesting that introverts are more sensitive to punishment and frustrative non-reward than are extraverts; and that neuroticism reflects the extent of sensitivity to both reward and punishment.

I do not wish to consider this research area in detail. However, I do point out that insofar as sensitivity to reward and punishment, and the ability to learn and be influenced by reward and punishment, may be important in personality, and are closely involved in emotion according to the theory developed here, there may be close links between the neural bases of emotion, to be described in Chapter 4, and personality.

## 4 The neural bases of emotion

In this chapter the neural bases of emotion are considered. Particular attention is paid to research in non-human primates as well as in humans, partly because the developments in primates in the structure and connections of neural systems involved in emotion such as the amygdala and orbitofrontal cortex make studies in primates particularly important for understanding emotion in humans.

### 4.1 Introduction

Some of the main brain regions implicated in emotion will now be considered in the light of the introduction given in Chapter 3 on the nature and functions of emotion. These brain regions include the amygdala, orbitofrontal cortex, and basal forebrain areas including the hypothalamus. Some of these brain regions are indicated in Figs 4.1 and 4.2. Particular attention is paid to the functions of these regions in primates, for in primates the neocortex undergoes great development and provides major inputs to these regions, in some cases to parts of these structures thought not to be present in non-primates. An example of this is the projection from the primate neocortex in the anterior part of the temporal lobe to the basal accessory nucleus, of the amygdala (see below). Studies in primates are thus particularly relevant to understanding the neural basis of emotion in the human.

The way in which recent studies in primates indicate that the neural processing of emotion is organized is as follows (see Fig. 4.3). First, there are brain mechanisms that are involved in computing the reward value of primary (unlearned) reinforcers. The primary reinforcers include taste, touch (both pleasant touch and pain), and to some extent smell, and perhaps certain visual stimuli, such as face expression. Then some brain regions are concerned with learning associations between previously neutral stimuli, such as the sight of objects or of individuals' faces, with primary reinforcers. These brain regions include the amygdala and orbitofrontal cortex. For the processing of primary reinforcers, and especially for secondary reinforcers, the brain is organized to process the stimulus first to the object level (so that if the input is visual, the object can be recognized independently of its position on the retina, or size, or view), and then to determine whether the stimulus is rewarding or punishing. Once the relevant brain regions have determined whether the input is reinforcing, whether primary or secondary, the signal is passed directly to output regions of the brain, with no need to produce peripheral body or autonomic responses. The brain regions in which the reinforcing, and hence emotional, value of stimuli are represented interface to three main types of output system. The first is the autonomic and endocrine system, for producing such changes as increased heart rate and release of