BEHAVIOURAL AND BRAIN SCIENCES (2000) **23**, 177-234
Printed in the United States of America

Précis of *The Brain and Emotion*

Edmund T. Rolls
Department of Experimental Psychology, University of Oxford, Oxford
OX1 3UD, England.
Edmund.Rolls@psy.ox.ac.uk

Abstract: The topics treated in *The Brain and Emotion* include the definition, nature and functions of emotion (Chapter 3), the neural bases of emotion (Chapter 4), reward, punishment and emotion in brain design (Chapter 10), a theory of consciousness and its application to understanding emotion and pleasure (Chapter 9), and neural networks and emotion-related learning (Appendix). The approach is that emotions can be considered as states elicited by reinforcers (rewards and punishers). This approach helps with understanding the functions of emotion, and with classifying different emotions; and in understanding *what* information processing systems in the brain are involved in emotion, and *how* they are involved. The hypothesis is developed that brains are designed around reward and punishment evaluation systems, because this is the way that genes can build a complex system that will produce appropriate but flexible behavior to increase fitness (Chapter 10). By specifying goals rather than particular behavioral patterns of responses, genes leave much more open the possible behavioral strategies that might be required to increase fitness. The importance of reward and punishment systems in brain design also provides a basis for understanding brain mechanisms of motivation, as described in Chapters 2 for appetite and feeding, 5 for brain-stimulation reward, 6 for addiction, 7 for thirst, and 8 for sexual behavior.

Keywords: amygdala; brain evolution; consciousness ; dopamine; emotion; hunger; orbitofrontal cortex; punishment; reward; taste

1. Introduction

What are emotions? Why do we have emotions? What are the rules by which emotion operates? What are the brain mechanisms of emotion, and how can disorders of emotion be understood? Why does it feel like something to have an emotion?

What motivates us to work for particular rewards such as food when we are hungry, or water when we are thirsty? How do these motivational control systems operate to ensure that we eat approximately the correct amount of food to maintain our body weight or to replenish our thirst? What factors account for the overeating and obesity which some humans show?

Why is the brain built to have reward, and punishment, systems, rather than in some other way? Raising this issue of brain design and why we have reward and punishment systems, and emotion and motivation, produces a fascinating answer based on how genes can direct our behavior to increase fitness. How does the brain produce behavior by using reward, and punishment, mechanisms? These are some of the questions considered in *The Brain and Emotion* (Rolls, 1999).

The brain mechanisms of both emotion and motivation are considered together. The examples of motivated behavior described are hunger (Chapter 2), thirst (Chapter 7), and sexual behavior (Chapter 8). The reason that both emotion and motivation are treated is that both involve rewards and punishments as the fundamental solution of the brain for interfacing sensory systems to action selection and execution systems. Computing the reward and punishment value of sensory stimuli and then using selection between different rewards and avoidance of punishments in a common reward-based currency appears to be the general solution that brains use to produce appropriate behavior. The behavior selected is appropriate in that it is based on the sensory systems and reward decoding that our genes specify (through the process of natural selection) in order to maximise fitness (reproductive potential).

The book provides a modern neuroscience-based approach to information processing in the brain, and deals especially with the information processing involved in emotion (Chapter 4), hunger, thirst and sexual behavior (Chapters 2, 7 and 8), and reward (Chapters 5 and 6). The book though links this analysis to the wider context of the nature of emotions, their functions (Chapter 3), how they evolved (Chapter 10), and the larger issue of why emotional and motivational feelings and consciousness might arise in a system organised like the brain (Chapter 9).

The Brain and Emotion is thus intended to uncover some of the important principles of brain function and design. The book is also intended to show that the way in which the brain functions in motivation and emotion can be seen to be the result of natural selection operating to select genes which optimise our behavior by building into us the appropriate reward and punishment systems and the appropriate rules for the operation of these systems.

A major reason for investigating the actual brain mecha-

EDMUND ROLLS is a Professor of Experimental Psychology at University of Oxford, and Associate. Director of the Medical Research Council Interdisciplinary Research Centre for Cognitive Neuroscience at Oxford University. His recent books include *Neural networks and brain function* (with A. Treves, Oxford University Press, 1998) and *An introduction to connectionist modelling of cognitive function* (with P. McLeod and K. Plunkett, Oxford University Press, 1998).

nisms that underlie emotion and motivation, and reward and punishment, is not only to understand how our own brains work, but also to have the basis for understanding and treating medical disorders of these systems (such as altered emotional behavior after brain damage, depression, anxiety and addiction). It is because of the intended relevance to humans that emphasis is placed on research in non–human primates. It turns out that many of the brain systems involved in emotion and motivation have undergone considerable development in primates. For example, the temporal lobe has undergone great development in primates, and a number of systems in the temporal lobe are either involved in emotion (e.g. the amygdala), or provide some of the main sensory inputs to brain systems involved in emotion and motivation. The prefrontal cortex has also undergone considerable development in primates: one part of it, the orbitofrontal cortex, is very little developed in rodents, yet is one of the major brain areas involved in emotion and motivation in primates, including humans. The elaboration of some of these brain areas has been so great in primates that even evolutionarily old systems such as the taste system appear to have been reconnected (compared to rodents) to place much more emphasis on cortical processing, taking place in areas such as the orbitofrontal cortex (see Chapter 2). The principle of the stage of sensory processing at which reward value is extracted and made explicit in the representation may even have changed between rodents and primates, for example in the taste system (see Chapter 2). In primates, there has also been great development of the visual system, and this itself has had important implications for the types of sensory stimuli that are processed by brain systems involved in emotion and motivation. One example is the importance of facial identity and facial expression decoding, which are both critical in primate emotional behavior, and provide a central part of the foundation for much primate social behavior.

2. A Theory of Emotion, and some Definitions

Emotions can usefully be defined as states elicited by rewards and punishments, including changes in rewards and punishments (see also Rolls 1986a; 1986b; 1990). A reward is anything for which an animal will work. A punishment 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 pleasant touch, praise, or winning a large sum of money. Another example of an emotion might be fear produced by the sound of a rapidly approaching bus, or the sight of an angry expression on someone's face. We will work to avoid such stimuli, which are punishing. Another example would 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 would be relief, produced by the omission or termination of a punishing stimulus such as 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 punishments received, omitted, or terminated. A diagram summarizing some of the emotions associated with the delivery of reward or punishment or a stimulus associated with them, or with the omission of a reward or punishment, is shown in Fig.1.

Before accepting this approach, 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 below. 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 (see Strongman 1996) have in common that part of the process involves "appraisal" (e.g. Frijda 1986; Lazarus 1991; Oatley and Jenkins 1996). In all these theories the concept of appraisal presumably involves assessing whether something is rewarding or punishing. The description in terms of reward or punishment adopted here seems more tightly and operationally specified. I next consider a slightly more formal definition than rewards or punishments, in which the concept of reinforcers is introduced, and show how there has been a considerable history in the development of ideas along this line.

The proposal that emotions can be usefully seen as states produced by instrumental reinforcing stimuli follows earlier work by Millenson (1967), Weiskrantz (1968), Gray (1975; 1987) and Rolls (1986a; 1986b; 1990). (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.) Some stimuli are unlearned 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 occurs via a process like classical conditioning. 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 decreases 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 (the conditioned stimulus) that has previously been associated with an electrical shock (the primary reinforcer).

The converse reinforcement contingencies produce the opposite effects on behavior. The omission or termination of a positive reinforcer ("extinction" and "time out" respectively, sometimes described as "punishing") decreases the probability of responses. Responses followed by the omission or termination of a negative reinforcer increase in probability, this pair of negative reinforcement operations being termed "active avoidance" and "escape" respectively (see further Gray 1975; Mackintosh 1983).

This foundation has been developed (see also Rolls 1986a; 1986b; 1990) to show how a very wide range of emotions can be accounted for, as a result of the operation of a number of factors, including the following:

1. The *reinforcement contingency* (e.g. whether reward or punishment is given, or withheld) (see Fig. 1).

2. The *intensity* of the reinforcer (see Fig. 1).

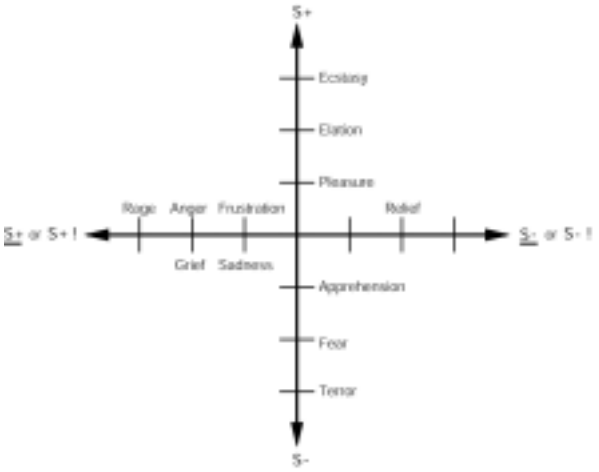

Fig. 1. Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the centre of the diagram, on a continuous scale. The classification scheme created by the different reinforcement contingencies consists of (1) the presentation of a positive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

3. 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 punisher, allowing states such as conflict and guilt to arise.)

4. Emotions elicited by stimuli associated with *different primary reinforcers* will be different.

5. Emotions elicited by *different secondary reinforcing stimuli* will be different from each other (even if the primary reinforcer is similar).

6. The emotion elicited can depend on whether an *active or passive behavioral response* is possible. (For example, if an active behavioral response can occur to the omission of a positive reinforcer, then anger might be produced, but if only passive behavior is possible, then sadness, depression or grief might occur.)

By combining these six factors, it is possible to account for a very wide range of emotions (for elaboration see Rolls, 1990 and *The Brain and Emotion*). It is also worth noting that emotions can be produced just as much by the recall of reinforcing events as by external reinforcing stimuli; that cognitive processing (whether conscious or not) is important in many emotions, for very complex cognitive processing may be required to determine whether or not environmental events are reinforcing. Indeed, emotions normally consist of cognitive processing which analyses the stimulus, and then determines its reinforcing valence; and then an elicited mood change if the valence is positive or negative. In that an emotion is produced by a stimulus, philosophers say that emotions have an object in the world, and that emotional states are intentional, in that they are about something. We note that a mood or affective state may occur in the absence of an external stimulus, as in some types of depression, but that normally the mood or affective state is produced by an external stimulus, with the whole process of stimulus representation, evaluation in terms of reward or punishment, and the resulting mood or affect being referred to as emotion.

Three issues receive discussion here (see further Rolls 1999). One is that rewarding stimuli such as the taste of food are not usually described as producing emotional states (though there are cultural differences here!). It is useful here to separate rewards related to internal homeostatic need states associated with (say) hunger and thirst, and to note that these rewards are not normally described as producing emotional states. In contrast, the great majority of rewards and punishers are external stimuli not related to internal need states such as hunger and thirst, and these stimuli do produce emotional responses. An example is fear produced by the sight of a stimulus which is about to produce pain.

A second issue is that philosophers usually categorize fear in the example as an emotion, but not pain. The distinction they make may be that primary (unlearned) reinforcers do not produce emotions, whereas secondary reinforcers (stimuli associated by stimulus-reinforcement learning with primary reinforcers) do. They describe the pain as a sensation. But neutral stimuli (such as a table) can produce sensations when touched. It accordingly seems to be much more useful to categorise stimuli according to whether they are reinforcing (in which case they produce emotions), or are not reinforcing (in which case they do not produce emotions). Clearly there is a difference between primary reinforcers and learned reinforcers; but this is most precisely caught by noting that this is the difference, and that it is whether a stimulus is reinforcing that determines whether it is related to emotion.

A third issue is that, as we are about to see, emotional states (i.e. those elicited by reinforcers) have many functions, and the implementations of only some of these functions by the brain are associated with emotional feelings (Rolls 1999), including evidence for interesting dissociations in some patients with brain damage between actions performed to reinforcing stimuli and what is subjectively reported. In this sense it is biologically and psychologically useful to consider emotional states to include more than those states associated with feelings of emotion.

3. The Functions of Emotion

The functions of emotion also provide insight into the nature of emotion. These functions, described more fully elsewhere (Rolls 1990; 1999), can be summarized as follows:

1. The *elicitation of autonomic responses* (e.g. a change in heart rate) *and endocrine responses* (e.g. the release of adrenaline). These prepare the body for action.

2. *Flexibility of behavioral responses to reinforcing stimuli.* Emotional (and motivational) states allow a simple interface between sensory inputs and action systems. The essence of this idea is that goals for behavior are specified by reward and punishment evaluation. When an environmental stimulus has been decoded as a primary reward or punishment, or (after previous stimulus-reinforcer association learning) a secondary rewarding or punishing stimulus, then it becomes a goal for action. The animal can then perform any action (instrumental response) to obtain the reward, or to avoid the punisher. Thus there is flexibility of action, and this is in contrast with stimulus-response, or habit, learning in which a particular response

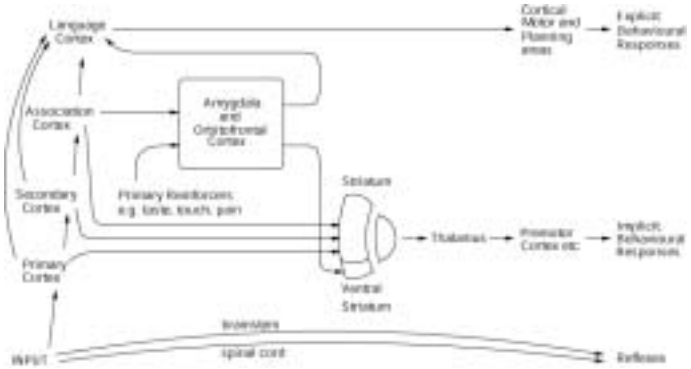

Fig. 2. Summary of the organisation of some of the brain mechanisms underlying emotion, showing dual routes to the initiation of action in response to rewarding and punishing, that is emotion-producing, stimuli. The inputs from different sensory systems to brain structures such as the orbitofrontal cortex and amygdala allow these brain structures to evaluate the reward- or punishment-related value of incoming stimuli, or of remembered stimuli. The different sensory inputs allow evaluations within the orbitofrontal cortex and amygdala based mainly on the primary (unlearned) reinforcement value for taste, touch and olfactory stimuli, and on the secondary (learned) reinforcement value for visual and auditory stimuli. In the case of vision, the 'association cortex' which sends representations of objects to the amygdala and orbitofrontal cortex is the inferior temporal visual cortex. One route for the outputs from these evaluative brain structures is via projections directly to structures such as the basal ganglia (including the striatum and ventral striatum) to allow implicit, direct behavioral responses based on the reward or punishment-related evaluation of the stimuli to be made. The second route is via the language systems of the brain, which allow explicit (verbalizable) decisions involving multistep syntactic planning to be implemented. After *The Brain and Emotion*, Fig. 9. 4.

to a particular stimulus is learned. It also contrasts with the elicitation of species-typical behavioral responses by sign releasing stimuli (such as pecking at a spot on the beak of the parent herring gull in order to be fed, Tinbergen (1951), where there is inflexibility of the stimulus and the response, and which can be seen as a very limited type of brain solution to the elicitation of behavior). The emotional route to action is flexible not only because any action can be performed to obtain the reward or avoid the punishment, but also because the animal can learn in as little as one trial that a reward or punishment is associated with a particular stimulus, in what is termed "stimulus-reinforcer association learning".

To summarize and formalize, two processes are involved in the actions being described. The first is stimulus-reinforcer association learning, and the second is instrumental learning of an operant response made to approach and obtain the reward or to avoid or escape from the punisher. Emotion is an integral part of this, for it is the state elicited in the first stage, by stimuli which are decoded as rewards or punishers, and this state has the property that it is motivating. The motivation is to obtain the reward or avoid the punisher, and animals must be built to obtain certain rewards and avoid certain punishers. Indeed, primary or unlearned rewards and punishers are specified by genes which effectively specify the goals for action. This is the solution which natural selection has found for how genes can influence behavior to promote fitness (as measured by reproductive success), and for how the brain could interface sensory systems to action systems.

Selecting between available rewards with their associated costs, and avoiding punishers with their associated costs, is a process which can take place both implicitly (unconsciously), and explicitly using a language system to enable long-term plans to be made (Rolls 1999). These many different brain systems, some involving implicit evaluation of rewards, and others explicit, verbal, conscious, evaluation of rewards and planned long-term goals, must all enter into the selector of behavior (see Fig. 2). This selector is poorly understood, but it might include a process of competition between all the competing calls on output, and might involve the basal ganglia in the brain (see Fig. 2 and Rolls 1999).

3. Emotion is *motivating*, as just described. For example, fear learned by stimulus-reinforcement association provides the motivation for actions performed to avoid noxious stimuli.

4. *Communication*. Monkeys for example 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 behavior of other animals. This aspect of emotion was emphasized by Darwin (1872), and has been studied more recently by Ekman (1982; 1993). He reviews evidence that humans can categorize facial expressions into the categories happy, sad, fearful, angry, surprised and disgusted, and that this categorization may operate similarly in different cultures. He also describes how the facial muscles produce different expressions. Further investigations of the degree of cross-cultural universality of facial expression, its development in infancy, and its role in social behavior are described by Izard (1991) and Fridlund (1994). As shown below, there are neural systems in the amygdala and overlying temporal cortical visual areas which are specialized for the face-related aspects of this processing.

5. *Social bonding*. Examples of this are the emotions associated with the attachment of the parents to their young, and the attachment of the young to their parents.

6. The current mood state can affect the *cognitive evaluation of events or memories* (see Oatley and Jenkins 1996). This may facilitate continuity in the interpretation of the reinforcing value of events in the environment. A hypothesis that backprojections from parts of the brain involved in emotion such as the orbitofrontal cortex and amygdala implement this is described in *The Brain and Emotion*.

7. Emotion may facilitate the *storage of memories*. One way 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 storing many details of the prevailing situation when a strong reinforcer is delivered may be useful in generating appropriate behavior in situations with some similarities in the future. This function may be implemented by the relatively nonspecific 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 Chapter 4 and Rolls and Treves 1998). 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. A third way 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 for perceptual representations or analyzers to be built which are different from each other if they are associated with different reinforcers, and for these to be less likely to be built if they have no association with reinforcement. 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 by Rolls and Treves (1998).

8. Another function of emotion is that by enduring for minutes or longer after a reinforcing stimulus has occurred, it may help to produce *persistent and continuing motivation and direction of behavior*, to help achieve a goal or goals.

9. Emotion may trigger the *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 (Rolls and Treves 1998).

4. Reward, Punishment and Emotion in Brain Design: an Evolutionary Approach

The theory of the functions of emotion is further developed in Chapter 10. Some of the points made help to elaborate greatly on 3.2 above. In Chapter 10, the fundamental question of why we and other animals are built to use rewards and punishments to guide or determine our behavior is considered. Why are we built to have emotions, as well as motivational states? Is there any reasonable alternative around which evolution could have built complex animals? In this section I outline several types of brain design, with differing degrees of complexity, and suggest that evolution can operate to influence action with only some of these types of design.

4.1 Taxes

A simple design principle is to incorporate mechanisms for *taxes* into the design of organisms. Taxes consist at their simplest of orientation towards stimuli in the environment, for example the bending of a plant towards light which results in maximum light collection by its photosynthetic surfaces. (When just turning rather than locomotion is possible, such responses are called tropisms.) With locomotion possible, as in animals, taxes include movements towards sources of nutrient, and movements away from hazards such as very high temperatures. The design principle here is that animals have through a process of natural selection built receptors for certain dimensions of the wide range of stimuli in the environment, and have

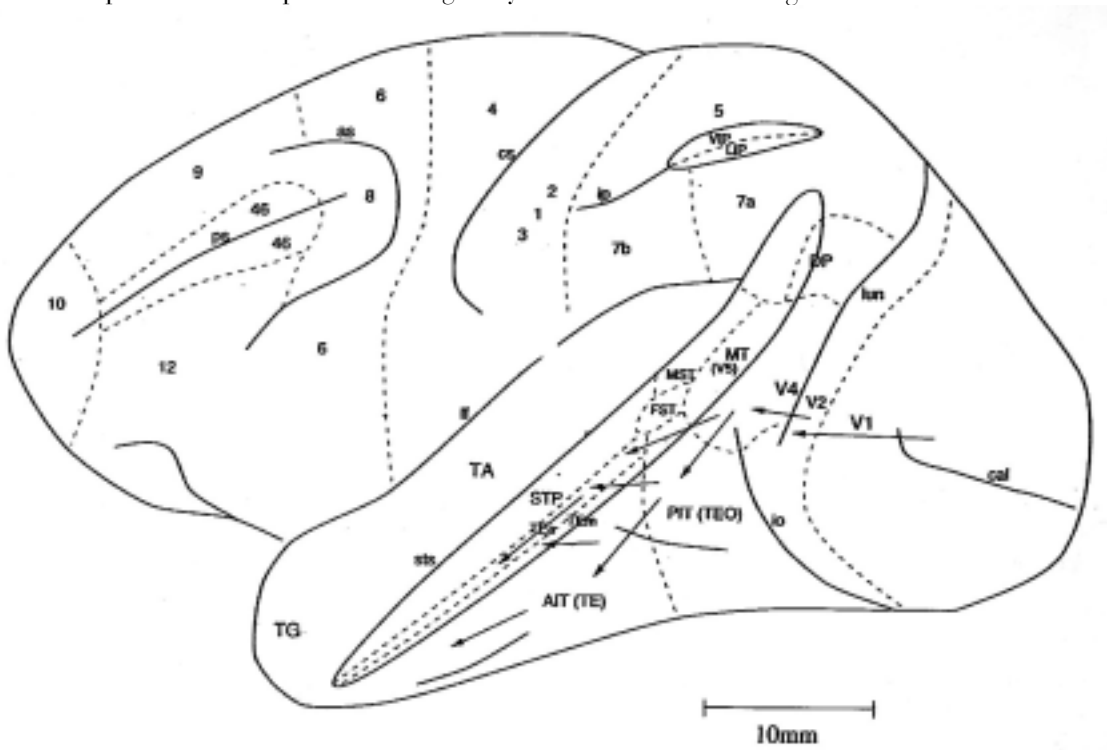

Fig. 3. Some of the pathways involved in emotion described in the text are shown on this lateral view of the brain of the macaque monkey. Connections from the primary taste and olfactory cortices to the orbitofrontal cortex and amygdala are shown. Connections are also shown in the 'ventral visual system' from V1 to V2, V4, the inferior temporal visual cortex, etc., with some connections reaching the amygdala and orbitofrontal cortex. In addition, connections from the somatosensory cortical areas 1, 2 and 3 that reach the orbitofrontal cortex directly and via the insular cortex, and that reach the amygdala via the insular cortex, are shown. as, arcuate sulcus; cal, calcarine sulcus; cs, central sulcus; lf, lateral (or Sylvian) fissure; lun, lunate sulcus; ps, principal sulcus; io, inferior occipital sulcus; ip, intraparietal sulcus (which has been opened to reveal some of the areas it contains); sts, superior temporal sulcus (which has been opened to reveal some of the areas it contains). AIT, anterior inferior temporal cortex; FST, visual motion processing area; LIP, lateral intraparietal area; MST, visual motion processing area; MT, visual motion processing area (also called V5); PIT, posterior inferior temporal cortex; STP, superior temporal plane; TA, architectonic area including auditory association cortex; TE, architectonic area including high order visual association cortex, and some of its subareas TEa and TEm; TG, architectonic area in the temporal pole; V1 – V4, visual areas 1 – 4; VIP, ventral intraparietal area; TEO, architectonic area including posterior visual association cortex. The numerals refer to architectonic areas, and have the following approximate functional equivalence: 1, 2, 3, somatosensory cortex (posterior to the central sulcus); 4, motor cortex; 5, superior parietal lobule; 7a, inferior parietal lobule, visual part; 7b, inferior parietal lobule, somatosensory part; 6, lateral premotor cortex; 8, frontal eye field; 12, part of orbitofrontal cortex; 46, dorsolateral prefrontal cortex. (From *The Brain and Emotion*, Fig. 4. 1.)

linked these receptors to mechanisms for particular responses in such a way that the stimuli are approached or avoided.

4.2 Reward and punishment

As soon as we have approach toward stimuli at one end of a dimension (e.g. a source of nutrient) and away from stimuli at the other end of the dimension (in this case lack of nutrient), we can start to wonder when it is appropriate to introduce the terms rewards and punishers for the stimuli at the different ends of the dimension. By convention, if the response consists of a fixed reaction to obtain the stimulus (e.g. locomotion up a chemical gradient), we shall call this a taxis, not a reward. On the other hand, if an arbitrary operant response can be performed by the animal in order to approach the stimulus, then we will call this rewarded behavior, and the stimulus the animal works to obtain is a reward. (The operant response can be thought of as any arbitrary action the animal will perform to obtain the stimulus.) This criterion, of an arbitrary operant response, is often tested by bidirectionality. For example, if a rat can be trained to either raise or lower its tail, in order to obtain a piece of food, then we can be sure that there is no fixed relationship between the stimulus (e.g. the sight of food) and the response, as there is in a taxis.

The role of natural selection in this process is to guide animals to build sensory systems that will respond to dimensions of stimuli in the natural environment along which actions can lead to better ability to pass genes on to the next generation, that is to increased fitness. The animals must be built by such natural selection to make responses that will enable them to obtain more rewards, that is to work to obtain stimuli that will increase their fitness. Correspondingly, animals must be built to make responses that will enable them to escape from, or learn to avoid, stimuli that will reduce their fitness. There are likely to be many dimensions of environmental stimuli along which responses can alter fitness. Each of these dimensions may be a separate reward-punishment dimension. An example of one of these dimensions might be food reward. It increases fitness to be able to sense nutrient need, to have sensors that respond to the taste of food, and to perform behavioral responses to obtain such reward stimuli when in that need or motivational state. Similarly, another dimension is water reward, in which the taste of water becomes rewarding when there is body fluid depletion (see Chapter 7).

With many reward/punishment dimensions for which actions may be performed (see Table 10.1 of *The Brain and Emotion* for a non-exhaustive list!), a selection mechanism for actions performed is needed. In this sense, rewards and punishers provide a *common currency* for inputs to response selection mechanisms. Evolution must set the magnitudes of each of the different reward systems so that each will be chosen for action in such a way as to maximize overall fitness. Food reward must be chosen as the aim for action if a nutrient is depleted; but water reward as a target for action must be selected if current water depletion poses a greater threat to fitness than the current food depletion. This indicates that each reward must be carefully calibrated by evolution to have the right value in the common currency for the competitive selection process. Other types of behavior, such as sexual behavior, must

Fig. 4. Diagrammatic representation of some of the connections described in the text. V1 - striate visual cortex. V2 and V4 - cortical visual areas. In primates, sensory analysis proceeds in the visual system as far as the inferior temporal cortex and the primary gustatory cortex; beyond these areas, for example in the amygdala and orbitofrontal cortex, the hedonic value of the stimuli, and whether they are reinforcing or are associated with reinforcement, is represented (see text). The gate function refers to the fact that in the orbitofrontal cortex and hypothalamus the responses of neurons to food are modulated by hunger signals. After *The Brain and Emotion*, Fig. 4.

be selected sometimes, but probably less frequently, in order to maximise fitness (as measured by gene transmission into the next generation). Many processes contribute to increasing the chances that a wide set of different environmental rewards will be chosen over a period of time, including not only need-related satiety mechanisms which decrease the rewards within a dimension, but also sensory-specific satiety mechanisms, which facilitate switching to another reward stimulus (sometimes within and sometimes outside the same main dimension), and attraction to novel stimuli. Finding novel stimuli rewarding, is one way that organisms are encouraged to explore the multidimensional space in which their genes are operating.

The above mechanisms can be contrasted with typical engineering design. In the latter, the engineer defines the requisite function and then produces special-purpose design features which enable the task to be performed. In the case of the animal, there is a multidimensional space within which many optimisations to increase fitness must be performed. The solution is to evolve reward / punishment systems tuned to each dimension in the environment which can increase fitness if the animal performs the appropriate actions. Natural selection guides evolution to find these dimensions. In contrast, in the engineering design of a robot arm, the robot does not need to tune itself to find the goal to be performed. The contrast is between design by evolution which is 'blind' to the purpose of the animal, and design by a designer who specifies the job to be performed (cf Dawkins 1986). Another contrast is that for the animal the space will be high-dimensional, so that the most appropriate reward for current behavior (taking into account the costs of obtaining each reward) needs to be selected, whereas for the robot arm, the function to perform at any one time is specified by the designer. Another contrast is that the behavior (the operant response) most appropriate to obtain the reward must be selected by the animal, whereas the movement

to be made by the robot arm is specified by the design engineer.

The implication of this comparison is that operation by animals using reward and punishment systems tuned to dimensions of the environment that increase fitness provides a mode of operation that can work in organisms that evolve by natural selection. It is clearly a natural outcome of Darwinian evolution to operate using reward and punishment systems tuned to fitness-related dimensions of the environment, if arbitrary responses are to be made by the animals, rather than just preprogrammed movements such as tropisms and taxes. Is there any alternative to such a reward / punishment based system in this evolution by natural selection situation? It is not clear that there is, if the genes are efficiently to control behavior. The argument is that genes can specify actions that will increase fitness if they specify the goals for action. It would be very difficult for them in general to specify in advance the particular responses to be made to each of a myriad of different stimuli. This may be why we are built to work for rewards, avoid punishers, and to have emotions and needs (motivational states). This view of brain design in terms of reward and punishment systems built by genes that gain their adaptive value by being tuned to a goal for action offers I believe a deep insight into how natural selection has shaped many brain systems, and is a fascinating outcome of Darwinian thought.

This approach leads to an appreciation that in order to understand brain mechanisms of emotion and motivation, it is necessary to understand how the brain decodes the reinforcement value of primary reinforcers, how it performs stimulus-reinforcement association learning to evaluate whether a previously neutral stimulus is associated with reward or punishment and is therefore a goal for action, and how the representations of these neutral sensory stimuli are appropriate as an input to such stimulus-reinforcement learning mechanisms. It is to these fundamental issues, and their relevance to brain design, that much of the book is devoted. How these processes are performed by the brain is considered for emotion in Chapter 4, for feeding in Chapter 2, for drinking in Chapter 7, and for sexual behavior in Chapter 8.

5. The Neural Bases of Emotion (Chapter 4)

Some of the main brain regions implicated in emotion will now be considered, in the light of this theory of the nature and functions of emotion. The description here is abbreviated, focussing on the main conceptual points. More detailed accounts of the evidence, and references to the original literature, are provided by Rolls (1990; 1992b; 1996; 1999). The brain regions discussed include the amygdala and orbitofrontal cortex. Some of these are indicated in Figs. 3 and 4. 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).

5.1 Overview

A schematic diagram introducing some of the concepts useful for understanding the neural bases of emotion is provided in Fig. 2, and some of the pathways are shown on a lateral view of a primate brain in Fig. 3 and schematically in Fig. 4.

5.1.1. Primary, unlearned, rewards and punishers. For primary reinforcers, the reward decoding may occur only after several stages of processing, as in the primate taste system, in which reward is decoded only after the primary taste cortex. By decoding I mean making explicit some aspect of the stimulus or event in the firing of neurons. A decoded representation is one in which the information can be read easily, for example by taking a sum of the synaptically weighted firing of a population of neurons. This is described in the Appendix, together with the type of learning important in many learned emotional responses, pattern association learning between a previously neutral, e.g. visual, stimulus and a primary reinforcer such as a pleasant touch. Processing as far as the primary taste cortex (see Fig. 4) represents what the taste is, whereas in the secondary taste cortex, in the orbitofrontal cortex, the reward value of taste is represented. This is shown by the fact that when the reward value of the taste of food is decreased by feeding it to satiety, the responses of neurons in the orbitofrontal cortex, but not at earlier stages of processing in primates, decrease their responses as the reward value of the food decreases (as described in Chapter 2: see also Rolls 1997). The architectural principle for the taste system in primates is that there is one main taste information processing stream in the brain, via the thalamus to the primary taste cortex, and the information about the identity of the taste in the primary cortex is not contaminated with modulation by how good the taste is, produced earlier in sensory processing. This enables the taste representation in the primary cortex to be used for purposes which are not reward-dependent. One example might be learning where a particular taste can be found in the environment, even when the primate is not hungry so that the taste is not pleasant.

Another primary reinforcer, the pleasantness of touch, is represented in another part of the orbitofrontal cortex, as shown by observations that the orbitofrontal cortex is much more activated (measured with functional magnetic resonance imaging, fMRI) by pleasant than neutral touch than is the primary somatosensory cortex (Francis et al. 1999) (see Fig. 4). Although pain may be decoded early in sensory processing in that it utilizes special receptors and pathways, some of the affective aspects of this primary negative reinforcer are represented in the orbitofrontal cortex, in that damage to this region reduces some of the affective aspects of pain in humans.

5.1.2. The representation of potential secondary (learned) reinforcers. For potential secondary reinforcers (such as the sight of a particular object or person), analysis goes up to the stage of invariant object representation (in vision, the inferior temporal visual cortical areas, see Wallis and Rolls 1997 and Figs. 3 and 4) before reward and punishment associations are learned. The utility of invariant representations is to enable correct generalisation to other instances (e.g. views, sizes) of the same or similar objects, even when a reward or punishment has been associated with one instance previously. The representation of the object is (appropriately)

in a form which is ideal as an input to pattern associators which allow the reinforcement associations to be learned. The representations are appropriately encoded in that they can be decoded in a neuronally plausible way (e.g., using a synaptically weighted sum of the firing rates, i.e., inner product decoding as described in the Appendix); they are distributed so allowing excellent generalisation and graceful degradation; and they have relatively independent information conveyed by different neurons in the ensemble, providing very high capacity and allowing the information to be read off very quickly, in periods of 20-50 ms (see Rolls and Treves 1998, Chapter 4 and the Appendix). The utility of representations of objects that are independent of reward associations (for vision in the inferior temporal cortex) is that they can be used for many functions independently of the motivational or emotional state. These functions include recognition, recall, forming new memories of objects, episodic memory (e.g., to learn where a food is located, even if one is not hungry for the food at present), and short term memory (see Rolls and Treves 1998).

An aim of processing in the ventral visual system is to help select the goals (e.g., objects with reward or punishment associations) for actions. I thus do not concur with Milner and Goodale (1995) that the dorsal visual system is for the control of action, and the ventral visual system is for "perception" (e.g., perceptual and cognitive representations). The ventral visual system projects via the inferior temporal visual cortex to the amygdala and orbitofrontal cortex, which then determine using pattern association the reward or punishment value of the object, as part of the process of selecting which goal is appropriate for action. Some of the evidence for this described in Chapter 4 is that large lesions of the temporal lobe (which damage the ventral visual system and some of its outputs, such as the amygdala) produce the Kluver-Bucy syndrome, in which monkeys select objects indiscriminately, independently of their reward value, and place them in their mouths. The dorsal visual system helps with executing those actions, for example, with grasping the hand appropriately to pick up a selected object. (This type of sensori-motor operation is often performed implicitly, i.e. without conscious awareness.) Insofar as explicit planning concerning future goals and actions requires knowledge of objects and their reward or punishment associations, it is the ventral visual system that provides the appropriate visual input.

In non-primates, including, for example, rodents, the design principles may involve less sophisticated features, because the stimuli being processed are simpler. For example, view invariant object recognition is probably much less developed in non-primates: the recognition that is possible is based more on physical similarity in terms of texture, colour, simple features etc. (see Rolls and Treves 1998, section 8.8). It may be because there is less sophisticated cortical processing of visual stimuli in this way that other sensory systems are also organised more simply, for example, with some (but not total, only perhaps 30%) modulation of taste processing by hunger early in sensory processing in rodents (see Scott et al. 1995). Moreover, although it is usually appropriate to have emotional responses to well-processed objects (e.g., the sight of a particular person), there are instances, such as a loud noise or a pure tone associated with punishment, where it may be possible to tap off a sensory representation early in sensory processing that can be used to produce emotional responses. This may occur in rodents, where the subcortical auditory system provides afferents to the amygdala (see Chapter 4 on emotion).

Especially in primates, the visual processing in emotional and social behavior requires sophisticated representation of individuals, and for this there are many neurons devoted to face processing (see Wallis and Rolls 1997). In macaques, many of these neurons are found in areas TEa and TEm in the ventral lip of the anterior part of the superior temporal sulcus. In addition, there is a separate system that encodes facial gesture, movement, and view, as all are important in social behavior, for interpreting whether specific individuals, with their own reinforcement associations, are producing threats or appeasements. In macaques, many of these neurons are found in the cortex in the depths of the anterior part of the superior temporal sulcus.

5.1.3. Stimulus-reinforcement association learning. After mainly unimodal processing to the object level, sensory systems then project into convergence zones. Those especially important for reward, punishment, emotion and motivation, are the orbitofrontal cortex and amygdala, where primary reinforcers are represented. These parts of the brain appear to be especially important in emotion and motivation not only because they are the parts of the brain where the primary (unlearned) reinforcing value of stimuli is represented in primates, but also because they are the regions that learn pattern associations between potential secondary reinforcers and primary reinforcers. They are thus the parts of the brain involved in learning the emotional and motivational value of stimuli.

5.1.4. Output systems. The orbitofrontal cortex and amygdala have connections to output systems through which different types of emotional response can be produced, as illustrated schematically in Fig. 2. The outputs of the reward and punishment systems must be treated by the action system as being the goals for action. The action systems must be built to try to maximise the activation of the representations produced by rewarding events and to minimise the activation of the representations produced by punishers or stimuli associated with punishers. Drug addiction produced by psychomotor stimulants such as amphetamine and cocaine can be seen as activating the brain at the stage where the outputs of the amygdala and orbitofrontal cortex, which provide representations of whether stimuli are associated with rewards or punishers, are fed into the ventral striatum and other parts of the basal ganglia as goals for the action system.

After this overview, a summary of some of the points made about some of the neural systems involved in emotion discussed in *The Brain and Emotion* follows.

5.2 The Amygdala

5.2.1. Connections and neurophysiology. Some of the connections of the primate amygdala are shown in Figs. 3 and 4 (see further *The Brain and Emotion*, Figs. 4.11 and 4.12). It receives information about primary reinforcers (such as taste and touch). It also receives inputs about stimuli (e.g., visual ones) that can be associated by learning with primary reinforcers. Such inputs come mainly from the inferior temporal

visual cortex, the superior temporal auditory cortex, the cortex of the temporal pole, and the cortex in the superior temporal sulcus. These inputs in primates thus come mainly from the higher stages of sensory processing in the visual (and auditory) modalities, and not from early cortical processing areas.

Recordings from single neurons in the amygdala of the monkey have shown that some neurons do respond to visual stimuli, and with latencies somewhat longer than those of neurons in the temporal cortical visual areas, consistent with the inputs from the temporal lobe visual cortex; and in some cases the neurons discriminate between reward-related and punishment-associated visual objects (see Rolls 1999). The crucial site of the stimulus-reinforcement association learning which underlies the responses of amygdala neurons to learned reinforcing stimuli is probably within the amygdala itself, and not at earlier stages of processing, for neurons in the inferior temporal cortical visual areas do not reflect the reward associations of visual stimuli, but respond to visual stimuli based on their physical characteristics (see Rolls 1990; 1999). The association learning in the amygdala may be implemented by associatively modifiable synapses (see Rolls and Treves 1998) from visual and auditory neurons onto neurons receiving inputs from taste, olfactory or somatosensory primary reinforcers. Consistent with this, Davis (1992) has found in the rat that at least one type of associative learning in the amygdala can be blocked by local application to the amygdala of a NMDA receptor blocker, which blocks long-term potentiation (LTP), a model of the synaptic changes which underlie learning (see Rolls and Treves 1998). Consistently, the learned incentive (conditioned reinforcing) effects of previously neutral stimuli paired with rewards are mediated by the amygdala acting through the ventral striatum is that amphetamine injections into the ventral striatum enhanced the effects of a conditioned reinforcing stimulus only if the amygdala was intact (see Everitt and Robbins 1992). The lesion evidence in primates is also consistent with a function of the amygdala in reward and punishment-related learning, for amygdala lesions in monkeys produce tameness, a lack of emotional responsiveness, excessive examination of objects, often with the mouth, and eating of previously rejected items such as meat. There is evidence that amygdala neurons are involved in these processes in primates, for amygdala lesioning with ibotenic acid impairs the processing of reward–related stimuli, in that when the reward value of a set of foods was decreased by feeding it to satiety (i.e. sensory–specific satiety), monkeys still chose the visual stimuli associated with the foods with which they had been satiated (Malkova et al. 1997).

Further evidence that the primate amygdala does process visual stimuli derived from high order cortical areas and of importance in emotional and social behavior is that a population of amygdala neurons has been described that responds primarily to faces (Leonard et al. 1985; see also Rolls 1992a ; 1992b; 1999). Each of these neurons responds to some but not all of a set of faces, and thus across an ensemble conveys information about the identity of the face. These neurons are found especially in the basal accessory nucleus of the amygdala (Leonard et al. 1985), a part of the amygdala that develops markedly in primates (Amaral et al. 1992). This part of the amygdala receives inputs from the temporal cortical visual areas in which populations of neurons respond to the identity of faces, and to face expression (see Rolls and Treves 1998; Wallis and Rolls 1997). This is probably part of a system which has evolved for the rapid and reliable identification of individuals from their faces, and of facial expressions, because of their importance in primate social behavior (see Rolls 1992a; 1999).

Although Le Doux's (1992; 1994; 1996) model of emotional learning emphasizes subcortical inputs to the amygdala for conditioned reinforcers, this applies to very simple auditory stimuli (such as pure tones). In contrast, a visual stimulus will normally need to be analyzed to the object level (to the level e.g., of face identity, which requires cortical processing) before the representation is appropriate for input to a stimulus-reinforcement evaluation system such as the amygdala or orbitofrontal cortex. Similarly, it is typically to complex auditory stimuli (such as a particular person's voice, perhaps making a particular statement) that emotional responses are elicited. The point here is that *emotions are usually elicited to environmental stimuli analyzed to the object level (including other organisms), and not to retinal arrays of spots or pure tones*. Thus cortical processing to the object level is required in most normal emotional situations, and these cortical object representations are projected to reach multimodal areas such as the amygdala and orbitofrontal cortex where the reinforcement label is attached using stimulus-reinforcer pattern association learning to the primary reinforcers represented in these areas. Thus while LeDoux's (1996) approach to emotion focusses mainly on fear responses to simple stimuli such as tones implemented considerably by subcortical processing, *The Brain and Emotion* considers how in primates including humans most stimuli, which happen to be complex and require cortical processing, produce a wide range of emotions; and in doing this addresses the functions in emotion of the highly developed temporal and orbitofrontal cortical areas of primates including humans, areas which are much less developed in rodents. When the learned association between a visual stimulus and reinforcement was altered by reversal (so that the visual stimulus formerly associated with juice reward became associated with aversive saline and vice versa), it was found that 10 of 11 neurons did not reverse their responses (and for the other neuron the evidence was not clear, see Rolls 1992b). In contrast, neurons in the orbitofrontal cortex do show very rapid reversal of their responses in visual discrimination reversal. It has accordingly been proposed that during evolution with the great development of the orbitofrontal cortex in primates, it (as a rapid learning system) is involved especially when repeated relearning and re-assessment of stimulus-reinforcement associations is required, as described below, rather than during initial learning, in which the amygdala may be involved.

Some amygdala neurons that respond to rewarding visual stimuli also respond to relatively novel visual stimuli; this may implement the reward value which novel stimuli have (see Rolls 1999).

The outputs of the amygdala (Amaral et al. 1992) include projections to the hypothalamus and also directly to the autonomic centres in the medulla oblongata, providing one route for cortically processed signals to reach the brainstem and produce autonomic responses. A further interesting output of the amygdala is to the ventral striatum including the nucleus accumbens, for via this route information processed in the

amygdala could gain access to the basal ganglia and thus influence motor output (see Fig. 2 and Everitt and Robbins 1992). In addition, mood states could affect cognitive processing via the amygdala's direct backprojections to many areas of the temporal, orbitofrontal, and insular cortices from which it receives inputs.

5.2.2. Human neuropsychology of the amygdala. Extending the findings on neurons in the macaque amygdala that responded selectively for faces and social interactions (Leonard et al, 1995; Brothers and Ring, 1993), Young et al. (1995; 1996) have described a patient with bilateral damage or disconnection of the amygdala who was impaired in matching and identifying facial expression but not facial identity. Adolphs et al. (1994) also found facial expression but not facial identity impairments in a patient with bilateral damage to the amygdala. Although in studies of the effects of amygdala damage in humans greater impairments have been reported with facial or vocal expressions of fear than with some other expressions (Adolphs et al. 1994; Scott et al. 1997), and in functional brain imaging studies greater activation may be found with certain classes of emotion–provoking stimuli (e.g., those that induce fear rather than happiness, Morris et al. 1996), I suggest in *The Brain and Emotion* that it is most unlikely that the amygdala is specialised for the decoding of only certain classes of emotional stimuli, such as fear. This emphasis on fear may be related to the research in rats on the role of the amygdala in fear conditioning (LeDoux 1992; 1994). Indeed, it is quite clear from single neuron studies in non–human primates that some amygdala neurons are activated by rewarding and others by punishing stimuli (Ono and Nishijo 1992; Rolls 1992a; 1992b; Sanghera et al. 1979; Wilson and Rolls 1993), and others by a wide range of different face stimuli (Leonard et al. 1985). Moreover, lesions of the macaque amygdala impair the learning of both stimulus–reward and stimulus–punisher associations. Further, electrical stimulation of the macaque and human amygdala at some sites is rewarding, and humans report pleasure from stimulation at such sites (Halgren 1992; Rolls 1975; 1980; Sem-Jacobsen 1968; 1976). Thus any differences in the magnitude of effects between different classes of emotional stimuli which appear in human functional brain imaging studies (Davidson and Irwin 1999; Morris et al. 1996) or even after amygdala damage (Adolphs et al. 1994; Scott et al. 1997) should not be taken to show that the human amygdala is involved in only some emotions. Indeed, in current fMRI studies we are finding that the human amygdala is activated perfectly well by the pleasant taste of a sweet (glucose) solution (in the continuation of studies reported by Francis et al. 1999), showing that reward-related primary reinforcers do activate the human amygdala.

5.3. The Orbitofrontal Cortex

5.3.1. Connections and neurophysiology of the orbitofrontal cortex. The orbitofrontal cortex receives inputs from the primary taste cortex in the insula and frontal operculum, the primary olfactory (pyriform) cortex, and the primary somatosensory cortex (see Figs. 3 and 4). Neurons in the orbitofrontal cortex, which contains the secondary and tertiary taste and olfactory cortical areas, respond to the reward value of taste and olfactory stimuli, in that they respond to the taste and odor of food only when the monkey is hungry. Moreover, sensory-specific satiety for the reward of the taste or the odor of food is represented in the orbitofrontal cortex, and is computed here at least for the taste of food. In addition, some orbitofrontal cortex neurons combine taste and olfactory inputs to represent flavor, and the principle by which this flavor representation is formed is by olfactory-to-taste association learning. Inputs from the oral somatosensory system produce a representation of the fat content of food in the mouth (Rolls et al, 1999; the activation of these neurons is also decreased by feeding to satiety), and more generally of food texture, and also of astringency. FMRI studies in humans show that the orbitofrontal cortex is also activated more by pleasant touch than by neutral touch, relative to the somatosensory cortex (Francis et al. 1999). Thus, there is a rich representation of primary (unlearned) reinforcers in the orbitofrontal cortex, including taste and somatosensory primary reinforcers, and of odor, which is in this case partly secondary (learned). The representation is rich in that there is much information that can be easily read from the neuronal code (see Rolls and Treves 1998) about exactly which taste, touch, or odor is being delivered. It is important that reinforcers be represented in a way which encodes the details of which reinforcer has been delivered, for it is crucial that organisms work for the correct reinforcer as appropriate (e.g., for food when hungry, and for water when thirsty), and that they switch appropriately between reinforcers (using for example the principle of sensory-specific satiety, for which a representation of the sensory details of the reinforcer is needed).

The primate orbitofrontal cortex also receives inputs from the inferior temporal visual cortex, and is involved in stimulus-reinforcer association learning, in that neurons in it learn visual stimulus to taste reinforcer associations in as little as one trial. Moreover, and consistent with the effects of damage to the orbitofrontal cortex which impair performance on visual discrimination reversal, Go/NoGo tasks, and extinction tasks (in which the lesioned macaques continue to make behavioral responses to previously rewarded stimuli), orbitofrontal cortex neurons reverse visual stimulus reinforcer associations in as little as one trial. Moreover, a separate population of orbitofrontal cortex neurons responds only on non-reward trials (Thorpe et al. 1983). There is thus the basis in the orbitofrontal cortex for rapid learning and updating by relearning or reversing stimulus-reinforcer (sensory-sensory, e.g. visual to taste) associations. In the rapidity of its relearning / reversal, the primate orbitofrontal cortex may effectively replace and perform better some of the functions performed by the primate amygdala. In addition, some visual neurons in the primate orbitofrontal cortex respond to the sight of faces. These neurons are likely to be involved in learning which emotional responses are currently appropriate to particular individuals, and in making appropriate emotional responses given the facial expression (see Rolls 1996).

The evidence thus indicates that the primate orbitofrontal cortex is involved in the evaluation of primary reinforcers, and also implements a mechanism which evaluates whether a reward is expected, and generates a mismatch (evident as a firing of the non-reward neurons) if reward is not obtained when it is expected (Thorpe et al. 1983; Rolls 1990; 1996; 1999).

These neuronal responses provide further evidence that the orbitofrontal cortex is involved in emotional responses, particularly when these involve correcting previously learned reinforcement contingencies, in situations which include those usually described as involving frustration.

5.3.4. Human neuropsychology of the orbitofrontal cortex. It is of interest and potential clinical importance that a number of the symptoms of frontal lobe damage in humans appear to be related to this type of function, of altering behavior when stimulus-reinforcement associations alter. Thus, humans with ventral frontal lobe damage can show impairments in a number of tasks in which an alteration of behavioral strategy is required in response to a change in environmental reinforcement contingencies (Damasio 1994; see Rolls 1990; 1996; 1999). Some of the personality changes that can follow frontal lobe damage may be related to a similar type of dysfunction. For example, the euphoria, irresponsibility, lack of affect, and lack of concern for the present or future which can follow frontal lobe damage may also be related to a dysfunction in altering behavior appropriately in response to a change in reinforcement contingencies.

Some of the evidence that supports this hypothesis is that when the reinforcement contingencies unexpectedly reversed in a visual discrimination task performed for points, patients with ventral frontal lesions made more errors in the reversal (or in a similar extinction) task, and completed fewer reversals, than control patients with damage elsewhere in the frontal lobes or in other brain regions (Rolls et al. 1994). The impairment correlated highly with the socially inappropriate or disinhibited behavior of the patients, and also with their subjective evaluation of the changes in their emotional state since the brain damage. The patients were not impaired in other types of memory task, such as paired associate learning. Bechara and colleagues also have findings which are consistent with these in patients with frontal lobe damage when they perform a gambling task (Bechara et al. 1994; 1997; 1996; see also Damasio 1994). The patients could choose cards from two piles. The patients with frontal damage were more likely to choose cards from a pile which gave rewards with a reasonable probability but also had occasional very heavy penalties. The net gains from this pile were lower than from the other pile. In this sense, the patients were not affected by the negative consequences of their actions: they did not switch from the pile of cards which though providing significant rewards also led to large punishments being incurred.

To investigate the possible significance of face-related inputs to the orbitofrontal visual neurons described above, the responses of the same patients to faces were also tested. Tests of face (and also voice) expression decoding were included, because these are ways in which the reinforcing quality of individuals are often indicated. The identification of facial and vocal emotional expression were found to be impaired in a group of patients with ventral frontal lobe damage who had socially inappropriate behavior (Hornak et al. 1996). The expression identification impairments could occur independently of perceptual impairments in facial recognition, voice discrimination, or environmental sound recognition. This provides a further basis for understanding the functions of the orbitofrontal cortex in emotional and social behavior, in that processing of some of the signals normally used in emotional and social behavior is impaired in some of these patients. Imaging studies in humans show that parts of the prefrontal cortex can be activated when mood changes are elicited, but it is not established that some areas are concerned only with positive or only with negative mood (Davidson and Irwin 1999). Indeed this seems unlikely in that the neurophysiological studies show that different individual neurons in the orbitofrontal cortex respond to either some rewarding or some punishing stimuli, and that these neurons can be intermingled.

5.4. Output systems for Emotion

I distinguish three main output systems for emotion, illustrated schematically in Fig. 2. Consideration of these different output systems helps to elucidate the functions of emotion. The first system produces autonomic and endocrine outputs, important in optimizing the body state for different types of action, including fight, flight, feeding and sex. The pathways include brainstem and hypothalamic connections for autonomic and endocrine responses to unlearned stimuli, and neural systems in the amygdala and orbitofrontal cortex for similar responses to learned stimuli. Operating at the same level as this system are brainstem pathways for unlearned responses to stimuli, including reflexes.

The second and third routes are for actions, that is, arbitrary behavioral responses, performed to obtain, avoid or escape from reinforcers. The first action route is via the brain systems that have been present in nonhuman primates such as monkeys, and to some extent in other mammals, for millions of years, and can operate implicitly. These systems include the amygdala and, particularly well-developed in primates, the orbitofrontal cortex. They provide information about the possible goals for action based on their decoding of primary reinforcers taking into account the current motivational state, and on their decoding of whether stimuli have been associated by previous learning with reinforcement. A factor which affects the computed reward value of the stimulus is whether that reward has been received recently. If it has been received recently but in small quantity, this may increase the reward value of the stimulus. This is known as incentive motivation or the "salted peanut" phenomenon. The adaptive value of such a process is that this positive feedback or potentiation of reward value in the early stages of working for a particular reward tends to lock the organism onto the behavior being performed for that reward. This makes action selection much more efficient in a natural environment, for constantly switching between different types of behavior would be very costly if all the different rewards were not available in the same place at the same time. The amygdala is one structure that may be involved in this increase in the reward value of stimuli early on in a series of presentations, in that lesions of the amygdala (in rats) abolish the expression of this reward incrementing process which is normally evident in the increasing rate of working for a food reward early on in a meal (Rolls and Rolls 1982). The converse of incentive motivation is sensory-specific satiety, in which receiving a reward for some longer time decreases the reward value of that stimulus, which has the adaptive function of facilitating switching to another reward stimulus.

After the reward value of the stimulus has been assessed in these ways, behavior is then initiated based on approach towards or withdrawal from the stimulus. A critical aspect of the

behavior produced by this type of system is that it is aimed directly towards obtaining a sensed or expected reward, by virtue of connections to brain systems such as the basal ganglia which are concerned with the initiation of actions (see Fig. 2). The expectation may of course involve behavior to obtain stimuli associated with reward, and the stimuli might even be present in a chain. The costs (or expected punishments) of the action must be taken into account. Indeed, in the field of behavioral ecology, animals are often thought of as performing optimally on some cost-benefit curve (see e.g. Krebs and Kacelnik 1991). Part of the value of having the computation expressed in this reward-minus-cost form is that there is then a suitable "currency", or net reward value, to enable the animal to select the behavior with highest current net reward gain (or minimal aversive outcome).

The second route for action to emotion-related stimuli in humans involves a computation with many "if...then" statements, to implement a plan to obtain a reward or to avoid a punisher. In this case, the reward may actually be *deferred* as part of the plan, which might involve not obtaining an immediate reward, but instead working to obtain a second more highly valued reward, if this is thought to be an optimal overall strategy in terms of resource use (e.g., time). In this case, syntax is required, because the many symbols (e.g., names of people) that are part of the plan must be correctly linked or bound. Such linking might be of the form: "If A does this, then B is likely to do this, and this will cause C to do this ...". The requirement of syntax for this type of planning implies that a language system in the brain is involved (see Fig. 2). (A language system is defined here as a system performing syntactic operations on symbols.) Thus the explicit language system in humans may allow working for deferred rewards by enabling use of an individual, one-off (i.e. one-time), plan appropriate for each situation. Another building block for such planning operations in the brain may be the type of short term memory in which the prefrontal cortex is involved. In non-human primates this short term memory might be for example of where in space a response has just been made. A development of this type of short term response memory system in humans to enable multiple short term memories to be held active correctly, preferably with the temporal order of the different items in the short term memory coded correctly, may be another building block for the multiple step "if .... then" type of computation forming a multiple step plan. Such short term memories are implemented in the (dorsolateral and inferior convexity) prefrontal cortex of non-human primates and humans (see Goldman-Rakic 1996; Petrides 1996), and the impairment of planning produced by prefrontal cortex damage (see Shallice and Burgess 1996) may be due to damage to a system of the type just described founded on short term or working memory systems.

While discussing the prefrontal cortex, we should note that when Damasio (1994) suggests that reason and emotion are closely linked as processes because they may both be impaired in patients with frontal lobe damage, this could be a chance association because the brain damage frequently affects both the orbitofrontal and the more dorsolateral areas of the prefrontal cortex, which are adjacent. (Indeed, some evidence for a dissociation of the functions of these areas in some patients with more restricted damage is actually presented by Damasio (1994) on page 61, and by Bechara et al. (1998)). The alternative I propose in *The Brain and Emotion* (and in Rolls and Treves 1998 Chapters 7 and 10), is that the orbitofrontal cortex, which receives inputs about what stimuli are present (from the ventral visual system, and from the taste and somatosensory systems) allows the reinforcing value of stimuli to be evaluated, and is therefore involved in emotion; whereas in contrast the more dorsolateral prefrontal cortex receives inputs from the "where" parts of the (dorsal) visual system, and is concerned with planning and executing actions based on modules for which a foundation is provided by neural networks for short term, working, memory.

These three systems do not necessarily act as an integrated whole. Indeed, in so far as the implicit system may be for immediate goals and the explicit system is computationally appropriate for deferred longer term goals, they will not always indicate the same action. Similarly, the autonomic system does not use entirely the same neural systems as those involved in actions, and therefore autonomic outputs will not always be an excellent guide to the emotional state of the animal, which the above arguments in any case indicate is not unitary, but has at least three different aspects (autonomic, implicit and explicit). Also, the costs and benefits and therefore the priorities that animals will place on achieving different goals will depend on the primary reinforcer involved. These arguments suggest that multiple measures are likely to be relevant when assessing the impact of different factors on welfare. It is likely to be important to measure not only autonomic changes, but also preference rankings between different reinforcers, and how hard different reinforcers will be worked for.

5.5. The role of dopamine in reward, addiction, and the initiation of action

The dopamine pathways in the brain arise in the midbrain, projecting from the A10 cell group in the ventral tegmental area to the nucleus accumbens, orbitofrontal cortex, and some other cortical areas; and from the A9 cell group to the striatum (which is part of the basal ganglia, see Cooper et al. 1996; Rolls 1999). Dopamine is involved in the reward produced by stimulation of some brain sites, notably the ventral tegmental area where the dopamine cell bodies are located. This self-stimulation depends on dopamine release in the nucleus accumbens. Self-stimulation at some other sites does not depend on dopamine. The self-administration of psychomotor stimulants such as amphetamine and cocaine depends on the activation of a dopaminergic system in the nucleus accumbens, which receives inputs from the amygdala and orbitofrontal cortex.

The dopamine release produced by these behaviors may be rewarding because it is influencing the activity of an amygdalo-striatal (and in primates also possibly orbitofrontal-striatal) system involved in linking the amygdala and orbitofrontal cortex, which can learn stimulus-reinforcement associations, to output systems. In a whole series of studies, Robbins et al. (1989) showed that conditioned reinforcers (for food) increase the release of dopamine in the nucleus accumbens and that dopamine-depleting lesions of the nucleus accumbens attenuate the effect of conditioned (learned) incentives on behavior.

Although the majority of the studies have focussed on rewarded behavior, there is also evidence that dopamine can be released by stimuli that are aversive. For example, Rada et al. (1998) showed that dopamine was released in the nucleus accumbens when rats worked to escape from aversive hypothalamic stimulation (see also Hoebel 1997; Leibowitz and Hoebel 1998). Also, Gray et al. (1997) (see also Abercrombie et al. 1989; Thierry et al. 1976) describe evidence that dopamine can be released in the nucleus accumbens during stress, unavoidable foot shock, and in response to a light or tone associated by Pavlovian conditioning with foot shock which produces fear. Because of these findings, it is suggested that the release of dopamine is actually more related to the initiation of active behavioral responses, such as active avoidance of punishment, or working to obtain food, than to the delivery of reward *per se* or of stimuli that signal reward. Although the most likely process to enhance the release of dopamine in the ventral striatum is an increase in the firing of dopamine neurons, an additional possibility is the release of dopamine by a presynaptic influence on the dopamine terminals in the nucleus accumbens.

What signals could make dopamine neurons fire? Some of the inputs to the dopamine neurons in the midbrain come from the head of the caudate nucleus where a population of neurons starts to respond in relation to a tone or light signalling in a visual discrimination task that a trial is about to begin, and stops responding after the reward is delivered or as soon as a visual stimulus is shown which indicates that reward cannot be obtained on that trial and that saline will be obtained if a response is made (Rolls et al 1983; Rolls and Johnstone, 1992). Similar neurons are also found in the ventral striatum (Williams et al. 1993). The responses of midbrain dopamine neurons described by Schultz et al. (1995; 1996; 1998) are somewhat similar to these cue-related striatal neurons which appear to receive their input from the overlying prefrontal cortex, and it is suggested that this is because the dopamine neurons are influenced by these striatal neurons with activity related to the initiation of action.

On the basis of these types of evidence, the hypothesis is proposed that the activity of dopamine neurons and dopamine release is more related to the initiation of action or general behavioral activation, and the appropriate threshold setting within the striatum (see Chapter 4 section 4 and Rolls and Treves 1998), than to reward *per se*, or a teaching signal about reward (cf. Schultz et al. 1995; Houk et al. 1995). The investigation of Mirenowicz and Schultz (1996) did not address this issue directly in that it was when the monkey had to disengage from a trial and make no touch response when a stimulus associated with an aversive air puff was delivered that dopamine neurons generally did not respond, and the task was thus formally very similar to the Go/NoGo task of Rolls, Thorpe and Maddison (1983) in which they described similar neurons in the head of the caudate that responded when the monkey was engaged in the task. One way to test whether the release of dopamine in this system means "Go" rather than "reward" would be to investigate whether the dopamine neurons fire, and dopamine release occurs and is necessary for, behavior such as active avoidance of a strong punishing, arousing, stimulus. It is noted in any case that if the release of dopamine does turn out to be related to reward, then it apparently does not represent all the sensory specificity of a particular reward or goal for action. Indeed, one of the main themes of *The Brain and Emotion* is that there is clear evidence on how with exquisite detail rich representations of different types of primary reinforcer, including taste and somatosensory reinforcers, are decoded by and present in the orbitofrontal cortex and amygdala, and the structures to which they project including the lateral hypothalamus and ventral striatum (Williams et al. 1993). Further, the same brain systems implement stimulus-to-primary reinforcer learning. In contrast, it is doubtful whether reward *per se* is represented in the firing of dopamine neurons; and even if it is, they do not carry the full sensory quality of orbitofrontal cortex neurons; and must in any case be driven by inputs already decoded for reward vs punishment in the orbitofrontal cortex and amygdala.

Given that the ventral striatum has inputs from the orbitofrontal cortex as well as the amygdala, and that some primary rewards are represented in the orbitofrontal cortex, the dopaminergic effects of psychomotor stimulant drugs (such as amphetamine and cocaine) may produce their effects in part because they are facilitating transmission in a primary reward-to-action pathway which is currently biassed towards reward by the inputs to the ventral striatum. In addition, at least part of the reason that such drugs are addictive may be that they activate the brain at the stage of processing after the one at which reward or punishment associations have been learned, where the signal is normally interpreted by the system as indicating "select actions to achieve the goal of making these striatal neurons fire" (see Fig. 2 and Rolls 1999).

6. Role of Peripheral Factors in Emotion

The James-Lange theory postulates that certain stimuli produce bodily responses, including somatic and autonomic responses, and that it is the sensing of these bodily changes that gives rise to the *feeling* of emotion (James 1884; Lange 1885). This theory is encapsulated by the statement: "I feel frightened because I am running away". This theory has gradually been weakened by the following evidence: (1) There is not a particular pattern of autonomic responses that corresponds to every emotion. (2) Disconnection from the periphery (e.g. after spinal cord damage or damage to the sympathetic and vagus autonomic nerves) does not abolish behavioral signs of emotion or emotional feelings (see Oatley and Jenkins 1996). (3) Emotional intensity can be modulated by peripheral injections of, for example, adrenaline (epinephrine) which produce autonomic effects, but it is the cognitive state as induced by environmental stimuli, and not the autonomic state, that produces an emotion, and determines what the emotion is. (4) Peripheral autonomic blockade with pharmacological agents does not prevent emotions from being felt (Reisenzein 1983). The James-Lange theory, and theories which are closely related to it in supposing that feedback from parts of the periphery (such as the face or body, as in A.Damasio's (1994) somatic marker hypothesis), leads to emotional feelings, also have however the major weakness that they do not give an adequate account of which stimuli produce the peripheral change that is postulated to eventually lead to emotion. That is, these theories do not provide an account of the rules by which only some environmental stimuli produce emotions, or how neurally only such

stimuli produce emotions.

Another problem with such bodily mediation theories is that introducing bodily responses, and then sensing of these body responses, into the chain by which stimuli come to elicit emotions would introduce noise into the system. Damasio (1994) may partially circumvent this last problem in his theory by allowing central representations of somatic markers to become conditioned to bodily somatic markers, so that after the appropriate learning, a peripheral somatic change may not be needed. However, this scheme still suffers from noise inherent in producing bodily responses, in sensing them, and in conditioning central representations of the somatic markers to the bodily states. Even if Damasio were to argue that the peripheral somatic marker and its feedback can be bypassed using conditioning of a representation (in e.g., the somatosensory cortex) he would apparently still wish to argue that the activity in the somatosensory cortex is important for the emotion to be appreciated or to influence behavior. (Without this, the somatic marker hypothesis would vanish.) The prediction would apparently be that if an emotional response or decision were produced to a visual stimulus, this would necessarily involve activity in the somatosensory cortex or other brain region in which the "somatic marker" would be represented. Damasio (1994) actually sees bodily markers as helping to make emotional decisions because they perform a bodily integration of all the complex issues that may be leading to indecision in the conscious rational processing system of the brain. This prediction could be tested (for example, in patients with somatosensory cortex damage), but it seems most unlikely that an emotion produced by an emotion-provoking visual stimulus would *require* activity in the somatosensory cortex. Damasio in any case effectively sees computation by the body of what the emotional response should be as one way in which emotional decisions are taken. In this sense, Damasio (1994) suggests that we should take it as an error that the rational self takes decisions, and replace this with a system in which the body resolves the emotional decision. In contrast, the theory developed in *The Brain and Emotion* is that in humans both the implicit and the explicit systems can be involved in taking emotional decisions; that they do not necessarily agree as these two systems respectively perform computation of immediate rewards, and deferred longer-term rewards achievable by multistep planning; that peripheral factors are useful in preparing the body for action but do not take part in decisions; and that in any case the interesting part of emotional decisions is how the reward or punishment value of stimuli is decoded by the brain, and routed to action systems, which is what much of *The Brain and Emotion* is about.

7. Conclusions

Although this précis has focussed on the parts of the book about emotion, and rather little on those parts concerned with hunger, thirst, brain-stimulation reward, and sexual behavior, which provide complementary evidence, or on the issue of subjective feelings and emotion, some of the conclusions reached in the book are as follows, and comments on all aspects of the book are invited:

(1) Emotions can be considered as states elicited by reinforcers (rewards and punishers). This approach helps with understanding the functions of emotion, and with classifying different emotions (Chapter 3); and in understanding *what* information processing systems in the brain are involved in emotion, and *how* they are involved (Chapter 4).

(2) The hypothesis is developed that brains are designed around reward and punishment evaluation systems, because this is how genes can build a complex system that will produce appropriate but flexible behavior to increase fitness (Chapter 10). By specifying goals, rather than particular behavioral patterns of responses, genes leave much more open the possible behavioral strategies that might be required to increase fitness. This view of the evolutionarily adaptive value for genes to build organisms using reward and punishment decoding and action systems in the brain (leading thereby to brain systems for emotion and motivation) places this thinking squarely in line with that of Darwin.

(3) The importance of reward and punishment systems in brain design helps us to understand the significance and importance not only of emotion, but also of motivational behavior, which frequently involves working to obtain goals that are specified by the current state of internal signals to achieve homeostasis (see Chapter 2 on hunger and Chapter 7 on thirst) or that are influenced by internal hormonal signals (Chapter 8 on sexual behavior).

(4) In Chapters 2 (on hunger) and 4 (on emotion) some of what may be the fundamental architectural and design principles of the brain for sensory, reward, and punishment information processing in primates including humans is outlined. These architectural principles include the following:

For potential secondary reinforcers, cortical analysis is to the level of invariant object identification before reward and punishment associations are learned, and the representations produced in these sensory systems of objects are in the appropriate form for stimulus-reinforcer pattern association learning. This requirement can be seen as shaping the evolution of some sensory processing streams. The potential secondary reinforcers for emotional learning thus originate mainly from high order cortical areas, not from subcortical regions.

For primary reinforcers, the reward decoding may occur after several stages of processing, as in the primate taste system, in which reward is decoded only after the primary taste cortex.

In both cases this allows the use of the sensory information by a number of different systems, including brain systems for learning, independently of whether the stimulus is currently reinforcing, that is a goal for current behavior.

The reward value of primary and secondary reinforcers is represented in the orbitofrontal cortex and amygdala, where there is a detailed and information rich representation of taste, olfactory, somatosensory and visual rewarding (and punishing) stimuli.

Another design principle is that the outputs of the reward and punishment systems must be treated by the action system as being the goals for action. The action systems must be built to try to maximise the activation of the representations produced by rewarding events, and to minimise the activation of the representations produced by punishers or stimuli associated with punishers. Drug addiction produced by psychomotor stimulants such as amphetamine and cocaine can be seen as activating the brain at the stage where the outputs of the amygdala and orbitofrontal cortex, which provide representa-

tions of whether stimuli are associated with rewards or punishers, are fed into the ventral striatum as goals for the action system.

(5) Especially in primates, the visual processing in emotional and social behavior requires sophisticated representation of individuals, and for this there are many neurons devoted to invariant face identity processing. In addition, there is a separate system that encodes facial gesture, movement, and view. All are important in social behavior, for interpreting whether a particular individual, with his or her own reinforcement associations, is producing threats or appeasements.

(6) After mainly unimodal cortical processing to the object level, sensory systems then project into convergence zones. The orbitofrontal cortex and amygdala are especially important for reward and punishment, emotion and motivation, not only because they are the parts of the brain where in primates the primary (unlearned) reinforcing value of stimuli is represented, but also because they are the parts of the brain that perform pattern association learning between potential secondary reinforcers and primary reinforcers.

(7) The reward evaluation systems have tendencies to self-regulate, so that on average they can operate in a common currency which leads on different occasions, often depending on modulation by internal signals, to the selection of different rewards.

(8) A principle that assists the selection of different behaviors is sensory-specific satiety, which builds up when a reward is repeated for a number of minutes. A principle that helps behavior to lock on to one goal for at least a useful period is incentive motivation, the process by which there is potentiation early on in the presentation of a reward. There are probably simple neurophysiological bases for these time-dependent processes in the reward (as opposed to the early sensory) systems which involve neuronal habituation and facilitation respectively.

(9) With the advances made in the last 30 years in understanding the brain mechanisms involved in reward and punishment, and emotion and motivation, the basis for addiction to drugs is becoming clearer, and it is hoped that there is now a foundation for improving the understanding of depression and anxiety and their pharmacological and non-pharmacological treatment, in terms of the particular brain systems that are involved in these emotional states (Chapter 6).

(10) Although the architectural design principles of the brain to the stage of the representation of rewards and punishments seem apparent, it is much less clear how selection between the reward and punishment signals is made, how the costs of actions are taken into account, and how actions are selected. Some of the putative processes, including the principles of operation of the basal ganglia and the functions of dopamine, are outlined in Chapters 4 and 6, but much remains to be understood. The dopamine system may not code for reward; but instead its activity may be more related to the initiation of action, and feedback from the striatum.

(11) In addition to the implicit system for action selection, there is in humans also an explicit system that can use language to compute actions to obtain deferred rewards using a one-time plan. The language system allows one-off multistep plans which require the syntactic organisation of symbols to be formulated in order to obtain rewards and avoid punishments. There are thus two separate systems for producing actions to rewarding and punishing stimuli in humans. These systems may weight different courses of action differently, in that each can produce behavior for different goals (immediate vs deferred).

(12) It is possible that emotional feelings, part of the much larger problem of consciousness, arise as part of a process that involves thoughts about thoughts, which have the adaptive value of helping to correct multistep plans where credit assignment for each step is required. This is the approach described in Chapter 9, but there seems to be no clear way to choose which theory of consciousness is moving in the right direction, and caution must be exercised here.

Acknowledgements. The author has worked on some of the experiments described here with G. C. Baylis, L. L. Baylis, M. J. Burton, H. C. Critchley, M. E. Hasselmo, C. M. Leonard, F. Mora, D. I. Perrett, M. K. Sanghera, T. R. Scott, S. J. Thorpe, and F. A. W. Wilson, and their collaboration, and helpful discussions with or communications from M. Davies and C. C. W. Taylor (Corpus Christi College, Oxford), and M. S. Dawkins, are sincerely acknowledged. Some of the research described was supported by the Medical Research Council.

References

Abercrombie, E. D., Keefe, K. A., DiFrischia, D. S. and Zigmond, M. J. (1989) Differential effect of stress on in vivo dopamine release in striatum, nucleus accumbens, and medial frontal cortex. *Journal of Neurochemistry* 52: 1655-1658.

Adolphs, R., Tranel, D., Damasio, H. and Damasio, A. (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372: 669-672.

Amaral, D. G., Price, J. L., Pitkanen, A. and Carmichael, S. T. (1992) Anatomical organization of the primate amygdaloid complex. In: *The Amygdala*, pp. 1-66, ed. J. P. Aggleton. Wiley-Liss.

Bechara, A., Damasio, A. R., Damasio, H. and Anderson, S. W. (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50: 7-15.

Bechara, A., Damasio, H., Tranel, D. and Anderson, S. W. (1998) Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience* 18: 428-437.

Bechara, A., Damasio, H., Tranel, D. and Damasio, A. R. (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275: 1293-1295.

Bechara, A., Tranel, D., Damasio, H. and Damasio, A. R. (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex* 6: 215-225.

Brothers, L. and Ring, B. (1993) Mesial temporal neurons in the macaque monkey with responses selective for aspects of socal stimuli. *Behavioural Brain Research* 57: 53-61.

Cooper, J. R., Bloom, F. E. and Roth, R. H. (1996) *The Biochemical Basis of Neuropharmacology*, 7th ed. Oxford University Press.

Damasio, A. R. (1994) *Descartes' Error*. Putnam.

Darwin, C. (1872) *The Expression of the Emotions in Man and Animals*, 3rd ed. University of Chicago Press.

Davidson, R. J. and Irwin, W. (1999) The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences* 3: 11-21.

Davis, M. (1992) The role of the amygdala in conditioned fear. In: *The Amygdala*, pp. 255-305, ed. J. P. Aggleton. Wiley-Liss.

Dawkins, R. (1986) *The Blind Watchmaker*. Longman.

Ekman, P. (1982) *Emotion in the Human Face*, 2nd ed. Cambridge University Press.

Ekman, P. (1993) Facial expression and emotion. *American Psychologist* 48: 384-392.

Everitt, B. and Robbins, T. W. (1992) Amygdala-ventral striatal interactions and reward-related processes. In: *The Amygdala*, pp. 401-430, ed. J. P. Aggleton. Wiley.

Francis, S., Rolls, E. T., Bowtell, R., McGlone, F., O'Doherty, J., Browning, A., Clare, S. and Smith, E. (1999) The representation of the pleasantness of touch in the human brain, and its relation to taste and olfactory areas. *NeuroReport* 10: 453-460.

Fridlund, A. J. (1994) *Human Facial Expression: An Evolutionary View*. Academic Press.

Frijda (1986) *The Emotions*. Cambridge University Press.

Goldman-Rakic, P. S. (1996) The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society B* 351: 1445-1453.

Gray, J. A. (1975) *Elements of a Two-Process Theory of Learning*. Academic Press.

Gray, J. A. (1987) *The Psychology of Fear and Stress*, 2nd ed. Cambridge University Press.

Gray, J. A., Young, A. M. J. and Joseph, M. H. (1997) Dopamine's role. *Science* 278: 1548-1549.

Halgren, E. (1992) Emotional neurophysiology of the amygdala within the context of human cognition. In: *The Amygdala*, pp. 191-228, ed. J. P. Aggleton. Wiley-Liss.

Hoebel, B. G. (1997) Neuroscience and appetitive behavior research: 25 years. *Appetite* 29: 119-133.

Hornak, J., Rolls, E. T. and Wade, D. (1996) Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia* 34: 247-261.

Houk, J. C., Adams, J. L. and Barto, A. C. (1995) A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In: *Models of Information Processing in the Basal Ganglia*, pp. 249-270, eds. J. C. Houk, J. L. Davies and D. G. Beiser. MIT Press.

Izard, C. E. (1991) *The Psychology of Emotions*. Plenum.

James, W. (1884) What is an emotion? *Mind* 9: 188-205.

Krebs, J. R. and Kacelnik, A. (1991) Decision Making. In: *Behavioural Ecology*, pp. 105-136, eds. J. R. Krebs and N. B. Davies. Blackwell.

Lange, C. (1885) The emotions. In: *The Emotions*, ed. E. Dunlap. Williams and Wilkins.

Lazarus, R. S. (1991) *Emotion and Adaptation*. Oxford University Press.

LeDoux, J. E. (1992) Emotion and the amygdala. In: *The Amygdala*, pp. 339-351, ed. J. P. Aggleton. Wiley-Liss.

LeDoux, J. E. (1994) Emotion, memory and the brain. *Scientific American* 270: 32-39.

LeDoux, J. E. (1996) *The Emotional Brain*. Simon and Schuster.

Leibowitz, S. F. and Hoebel, B. G. (1998) Behavioral neuroscience and obesity. In: *The Handbook of Obesity*, pp. 313-358, eds. G. A. Bray, C. Bouchard and P. T. James. Dekker.

Leonard, C. M., Rolls, E. T., Wilson, F. A. W. and Baylis, G. C. (1985) Neurons in the amygdala of the monkey with responses selective for faces. *Behavioural Brain Research* 15: 159-176.

Mackintosh, N. J. (1983) *Conditioning and Associative Learning*. Oxford University Press.

Malkova, L., Gaffan, D. and Murray, E. A. (1997) Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *Journal of Neuroscience* 17: 6011-6020.

Millenson, J. R. (1967) *Principles of Behavioral Analysis*. MacMillan.

Milner, A. D. and Goodale, M. A. (1995) *The Visual Brain in Action*. Oxford University Press.

Mirenowicz, J. and Schultz, W. (1996) Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 279: 449-451.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J. and Dolan, R. J. (1996) A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383: 812-815.

Oatley, K. and Jenkins, J. M. (1996) *Understanding Emotions*. Backwell.

Ono, T. and Nishijo, H. (1992) Neurophysiological basis of the Kluver-Bucy syndrome: responses of monkey amygdaloid neurons to biologically significant objects. In: *The Amygdala*, pp. 167-190, ed. J. P. Aggleton. Wiley-Liss.

Petrides, M. (1996) Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philosophical Transactions of the Royal Society* B 351: 1455-1462.

Rada, P., Mark, G. P. and Hoebel, B. G. (1998) Dopamine in the nucleus accumbens released by hypothalamic stimulation-escape behavior. *Brain Research* 782: 228-234.

Reisenzein, R. (1983) The Schachter theory of emotion: Two decades later. *Psychological Bulletin* 94: 239-264.

Robbins, T. W., Cador, M., Taylor, J. R. and Everitt, B. J. (1989) Limbic-striatal interactions in reward-related processes. *Neuroscience and Biobehavioral Reviews* 13: 155-162.

Rolls, B. J. and Rolls, E. T. (1982) *Thirst*. Cambridge University Press.

Rolls, E. T. (1975) *The Brain and Reward*. Pergamon.

Rolls, E. T. (1986a) Neural systems involved in emotion in primates. In: *Emotion: Theory, Research, and Experience. Vol. 3. Biological Foundations of Emotion*, pp. 125-143, eds. R. Plutchik and H. Kellerman. Academic Press.

Rolls, E. T. (1986b) A theory of emotion, and its application to understanding the neural basis of emotion. In: *Emotions. Neural and Chemical Control*, pp. 325-344, ed. Y. Oomura. Karger.

Rolls, E. T. (1990) A theory of emotion, and its application to understanding the neural basis of emotion. *Cognition and Emotion* 4: 161-190.

Rolls, E. T. (1992a) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society* 335: 11-21.

Rolls, E. T. (1992b) Neurophysiology and functions of the primate amygdala. In: *The Amygdala*, pp. 143-165, ed. J. P. Aggleton. Wiley-Liss.

Rolls, E. T. (1996) The orbitofrontal cortex. *Philosophical Transactions of the Royal Society B* 351: 1433-1444.

Rolls, E. T. (1997) Taste and olfactory processing in the brain and its relation to the control of eating. *Critical Reviews in Neurobiology* 11: 263-287.

Rolls, E. T. (1999) *The Brain and Emotion*. Oxford University Press.

Rolls, E. T., Burton, M. J. and Mora, F. (1980) Neurophysiological analysis of brain-stimulation reward in the monkey. *Brain Research* 194: 339-357.

Rolls, E. T., Hornak, J., Wade, D. and McGrath, J. (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry* 57: 1518-1524.

Rolls, E. T. and Johnstone, S. (1992) Neurophysiological analysis of striatal function. In: *Neuorphychological Disorders Associated with Subcortical Lesions*, pp. 61-97, eds. G. Vallar, S. F. Cappa and C. W. Wallesch. Oxford University Press.

Rolls, E. T., Thorpe, S. J. and Maddison, S. P. (1983) Responses of striatal neurons in the behaving monkey: I. Head of the caudate nucleus. *Behavioural Brain Research* 7: 179-210.

Rolls, E. T. and Treves, A. (1998) *Neural Networks and Brain Function*. Oxford University Press.

Rolls, E. T., Critchley, H. D., Browning, A. S., Hernadi, A. and Lenard, L. (1999) Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex. *Journal of Neuroscience* 19: 1532-1540.

Sanghera, M. K., Rolls, E. T. and Roper-Hall, A. (1979) Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Experimental Neurology* 63: 610-626.

Schultz, W. (1998) Predictive reward signal of dopamine neurons. *Journal of Neurophysiology* 80: 1-27.

Schultz, W., Romo, R., Ljunberg, T., Mirenowicz, J., Hollerman, J. R. and Dickinson, A. (1995) Reward-related signals carried by dopamine neurons. In: *Models of Information Processing in the Basal Ganglia*, pp. 233-248, eds. J. C. Houk, J. L. Davis and D. G. Beiser. MIT Press.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P. and Johnson, M. (1997) Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385: 254-257.

Scott, T. R., Yan, J. and Rolls, E. T. (1995) Brain mechanisms of satiety and taste in macaques. *Neurobiology* 3: 281-292.

Sem-Jacobsen, C. W. (1968) *Depth-electrographic stimulation of the human brain and behavior: from fourteen years of studies and treatment of Parkinson's Disease and mental disorders with implanted electrodes*. C.C. Thomas.

Sem-Jacobsen, C. W. (1976) Electrical stimulation and self-stimulation in man with chronic implanted electrodes. Interpretation and pitfalls of results. In: *Brain-Stimulation Reward*, pp. 505-520, eds. A. Wauquier and E. T. Rolls. North-Holland.

Shallice, T. and Burgess, P. (1996) The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society* B 351: 1405-1411.

Strongman, K. T. (1996) *The Psychology of Emotion*, 4th ed. Wiley.

Thierry, A. M., Tassin, J. P., Blanc, G. and Glowinski, J. (1976) Selective activation of mesocortical DA system by stress. *Nature* 263: 242-244.

Thorpe, S. J., Rolls, E. T. and Maddison, S. (1983) Neuronal activity in the orbitofrontal cortex of the behaving monkey. *Experimental Brain Research* 49: 93-115.

Tinbergen, N. (1951) *The Study of Instinct*. Oxford University Press.

Wallis, G. and Rolls, E. T. (1997) Invariant face and object recognition in the visual system. *Progress in Neurobiology* 51: 167-194.

Weiskrantz, L. (1968) Emotion. In: *Analysis of Behavioural Change*, pp. 50-90, ed. L. Weiskrantz. Harper and Row.

Williams, G. V., Rolls, E. T., Leonard, C. M. and Stern, C. (1993) Neuronal responses in the ventral striatum of the behaving macaque. *Behavioural Brain Research* 55: 243-252.

Wilson, F. A. W. and Rolls, E. T. (1993) The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. *Experimental Brain Research* 93: 367-382.

Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P. and Hanley, J. R. (1995) Face processing impairments after amygdalotomy. *Brain* 118: 15-24.

Young, A. W., Hellawell, D. J., Van de Wal, C. and Johnson, M. (1996) Facial expression processing after amygdalotomy. *Neuropsychologia* 34: 31-39.