BEHAVIORAL 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 (Ch. 3); the neural bases of emotion (Ch. 4); reward, punishment, and emotion in brain design (Ch. 10); a theory of consciousness and its application to understanding emotion and pleasure (Ch. 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, 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 (Ch. 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 the 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 quench our thirst? What factors account for the overeating and obesity that 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 1999a).

The brain mechanisms of both emotion and motivation are considered together. The examples of motivated behavior described are hunger (Ch. 2), thirst (Ch. 7), and sexual behavior (Ch. 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 (Ch. 4), hunger, thirst, and sexual behavior (Chs. 2, 7, and 8), and reward (Chs. 5 and 6). The book though links this analysis to the wider context of the nature of emotions, their functions (Ch. 3), how they evolved (Ch. 10), and the larger issue of why emotional and motivational feelings and consciousness might arise in a system organised like the brain (Ch. 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 that optimise our behavior by building into us the appropriate reward and punishment systems and the appropriate rules for the operation of these systems.

EDMUND ROLLS is a Professor of Experimental Psychology at the 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).

© 2000 Cambridge University Press 0140-525X/00 $12.50

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

A major reason for investigating the actual brain mechanisms that underlie emotion and motivation, and reward and punishment, is not only to understand how our own brains work, but also to have a 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 nonhuman 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 Ch. 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 Ch. 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 Figure 1.

Before accepting this approach, we should consider whether there are any exceptions to the proposed rule. Are there 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 & 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 that, 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 that 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 are therefore 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).
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 re-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

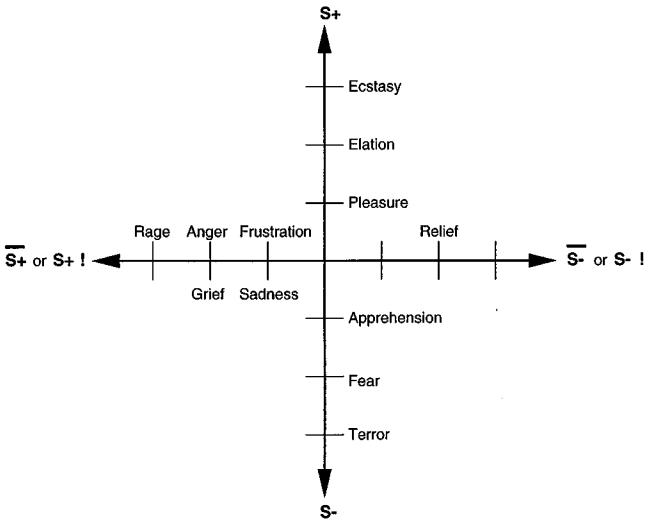


Figure 1. Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the centre of the diagram on a continuous scale. The classification scheme created by the different reinforcement contingencies consists of (1) the presentation of a positive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

ward 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; 1999a). It is also worth noting that emotions can be produced just as much by the recall of reinforcing events as by external reinforcing stimuli. 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.Because 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 referred to as *emotion.*

Three issues receive discussion here (see further, Rolls 1999a). 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 that 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 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 1999a). Indeed there is 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 as including 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 elsewhere more fully (Rolls 1990; 1999a), 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 to a particular stimulus is learned. It also contrasts with the elicitation of species-typical behavioral responses by sign-releasing stim-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

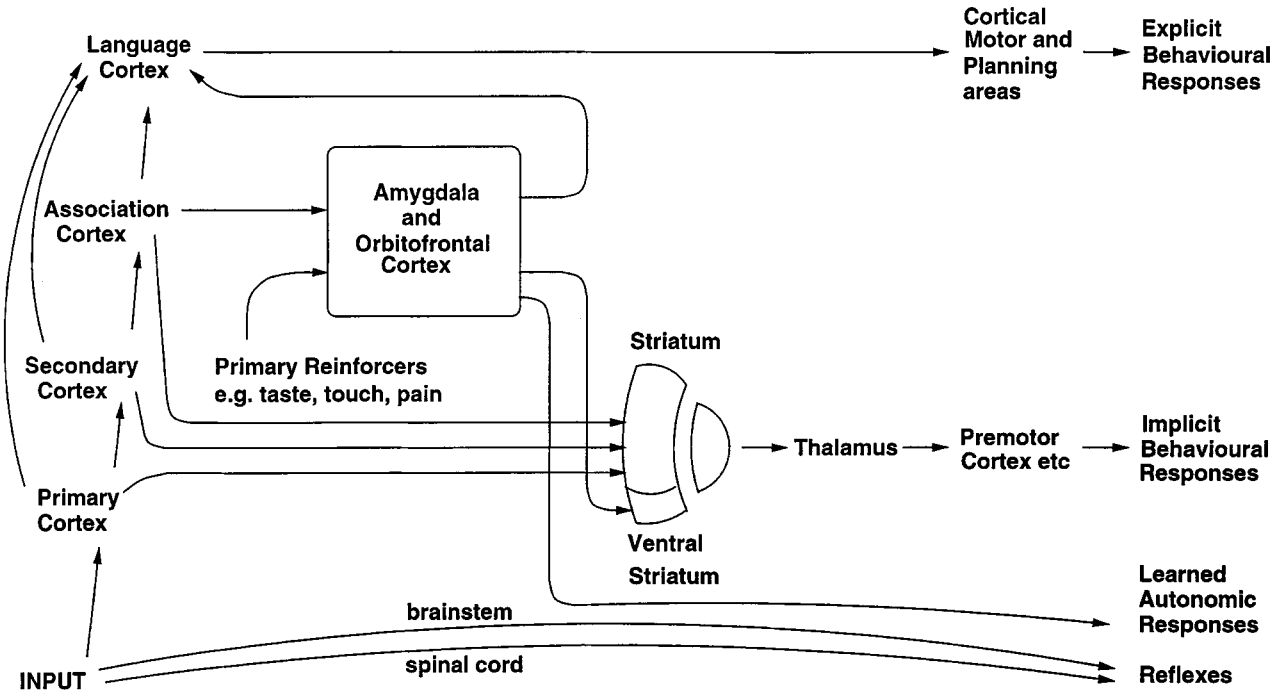


Figure 2. Summary of the organisation of some of the brain mechanisms underlying emotion, showing dual routes to the initiation of action in response to rewarding and punishing, that is, emotion-producing, stimuli. The inputs from different sensory systems to brain structures such as the orbitofrontal 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 (verbalisable) decisions involving multistep syntactic planning to be implemented. (After *The brain and emotion,* Fig. 9. 4.)

uli (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, 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 that 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 that effectively specify the goals for action. This is the solution that 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 benefits, and avoiding punishers with their associated costs, is a process that can take place both implicitly (unconsciously) and explicitly using a language system to enable long-term plans to be made (Rolls 1999a). Many different brain systems, some involving implicit evaluation of rewards and others involving explicit, verbal, and 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 among all the competing calls on output and might involve the basal ganglia in the brain (see Fig. 2 and Rolls 1999a).

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 develop-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

ment 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 that 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 & Jenkins 1996). This may facilitate continuity in the interpretation of the reinforcing value of events in the environment. A hypothesis described in *The Brain and Emotion* states that backprojections from parts of the brain involved in emotion such as the orbitofrontal cortex and amygdala implement this.

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 Ch. 4 and Rolls & 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 that 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 & 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 section 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 flexibly 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 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 towards 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 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

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

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 Ch. 7).

With many reward/punishment dimensions for which actions may be performed (see Table 10.1 of *The brain and emotion* for a nonexhaustive 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 different reward system 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 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 de-

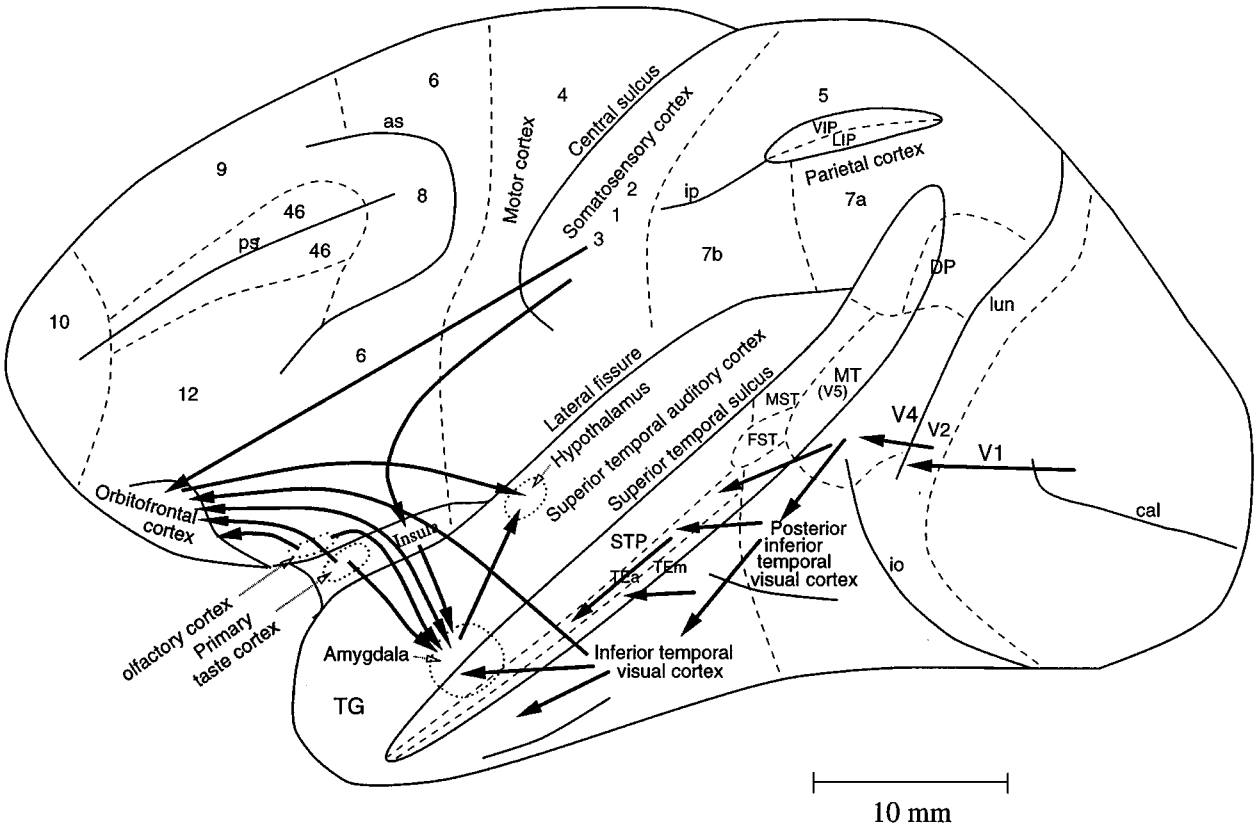


Figure 3. Some of the pathways involved in emotion described in the text are shown on this lateral view of the brain of the macaque monkey. Connections from the primary taste and olfactory cortices to the orbitofrontal cortex and amygdala are shown. Connections are also shown in the "ventral visual system" from V1 to V2, V4, the inferior temporal visual cortex, and so on, 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.)

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

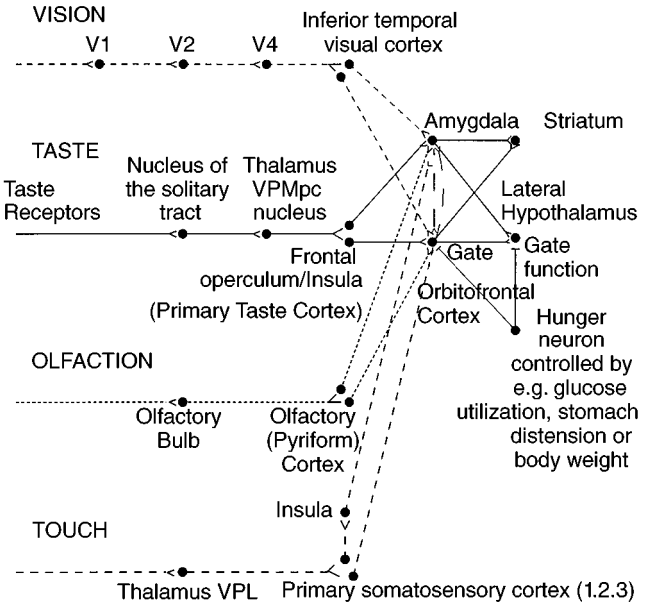


Figure 4. Diagrammatic representation of some of the connections described in the text. V1 – striate visual cortex. V2 and V4 – cortical visual areas. In primates, sensory analysis proceeds as far as the inferior temporal visual cortex and the primary gustatory cortex; beyond these areas, for example, 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. 2.)

sign features that 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 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 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, 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

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; 1999a). The brain regions discussed include the amygdala and orbitofrontal cortex. Some of these are indicated in Figures 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 nonprimates. 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 Figure 2, and some of the pathways are shown on a lateral view of a primate brain in Figure 3 and schematically in Figure 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

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

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, 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 Ch. 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 that 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 & 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 that is ideal as an input to pattern associators that 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, that is, inner-product decoding as described in the Appendix); they are distributed allowing excellent generalisation and graceful degradation. 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 msec (see Rolls & Treves 1998, Ch. 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 & 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 sensorimotor 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 nonprimates, 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 nonprimates: The recognition that is possible is based more on physical similarity in terms of texture, colour, simple features, and so on (see Rolls & Treves 1998, sect. 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, perhaps only 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 Ch. 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 & 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

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

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 therefore 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 Figure 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 Figures 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, 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 1999a). The crucial site of the stimulus-reinforcement associative learning that 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; 1999a). The association learning in the amygdala may be implemented by associatively modifiable synapses (see Rolls & 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 that underlie learning (see Rolls & 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 in that amphetamine injections into the ventral striatum enhanced the effects of a conditioned reinforcing stimulus only if the amygdala was intact (see Everitt & 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 & Treves 1998; Wallis & Rolls 1997). This is probably part of a system that 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; 1999a).

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

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

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 so, addresses the functions of emotion of the highly developed temporal and orbitofrontal cortical areas of primates including humans, areas that 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 primate amygdala 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 reassessment 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 that novel stimuli have (see Rolls 1999a).

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 & 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 (Brothers & Ring 1993; Leonard et al. 1985), 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 nonhuman primates that some amygdala neurons are activated by rewarding and others by punishing stimuli (Ono & Nishijo 1992; Rolls 1992a; 1992b; Sanghera et al. 1979; Wilson & 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; Rolls et al. 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 & 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. In the monkey, 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

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

representation is rich in that there is much information that can be easily read from the neuronal code (see Rolls & Treves 1998) about exactly which taste, touch, or odor is being delivered. It is important that reinforcers be represented in a way that 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, 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 that 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 nonreward trials (Thorpe et al. 1983). There is therefore 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, implements a mechanism that evaluates whether a reward is expected, and generates a mismatch (evident as a firing of the nonreward neurons) if reward is not obtained when it is expected (Thorpe et al. 1983; Rolls 1990; 1996; 1999a). 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 that include those usually described as involving frustration.

5.3.2. 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, that is, 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; 1999a). 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 that 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 were 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 that 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 different decks. The patients with frontal damage were more likely to choose cards from decks that gave rewards with a reasonable probability but also occasionally had very heavy penalties. The net gains from these decks were lower than from the other decks. In this sense, the patients were not affected by the negative consequences of their actions: they did not switch from the decks of cards that, although 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 is 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 & Irwin 1999). Indeed this seems unlikely because 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 Figure 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 un-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

learned 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 that 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 nut" 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, because 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 & 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 that 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 & 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.) Therefore 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 nonhuman 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 nonhuman primates and humans (see Goldman-Rakic 1996; Petrides 1996); and the impairment of planning produced by prefrontal cortex damage (see Shallice & 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 even though 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 & Treves 1998, Chs. 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, insofar 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 emo-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

tional 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 among 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 1999a). 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 an 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 & Hoebel 1998). Also, Gray et al. (1997) (and 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 that 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 that 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 & 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) and Schultz (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 Ch. 4, sect. 4, and Rolls & Treves 1998), than to reward *per se*, or a teaching signal about reward (cf. Houk et al. 1995; Schultz et al. 1995). The investigation of Mirenowicz and Schultz (1996) did not address this issue directly: 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 et al. (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 behavior is necessary 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 versus 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 cur-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

rently biased 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 1999a).

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 & 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 closely related to it in supposing that feedback from parts of the periphery (such as the face or body, as in 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, for example, 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, with classifying different emotions (Ch. 3), and in understanding *what* information-processing systems in the brain are involved in emotion, and *how* they are involved (Ch. 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 (Ch. 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 Ch. 2 on hunger and Ch. 7 on

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

thirst) or that are influenced by internal hormonal signals (Ch. 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 are 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 representations 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 associative 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 that 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 that 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 (Ch. 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 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 versus 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, so caution must be exercised here.

ACKNOWLEDGMENTS

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; 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.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Is reward an emotion?

Ralph Adolphs

Department of Neurology, Division of Cognitive Neuroscience, University of Iowa College of Medicine, Iowa City, IA 52242. **ralph-adolphs@uiowa.edu**

Abstract: *The brain and emotion* treats emotions as states elicited by reinforcers (reward or punishment), but it is unclear how this view can do justice to the diversity of emotions. It is also unclear how such a view distinguishes emotions from states such as hunger and thirst. A complementary approach to understanding emotions may begin by considering emotions as aspects of social cognition.

Although Edmund Rolls's new book, *The brain and emotion,* ranges widely, its focus and expertise are centered on the topic that Rolls himself has pioneered: single-unit neurophysiology of reward mechanisms in the amygdala and orbitofrontal cortex of animals. Much of the book concerns the analysis of responses to a variety of sensory stimuli that are rewarding or aversive to the animal, and the corresponding question of how the reinforcing properties of stimuli are represented in the brain is also treated in some detail. In this *BBS* commentary, I would like to explore just one of the important issues that the book raises: How do our concepts of reward and punishment, which rely essentially on the notion of behavioral reinforcement, relate to the concept of an emotion? In a nutshell, can emotions such as happiness, fear, anger, sadness, as well as embarrassment, love or awe, be understood from within the framework of reward and punishment? Rolls evidently believes that they can, but I am not so sure.

Attempts to produce a taxonomy of emotions have tended to fall between two extremes. At one extreme are those stimulated by the classical experiments of Schacter and Singer (1962): all emotions share in common something like emotional arousal and perhaps all negative emotions share in common aversion, but what distinguishes individual emotions is their cognitive content, which is presumed to be separate from their motivational content. At the other extreme are theories like those of Paul Ekman (1992; 1993), postulating that there are specific neural systems for each specific, basic emotion – although people vary in what they consider to be the basic emotions (see, e.g., the scheme proposed by Panksepp 1998a). As expected, most people agree that emotions draw upon some underlying dimensions, but also acknowledge that different emotions correspond to different patterns of stimuli, behaviors, and experience, and will rely on partly distinct neural systems.

Rolls tends towards the former, more reductionist, end of the spectrum, as illustrated for instance in his Figure 3.1 (p. 63 of the book and Figure 1, p. 179 of the target article). According to his scheme, emotions fall on a 2-dimensional space with axes specified by either presenting or withholding stimuli with positive or negative reinforcement contingencies. This leads him to put together on the same axis some emotions one might intuitively consider quite different, for example, frustration, sadness, anger, grief, and rage.

Although the attempt to find among emotions some simpler, underlying dimensions is certainly worthwhile, and I do not deny that there are dimensions that capture a portion of the variance in measures that assess emotion, there are two major problems with the whole approach. First, common experience suggests that there are more emotions than any current scheme of underlying dimensions would permit. Fear, anger, happiness, love, awe, jealousy, embarrassment, guilt, disgust, and grief all differ qualitatively, so that it is implausible that they should differ only with respect to the magnitude of some shared dimensions. Second, there is now good evidence from neuroscience to suggest that at least some of the emotions we would consider distinct do in fact engage distinct neural structures. Some examples are the disproportionate involvement of the amygdala in fear (Adolphs et al. 1995; Le Doux 1996) and the disproportionate involvement of the insula in disgust (Phillips et al. 1997) (see also Panksepp [1998a] for a recent inventory of some specific emotional neural systems). On the face of it, knowledge of how fear and disgust might be related to reward and punishment would seem insufficient. They are both aversive emotional states, but they are also very clearly different in terms of the sets of stimuli that normally evoke them, in terms of the sets of behavioral responses engendered, in terms of the conscious experience of the emotions, and, in fact, in terms of their neural underpinnings. What is needed is a story that accounts for their differences.

An analogous problem arises in distinguishing bona fide emotional states from states that are also related to reward and punishment, and that likely share some of the same circuitry as that which subserves emotion, but that are not emotions. Rolls himself provides us with the examples, because they are the states whose neurophysiology he discusses in detail: thirst, hunger, and so on. It seems to me that one ingredient which distinguishes at least a large number of those states we normally consider emotions from those we normally do not – such as thirst, pain, and hunger – is the former's relevance to social behavior and key role in social communication. I would thus like to suggest an exploration of the neurobiology of emotion that takes a starting point complementary to the one that Rolls describes: begin with findings from ethology and from comparative, developmental, and social psychology. There is no shortage of this literature; in fact, most writings on the topic of emotion, until very recently, came from the domain of social psychology (I resist citation here, since it would have to be very incomplete).

Taking the approach I suggest seriously requires detailed knowledge of ethology, of developmental and social psychology, and of related fields. Some such syntheses are now emerging; for instance in books by Brothers (1997), Damasio (1994), Panksepp (1998a), and Schore (1994), to name a few recent ones. No doubt, much about our commonsense categories for emotions will need to be revised (cf. Griffiths 1997), but it seems premature to collapse all emotions into concepts of reward and punishment. This is not to say that I think the framework Rolls describes is useless; quite the contrary, I believe that we need to investigate emotion *both* from the point of view of reward and punishment, as well as from the perspective of social cognition. The former is probably more suited to exploring emotion's neural underpinnings in nonprimate animals, while only the latter can give us the full richness of emotion found in humans.

Emotions or emotional feelings?

Murat Aydede

Department of Philosophy, The University of Chicago, Chicago, IL 60637.
m-aydede@uchicago.edu **humanities.uchicago.edu/faculty/aydede/**

Abstract: I criticize Rolls's account of what makes emotional states conscious.

It turns out that Rolls's answer to Nagel's (1974) question, "What is it like to be a bat?" is brusque: there is *nothing* it is like to be a bat – provided that bats don't have a linguistically structured internal representational system that enables them to think about their first-order thoughts which are also linguistically structured. For phenomenal consciousness, a properly functioning system of higher-order linguistic thought (HOLT) is necessary (Rolls 1999a,

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

p. 262). By this criterion, not only bats, but also a great portion of the animal kingdom, perhaps all animal species except humans, turn out to lack phenomenal consciousness. Indeed, even human babies, and perhaps infants before the early stages of acquiring their first language, are likely to lack such consciousness, if one considers the level of conceptual sophistication required by the HOLT hypothesis. In order to have a higher-order thought, one needs to have the concept of a *thought* in addition to the (linguistically structured) representational resources to articulate the conceptual content of the lower-order thought. Indeed, Rolls believes (p. 262) that phenomenal consciousness may be quite a late arrival in the history of evolution of the mind/brain: certainly much later than the ability to think (in a linguistically structured internal medium), because it requires the ability to think not only about one's physical environment, but also about one's own thoughts. Many would take such consequences as a *reductio* of Rolls's thesis about consciousness; for surely the thesis seems to overintellectualize phenomenal consciousness.

But Rolls does not deny that animals have emotions, only that they have emotional feelings. The latter are conscious, whereas the former are not. There is nothing it is like to be in an emotional state unless one is endowed with a HOLT system. But why does Rolls feel he must embrace this startling and implausible conclusion? After telling us that emotions are brain states caused by positively or negatively reinforcing stimuli, including changes in such stimuli, Rolls offers a fascinating tour of the physiology of reward and motivation in which he makes quite elaborate and specific proposals as to where and how in the brain such states are likely to occur (primarily in the amygdala and orbitofrontal cortex). But Rolls's actual "definition" of emotion is only part of the story. Later (especially in sects. 4.6–4.9, 9.3, 10.3, among others), he implicitly supplements this definition by telling us that emotions are also the indirect causes of certain types of motivated behavior. In particular, they are (in humans) inputs to two brain systems that decide on the behavioral output by computing the reward value of each behavioral choice against its odds before outputting to motor areas. One system, which humans share with other primates, consists of the basal ganglia and its structures (including, perhaps, the inferior temporal visual cortex – cf. p. 285). The output of this system is processed in the premotor cortex and then fed into the motor system before the actual ensuing behavioral response. Rolls calls the behavior elicited by this route "implicit behavior." The other system, which is perhaps specific to humans, is the language cortex, which receives the outputs of amygdala and orbitofrontal cortex, processes them in a syntactically structured symbolic medium before initiating action through the cortical motor and planning areas. Behavior from this route is dubbed "explicit behavior." (In addition to these two, there are also reflex circuits between relatively unprocessed sensory inputs and motor reactions.)

So the core of Rolls's account is a straightforward functionalist (in fact, psychofunctionalist – see Block 1980) characterization of emotional states. Hence we can explicitly rewrite Rolls's actual definition thus: emotional states are those states (mainly realized in the amygdala and orbitofrontal cortex) that play the above-specified type of functional/causal role (to be specified more fully and explicitly of course) in the central neural economy of the brain. But then Rolls asks: Why should there be anything it is like to be in such states? Why should such states feel like anything at all? (I am of course delighted to see a scientist who thinks that he should have an answer to this question. So I truly applaud Rolls's attempt to give an answer – albeit a tentative and cautious one.) Feeling that his account of emotions seems to leave out the qualia of emotions, Rolls proposes the HOLT hypothesis.

But I am puzzled by why Rolls thinks that his helps. Before explaining my misgivings, a few clarifying remarks (or exploratory speculations) about the hypothesis are in order. As Rolls points out, the HOLT account closely resembles Rosenthal's HOT theory of state consciousness, though it adds the requirement that thoughts are realized in a syntactically structured (language-like) representational medium, or "mentalese," as per the Language of Thought Hypothesis (LOTH – for a presentation of which, see Aydede 1998). After distinguishing between first- and higher-order thoughts (all realized in mentalese), Rolls insists that second-order thoughts (i.e., thoughts about first-order thoughts) are required for emotions to become conscious, that is, to turn them into emotional feelings. This is a bit puzzling because, as characterized by Rolls, emotional states are obviously not thoughts realized in mentalese. Rolls sometimes talks about the firing of a bunch of (specific) neurons in the orbitofrontal cortex as a representation of the reward value of, say, a particular taste – intuitively what we might otherwise call, its pleasantness, or the lack thereof. But, as far as I can tell, these representations are not privy to the computations of the explicit linguistic system. (However, see sect. 4.6.3, where Rolls seems to suggest, somewhat confusingly, that the computation of reward value in the linguistic system might depend on activity in the orbitofrontal cortex and amygdala, which are not identified by Rolls as sites of the higher linguistic cortex.) So it is not clear how these representations are supposed to be made conscious by shining the light of higher-order thought on them.

Perhaps all Rolls means to say is this: in order for emotional states (say, orbitofrontal representations of reward value) to become conscious they must be the target of a first-order thought; nothing more is required. On this interpretation, all that the organism needs to do is to think about such representations – perhaps something like the internal version of "Oh, this is pleasant" where "pleasant" is represented by a mentalese predicate, and "this" is a Mentalese quasi-demonstrative referring to – what? It cannot be the orbitofrontal representation, for *that* is not pleasant. What is pleasant is presumably the taste, which is processed, according to Rolls, independently of and prior to its reward value, in the primary taste cortex. So perhaps, it refers to a purely sensory (i.e., affectively neutral) representation of that particular taste, say in the primary taste cortex. But this does not seem right either. How could an affectively neutral sensation be pleasant? But then what is the first-order thought supposed to be about? What these questions/reflections seem to indicate is that if a version of HOT theory is claimed to account for emotional feelings, then this account is absent from Rolls's book.

But suppose that such an account were provided. Surely it would be a causal/functional one, in that the postulated higher-order states will be brain states causally connected, perhaps in quiet complicated ways, to the sensory and affective states realized in the relevant brain sites. So, for instance, suppose whenever I am in those emotional states, I am causally prompted to have thoughts (brain states realized up in the linguistic cortex) about my emotional states (just further brain states down in the orbitofrontal cortex and amygdala) that – *voilà* – transform the latter into emotional feelings, full-blown phenomenal states in all their glory. How is this supposed to be less mysterious? Why should there be anything it is like to be in a state playing *this* (more sophisticated) causal/functional role? Contrary to Rolls, as things stand, I do not see any advance here.

Nevertheless, I am no pessimist or mysterian about phenomenal consciousness. On the contrary, I believe that the mystery can probably be solved by more or less the same naturalistic or scientific methods that Rolls so skillfully employs throughout the book in uncovering some of the brain mechanisms of emotions. I even believe that he is right in thinking that at some point in our analysis of phenomenal consciousness we need to bring in some form of higher-order mental state theory (this could be a HOT or HO-Perception account in the way introspection has been traditionally conceived, i.e., as a kind of inner *sense*). But to solve the mystery of affective/emotive qualia will require at least two things. First, we need to account for such qualia not representationally, but rather in purely psychofunctional terms, namely as *ways* of processing incoming sensory information en route to setting behavioral parameters. I believe Rolls's research gives substantial support to this kind of approach. Second, we need to tell a naturalistic story about our peculiar first-person epistemic access to them, a

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

story which does justice to the subjectivity of the mental. This is where going higher-order is likely to play an essential role, especially when we bring in some of the resources and peculiarities of indexical and reflexive reference to such an account. For exploration of some of these themes, see Dretske (1995), Lycan (1996), Rey (1997), Tye (1995), and Aydede (in preparation).

ACKNOWLEDGMENT
Many thanks to Philip Robbins for his comments and corrections.

Are emotions so simple?

Aaron Ben-Ze'ev
Department of Philosophy, University of Haifa, Haifa 31905, Israel.
benzeev@research.haifa.ac.il

Abstract: Rolls's book, *The brain and emotion* is an important and valuable contribution to our understanding of the brain mechanisms that underlie emotional processes. Its explanatory value is less obvious when it comes to psychological and philosophical issues concerning the nature of emotions.

The nature of emotions. Rolls defines emotion as "states elicited by rewards and punishers" (p. 60). In one sense, Rolls's definition is too broad: there are many other states that are elicited by rewards and punishers and which are not emotions. For example, when I decide to clean my car, this decision and the actual cleaning are states elicited by rewards and punishers but these are not emotional states. In another sense, Rolls's definition is too narrow: it refers only to the causes of emotions and not to their nature. I would prefer an approach that characterizes typical features of emotions rather than defines the "essence" of emotions. There is always the danger of superficiality in the latter approach.

Rolls's emphasis on the element of reward and punishment reveals, however, an important feature of emotions: the positive or negative evaluation underlying each emotional state. Emotions typically occur when *we perceive positive or negative significant changes in our personal situation* (Ben-Ze'ev 2000). Spinoza, for example, claims that when we undergo great change, we pass to a greater or lesser perfection, and these changes are expressed in emotions. As we change for the better we are happy and for the worse unhappy (Spinoza 1677:IIIp6; IIIdef.aff.; Vp39s).

Classifying the emotions. Rolls's classification of emotions (sect. 3.1.3) is problematic for various reasons. First, it does not distinguish between emotions and other affective states. In addition to emotions, the affective realm includes other phenomena such as sentiments, moods, affective disorders, and affective traits (Ben-Ze'ev 2000, Ch. 4). Second, Rolls's classification is intended to refer to all different emotions, but actually it refers to only few emotions and many nonemotional states. Social emotions, such as envy, pleasure-in-others'-misfortune, regret, pride, and shame, are not mentioned in the index, and hardly, if at all, in the text. Third, Rolls considers emotional intensity to be a factor in classifying the emotions. However, emotional intensity is not a useful tool for this purpose as each emotion can be more or less intense depending on the different circumstances. Accordingly, I do not believe – as is implied in the diagram – that anger is always, or even typically, more intense than sadness, or that it is on the same level of intensity that grief is.

The functions of emotions. Rolls's discussion of the functions of emotions is illuminating – although we may reduce the ten functions into a more limited number of basic functions. I have some doubts concerning the ninth function, namely, "to produce persistent motivation and direction of behaviour" (p. 70). Typical emotions are essentially transient states. The association of emotional intensity with change causes the intensity to decrease steadily due to the transient nature of changes. This association is a natural mechanism enabling the system to return within a relatively short period to normally functioning. In light of these considerations, one may wonder whether we can characterize one function of emotions as that of producing persistent motivation. However, if our time-scale is not one of months or years, but of minutes or hours, then Rolls's claim makes sense. Moreover, emotional values, or values whose fulfillment or violation will generate significant emotional intensity, can no doubt produce persistent motivation and direction of behavior as Rolls suggests.

Subjective feelings. Rolls's explanation of subjective feelings or qualia is odd. He argues that such feelings are possible only in "(linguistically based) higher-order thought processing" (p. 251); "it is a property of the higher-order thought system that it feels like something when it is operating" (p. 253). Feelings are even not part of all thoughts but only those involving thoughts about thoughts. Hence, non-human animal behavior may be very similar to human behavior, "but would not imply qualia" (p. 252). It is hard to understand why Rolls wants to go against the evidence of evolution and claim that first we have linguistic thought capacities and only then feelings. Does Rolls really think that neonates, dogs, and cats do not have feelings of pain and pleasure? When neonates cry do they not have some disagreeable feelings? This over-intellectualization of the mind is not warranted. I believe that the feeling dimension is a primitive mode of consciousness associated with our own state. It is the lowest level of consciousness; unlike higher levels of awareness, such as those found in perception, memory, and thinking, the feeling dimension has no meaningful cognitive content. It expresses our own state, but is not in itself directed at this state or at any other object. In light of its importance in identifying our own state, it is plausible that mental life begins – from both an evolutionary and personal viewpoint – with states of feeling; life at this state is a succession of agreeable and disagreeable sensations. Later on, when intentional capacities are developed, the feeling dimension usually becomes part of a complex mental state which also includes the intentional dimension (Ben-Ze'ev 2000, Ch. 3).

Consciousness, higher-order thought, and stimulus reinforcement

José Luis Bermúdez
Department of Philosophy, University of Stirling, Stirling FK9 4LA, Scotland.
jb10@stir.ac.uk **www.stir.ac.uk/philosophy/cnw/webpage1.htm**

Abstract: Rolls defends a higher-order thought theory of phenomenal consciousness, mapping the distinction between conscious and non-conscious states onto a distinction between two types of action and corresponding neural pathways. Only one type of action involves higher-order thought and consequently consciousness. This account of consciousness has implausible consequences for the nature of stimulus-reinforcement learning.

According to higher-order thought (HOT) theories of consciousness, conscious states are representational states upon which HOTs are (or can be) directed (Rosenthal 1991). It is natural and common to object to such theories on the grounds that they rule out the possibility of consciousness for large sections of the animal kingdom (Dretske 1995). Irrespective of where exactly one draws the line between those creatures capable of HOT and those not, consciousness has seemed to many to extend further down the phylogenetic (and indeed ontogenetic) ladder than HOT (Bermúdez 1998a; Dawkins 1998). Although this is often presented as self-evident, it can be defended on the following grounds. The most plausible model we possess for explaining the vast majority of animal behaviour is that provided by conditioning theory (Dickenson 1980). The basic principle of conditioning theory is that certain patterns of behaviour are reinforced by being associated with primary positive reinforcers, and inhibited by being associated with primary negative reinforcers. But learning through condi-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

tioning works because primary reinforcers have qualitative aspects. It is impossible to divorce pain's being a negative reinforcer from its feeling the way it does. It is impossible to divorce soothing vocalizations being positive reinforcers from their sounding the way they do. The success of stimulus-reinforcement models of learning therefore entails the falsity of higher-order thought theories of consciousness.

One of the many reasons Rolls's *The brain and emotion* is interesting is that it offers the tools for an empirical response to this line of objection to HOT theories. His position, briefly outlined, is the following. He distinguishes two types of reward/punishment-based actions, to each of which there corresponds a distinct neural pathway. Many actions can be performed relatively automatically. Rolls hypothesises that actions of this type are under the control of phylogenetically primitive brain systems like the basal ganglia. These systems "control behaviour in relation to previous associations of stimuli with reinforcement" (p. 256), yielding direct behavioural responses based on assessment of the reinforcement-related value of the stimulus. Such assessment is carried out in the amygdala and (for primates) in the orbito-frontal cortex. Alternatively, however, in certain higher animals the outputs of the amygdala and orbito-frontal cortex can feed into the language areas of the brain and thence into the cortical motor and planning areas. The operation of these systems involves HOT (in the form of the syntactic manipulation of symbols) and conscious control. It is, on Rolls's view, only when sensory input is processed in these areas that it becomes conscious. The consequence, as he readily admits, is that creatures lacking these processing capabilities will not be conscious – "Raw sensory feels, and subjective states associated with emotional and motivational states, may not necessarily arise first in evolution" (p. 262).

There are two separate questions to be raised for this account. The first is whether Rolls gives a plausible account of the operation of stimulus-reinforcement responses in non-linguistic creatures. The second is whether he gives a plausible account of stimulus-reinforcement responses in language-using creatures. I will concentrate on the second, but it has clear implications for the first.

Rolls says: "much complex animal, including human, behaviour can take place using the implicit, non-conscious, route to action" (p. 261). Given his account of how qualia arise (in the context of the conscious processing involved in higher-order planning) it looks as if the sensory causes of such behavior cannot be conscious. Suppose that as a result of associative learning a certain auditory stimulus has become a secondary reinforcer, Rolls's theory appears to entail that whenever I respond behaviourally, and without HOT, to that secondary reinforcer, the auditory stimulus must be non-conscious. This is so because in such situations the sensory input does not go through the cortical motor and planning areas where HOT might be brought to bear (see Fig. 9.4, for example).

This is highly counter-intuitive. There are many situations every day in which we all seem to respond to secondary reinforcers in a way that does not involve HOT and yet in which the secondary reinforcers are consciously experienced. Examples range from eating a meal to listening to music. Is not the existence of this everyday phenomenon a straightforward counter-example to Rolls's theory?

Rolls might respond in either of two ways. He might (1) maintain that in such situations the relevant sensory information *does* input to the motor and planning areas – and that is why it is conscious, or (2) suggest that sensory input to direct behavioural responses can be conscious in creatures capable of HOT, even when it does not feed into the systems that engage in higher-order thought. Neither of these strategies seems satisfactory. Strategy (1) seems to undercut the motivation for identifying the two distinct types of action and corresponding neural pathways – as well as being incompatible with the way in which he sets up the distinction between two types of action in the first place. Strategy (2) seems to undercut the idea that sensory input becomes conscious in virtue of featuring in higher-order planning, and in turn to weaken the idea that phenomenal consciousness cannot emerge in evolution before the capacity for HOT.

A dispositionalist version of the HOT theory might seem to offer a way of developing strategy (2). That is, a state might become conscious not in virtue of actually feeding into higher-order planning, but rather in virtue of its potential for feeding into such planning. Rolls mentions the possibility of such an approach (p. 248). This modification of the theory, however, would entail abandoning the account of how qualia emerge. It is also rather implausible: Why should what might or might not happen to sensory information further downstream affect whether it is conscious or not? It is more plausible that the order of explanation is the other way round, so that it is only if a state is conscious that it can feed into planning and the deliberate initiation of action. An important part of the functional role of phenomenal consciousness seems to be to make information available for higher-order planning and reflection. This is consistent with what we know from blindsight patients and perceptual masking experiments. Certainly it seems more likely that blindsight patients cannot deliberately plan actions involving objects in their blindfields because the information they possess about those objects is non-conscious, than that the information such patients possess about objects in their blindfield is non-conscious because they cannot deliberately plan actions involving those objects (Bermúdez 1998b; Van Gulick 1994). If this is right, then our account of phenomenal consciousness cannot rest solely upon the operation of higher-order thought.

The upshot of all this, I think, is to cast doubt, not on Rolls's general distinction between two types of action and corresponding neural pathways, but on his proposal to use that distinction to distinguish conscious from non-conscious states. This in turn casts doubt upon his provocative suggestions, first, that phenomenal consciousness does not emerge in evolution before the capacity for higher-order thought and, second, that stimulus-reinforcement learning is independent of phenomenal consciousness.

Conceptualizing motivation and emotion

Ross Buck

Department of Communication Sciences, University of Connecticut, Storrs, CT 06269-1085. **buck@uconnvm.uconn.edu**
wattlab.coms.uconn.edu/faculty/buck.htm/

Abstract: Motivation and emotion are not clearly defined and differentiated in Rolls's *The brain and emotion,* reflecting a widespread problem in conceptualizing these phenomena. An adequate theory of emotion cannot be based upon reward and punishment alone. Basic mechanisms of arousal, agonistic, and prosocial motives-emotions exist in addition to reward-punishment systems.

Rolls presents a detailed account of brain mechanisms of emotion, emphasizing those aspects involving reward and punishment, and exploring mechanisms ranging from those of drug addiction and sperm competition, to consciousness. My review concentrates on two points concerning the conceptualization of emotion: the differentiation of motivation and emotion, and the feasibility of constructing an adequate theory of emotion based upon reward-punishment alone.

A recurring problem with conceptualizing emotion concerns the differentiation of motivation from emotion. In this book the relationship between motivation and emotion is not made entirely clear, and the reader must search out the author's assumptions. Rolls defines "motivated behaviour" as "present when an animal . . . [performs] an arbitrary operant response to obtain a reward or to escape from or avoid a punishment" (p. 3). In his definition, motivation requires learning, either classical conditioning or instrumental learning, and behavior reflecting reflexes or instincts would appear not to qualify as motivated behavior. This definition is less general than the common definition of motivation in terms of the arousal and direction of behavior (Kleinginna & Kleinginna 1981). Rolls defines "emotions" as "states elicited by rewards and

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

punishers, including changes in rewards and punishers" (p. 60). It appears from these definitions that motivation would be one of the states included under the term "emotion." At another point, Rolls lists motivation as one of the functions of emotion. He writes, "emotion affects motivation" (p. 68). That is, "fear learned by stimulus-reinforcer association formation provides the motivation for actions performed to avoid noxious stimuli . . . [and] positive reinforcers elicit motivation, so that we will work to obtain the rewards" (p. 68).

In elaborating upon his definition of emotion, Rolls notes that "there are some rewarding stimuli that some may wish to exclude from those that cause emotional states" (p. 65), but he does not offer specific exclusion or inclusion criteria. He writes, "when positively reinforcing stimuli (such as the taste of food or water) are relevant to a drive state produced by a change in the internal milieu (such as hunger or thirst), then we do not normally classify these stimuli as emotional, though they do produce pleasure" (p. 65). He notes, however, that some may have a greater emotional response to savoring food than others. Elsewhere, he distinguishes motivational and emotional qualia (defined as "raw sensory feel"), giving the example of hunger for the former and pleasure induced by touch for the latter (p. 251). Rolls notes that sex constitutes something of a special case: "such stimuli may be made to be rewarding . . . partly because of the internal hormonal state. Does this mean that we wish to exclude such stimuli from the class that we call emotion-provoking . . . ?" (p. 65). He acknowledges that there is no clear answer to this question, although unlike many emotion theorists it appears that Rolls's answer is clearly "No," because he includes an excellent chapter on sex.

Rolls's differentiation of motivation and emotion is generally consistent with the customary way that these terms are distinguished, but it also reflects the fundamental incoherence of this common view. I suggest it is not possible to distinguish coherently between motivation and emotion because they are aspects of the same phenomenon, two sides of the same coin, which by definition always occur together. I have defined motivation as a potential inherent in a system of behavior control, and emotion as the manifestation or "read-out" of motivational potential. An analogy is the relationship of energy and matter. Energy is a potential that is not seen; rather, it is manifested in matter: in heat, light, force. Analogously, motivation is not seen but is manifested in emotion: in arousal, expression, experience. In this view, instincts and even simple reflexes qualify as primary motivational-emotional systems (*primes*), existing in even the simplest single-celled creatures (Buck 1985).

My second point concerns the adequacy of reward and punishment alone in accounting for emotional phenomena. Rolls criticizes other approaches to emotion on the grounds that his description in terms of reward and punishment terminology is "more precisely and operationally specified" (p. 61). Indeed, his theory is grounded in the relatively explicit and widely known terminology of classical conditioning and instrumental learning as well as brain research. Also, his explication of reward-punishment systems in the brain and their implications for emotion may be correct as far as it goes. However, such an approach cannot alone account for motivation-emotion. In his book Rolls discounts or overlooks some sources of contradictory information. For example, he disagrees with LeDoux's (1996) emphasis on the role of subcortical inputs to the amygdalae in fear conditioning, what LeDoux termed the "low road." Instead, Rolls emphasizes the role of cortical inputs to the amygdalae in associating stimuli with primary reinforcers, both rewards and punishers, although LeDoux agreed with the importance of this "high road" as well. Moreover, Rolls does not consider Panksepp's (1998a) work on, for example, attachment emotions.

I suggest that there are three fundamental bases of motivation-emotion, or primes, which cannot be reduced to reward-punishment. These are arousal, agonistic, and prosocial primes. All creatures manifest approach and avoidance behaviors, but they also manifest changes in activity-quiescence, agonistic competition, and prosocial cooperation. There are numerous examples of colonial microorganisms – even simple prokaryotic myxobacteria – that can either separate and live "competitively" as single individuals, or "cooperatively" aggregate with differentiated parts and functions to form a more complex multicelled organism (Lackie 1986; Losick & Kaiser 1997). The latter fundamental cooperation – arguably involving emotional communication based upon communicative genes – has important implications for evolutionary theory regarding empathy and altruism (see Buck & Ginsburg 1991; 1997).

These fundamental motivational-emotional systems or primes are based upon "informational molecules" evolved thousands of millions of years before the evolution of the brain. They are, in effect, cognitive systems – ways of knowing – in their own right. I suggest that reward-punishment systems are associated most closely with happiness, sadness, and anxiety; agonistic systems with fear, anger, and disgust; and prosocial system with a panoply of attachment-related emotions including love (see Buck 1999).

Roads not taken: The case for multiple functional-level routes to emotion

Tim Dalgleish

Medical Research Council, Cognition and Brain Sciences Unit, Cambridge CB2 2EF, United Kingdom. **tim.dalgleish@mrc-cbu.cam.ac.uk**

Abstract: This review focuses on the theory of emotion outlined in Chapter 3 of Rolls's *The brain and emotion.* It is proposed that Rolls's emphasis on a relatively simple neurobiologically derived emotion scheme does not allow him to present a comprehensive account of emotion. Consequently, high-level cognitive processes, such as appraisal, end up being retained in the theory despite Rolls's skepticism about their utility. An argument is put forward that the concept of appraisal in the emotion literature is more than semantic convention and actually allows us to talk about multiple functional-level routes to the generation of emotion – a characteristic of the latest generation of theories in the cognition-emotion literature.

Rolls offers us a wide-ranging and timely theoretical analysis of the burgeoning research on affective neuroscience. The book offers a comprehensive assessment of the basic brain and behaviour literature associated with emotion, while at the same time offering a wealth of new ideas in domains as disparate as consciousness, morality, literature, and neurotransmitters.

A cornerstone of the book is Rolls's theory of emotion that is presented in Chapter 3. In this review I confine myself to two separate but related comments concerning the theory. In brief, Rolls proposes that emotions can be defined as "states produced by stimuli which can be shown to be instrumental reinforcers" (p. 62). Different emotions, he argues, can then be defined and classified according to whether the reinforcer is negative or positive and by the reinforcement contingency that presides; that is, presence of a reinforcer, omission of a reinforcer, or the termination of a reinforcer. The strength of the reinforcement contingency determines the intensity of the emotion (see Fig. 3.1, p. 63). So, for example, the omission of a positive reinforcer can lead to sadness or, more intensely, grief.

This basis of Rolls's theory has an elegant simplicity and he favourably contrasts his ideas with the proposals of cognitive appraisal theorists (e.g., Frijda 1986; Oatley & Johnson-Laird 1987) in which emotions are a function of a cognitive evaluation of the meaning of a given event in terms of the organism's goals or current concerns. Rolls prefers his own system to the idea of appraisals because it "simply seems much more precisely and operationally specified" (p. 61). I will try to argue that Rolls does away with the idea of appraisals only to reintroduce them again by the back door. Furthermore, I will suggest that appraisal can be usefully thought of as a separate functional-level route to the generation of emotion, contrasting with emotions that are derived from learned associations.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Rolls justifies his sidestepping of appraisal theory in terms of the versatility of his own emotion scheme. The crucial question here is how well any theory of emotion can account for the range of different emotions that is possible. Rolls's basic classification scheme, as he emphasises, can account for a wide range of emotions. This range is further increased by the introduction of other ground rules for the production of emotions (to augment the simple parameters described above), namely: (1) any environmental stimulus can have a number of different reinforcement associations; and (2), emotions can differ as a function of their original primary reinforcers, even if the secondary reinforcers are identical.

The taxonomy of emotions accounted for by this basic set of assumptions is indeed impressive. However, as Rolls implicitly acknowledges, it is not complete and so two further rules are included to enable a more comprehensive account of the variety of emotions. First, emotions are proposed to differ as a function of whether the cognitive evaluations associated with the perceptions of the eliciting stimuli are different. Second, emotions are proposed to vary as a function of how the environment at the time constrains the type of behavioural response that can be made (p. 64).

The latter two ground rules look very familiar to an appraisal theorist such as myself! Essentially, the proposal is that for a complete taxonomy of emotions there needs to be scope for an appraisal of the meaning to the individual of the eliciting stimulus and for an appraisal of the opportunities provided by the environmental milieu. With this analysis, Rolls leaves us with a simple reinforcement contingency scheme allied to an appraisal system that is necessary for certain types of emotion and/or fine tuning of more primary emotional responses. This is not altogether different from appraisal theories that emphasise primary and secondary levels of appraisal (see Scherer 1999, for a review) and it certainly begs the question of whether Rolls's system is "more precisely and operationally specified" than appraisal-based accounts.

However, it is possible that there is a more interesting issue lurking in the shadows here. More recent cognitive accounts of emotion (Johnson & Multhaup 1992; Leventhal & Scherer 1987; Power & Dalgleish 1997; Teasdale & Barnard 1993) emphasise the utility of thinking of more than one functional route to the generation of emotions. For example, in the SPAARS model of Power and Dalgleish (1997) an associative route to emotions is proposed that is based on previously learned contingencies between stimuli and emotional responses as a function of parameters such as reward and punishment. In addition, an appraisal route to emotions is outlined that is a function of "on-line" evaluation of the meaning of a current stimulus in terms of the organism's active goals and concerns. It seems plausible that Rolls's scheme for emotions goes against the flow of recent cognitive theorizing and conflates these two routes. This requires him to propose a set of ground rules, some of which concern previously learned/conditioned contingencies and some of which concern aspects of on-line evaluation. Separating these routes to emotion at a functional-level has neural plausibility (e.g., Izard 1993; LeDoux 1995), as is clear from the literature reviewed in the rest of the book, and provides a more comprehensive account of the generation of mixed feelings and emotional conflict than is offered by a scheme that articulates a single functional-level route, however elaborate.

In summary, I have argued in this review that, despite the emphasis on a simple neurobiological scheme for emotion, Rolls finds himself (inevitably in my view) reintroducing the concept of appraisal in order to make his account of emotions a comprehensive one. I have further tried to suggest that such labeling of something as an appraisal process might be more than just a semantic preference and may reveal something fundamental about the way emotions are elicited in the mind under certain conditions.

Affect programs, intentionality, and consciousness

Craig DeLancey

Department of Philosophy, Program of Cognitive Science, Indiana University, Bloomington, IN 47405. **cdelance@alumni.indiana.edu**
www.cs.indiana.edu/hyplan/cdelance

Abstract: I express two concerns with the theory of emotion that Rolls provides: (1) rewards and punishers alone fail to explain the basic emotions; (2) Rolls needs to clarify his notion of the intentionality of emotions. I also criticize his theory of consciousness, arguing that it fails to explain qualia, and that ironically it is emotions which make this most evident.

Rolls's excellent book offers a parsimonious approach to understanding emotions and a compelling synthesis of neuroscientific evidence in defense of this position; it also offers a theory of consciousness. I have two concerns, and some criticisms.

My primary concern is that the characterization of emotions as "states elicited by rewards and punishers" is so broad that it can support a view that all emotions are of a generic kind, and that nothing significant and essentially affective distinguishes emotions like fear or anger from states like pleasure or relief. Many of us who believe in some version of the affect program theory, for example, hold that some emotions have as a constituent motor "programs," including perhaps facial expressions and other expressive behaviors but also perhaps relational activities. There are a number of reasons to believe this. For example, just as there are primary reinforcers (which need not be learned but arise from our inherited biological structure), there appear to be primary actions (like flight, attack, grooming, and so on). Also, this view is part of a reasonable phylogenetic theory, in which some emotions evolved from inherited capabilities for particular actions, and furthermore some of the physiological responses accompanying these emotions may have evolved to facilitate these actions. Allowing for action programs can explain the existence and even utility of certain kinds of emotional behaviors which are irrational and so presumably should not arise, on a simple cost-benefit analysis: expressive behaviors (e.g., kicking a tree when you are mad at your boss); post-functional behaviors (e.g., fleeing farther than you know is necessary from a rattlesnake); and akratic behaviors (e.g., avoiding a medical checkup because you fear the results). Post-functional activities, for example, should be unpredictable if emotions were guided only by avoiding punishment or getting reward, because the action should cease when either is accomplished. Frank (1988) has provided an analysis of how some of these behaviors can indirectly result in outcomes which are beneficial

The claim that some emotions are in part constituted by motor programs, or other theories that there are fundamentally distinct emotions, may be compatible with Rolls's framework, but Rolls appears to want to deny such approaches. Thus, for example, he downplays evidence that the amygdala plays a greater role in fear than in some other emotions, in support of a more generic framework where the amygdala processes primary reinforcers and some secondary reinforcers (pp. 102ff).

My second concern is that Rolls needs to clarify his notions about the intentionality of emotions. Rolls argues that emotions are normally object-directed intentional states. I believe he is correct about this, but humans can have emotions not only toward concrete objects ("Eric is afraid of that snake") but also toward events or states of affairs ("Adam fears that he will flunk the exam"). This means that we need to explain not only how we can represent and recognize concrete objects invariantly, but also how we can represent and recognize events or states of affairs. Furthermore, in distinguishing mood from emotion by claiming that emotions are moods with intentional objects (p. 62), Rolls allows intentional states toward concrete objects; but elsewhere (p. 261) he uses "intentional" in a non-standard way, to mean "states with intentions, beliefs, and desires." This latter formulation would make most instances of emotions into moods, so presumably it is

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

a mistake; it betrays his very strong emphasis on linguistic forms of thought. This emphasis is surprising both for a connectionist, and for an expert on emotion. It may also explain why he downplays the role of subcortical pathways for eliciting emotions (e.g., p. 104).

My criticisms are of Rolls's theory of consciousness. It cannot explain phenomenal experience; ironically, emotions offer the best evidence for this. Affective experiences can be very different (happiness is quite different in its experience than is fear) and can admit of wide degrees of intensity (rage is much more intense than annoyance). There is no place for either property in the higher order linguistic thought (HOLT) theory of consciousness that Rolls endorses. If I understand him correctly, qualia come in as first-order thoughts, which are conscious when second-order thoughts about them occur. This approach offers a reasonable explanation of some notions of consciousness (e.g., reflective cogitation about one's symbolic cognitive states), but not of phenomenal experience. Rolls recognizes the distinction between phenomenal experience and other explicitly functional notions of consciousness (pp. 244–45), but he also fails to respect it (thus, he erroneously accuses Chalmers of inconsistency by comparing Chalmers's discussion of awareness with Chalmers's discussion of phenomenal experience; see pp. 248–49 and the footnote on p. 249). His arguments that HOLT accounts for qualia can be read as playing on this ambiguity; for example, he argues that the HOLT system can use sensory information to plan, and so sensory qualia should be conscious (p. 251). But, for such use, sensory states need only be conscious in that some symbol of them is active in the syntactic system; this provides no explanation of why they would have the phenomenal nature they do. Symbols, by definition, are merely tokened or not; thus, there is no magnitude for intensity to be had. Also, *prima facie*, thinking about 2 and thinking about 3 should be no different than feeling fear and feeling anger (all of them are symbols tokened in the system and can act to help shape reflective cogitation like planning), whereas we know the former are not significantly distinct experiences and the latter are.

Furthermore, Rolls argues that the symbols of the HOLT system must be "grounded" in the world (p. 251) (he is attempting to avoid the consequence that, say, a C compiler written in C and compiling its own source code is therefore conscious). But this will amount to an extraordinary kind of externalism concerning phenomenal experience; systems will be conscious not because of their individual structure alone but also because of their history and context. Finally, HOLT is likely inadequate not just as an account of phenomenal experience, but also of self-awareness; Panksepp (1998a, p. 308) has observed that the coherent affects, intentions, and activity of split-brain patients strongly counsels against associating conscious awareness with any lateralized ability such as language, and rather suggests that the core of such awareness lies in subcortical affective and motor processes.

Emotional networks: The heart of brain design

John C. Fentress

Department of Psychology, University of Oregon, Eugene, OR 97403.
fentress@is.dal.ca

Abstract: The concept of emotion as defined by Rolls is based upon reinforcement mechanisms and their underlying neural networks. He shows how these networks process signals at many levels, through both separate and convergent pathways essential for adaptive action. While many behavioral issues related to emotion are omitted from his review, he succeeds admirably in summarizing both the "current state of the art" in single unit analyses and in pointing out how future research directions may be crafted.

My reading of Rolls's book is that of an ethologist, with a strong interest in the temporal structure of behavior and its underlying brain and developmental substrates. In ethology, motivation is often described operationally in terms of changes in responsiveness to specific environmental stimuli (Hinde 1970). These changes alter the selection and performance properties of individual actions and their rules of combination in time (Fentress 1990; 1991). In the present book Rolls reviews a number of underlying brain operations that are highly relevant to these behaviorally defined concerns. He views emotional behavior, with its base in the balance between rewards and punishments, as being "at the heart of brain design." In a broad sense he echoes Darwin's (1872) concern with the expression of emotions, but with the analytical tools of modern techniques for single unit analyses of individual networks and their combinatorial operations.

The text is not comparative in the ethological sense, but focuses upon the primate brain as a model for human emotional expression. Nor is this the place to find a detailed phenomenological account of rules by which often conflicting action tendencies are integrated. Rather, Rolls takes the framework of objectively defined rewards and punishments to demonstrate, for individual classes of behavior, the various stages of neuronal processing that together make effective action possible. He does this both in the sense of action coherence and flexibility in performance. With respect to the latter, Rolls borrows heavily from classical learning theory in which associations between sensory and response events are mediated through positive and negative reinforcers. In the latter part of the book (especially Ch. 8 on sexual behavior) he attempts to anchor these associative and motivational networks into issues of adaptive value and genetic selection.

Rather than review the whole book (which I admire) I shall concentrate upon a few themes and problems that I believe are broadly relevant. One of the fundamental questions in brain-behavior design is how individual events (actions, neural pathways) are both kept separate and combined in integrated action. A related question is how these events establish an adaptive balance between stability and flexibility in performance properties. Each question can be examined at a number of complementary levels, and across multiple time frames. Often the answers are both time frame and level dependent.

In many respects Rolls handles each of these questions extremely well. By documenting stages of processing for various motivational systems, such as hunger (Ch. 2) and thirst (Ch. 7), he skillfully dissects the multidimensional nature of complex behavior, with its separate pathways and converging operations. He shows how basic (e.g., sensory) processes can remain invariant in their properties while "higher-order" processing stages can be modulated reversibly in time, and also demonstrates how associative mechanisms can lead to long lasting changes in performance. For the ethologist much of this material may appear to be presented in a dense form, and I suspect that additional summary diagrams would have been helpful. This is particularly true for students new to the area. However, the effort in reading produces its rewards. Rolls demonstrates, through a number of lovely examples largely based upon work in his laboratory (nearly 100 references to Rolls as first author alone!), how integrative dynamics in the brain relevant to whole organism behavior can be traced in a systematic manner.

My personal favorite chapter is number six. Here Rolls examines pharmacological and chemical substrates of reward and its neural output systems, with particular reference to the basal ganglia. It is well worth thinking about these circuits in comparison to dynamics often seen at the intact, freely behaving, organism level. Take, for example, the following puzzle: brain circuits, as well as behaviorally defined operations *must* on the one hand be isolable (independent) and on the other hand joined together (integrated). This means that systems have both self-organizing (intrinsic) properties and rules of interaction that cross these systems. How can brains do both of these things? This is not a simple dichotomy. Rather it is a statement about the *relative* balance among intrinsically ordered and interactive processes that individually and collectively are *dynamically ordered* (change in time) – but in a con-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

strained way! Indeed, it is precisely here, with these two organizational polarities in action, that most ethological and many psychological models of motivation find their fundamental puzzles (Fentress 1991).

Without neurobiological data behavioral models must remain abstract (and incompletely satisfying). They can, for example, be expressed in "information space" terms. One such conceptualization is to view integrative systems as consisting of core excitatory processes and surround inhibitory processes. This means that a given system tends to block the expression of other systems. Further, going back to the early work of Tinbergen (1951) and others, it is clear that many expressive systems in behavior can be either broadly or narrowly focused (e.g., appetitive behavior versus consummatory acts, respectively). This dynamic focusing provides an alternative view to static hierarchical models, in that the very boundaries of a behavioral control system can be modeled in terms of broadening and narrowing cores, and variably extending lateral inhibitory pathways. The basic dynamic is that as systems become more strongly activated, the selectivity in both response to specific stimuli and details of action performance can become narrowed (focused), insulated from disruptive inputs, and more effective in blocking alternative forms of action (Fentress 1990). Given the overall dynamics, multiply defined systems can either operate independently, synergistically in combination, or antagonistically (via the shifting dimensions of core excitatory processes and surround inhibitory processes, behaviorally defined).

But do brains actually do this? Certainly it is known that receptive field properties are often of a center-surround nature, and that these properties can be dynamically ordered (Gilbert 1995). Through an impressive combination of raw data and modeling (cf. Rolls & Treves 1998: a valuable companion text on neural modeling) Rolls and his colleagues are among those who have shown how basal ganglia circuits might operate in adaptive action sequences. In brief, these circuits are seen to translate motivational and cognitive processes into action. To do so they both collect converging signals (e.g., from the cortex) and then separate these collections through lateral inhibitory pathways in the process of response selection. It would be interesting indeed to know whether these "core" and "surround" properties have different relative thresholds, for the behavioral data suggest that many behavioral systems are broadly focused during early and weak activation, only to become more tightly and narrowly structured in time, and with stronger activation (Fentress 1991; 1999).

Obviously the jump between behavioral models of intact organism actions and neurobiological data on individual circuits remains a large one, often involving at least tentative leaps of faith. However, careful reading of Rolls's book shows how the gap is becoming narrowed, at least in principle. What is particularly important is his explicit realization that behavioral and neurobiological data are necessary complements. As Hebb pointed out many years ago (Hebb 1949) it is problems of *organization* in behavior that sets the most important questions for brain operations, and brain operations that mediate the organizational processes seen at the behavioral level.

We are still left with many basic conceptual as well as analytical puzzles. For example, most dictionaries and laypeople define emotion in terms of its affective (consciously experienced) properties, and this can seem a long way from the mechanics of brain operation. To his credit Rolls attempts to grapple with issues of consciousness (Ch. 9), through such linguistic maneuvers as "thinking about thinking." Indeed, he links conscious experience within the linguistic system. I suspect this does injustice to the layers of conscious experience (e.g., being aware, aware that one is aware, caring that one is aware, etc.), but one can see where his arguments are going. I am personally not convinced that he or other recent authors (e.g., Crick 1994) have succeeded in joining the phenomenological aspects of awareness with its mechanical underpinnings, but at the very least they have shown how brain machinery can mediate the phenomenology of emotion and related mental states.

While the first chapters of this book are densely neurobiological, the latter chapters, beginning with chapter eight on sexual behavior, offer a rather abrupt departure in both substance and style. Indeed, chapter eight is largely sociobiological in its tone, with the explicit intent to show how genes can code for adaptive performance (defined in gene selection terms). Although I believe that I understand the goal of this review (to link genes to adaptive performance, and thus to emotional behavior) there is no real mention of developmental trajectories that mediate these presumed gene actions. These trajectories are themselves both isolable and interactive, with both flexibilities and constraints that we are just beginning to glimpse (Fentress 1991). It is not enough to say that genes work, and for the most part work well. I have a personal revulsion to classic separations of "innate" versus "learned," "primary" versus "secondary" reinforcers and the like because these dichotomous constructs completely bypass the puzzle of how genes and experience join together, starting with the embryo. A true evolutionary perspective cannot exist without attention to ontogenetic molding of brain and behavioral events (cf. Hall 1992).

Perhaps my main critique of Rolls's book is that some of its conceptual underpinnings approach the tautological: animals do some things more than others, and differential reinforcement can mediate these differences (but is also defined by the very behavioral consequences they are presumed to produce!). Genes are operated on by natural selection, and in animals genetic differences influence brain and behavior, in presumably adaptive ways. To say that animals do what they do because they are responsive to reinforcers which in turn have adaptive value, is more of a setting out of puzzles than a solution to these puzzles. However, neither a single author nor a single book can do everything. I suspect that the present contribution by Rolls will be applauded by many scientists interested in "the heart of" brain and behavior. It is my hope that my ethological friends will also work through the book to see how their interests and those of neuroscientists can converge in the future.

Emotion theory?

Nico H. Frijda

Faculty of Psychology, Amsterdam University, 1018 WB Amsterdam, The Netherlands. **pn_frijda@macmail.psy.uva.nl**

Abstract: The book contains a masterly review of Rolls's single-neuron research reflecting rewards. It places that research in the context of the neobehaviorist theory of emotions. That theory provides a useful first approximation to emotion-eliciting conditions but has little to tell about emotions as motivational states or response dispositions: nor does it give a rationale for what are considered to be primary rewarding stimuli.

The emotion theory that serves as background for neurological exploration is a version of the neo-behaviorist reinforcement theory of emotions of Millenson (1967), Mowrer (1960), and Gray (1987). Emotions are seen as the outcomes of positive and negative rewards or reinforcements, primary or secondary. Different emotions depend upon delivery, termination, or omission of these reinforcements, their intensity, and upon availability of responses. It is a sound theory, as far as it goes. In some form or other, the scheme is fairly generally accepted. In its basic assumptions it resembles cognitive emotion theory, in which "primary appraisal" is what turns stimuli into primary or secondary reinforcements, and "secondary appraisal" represents the cognitive processes responsible for detecting the various contingencies. Rolls's theory also shares with most current theories the adaptational view: emotions exist for dealing with major adaptational dilemmas. Pleasure and pain (or reward and punishment) serve as common currency in goal priority fixings. Emotions allow goal-oriented, that is, flexible behavior instigation by relevant events, instead of mere rigid fixed action patterns.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

The behaviorist scheme has a certain utility for ordering the data on emotion instigation, as well as for ordering neurobehavioral findings and guiding research. The book, for instance, provides suggestions for different functions of the establishment of emotional appraisals (centered around the amygdala) and the correction of existing appraisals (centered around the orbitofrontal cortex), with both types of processes showing different parameters. And of course the detailed and careful review of the admirable research tracing the effects of "rewards" through the cortex, as well as that concerning the processing of facial identity and facial expression, are fascinating, also for the emotion theorist.

At the same time, there is much that leaves such a theorist puzzled and dissatisfied. This applies first of all to the central notion of reward. Emotions are defined as "states elicited by rewards and punishments"; rewards and punishments are defined as stimuli the individual will work to obtain or to avoid. The states thus are defined in terms of their antecedents and (certain of) their consequents. No consequences are described, either of emotional states in general or of those corresponding to the different contingencies. No characterizations are given of the states themselves. This also applies to "reward." Neither reward nor the resulting states are defined in terms of process or of functional properties, other than a possible ultimate learning effect. How can an effect upon future learning serve as a common currency?

There is thus no indication of what it means to have received a reward when no response has as yet been evoked. It remains unclear how an emotion can be motivating and constitute a goal, after the eliciting stimulus has disappeared (as, for instance, in sadness); nor is there a discussion of possible neural consequences given the different reinforcement contingencies, or how and where they influence response selection.

In part this is due to a basic trait of the neobehaviorist model, namely, its being focused upon operant learning. No reference is made to the evidence for innate emotion-specific and species-specific behavioral programs and "expressive" behaviors, or their underpinnings in the brain. "How does the brain produce behavior?" remains an unanswered question, both in the emotion theory and in the review and interpretation of neuropsychological findings. This extends to the absence of discussion of the mechanisms for motivation and attention regulation. By "motivation" I here understand, as does Rolls, the states of latent response readiness, the orientation towards behavior to come and towards stimuli to obtain or avoid. For that, others developed the notions of tonic readiness for action (Pribram 1981), activation (Gray 1987), seeking (Panksepp 1998a), or "action readiness" (myself). Such notions seem indispensible for understanding the persistence in efforts at reaching goals, as well as for distinguishing the phenomena usually referred to as "emotions" from those of mere like or dislike (as evident in preferences or quantities consumed). They are also indispensible for understanding the occurrence of undirected, "excited" behaviors and unspecific physiological arousal, or the split between liking and wanting that Berridge (1999) has experimentally demonstrated. It also seems unsatisfactory to subsume all forms of non-response under the one heading of non-reward, instead of distinguishing, at the basic process level, true non-reward or extinction, punishment-instigated response suppression, response postponement, and response inhibition, which may (or may not) suggest brain provisions for effort and inhibition.

I also think that identifying "reward" (that is, pleasant events) as originating in primary reinforcing stimuli is unsatisfactory. Many rewarding conditions do not fit the notion of "stimuli." Instances include "altruism," "courtship," and "solving an intellectual problem" that occur in the (somewhat gratuitous) listing of primary reinforcers on pp. 272–73. Obviously, conditions for whatever in those domains corresponds to primary reinforcers are complex and of unknown neural structure. Moreover, many rewards may not come from evolutionary determined stimulus sensitivities, but from the completion of actions; this, too, would seem to require different sorts of neurological theorizing.

Furthermore, certain emotional states are independent of the occurrence of rewards or punishments, for instance, many moods and the joys of young animals that underlie rough-and-tumble play. Emotions may not just equal the instigations for solving adaptational dilemmas. They are, I would say, instigations for maintaining or changing relationships with the environment, whether for solving adaptational dilemmas, for fitting in with elementary social satisfactions, or for answering relational urges that just are part and parcel of certain biological autonomous systems.

These issues surrounding "reward," innate forms of behavioral instigation, and motivational phenomena are essential in understanding emotions, but remain mysteries in the emotion theories that are limited to a reinforcement basis. The lacunae are clear in the somewhat embarrassing chapter on sexual behaviour, reward, and brain function. It is for this reason that I think the book's title is not well chosen. The title of Rolls's previous book, *The brain and reward*, was both more modest and more appropriate, as it would have been for this volume.

Structuring an emotional world

Jordan Grafman

Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health MSC 1440, Bethesda, MD 20892-1440. **jgr@box-j.nih.gov** **intra.ninds.nih.gov/mnb/cns/index.html**

Abstract: Rolls emphasizes the role of emotion in behavior. My commentary provides some balance to that position by arguing that stored social knowledge dominates our behavior and controls emotional states, thereby reducing emotions to a subservient role in behavior.

Rolls has written a very stimulating book that led me to rethink my opinion about how emotions interact with behavior. As a result, I have chosen to focus my comments on a particular interest of mine – the relationship of emotion to knowledge stored in the human prefrontal cortex – that is discussed in several places in his book including Chapters 4, 9, and 10.

I do not take issue with much of Rolls's description of the role of emotions in behavior but I do dispute the importance of emotion as conceived by Rolls. Rolls shows that emotions are states, have adaptive value because of their labeling of appropriate and inappropriate actions through reinforcement and punishment, and are a motivational force in obtaining the goals of the organism. Rewards and punishments are seen as the goals for which the organism develops action plans. While these claims may hold for primates, there is room for argument if Rolls wants to extend this characterization of the role of emotions to humans. In a nutshell, my point of view is that emotions are prone to instability, given the multiplicity of cues in our environment eliciting excitation and inhibition. The activation of stored social knowledge stabilizes and structures the world so we may navigate through it without depending on the immediacy of emotional reaction (Grafman 1995; Partiot et al. 1996). Therefore, while emotion may color the canvas, the scene is painted by cognitive knowledge.

The human prefrontal cortex plays an important role in interacting with brain structures concerned with emotion. For example, as Rolls points out, there are direct connections between orbitofrontal cortex and the amygdala in the primate. Neuroanatomical connectivity while specified for the monkey (Barbas 1995) is much less specified for humans, which leaves a "neuroanatomical hole" in which a number of speculative inferences may fall. It is relevant to an understanding of the relationship between emotion and cognition that the human orbitofrontal cortex also has many connections to and from dorsolateral prefrontal cortex and other association areas besides limbic structures.

There is evidence from the human cognitive neuroscience literature that the prefrontal cortex can be subdivided into regions that are concerned with mechanistic versus social cognition, with adaptive versus predictive behavior, and with fine versus coarse

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

coding of knowledge (Grafman et al. 1995). My own view is that the prefrontal cortex stores knowledge that is abstracted from ongoing events (such as thematic features) and that this knowledge would include both semantic information and grammatical rules (Grafman 1995). Recent research suggests that the right prefrontal cortex stores information that needs to be abstracted across events (i.e., coarse coding) such as themes and morals (Nichelli et al. 1995). The left prefrontal cortex stores the individual event (fine coding) characteristics that make up a structured event complex (SEC) and the grammatical structure of events (Nichelli et al. 1995). The ventromedial prefrontal cortex appears concerned with storing domain specific social knowledge whereas the dorsolateral prefrontal cortex is concerned with storing domain specific non-social knowledge (Partiot et al. 1995). Frontopolar and anterior lateral prefrontal cortex appear particularly important for storing events occurring in ill-structured environments whereas medial prefrontal cortex appears particularly important for storing events in well-structured environments (Koechlin et al. 1999). It is my contention that these forms of distributed knowledge, when linked together, make the major contribution in managing our day to day plans, allowing for conceptual understanding, and structuring our social behavior.

Individual social experience is unique, although clearly culturally shaded, which makes the ventromedial prefrontal cortex the seat of the most individual of knowledge stores. I agree with Rolls that there is a genetic contribution permitting social knowledge that is dependent upon a sequence of events to be stored in the prefrontal cortex (i.e., the ability to abstract knowledge from trains of events). I would add, however, that the identity of the specific social memories stored in prefrontal cortex must be driven by experience and are not constrained by genetics. The encoding, consolidation, and retrieval of these social memories are selectively disrupted by focal lesions in the ventromedial prefrontal cortex (more on this below). How might this proposed social knowledge system interact with emotional states?

Emotions are states that are bound to, and modulated by, social knowledge. Since social knowledge activation quite frequently may have to be sustained from minutes to hours, this sustained activation must provide an inhibitory input to the neural structures governing emotional states in order to modulate the emotional input that is required to maintain the necessary motivation to attain a long term goal. The goal I am referring to, however, is an event that is stored as part of a structured social event complex in the ventromedial prefrontal cortex (Grafman 1995). It is the activation of a structured social event complex that inhibits, modulates, and otherwise dictates the intensity of the emotional states we experience. This inhibition serves to put off immediate gratification (that is dependent upon the magnitude and intent of immediate sensory input) in favor of long-term goals.

Contrary to one of Rolls's arguments, I believe there is little evidence to suggest that we need a ventral visual system to provide the appropriate input for planning future actions because, for example, blind humans do quite well in developing and executing social plans stored in ventromedial prefrontal cortex.

What happens when the prefrontal cortex is damaged in humans and its inhibition of emotional reaction is reduced? The patient becomes more likely to respond to environmental provocation from objects or faces or the ramifications of a single (as opposed to more complex) thought. This dependency on the moment has negative consequences that are characterized by increased aggression, inappropriate sexual content in social interactions, and a dramatic susceptibility to distraction (Dimitrov et al. 1996; Grafman et al. 1996; Rueckert & Grafman 1996; Sirigu et al. 1995; 1996). Patients begin to rely on the present because they have difficulty accessing the social plans and rules that insure that they remain on-task towards achieving one or more goals. Deficits in social behavior following frontal lobe lesions are not simply a response to changes in reinforcement contingencies. Patients with prefrontal lesions are also impaired in retrieving the stored rules of behavior which, when activated, provide a counterpoint to the vicissitudes of emotionally laden reinforcement contingencies. There is also evidence that some patients whose social behavior is inappropriate are nevertheless able to use verbal systems to articulate what the appropriate social behavior would be in a particular situation. This observation indicates that social behaviors are made up of a fabric of cognitive processes with each component process contributing to the whole of the behavior and subject to selective impairment (Sirigu et al. 1998). So although emotions can serve to modulate the activity of a structured social event complex, I would argue that emotions are more typically restrained by the execution of social behavior (which not only represent emotionally arousing activities but also more mundane plans such as preparing a routine breakfast or taking the bus to work – while these latter structured social event complexes include goals, they are hardly the stuff of instant reward and punishment scenarios).

Humans live in environments that vary between their being ill-structured and well-structured (Goel et al. 1997). Human thoughts and ideas likewise may vary between the ill-structured and the well-structured. Under most circumstances, humans strive to structure the world around them and their thoughts through planning and understanding. If emotion merely modulates the structuring of our behavior, have we over-emphasized its importance in social behavior in contemporary cognitive neuroscience relative to the role that knowledge stored in ventromedial prefrontal cortex plays in human social cognition?

As scientists, we often take risks and chances, and challenge the established point of view in order to evoke change. If the neural structures governing our emotions are intact, could we still determine that taking a risk is worthwhile when the structured social event complex that contains the knowledge that is required to make that determination is unavailable? My answer is, no. In my view, emotions are like fire and water, good servants but bad masters. The ventromedial (and dorsolateral) prefrontal cortex by storing knowledge based upon the coding of structured social event complexes insures that social knowledge can dominate our behavior and control emotional states which reduce emotions to a subservient but contributing role in civilized behavior.

Adaptive accounts of physiology and emotion

Alasdair I. Houston[a] and John M. McNamara[b]

[a]*School of Biological Sciences,* [b]*School of Mathematics, Centre for Behavioural Biology, University of Bristol, Bristol BS8 1VG, United Kingdom.*
{a.i.houston; john.mcnamara}@bristol.ac.uk

Abstract: Rolls discusses various adaptive explanations of physiological processes and the emotions. We give a critical analysis of some of these from the perspective of behavioural ecology. While agreeing with the approach adopted by Rolls, we identify topics that could have been better presented by making use of the existing literature.

Rolls presents a stimulating account of the physiological basis of emotions within an evolutionary context. A general strength of the book is a stress on the complexity of the environment on which selection has acted as compared to the environment of the laboratory. We believe that an integration of causal and functional accounts of behaviour can only be achieved if this distinction is kept in mind when constructing models. We cannot expect animals to behave optimally in any particular environment that we provide for them. What we might expect is that natural selection will have produced rules that perform well in the range of environments in which evolution has occurred (e.g., Houston & McNamara 1989; McNamara 1996; McNamara & Houston 1980). Such rules may perform poorly in some environments – the preference for sweet but non-nutritive substances (Rolls, pp. 69–70) is a well known example. As Rolls points out, in a complex environment, relatively fixed relationships between the world and behaviour are unlikely

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

to be adequate. The animal will need some general flexible procedure, and Rolls makes the case for emotional states as a part of this.

Although we agree with Rolls's general aims, we feel that his discussion of adaptive explanations is not always as thorough as it might be. Despite mentioning the danger of plausible stories (p. 219), there are places where little or no justification is given for an adaptive explanation (e.g., Table 10.1). There are also cases in which an offered explanation may be incorrect. For example, on p. 66, Rolls mentions sensitivity to a change in the rate or magnitude of reinforcers. He suggests that it is adaptive to increase work rate when the rate of reinforcement is increased. It has been shown, however, that an increase in the availability of food should not necessarily result in an increase in the optimal level of foraging effort (Abrams 1991; Houston & McNamara 1989; McNamara & Houston 1994).

The adaptive arguments proposed in the context of emotion need to be fleshed out before they can be evaluated. On p. 69, it is suggested that communication is one of the functions of emotion: "Communicating emotional states may have survival value, for example by reducing fighting." There is a vast literature on this issue, none of which is mentioned by Rolls. The application of game theory to animal contests was originally seen to imply that animals should not communicate their intentions to their opponents (e.g., Caryl 1979). This position has been modified by subsequent work. Enquist (1985) demonstrated that the honest communication of intentions during a contest could be evolutionarily stable. There is now a general theory of honest signalling (Grafen 1990; see Hauser & Nelson 1991; Johnstone 1997, for reviews). Moving from the specific case of fighting to emotions in general, is it obvious that animals should accurately signal their emotions? Trivers (1971) raised the possibility that emotions play an important part in promoting and maintaining co-operative behavior. Frank (1988) develops a general account of emotions as honest indicators; for an example see Frank (1989). We believe that the time is ripe for the investigation of adaptive physiological mechanisms (see Houston & McNamara 1999 for further discussion). Rolls provides much of the basic information necessary for such an enterprise. As a simple example, consider the increase in motivation following a recently obtained reward. Rolls mentions that this positive feedback prevents the animal from rapidly switching between activities. Switching would be costly if it takes time or energy to change from performing one activity to performing another, so positive feedback can be seen as adaptive given switching costs. Positive feedback has been analysed from a causal point of view by McFarland and McFarland (1968) and Houston and Sumida (1985). A functional analysis of how costs should influence switching is provided by Larkin and McFarland (1978). The next step would be to predict the design of a motivational system that would respond appropriately to a range of environments in which the costs of switching differed.

Reinforcement, emotion, and consciousness

Carroll Izard[1]

Department of Psychology, University of Delaware, Newark, DE 19716.
izard@udel.edu

Abstract: Rolls presents a good integrative summary of the neural bases of emotions, adds new findings and insights, and takes a stance on controversial issues such as separate or distinct brain systems for processing emotion information and for planning and action. This commentary raises questions about his explanations of emotion activation, response to novelty, the evolution of emotions, and the phenomenal experience of emotions in human consciousness.

Drive, reinforcement, and emotions: What the book is about. Although Rolls's book features emotion in the title, it mainly concerns the traditional concepts of reward and reinforcement, and motivation as it relates to internal homeostatic processes or drives triggered by internal need-related signals such as glucose concentrations in blood plasma. When Rolls does turn to emotion, he still writes mainly about reinforcement, for he defines emotions as mental states elicited by reward and punishment. Or more formally (as he sees it), emotions derive from instrumental reinforcing stimuli.

In contrast to Damasio (1994), Rolls offers some evidence and argument for separate or distinct brain systems for assessing the reinforcement value of a stimulus or processing emotion information (orbitofrontal cortex) and for planning and executing actions (dorsolateral prefrontal cortex). In contrast to LeDoux (1996), he argues that the brain has to process stimuli (in neocortical systems) to the object level to learn reinforcement associations and generate emotion. Thus, according to Rolls, LeDoux's subcortical route (thalamoamygdala, "low road") to emotion (which some of us thought might help explain unlabeled emotion feelings) concerns exceptions to the general rule, particularly in humans and nonhuman primates.

Emotion activation. Appraisal, cognitive-behavioral, and biosocial theorists will likely find problems with Rolls's definition of emotions and his explanation of their activation. He explains emotion activation in terms of the valence of the reinforcer and the reinforcement contingency. He claims that this principle can explain many emotions though he provides formulas for only five, if you disregard synonyms. His explanation certainly seems to have the advantage of simplicity. Yet it may prove deceptively complex, considering the six possibilities for sequencing the presentation or withholding of positive and negative reinforcement and the different combinations of these sequences.

Still, these formulas do not comprise exclusive rules for the activation of any emotion. For a general example, Rolls's reinforcement rules do not explain changes in emotion or mood due to periodic changes in hormone levels. For a specific example, consider the rule that guilt results from the presentation of a stimulus that both rewards and punishes. Many challenging endeavors (e.g., climbing Mt. Everest) that offer numerous rewards and punishments may have nothing to do with guilt. Reports from those who succeed in reaching the peak make no mention of guilt experienced along the way. Research on non-egoistic helping behavior (Batson 1990), also suggests that the stark hedonism of Rolls's position may not capture significant components of human behavior. Moreover, research with children suggests that punishment in response to a child's enjoying something forbidden (e.g., bullying another child) may elicit fear or anger (not guilt) and have a poor chance of contributing to the development of conscience and moral behavior. Parental behavior that models empathy for the victim and directly induces guilt in the child works better (Hoffman 1975; Zahn-Waxler et al. 1979).

Rolls's theory of emotion activation does not address the elicitation of emotion through modeling or contagion. A number of studies with human children and young nonhuman primates have shown that parental modeling of an emotion response (e.g., the expression of fear) provides a means of learning to avoid a dangerous stimulus (fear) without ever directly experiencing it (Mineka & Cook 1984; Sorce et al. 1985). In the animal studies, fear conditioning occurred as feral reared macaque mothers modeled fear of a snake to their laboratory reared juveniles. That the laboratory setting probably provided few previous opportunities for modeling fear behavior makes it difficult to use reinforcement history to explain the results. Accepting Sacket's (1966) evidence of the juveniles' innate ability to decode and respond appropriately to adult expression of fear is a more parsimonious explanation and coheres with the concept of the primacy of emotions in evolution and contradicts Rolls's notion of language primacy (discussed later).

Although Rolls gives rather attractive descriptions of the reinforcement conditions that result in emotion activation, we might wonder why we would need emotion concepts at all if we could describe affective states of mind more objectively without them. The case for the more objective route might be stronger were it not possible to quibble with almost any of Rolls's formulas for

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

emotion activation. For example, it is reasonable, as he proposes, that anger might result from the omission of a positive reinforcement and the possibility of an active behavioral response. However, if the positive reinforcement were critical to health or well being and its accessibility remote, fear might occur. Research with human participants shows a great flexibility in emotion system responses to life's contingencies.

In summary, Rolls's theory of emotion activation certainly has the solid behavioristic (s-r) framework that has long served all those branches of behavioral science that depend on models involving classical and instrumental conditioning and the concepts of primary and secondary reinforcement. In emotion research, investigators use these concepts in animal models much more frequently than in work with humans. Yet, Rolls's hard science model has the advantage of offering new possibilities for quantifying and manipulating hypothesized antecedents of emotions (valence and sequence of reinforcing stimuli) and thus may attract a number of emotion researchers.

Emotional and non-emotional explanations of response to novelty. Rolls makes highly attractive as well as puzzling propositions in his treatment of animals' (particularly primates') responses to novelty. Having a long-standing concern with this system, conceived as the emotion of interest, I delighted in Rolls's explication of its neural substrates. He found neurons in the amygdala that respond to novel or relatively novel stimuli that have no association with reward. Whereas I would have rejoiced in having discovered the neural basis for the innate positive emotion that drives exploration and much of learning and creativity, Rolls declares that the initial response to novelty does not have the quality of an emotion because it is a response to a novel stimulus that has no reinforcement history. He does not consider the more parsimonious position that novelty serves as an innate activator of interest, which operates as emotion or motivation. Inability to incorporate such an important source of motivation into the conceptual framework for emotions seems a serious limitation of Rolls's reward-dependent theory.

Rolls's explication of response to novelty does not seem as clear or straightforward as other aspects of his theory. Although his theory dictates that the initial response to a novel stimulus cannot comprise emotion, the novel stimulus leads the monkey to "reach out for and explore the objects, and in this respect the novel stimuli are reinforcing" (p. 105). These seem like self-contradictory propositions and they provide no explanation of the critical initial response to novelty.

Rolls goes on to say that the interest-relevant "amygdala neurons . . . operate as filters which provide an output if a stimulus is associated with a positive reinforcer, or is positively reinforcing because of relative unfamiliarity." Of particular significance to those of us who define positive response to novelty in terms of an innate emotion called interest, Rolls goes on to say (p. 106) that "the functions of this output may be to influence the *interest* (emphasis added) shown in a stimulus." This is precisely the function some of us have ascribed to interest operating as an emotion (Izard 1977; Renniger & Wozniak 1985). Rolls makes persuasive arguments for the adaptive advantage of a mechanism that drives interest or exploration.

Meta-cognition, consciousness, and emotion feelings. Rolls also moves into the middle of a controversy in his theory of consciousness and its application to understanding emotion. He proposes a theory, admittedly speculative and philosophical, which explains consciousness as a function of higher order thoughts about lower order thoughts, or meta-cognition. Feelings emerge because it "feels like something for a machine with higher-order thoughts to be thinking about its own first- or lower-order thoughts" (p. 249). He thus makes consciousness and feeling dependent on language, or at least on meta-cognition involving syntactic manipulation of symbols. Rolls's theory, like a number of similar ones, may help explain introspective or reflective consciousness or self-consciousness.

However, Rolls's theory makes assumptions superfluous to the explanation of consciousness as subjective experience or emotion feelings. Why should the subjective experience of emotion require meta-cognition? Obvious adaptive advantages accrue from the generation of emotion experience as a direct function of sense perception and lower order cognition (e.g., feature detection). For example, fear (and avoidance behavior) activated by the mere detection of critical features of a biologically prepared stimulus (predator, rapidly looming object, dangerous height) might prove essential to survival.

Rolls's position implies that my feeling of joy at a long-delayed reunion with a good friend requires that I think about the thought (or experience) of seeing my friend. This circuitous route to emotion is imparsimonious. In denying direct awareness of feelings activated by the recognition of a friend's smiling face and requiring higher order thoughts about lower order thoughts to experience joy (or any other feeling), Rolls's theory seems to miss the spontaneity of emotional life. Though Rolls declares that the requisite meta-cognition involves only syntactic manipulations of symbols, and not necessarily lexical symbols, verbal language seems implied in his arguments on the necessity of thoughts about thoughts. In any case, this necessary condition means that all forms of thinking other than meta-cognition and all sense perception take place outside of consciousness as Rolls defines it.

Most philosophers from Husserl forward hold that all consciousness is consciousness of something and always posits or intends an object (McCulloch 1994). Thus, first order nonreflective thoughts and attentional processes enjoy consciousness (Chalmers 1996). If this were not true, we would have to assume a state other than consciousness for much cognition and action and for all things like the highly skilled and rapid actions of those in the performing arts and professional sports. The execution of movements like those required by a Chopin etude for piano or a Bach concerto for violin or striking a 150 kph ball with a bat or the center of a racquet allows no time for thoughts about the thought of making the next move. The intrusion of meta-cognition into such cerebellar-dominated and virtually automatic movements would doubtless disrupt or destroy the performance. So, reflective consciousness does not accompany these acts. Yet, if we do these acts truly outside of consciousness ("unconsciously"), how come we can recall aspects or even details of the performance at will?

Do you really have to think about the thought of having stubbed your toe in the dark (or any unanticipated trauma) before feeling pain and anger? Human infants express distress and anger to unanticipated pain long before they demonstrate syntactic manipulation of symbols (Izard et al. 1987). Children cannot comprehend syntactical operations until about 16 months, show them in speech production until about 18 months, correct their own speech (rudimentary meta-cognition?) until about 24 months (Bretherton et al. 1981) or demonstrate genuine meta-cognition until about age four or five years. Thus according to Rolls's theory, the young child has no conscious experiences and hence no feelings. Yet, much research shows that infants and toddlers express and detect a wide range of emotions and show clear signs of responding in terms of the motivation inherent in emotion (Izard et al. 1995; Termine & Izard 1988). Much other behavioral and psychological research makes an even stronger case for inferring emotional experiences in three- to five-year-old children. The adaptive advantage of the power of emotion to motivate responses to challenging contingencies seems to call for a parsimonious route to its activation and presence in consciousness.

To Rolls's credit, he repeatedly admits that he engaged in speculation about "the great mystery" that has puzzled philosophers for centuries – why and how do certain types of neural processing come to feel like something (p. 244). Or as others have put it, how does perceiving, thinking, and acting, or behaving like a cognitive agent result in consciousness? After studying Rolls's position and reviewing some others, one could reach the conclusion, perhaps implicit in Rolls's book, that this domain might yield equally as well, or better, to the methods of philosophy as to those of science.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Emotion, language, and evolution. Given his theory that consciousness and emotion feelings depend on meta-cognition, Rolls does not surprise us by surmising that the language system may have evolved before emotions. Two factors make this an unlikely scenario. First, since Darwin, investigators have continued to find evidence for the evolution of universal emotion expressions (Ekman et al. 1972; Izard 1971; 1994), and it seems quite plausible that expressive signals (facial and vocal) constituted the first "language" system in phylogeny. Rolls, like other theorists, specifies communication as one of the functions of emotions. The significance of emotional communication in mammalian social life and the importance of emotions as motivation constitute argument for their primacy in evolution. Both common observation and a considerable body of research show that in ontogeny emotional (expressive-signal) communication precedes language by as much as a year (Izard et al. 1995). Expressing emotions as a form of communication probably requires less complex brain systems than syntactic manipulation of symbols or verbal language. It seems plausible that selection pressures acted directly on emotions. Emotions can be considered, in part, as a set of special sensory systems with a very broad range of sensitivities to environmental contingencies and mechanisms for communicating about them.

Second, a number of other functions assigned to emotions by Rolls and other theorists (e.g., flexibility of behavioral responses, social bonding, regulation of perception, learning, and memory) seem sufficiently pivotal in mammalian life to give emotions primacy in evolution. It would probably be impossible to show, for example, that social bonding could occur or that the attachment behavioral systems operate in preverbal children on the basis of reward systems independent of emotions. Students of these processes explain them largely in terms of emotions.

NOTE

1. I am indebted to Carroll Izard II for comments and discussion on philosophical theories of consciousness and to Thomas R. Scott for a critical reading of the manuscript and helpful suggestions.

Emotion, representation, and consciousness

Leonard D. Katz

Department of Linguistics and Philosophy, Massachusetts Institute of Technology E39-245, Cambridge, MA 02139. **lkatz@mit.edu**

Abstract: Rolls's preliminary definitions of emotion and speculative restriction of consciousness, including emotional sentience, to humans, display behaviorist prejudice. Reinforcement and causation are not by themselves sufficient conceptual resources to define either emotion or the directedness of thought and motivated action. For any adequate definition of emotion or delimitation of consciousness, new physiology, such as Rolls is contributing to, and also the resources of other fields, will be required.

> The affects are those [psychobiological states] involving pleasure or distress those affected by which are changed in ways that affect their judgments and decisions.
>
> Aristotle, Rhetoric II,1:1378a20–22

> I mean by affects [psychobiological states] such as anger, fear, shame, and appetitive desire, all that normally involve conscious intrinsic pleasure or distress.
>
> Aristotle, Eudemian Ethics II,2:1220b12–14

> The affects are set apart by pleasure and distress.
>
> Aristotle, Eudemian Ethics II,4:1221b36–37

Rolls, like Aristotle in these passages, casts his net wide, including in the subject of his book and in his definitions of emotion (ignoring, as he does later, their tentative first refinement excluding drives and appetites, p. 65) the full range of affective states that have also, broadly speaking, cognitive and motivational roles.

After defining emotions roughly as "states elicited by rewards and punishers," (p. 61), Rolls shifts to "states produced by instrumental reinforcing stimuli" to provide "an operational definition of what causes an emotion." (p. 61) There will, however, be many organic states that rewards or reinforcing stimuli produce in organisms, such as fluid repletion, that are not themselves emotions. And, similarly, there are many relatively pure sensory and cognitive states so caused, including some the neural basis of which Rolls at length describes, that are not emotional, even in Rolls's broad intended use, even when they figure in the causation of states that are. We thus do not have a complete definition or even a sufficient condition for the intended class, but at most a partial functional specification of the causal role that the kind of states in question characteristically have in the psychological economy of organisms. More is needed to pick out the intended affective states from the ones that are more purely informative about the world or not psychological states at all. The cheap and easy way to do this is to bring in operationally undefined affective or motivational concepts, such as Aristotle's pleasure and pain or the related Aristotle-descended notions of appraisal and evaluation, which Rolls rejects as imprecise and operationally undefined. But it seems Rolls's preferred concepts, those of behaviorist psychology plus that of causation, are not up to the job.

Rolls's equation of intentional states' directedness upon objects with their being "produced by stimuli or objects" (p. 62) is a related point at which a facile behavioral-cum-causal analysis of a behaviorally unreduced psychological notion will not do. A child may love Father Christmas or fear the bogeyman without these emotions having been caused by their nonexistent objects (Brentano 1874). And even where the object does exist and plays some appropriate causal role, mere appeal to causation will not distinguish which item in an emotion's typically long and wide causal history the object of the emotion is. This must enter its causal history in just the right way, through a complex chain of processing and representations of the kind that Rolls's and related scientific work is so usefully beginning to explain, to become its intentional object. Nothing less than a fairly complete account of how this is done and of what makes a state play the distinctive role of an affective state in it – since just as not everything in the causal history of an emotion will be its object, not just anything or even any representational state an object causes will represent it – will allow us to dispense, at an advanced stage of the inquiry, with the unanalyzed notions of emotion, object of thought, and goal of action that Rolls apparently believes can be identified in behavioral and gross causal terms at its very start. Any such account will have to specify the level of generality at which the relevant representations function so that not "[t]he emotion-provoking stimulus" but the attainment of the object or class of rewards it represents can become "a goal for action" (p. 65) in an affectively motivated way.

The same behaviorist prejudice that lurks in the background of these definitions seems to infect also the discussion of consciousness in Chapter 9, but it is now restricted to such animals as do not have the symbolic, syntactic, and thinking-about-thought capacities that most adult humans do. To be sure, nonlinguistic animals (like infants and many stroke patients) cannot tell us about their experiences. But that amounts to little without the concurrence of neurological or developmental evidence, for which philosophical speculation is a poor substitute. The processors handling syntax are not the only ones capable of flexibly organizing complex sequences of behavior (think of those organizing instrumental motor behavior) or of representing its progress through time. Indeed, the cognitive maps we share with Tolman's rats are likely used to do this in a manner more accessible to our consciousness than the workings of the processors handling linguistic syntax are. And the indications of right hemisphere advantages in emotional domains and of emotional blunting and neglect after right but not left hemisphere strokes of right, more than left, hemisphere involvement in emotion should at least give one pause before ascribing the consciousness of emotion to the generally left hemisphere syntactic processors. Neurology should

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

provide abundant material for more specific testing of Rolls's and related suggestions, if they become specific enough to become testable – and guidance in making them so – as will comparative and developmental psychology, on which see Hauser, 1996, pp. 597–608. If we are to progress beyond Aristotle in our overall understanding of emotion, we must heed his advice that care in psychological-level functional characterization as much as in physiology is necessary to the progress of this science (Aristotle, De Anima I,1:403a25–b11).

Reinforcement and punishment: Dissociable systems for action and emotion?

Simon Killcross

Department of Psychology, University of York, Heslington, York, YO10 5DD, United Kingdom. **ask1@york.ac.uk** **www.york.ac.uk/depts/psych/**

Abstract: Rolls presents a theory of emotion based on the premise that emotions are evoked by events that are capable of being instrumental reinforcers and punishers. As support for this theory is drawn almost entirely from experiments in non-human primates, valuable insights into the relationship between punishment and reinforcement systems, and the nature of instrumentality, may have been overlooked.

Rolls's book presents an impressive battery of information concerning the study of reward processes in animals. When mentioning reward processes here I do so with the intention of highlighting what appears to be the central concern of *The brain and emotion.* Rolls provides an interesting and informative review of a large body of recent work examining the nature of brain responses to rewarding stimuli, much of it derived from his own experiments examining single-cell responding in primates. This work is without doubt of the highest quality and has produced a wealth of detail concerning the likely systems underlying the responses of various brain regions to stimuli that are reinforcing.

A central tenet of the book is that events that produce emotions are those that are reinforcing or punishing (or at least capable of being shown to be reinforcers or punishers). Whilst there is much to recommend this stance, it would appear to have several consequences. First, it is immediately apparent that because of the nature of Rolls's work with primates, much of the argument concerning the role of punishment in emotion must be made by analogy. With good cause there is little, if any, work on aversive motivational systems in primates, and notable exceptions (e.g., Mirenowicz & Schulz 1996) tend to use aversive treatments that are extremely mild, such as a puff of air to the hand.

It is a shame then that Rolls's decision to rely largely on primate literature means that much work looking at punishment systems in other laboratory animals receives little discussion. Perhaps Rolls considers this work to be of lesser importance with respect to human emotion. But surely in the absence of other information it is worth taking into account. That is especially important when one considers the nature of psychological theories examining appetitive and aversive motivational systems.

The tacit assumption of Rolls's theory is that the two systems share a more or less common neural substrate involving the orbitofrontal cortex, amygdala, and various output routes. Indeed there is little in Rolls's own work or in *The brain and emotion* to suggest that the two systems might rely in any way on different processes. However, there are clearly other data at odds with this position. Work examining the role of the amygdala in fear conditioning in non-primates has shown that lesions of the amygdala produce, by several measures, a complete abolition of this form of stimulus-reinforcement (to use Rolls's term) learning. However, no such effects are found in the appetitive domain. Lesions of the amygdala, in non-human primates and rats, appear to produce no deficits in the formation of simple associations between stimuli and rewards, but rather in higher-order manifestations of such associations such as those found in second-order conditioning and conditioned reinforcement procedures.

This difference is glossed over in sections dealing with the amygdala (Ch. 4) with the conclusion that "there is thus much evidence that the amygdala is involved in responses made to stimuli that are associated by learning with primary reinforcers" (p. 101). But perhaps there is good cause to think that there may be more to the observed differences in empirical findings than meets the eye.

One approach might be to suggest that the findings regarding amygdala lesions in non-primates is due to their lack of development of frontal regions, which in primates allow orbitofrontal mechanisms to subserve many of the functions perhaps usually devolved to the amygdala, that is, there is redundancy in the system if it is sufficiently well developed. Unfortunately this cannot account for the differences observed between appetitive and aversive tasks in non-primate species following amygdala damage. Why might the lack of frontal development manifest itself as amygdala-dependent deficits in aversive, but not appetitive, tasks?

A second approach might be to say that the current theories of amygdala function – namely that the amygdala is the site of formation of stimulus reinforcer associations – fails to capture fully the nuances of changes in emotional processing that result from damage to this region.

A positive consequence of Rolls's definition of emotion is that it brings into sharp relief the difference between feelings and emotions. The operational definition of emotional states as being created by events that can act as reinforcers or punishers immediately removes us from the clouding issues of what, if anything, the phenomenology of these emotional states might be. However, this approach merits further examination. Rolls considers emotions (as opposed to mood states) as things that take or have an object, and are thus intentional states (p. 62). This is consistent with the idea that emotionally-significant events are those that can act as the object of a goal-directed action, but one must here be explicit that one is then considering theories of instrumental action based on intentionality. In Tolman's (1959) terms, a response is the product of a means-end-readiness (belief or expectation) about an environmental contingency and a valence (desire) for a particular outcome. Hence emotional states come to be evoked by events that possess such a valence.

Rolls hints throughout his work that he is talking about goal-directed actions, or the product of action-outcome (A-O) associations rather than stimulus-response (S-R) habits. However, despite a willingness to employ definitions based on reinforcement contingency there is little explicit attempt to analyse behavioral tasks employed by primates in this way. What is a discrimination reversal if not a change in A-O contingencies? Once more there is an accumulation of evidence that such systems are not unique to primates (e.g., Balleine & Dickinson 1998) and again work in the field of emotion would benefit immensely from a greater discourse between primate and non-primate research.

The essential roles of emotion in cognitive architecture

Kevin B. Korb and Ann E. Nicholson

School of Computer Science and Software Engineering, Monash University, Clayton, VIC 3168 Australia. **{korb; annn}@csse.monash.edu.au** **www.csse.monash.edu.au/~annn/**

Abstract: Rolls's presentation of emotion as integral to cognition is a welcome counter to a long tradition of treating them as antagonists. His eduction of experimental evidence in support of this view is impressive. However, we find his excursion into the philosophy of consciousness less successful. Rolls gives syntactical manipulation the central role in consciousness (in stark contrast to Searle, for whom "mere" syntax inevitably falls short of consciousness), and leaves us wondering about the roles left for emotion after all.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Recent thinking has trended against a 2,500 year tradition in Western thought to regard emotion and cognition as antithetical. Socrates, for one, praised the reflective life of the mind (his injunction: "know thyself") and denigrated the body (and its emotions) as an interference. Christianity further magnified the distance between reflective cognition (soul) and degenerate emotion, until it was thought nearly axiomatic that emotion and cognition were related only in being antagonistic. In the last couple of decades, more and more dissonant voices are being heard however: in philosophy, Ronald De Sousa (1987) has argued that emotions are integral to cognition; in neuropsychology, Antonio Damasio (1994) has argued that a flat affect, caused by neurological damage, invariably leads to defective decision making, with people unable to value their options. And Rosalind Picard's recent book *Affective Computing* (1997) would have been thought ludicrously titled by most computer scientists not long ago.

Nevertheless, the traditional dichotomy between cognition and emotion is clearly reflected in the history of AI and it still remains dominant. AI paradigms such as "rational agents" (Russell & Norvig 1995) and connectionism (Rumelhart et al. 1986) make little or no use of emotions. Much of the recent AI work on affective agents has been aimed at building more engaging human-computer interfaces, rather than developing an underlying emotional intelligence (Bates 1994; Maes 1995). One of the main messages of Rolls's and related research is that emotions are intrinsic to intelligence: they are not some "lazy Susan" of spices to be sprinkled on top after the hard work of cognition has been completed, as most AI work assumes.

Rolls attempts a comprehensive survey of the neuroscience and psychology of emotion. His foundation is a neurophysiological story of how positive and negative reinforcement shape behaviour through operant conditioning, through association learning of secondary reinforcers and through such mechanisms as sensory-specific satiety – a diminution of reinforcement from repeated sampling of a single type of stimulus. Rolls draws upon a rich experimental literature, particularly in the study of feeding (Ch. 2) and drinking (Ch. 7) behaviour, demonstrating strong interconnections between cognition (identification, planning), reward, and emotions. Furthermore, this story is well balanced by the plausibility of the corresponding evolutionary "why" story, that is, of evolutionary explanations of the functions subserved by these brain mechanisms. (In passing, we point out that in dealing with sexual behaviour in Ch. 8 the balance is lost: the stories of evolutionary psychology and sociobiology dominate and the experimental tradition largely disappears. Perhaps the experimentalists simply find sexual behaviour unrewarding to study?)

In Chapter 3, "The nature of emotion," Rolls outlines his theory of emotions and compares it with contending theories. The key idea is that emotions mediate states bearing utility (e.g., conscious states of pain and pleasure) and actions intended to bring about or avoid such states (p. 60). We believe this is a fundamental and important point. It explains the defective decision making in the neurological cases of Damasio: if the target reward state is not immediately available, and if the mediating emotional state is absent, then there is no goal for planned actions to be directed towards. From this central role the functions of emotion commonly mentioned appear to flow naturally. Of the nine functions of emotion Rolls discusses, for example, the central one of enabling flexible behaviour, beyond reaction to immediate rewards, is nearly a direct consequence. So, too, the role of emotion in initiating decision making and action (motivation, arousal). This is, of course, connected to the elicitation of autonomic and hormonal responses that Rolls emphasizes. This is also related to the apparent role of emotion in rendering some objects salient, and in general in focusing attention (an aspect of emotion that Rolls oddly overlooks). The exact relation between these is, however, *not* clear: how exactly do arousal, attention, and the initiation of decision making relate? It would have been interesting to see what Rolls would make of this question; unfortunately, it is not raised.

Although architectural details relating emotional states, planning, beliefs, and so on, are mostly lacking, Rolls nevertheless presents a reasonably clear picture of the inter-relation of emotion and cognition in the first eight chapters. We believe it seriously undermines Griffiths' thesis (1997) that the simpler emotions ("affect programs") are isolated from cognition. This clear picture, unfortunately, unravels during Rolls's subsequent excursion into the philosophy of consciousness (Ch. 9).

Following Rosenthal's higher-order thought (HOT) theory (Rosenthal 1990) Rolls holds that the distinguishing feature of consciousness is that first-order thoughts (e.g., about the world) become objects of a higher-order thought (e.g., a belief about the first-order thought). But Rolls insists that the thoughts and meta-thoughts must be more than that: they must have an explicit syntax; the correct HOT must be HOLT (higher-order *linguistic* thought) in order to account for complex planning and the ability to correct such plans. In consequence, all of the above emotional processing, mediating reward and action, must be fundamentally pre-conscious or unconscious. And, its apparent role in complex planning is at least greatly vitiated. Emotional feelings are conscious, but that is only because the emotional states somehow become the object of linguistic representations. This line of thought leads Rolls to doubt that nonhuman animals are conscious, except perhaps primates, whose linguistic capacities are in doubt. It appears that the sharp duality of emotion and cognition has simply re-emerged as a duality between emotion and consciousness. It is ironic that the distinguishing feature for consciousness according to Rolls – syntactic (meta-)processing – is precisely what Searle (1980) argued was radically insufficient for consciousness.

We do not believe that HOLT is a well-considered theory of consciousness. It seems odd that, after arguing for a deeper role of emotions in planning and decision making than usual, Rolls dismisses as evidence of consciousness what are commonly taken as paradigmatic, such as suffering, joy, and so forth. Regarding emotional processing in nonhumans, Rolls says (p. 257) "animals are often thought of as performing optimally on some cost-benefit curve. . . . This does not at all mean that the animal performs a cost-benefit analysis." There are after all many possible ways of behaving *as if* doing a cost-benefit analysis. Quite right. By the very same token, however, evolution may well have produced any number of mechanisms other than an explicit syntax for representing the world and so also for representing such representations in the brain. If some nonhuman animal is using such representations to plan and meta-representations to correct its plans, it is clearly thinking. Denying then that it is conscious merely because *in humans* consciousness is normally accompanied by language appears unjustifiably anthropocentric.

A sharp duality between emotion and cognition, the implicit and explicit, non-symbolic and symbolic, and reactive and planned behaviour appears to be nearly ineluctably seductive, in the end seducing even Rolls, who set out to contest it. If we accept this new dualism, it leaves us with the question of what to do with emotions: if syntax is both necessary and sufficient for complex planning (i.e., emotional processing is not necessary), it must be sufficient also for the simpler planning using emotional processing. Whence then the need for emotions? What evolutionary story is left for their development? We prefer to suggest that syntax is *neither* necessary nor sufficient for complex planning.

In AI, it has become clear that some hybrid of the two parallel approaches to planning is necessary; that is, calculating optimal policies over a complex state space is often too complex, so relying on heuristic intervention has become of interest. Yet hybrid systems that handle both normative computations and reactive needs imposed by a hostile environment have been difficult to design. Incorporating the functions of emotion identified by Rolls (prior to Ch. 9) into AI architecture, with their role in the development of flexible responses, directing attention, and initiating decision-making appears to be an important challenge for all AI researchers, and not just those in the "affective agent" community.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

A taste of things to come

Jerald D. Kralik and Marc D. Hauser

Department of Psychology and Program in Neurosciences, Harvard University, Cambridge, MA 02138. **{bach; hauser}@wjh.harvard.edu**
www.wjh.harvard.edu/~mnkylab

Abstract: Rolls uses evolutionary theory and behavioral learning theory in his analysis of emotion. We believe that both theories are greatly underutilized, leaving an incomplete description of the nature of emotion and its neural foundation.

Two of the theoretical frameworks used by Rolls are evolutionary theory and behavioral learning theory. Both of these theories should play important roles in neuroscience, and Rolls should be commended for integrating them. We believe, however, that his use of both theories still falls far short of their potential contributions to the understanding of emotion and the brain.

Throughout the book, Rolls attempts to provide ultimate explanations for neural mechanisms. Although this use of evolutionary theory is important, there is almost no attempt to consider alternative hypotheses, or to propose specific tests that would enable researchers to assess the explanatory power of each hypothesis. For instance, evolutionary speculations are offered for many of the proposed functions of emotion (see sect. 3.1.5). It is suggested, for example, that grief and sadness may be adaptive in motivating an individual to stop responding to a reinforcer that may no longer be available (e.g., a lost family member). However, could the emotions of grief and sadness be adaptive for other reasons, and if so, how would one experimentally test between these alternatives? Consider, for example, the possibility that grief and sadness motivate an individual to increase responding to a reinforcer whose continued availability is uncertain (e.g., working to keep a mate from abandoning you or heightening parental care for a seriously sick infant). Moreover, could such emotions be a consequence of some other adaptation such as the affiliative emotional bonds that exist between kin? To test between these alternative hypotheses, one could propose comparing the neurophysiology of different species that have different mating systems, levels of parental care, or kin networks, exploring the ontogeny of emotional responses of individuals to attachment and loss, and running experiments to assess the fitness consequences of particular emotional states.

Rolls's evolutionary explanations are commonly summoned after the fact, providing de facto justification for a particular neural process. However, the power of Darwin's theory, and its subsequent adaptation by behavioral ecologists and evolutionary psychologists, has been to generate, a priori, specific hypotheses about the functional architecture of the brain in a comparative context. For example, consider work on the orbitofrontal cortex, a target structure in Rolls's analysis. Electrophysiological and lesion studies suggest that the primate orbitofrontal cortex decodes the reward value of stimuli, and rapidly learns (and relearns) associations between stimuli. Because the orbitofrontal cortex is much more developed in primates, one is then led to explore the socioecological factors that might have contributed to the need for rapid association learning in primates. Rolls suggests that dynamic foraging requirements (eating over 100 varieties of fruit) and dynamic social relationships may have been two such factors. The critical issue, though, is whether these socioecological hypotheses were generated first, helping to uncover the role of the orbitofrontal in rapid-learning, or whether these explanations were provided after the fact. The logical order of these ideas in the book suggests the latter. Surprisingly, Rolls does not consider, for example, research in behavioral ecology conducted over fifteen years ago that predicted, a priori, that there would be differences in the brains of frugivores and folivores, nor does he consider the comparative neuroanatomical findings that have confirmed this prediction (see Allman 1999; Clutton-Brock & Harvey 1980).

We agree, along with many other scientists, that evolutionary theory should play a leading role in uncovering the workings of the brain (for general reviews, see Allman 1999; Deacon 1997; Pinker 1997). Indeed some terrific work has already demonstrated the powerful influence socioecological factors have on brain structure and function, and how an understanding of these factors can help generate predictions about the brain. Such work includes predictions of relative hippocampal volume of voles, parasitic cowbirds, and cacheing birds, based on these species' particular mating systems and foraging strategies (reviewed in Sherry 1997; Hauser 2000). The same approach can also illuminate our understanding of the emotional systems of the brain. As an example, consider the functions of the primate orbitofrontal cortex. Because natural selection is most likely to have shaped significant new developments from already functioning older ones, we need to determine how the system was built to determine precisely how it is working. For a complete understanding of the relevant functions of the orbitofrontal cortex, then, we need to answer the following evolutionary questions. What part of the emotional system was already in place before the evolution of the orbitofrontal? Was rapid-learning conducted in any of these structures before the evolution of the orbitofrontal? When the orbitofrontal evolved, did it take or modify components of the original emotional system, or were only new components added? Which homeotic genes were replicated and modified to produce the orbitofrontal cortex? What particular developmental processes increased the size of the orbitofrontal and modified projections to and from it? Did the orbitofrontal evolve relatively independently, or did it come with other structures, such as the dorsolateral prefrontal? Did rapid-learning evolve in another way in some other species, without the evolution of the orbitofrontal? To answer these questions, comparative analyses must be conducted of (1) species with progressively evolved orbitofrontal cortices, (2) distantly related species with similar social, mating or foraging demands, and (3) closely related species with different social, mating or foraging demands.

Rolls appropriately emphasizes the importance of behavioral learning theory for understanding the neural basis of cognition, behavior, and emotion. Unfortunately, however, Rolls sometimes ignores earlier developments in behavioral theory. As an example, consider some of Rolls's definitions of operant behavior (see sects. 1.2 and 3.1.2, for instance). Rolls equates punishment, punishers, and negative reinforcers, and states that "a punishment is something an animal will work to escape or avoid." (p. 3). However, over 80 years of behavioral research shows the importance of distinguishing between many of such concepts that Rolls equates. Punishment is the *decrease* in the likelihood of a response as a result of the response being followed by an aversive stimulus (i.e., a "punisher"; see Dickinson 1994; Mazur 1997). In punishment, one ceases to do something in order to escape or avoid an aversive stimulus. Negative reinforcement, on the other hand, is the *increase* in the likelihood of a response as a result of the response being followed by the termination or omission of an aversive stimulus (in this case termed a "negative reinforcer"). Thus, in negative reinforcement, one does something to escape or avoid an aversive stimulus. By not using these well-established definitions in the book, numerous inaccuracies arise that confuse more than clarify (e.g., see quotation above). Moreover, distinctions supported by behavioral research, such as the difference between punishment and negative reinforcement, may be maintained at the neural level, and thus should not be abandoned by neuroscience.

Rolls's treatment of emotion is clearly important and is leading us toward a true understanding of the phenomena involved. However, the description of emotion itself in the book is inconsistent and vague. Several times emotions are defined as "states elicited by rewards and punishers" (p. 60), at least twice as "consist[ing] of cognitive processing which results in a decoded signal that an environmental event (or remembered event) is reinforcing, together with the mood state produced as a result" (p. 62), and at least once as "responses elicited by reinforcing signals" (p. 125), even though it is said that "response[s] . . . are produced by emotional states"

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

(p. 67), and that "there is no necessary link between performing actions and emotion" (p. 60). Are the reinforcer selection (i.e., the "decision" of which reinforcer to respond to) and response selection mechanisms considered to be part of emotion, or are they influenced by emotion? Autonomic and endocrine responses cannot be a part of the definition, if Rolls is to avoid a bodily theory of emotion. Further, what precisely is meant by "cognitive processing" in the definition? For example, consider a subordinate rhesus monkey's fear response to a conspecific face. The perceptual processing of the face would presumably not be part of the emotion, so does the cognitive processing mechanism(s) correspond to the firing of populations of orbitofrontal and amygdala neurons to the face? Or is there some further cognitive processing that corresponds to, for instance, activity in the dorsolateral prefrontal cortex or hippocampus? What precisely, then, is the resulting "mood state"? Given the strong arguments for the existence of mechanisms for reinforcer decoding, working memory, reinforcer selection, and response selection, the book has not made it clear why the concept of "state" need be invoked at all.

Reward: Wanted – a better definition

Irving Kupfermann

Center for Neurobiology and Behavior, Columbia University, New York, NY 10032. **ik7@columbia.edu**

Abstract: Rolls's book depends significantly on a definition relating emotion to reward and learning. This definition confuses two separable concepts, and may result in the exclusion of notions of emotion and motivation from lower animals that may possess limited learning capacities. A more useful definition might revolve around the notion that emotions are states that function to optimize the performance of behavior.

The main problem I see with Rolls's book concerns the definition of emotion and its relation to reward. In fact, the book might be more properly entitled "The vertebrate brain and reward." To understand the problem of definition, consider the following description. A food deprived animal has been inactive but now detects chemicals indicating the presence of food. It begins to locomote, periodically pausing and sampling the environment and sometimes changing directions. The behavior persists until the food source is reached. Contact with the food, evokes a complex autonomic response involving an increase in heart rate, blood pressure, and other autonomic responses. The food object is manipulated toward the mouth, at which time a series of bite/swallow responses occur. The rate and intensity of the bite/swallow responses initially increases, reaches an asymptotic value and then as the animal continues to ingest food, the rate and intensity of the feeding responses decrease until the animal no longer responds to food, and becomes inactive. It sounds as if this animal possesses a motivational state akin to that of hunger, and that the presence of food, evokes something akin to that of an aroused emotional state. But by the definitions developed by Rolls, it is inappropriate to analyze this situation in terms of emotions.

The animal I prefer to in the above example, is the mollusc *Aplysia* (Kupfermann 1974), an organism that exhibits numerous forms of behavioral plasticity of feeding behavior (Kupfermann et al. 1989), but which appears to have a very limited if not totally absent capacity to learn arbitrary responses when "rewarded" by motivationally relevant stimuli. Rolls seems to use the notion of reward and reinforcement interchangeably. It would seem to make more sense to use the term reward in its vaguely anthropomorphic sense as something "pleasurable or painful" and whose contingent presentation alters the probability of the occurrence of somewhat arbitrary responses. (I say somewhat arbitrary, instead of arbitrary, as used by Rolls, since the responses that can be modified by reward, are not as arbitrary as it might seem, but instead represent a relatively restricted class of responses. A given reward can alter the probability of some responses, but not others). The term reinforcement can then be reserved for stimuli which evoke behavioral responses that serve to maintain (or terminate) the stimuli, irrespective of whether the stimuli alter behavioral responses in the future.

By this definition, food stimuli are reinforcers for food-deprived *Aplysia.* Rolls rejects this definition since he wishes to eliminate taxes, tropisms, and reflexes. By limiting his definition to stimuli which can mediate learned responses, Rolls confuses emotional behavior and learning processes. This provides him the opportunity to review his extensive research on reward systems whether or not they are tied to emotional responses. Are we to believe that a severely memory and learning impaired individual no longer can experience emotions? Show this individual a snake and they will likely exhibit behavioral indices indicating the emotion of fear. And yet they may be totally incapable of learning and acquiring arbitrary responses to terminate the stimulus (for example pushing a "help" button).

Studies on invertebrates have clearly shown that elements of motivational states and/or emotional responses are features of almost all complex organisms. What then is the essence of such a primitive function. To my mind the key to emotion is that it is a state (not necessarily evoked by a stimulus, since it can occur spontaneously) that functions to optimize the performance of ongoing or likely-to-occur behavior, particularly behavior that is best executed quickly or energetically. These preparatory responses are needed to optimize the nervous system as well as the periphery (i.e., muscles and sensory apparatus) for the different behavioral requirements needed for various responses that can occur at various rates and with different combinations of movements. An intensively studied example is that of food-induced arousal in *Aplysia,* which involves the release of diverse modulatory substances at muscles and at central synapses (Kupfermann et al. 1997; Weiss et al. 1993). The actions of the modulators at central and peripheral sites serve to alter the dynamics of muscle contractions that occur at different rates for behaviors involving different responses all engaging the same set of muscles in the animal. This approach towards motivation and emotion is clearly related to the notions of Frijda (1986) and others who argue that emotions are related to a change in action-readiness. Rolls rejects this idea because certain well-learned and routine responses appear not to be associated with an emotional response. This objection is readily countered by restricting the action-readiness to those behaviors that require, or are likely to require, very strong actions or rapid switching between one type of behavior to another type.

The advantage of dissociating the concepts of emotion from those of learning is that it permits the separation of two related but not necessarily intertwined concepts. It allows the study of emotional behavior in lower animals and fosters mechanistic evolutionary comparison between all organisms. What Rolls may see as simple reflexes in lower animals often involves very sophisticated behavior patterns and behavioral "choices," and the complex behavior of higher animals conversely can involve some rather fixed and inflexible responses. Emotions, like other behavioral capacities, evolved from basic functions performed by very simple nervous systems.

Aside from the specific problems with the treatment of emotions there are also a number of general (albeit secondary) problems. Using the classification scheme of the author (Fig. 3.1), my particular emotional feeling state evoked by reading this book appears to reside in the quadrant defined along the dimensions of ecstasy, elation, and pleasure along one dimension, and of rage, anger, and frustration along the other dimension. On the one hand the book is a creative, stimulating, and provocative exposition of data and ideas; but on the other hand it is full of excessive repetition and has numerous virtually incomprehensible sentences that appear never to have been read by an editor (with the possible exception of the chapter on sex, which seems to be a noticeably easier read than the rest of the book).

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

On the behavioural interpretation of neurophysiological observation

Donald R. J. Laming

Department of Experimental Psychology, University of Cambridge, Cambridge, England CB2 3EB. **drjl@cus.cam.ac.uk**

Abstract: Examples of terror generated by an aircraft disaster, of human courtship behaviour, and of the application of laboratory techniques to the commercial training of animals suggest (1) that emotion is simply the subjective counterpart of (objective) motivation (so that separate brain mechanisms would be an embarrassment) and (2) the apparent involvement of reward and punishment is a consequence of the excessively narrow range of experimental procedures used and has no foundation in the design of the brain.

In the English language there is "motivation" and "emotion" and it is widely presumed that these two words categorise two different kinds of notion. Rolls (p. 60) proposes "that emotions are states [of mind] elicited by rewards and punishers" and later "A third function of emotion is that it is motivating" (p. 68). In this commentary I outline a much simpler relationship between motivation and emotion and, at the same time, a more profound one.

> At 7:13 A.M. on Thursday, 22nd August, 1985, British Airtours flight KT328 to Corfu was just taking off from Ringway Airport, Manchester. As the aircraft was gathering speed along the runway, seconds before actual take-off, the port engine exploded, puncturing the fuel pipe, wing tanks, and fuselage, and setting the aircraft on fire. The take-off was immediately aborted. But the fire spread to the interior of the aircraft within a matter of seconds, producing dense smoke and panic. Notwithstanding that the emergency services were alongside the aircraft within seconds of the explosion, 55 passengers were burnt beyond recognition; 82 escaped. (Davenport et al. 1985)

Testimony from survivors does not mention anything about "rewards and punishers" – the survivors were simply terrified. At the same time, they tried to escape – in fact, they panicked. This episode suggests this much simpler relationship between motivation and emotion: the circumstances of the aircraft disaster generated the motivation to escape, and that motivation was experienced by the passengers as terror. Speaking generally, *motivation* is a state that may be inferred from *objective* observation of a person's behaviour; *emotion* is the *subjective* experience of being motivated. If I observe an animal searching for food, that animal is motivated. If I am searching for food myself, I am hungry. The hunger is my subjective experience of being (objectively) motivated to search for food.

This idea has three immediate implications:

1. We do not need separate brain mechanisms for emotion – in fact, their discovery would be an embarrassment. The brain mechanism for an emotion is the brain mechanism for the corresponding motivation.

2. The question "Why do we have emotions?" needs no answer. If there be no feeling of emotion, there is no motivation and without motivation (to find food, to nurture the young) the species ceases to exist.

3. The idea that reward and punishment are mediated by specific brain mechanisms must also be discarded, but this is far from obvious and needs another example to put the point:

> A pretty fair-haired girl wearing a clinging grey shift dress and Doc Martens boots is walking down Great Western Road, Paddington, on a hot summer afternoon. As the camera follows her it repeatedly swings round to catch the many other people in Great Western Road, mostly young men, who turn round to watch the girl as she walks by. Two workmen look out of an upstairs window and one draws the other's attention to the girl. Four young men seated at a pavement cafe turn round in their seats as she passes. Another young man coming the other way turns round as he passes the girl and visibly says "Wow!" The girl smiles back. (Bromhall 1994)

This was the title sequence to the first program in a TV series entitled *The Sexual Imperative* and it demonstrates *two* views of young heterosexual men watching a girl walking down the street. There is the *personal view* which each has of his own internal feelings ("She's a very nice girl. I'd like to date her!" – and the viewer will also have his personal view) and there is the *camera view* that everyone else, especially the camera, has of that young man's head turning to watch the girl as she walks by. Two views of the same behaviour – very different in character – nevertheless, they go together in point of time and place.

That video sequence generates these further points.

4. The head-turning is almost mechanical, as if the bystanders were so many rod puppets. There is no reward – just a quasi-mechanical response to the image of the girl. When the girl smiles back, that could be (and often is) the beginning of that long-drawn-out two-way interaction we call courtship. The notion to be taken on board is that each adult member of the species (any animal species) is equipped with a repertoire of innate behaviour patterns sufficient to ensure, by interaction between adults, the replacement of natural losses through death. Sexual motivation is the engagement of those behaviour patterns: love, jealousy, frustration, and so on, are the subjective emotional counterparts.

5. Quasi-mechanical behaviour is by no means specific to sex, nor to humankind. Breland and Breland (1961) reported some notable failures of laboratory techniques applied to the training of animals of some 38 different species for commercial purposes. They *could* train a pig to put a (wooden) penny in a piggy bank, but the behaviour subsequently broke down. A racoon could be trained likewise, but give him two coins and he would do no more than rub them together. If the stimulus situation engages some innate behaviour pattern, that is what you get, and rewards prove ineffective.

6. The idea (Rolls, Ch. 10) that reward and punishment are built into the structure of the brain is therefore misconceived. Reward and punishment have acquired the status that they have in this manner.

Suppose you are studying learning in a sub-human species. You cannot give your subjects a list of paired associates and must therefore build the relationship to be learned into a task that the animals can perform. To make them perform typically requires both that they be motivated (e.g., hungry) and the provision of some stimulus of corresponding motivational significance (i.e., food), obtainable following successful performance of the task. It is not that the animals do not learn without motivation and reward – they do (e.g., latent learning, Thistlethwaite 1951) – but that they will not perform, and without the performance the learning cannot be observed. The psychological function of reward is therefore to create a task of motivational significance which will engage some instinctive behaviour pattern (e.g., seeking food), a pattern subject to modification by what the animal has learned.

The question whether performance of the task is, or is not, "rewarding" is neither here nor there. It happens that nearly all the tasks used in animal laboratories (i.e., classical and operant conditioning paradigms) fit into a response-reward pattern, but many other things that animals do in their natural state do not. The function for which Rolls (Ch. 10) seeks a foundation in brain design is actually a characteristic of the very narrow range of tasks used in animal laboratories. The idea that psychologists' choice of experimental paradigms has representation in the *animals'* brains is too, too much!

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Emotion, cognition, and free representation

Eoghan Mac Aogáin

Linguistics Institute of Ireland, 31 Fitzwilliam Place, Dublin 2, Ireland.
eoghan@ite.ie **www.ite.ie**

Abstract: The representation of events, in primates at any rate, is a separate process from their emotional evaluation. The same holds for cognitive evaluation. Here too representation and evaluation are separate operations. Acknowledging the symmetry leads to the notion of free representation.

Although *The brain and emotion* (*Brain*) contains very few direct criticisms of alternative views, it is one of the most important critiques of cognitive science yet to appear. With its companion volume, LeDoux (1996), it gives us a comprehensive account of evaluative and emotional states that succeeds in treating them strictly on a par with cognitive states. It breaks out of the one-dimensional account of information that generally prevails in cognitive science, in which the phenomenon of intentionality (aboutness, information, representation, reference) is linked primarily, if not exclusively, to belief-forming processes.

Emotions become secondary phenomena in the one-dimensional account, mere primers or consequences of cognition. Now we have an enlarged notion of information that serves not only cognition but also a whole family of attitude-forming processes that have nothing to do with the formation of beliefs but only with the maximization of reward.

But there is an unresolved conflict in *Brain*, between extraverted and introverted treatments of emotion. The extraverted treatment, based on animal learning paradigms and expected values, shows us the emotional system locked into the constancies and contingencies of the external environment in order to maximize reward. Emotion, revealed in the amount of work an animal is prepared to do to bring about the valued state is a separate process from perception of that state (p. 47). Rolls demonstrates again and again, particularly in the chapters on hunger and thirst, that evaluation is delayed to ensure that the eventual emotional states are just as "informed" about the external environment as correct beliefs are.

The introverted view is based on the self-stimulation paradigm, and brings with it the idea that an animal who wants food may also be said to want a particular kind of internal event (p. 24, in italic). The latter may even be called the representation of reward (p. 32) in the brain. This is entirely consistent with Rolls's abstract notion of representation (p. 77), but it is nonetheless problematic. In the case of hunger, for example, I would prefer to say that the only thing represented is food. Hunger is the behavioural disposition, analysed in marvellous chemical and anatomical detail by Rolls, to seek the food out and eat it. But nothing in the execution of the disposition is a representation of anything but rather an effect of representation.

I expected Rolls to treat consciousness as an intentional state, as he did with emotion (p. 62), and to deal primarily with consciousness *of* things, the world of perception in particular. But the introverted paradigm prevails, and Rolls goes inward, virtually equating consciousness with the abstruse topic of qualia. The chapter on sex goes to the opposite extreme, outwards and back into sociohistory, but does so for the same reason, namely that it largely abandons the extraverted perspective, rooted in the perception of the immediate environment. I was uneasy about the connectionist appendix also, since it reinstates the one-dimensional, cognitive model that is so effectively challenged in other parts of the book.

Perhaps Rolls's necessary preoccupation with emotion has resulted in a reduced notion of perception, expectation, memory and the other processes that are tuned not to reward but only to truth or reality. Contrary to what Rolls suggests (e.g., p. 86), expectations, even in rodent learning, are not keyed to stimuli or to physical properties of things but to events (see Rozeboom 1960). And in primates, and possibly other species as well, there is a clear separation of content-maintaining processes in perception from those that subsequently fix conviction and lead to behaviour. We can accept that perception is "impenetrable" to a degree (Pylyshyn 1999), but behaviour is generally tuned not to objects but to highly penetrable events. Thus we can consider appearances and hold back assent (Mac Aogáin 1999). This is a capacity likely shared by all species that show exploratory behaviour and are attracted to novelty. It is the exact counterpart, in cognition, of the object/reward separation that is central to Rolls's account of emotion.

By acknowledging the symmetry, both emotion and cognition appear as attitudes or behavioural dispositions, the one tuned to the attractiveness of things, the other to their likelihood. We might even consider belief as a kind of feeling, not unlike emotion. "An idea assented to," Hume noted, "*feels* different from a fictitious idea, that the fancy alone presents to us" (Hume 1738/1911, p. 99). Indeed it does, and very likely the feeling has its own physiology, although it is truth or likelihood that is now being appraised, not reward potential.

Nonetheless, a single notion of representation, common to both attitudes, will suffice. This is because of the separation of reference and attitude alluded to, equally complete on both dimensions, emotion, and cognition. When representation is "free" or "view-invariant" (p. 90) in this sense, including independence from attitudinal bindings, we don't need separate varieties of representation for emotion and cognition. As for the "representation" of the attitudes, rewards, values, and so on, there is no such event, only the activation of a behavioural disposition in response to earlier representation. A weaker, non-intentional term such as "binding," or "decoding" is sufficient to describe such things.

Intelligence and emotion

Eucaly Mogi

Brain-Operative Expression Team Brainway Group, Brain Science Institute (BSI), The Institute of Physical and Chemical Research (Riken), Saitama 351-0198, Japan. **eucaly@brainway.riken.go.jp**

Abstract: The explicit system for action selection integrates emotional information with the higher-order cognitive processes which culminate in the language system. Even the basic feels of emotion are what they are because they are integrated into the higher cognitive processes. The relation between emotion and intelligence would become increasingly important as the focus of brain science shifts to the integrative function of the prefrontal lobe.

We all know what emotions are. And the laymen's main concern usually tends toward the conscious feelings of emotion. As Rolls writes (presumably from his own experience at Oxford), at the end of a series of lectures on the brain mechanisms underlying emotion, an undergraduate is quite likely to feel that the most important aspect of emotion has not been properly accounted for, namely the subjective feeling (qualia) of emotions such as sadness, joy, anger. The problem of qualia, let alone emotional qualia, involves philosophical and epistemological difficulties and it is fitting that Rolls leaves the discussion of conscious emotional feeling until Chapter 9, after having done the more mundane but important task of discussing in great detail the neural mechanism underlying emotion, where the main concern is the flexible association of stimuli with their survival value.

In his reserved but bold presentation of a theory of consciousness and its application to understanding emotion (Ch. 9), it is interesting that Rolls puts forward the view that qualia, raw sensory, and emotional feels arise after having evolved a linguistic and semantically based higher-order thought system. There is a parallel between this and the view that basic elements of intentionality (such as a dog being "directed" to the visual image of a bone, or, in more simple terms, "seeing the bone") are dependent on the "original intentionality" which presupposes the higher order abil-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

ities such as to accept responsibility (Haugeland 1999). On first hearing, Haugeland's requiring the dog to be morally responsible in order to see a bone sounds like a fantastic but terribly wrong idea. However, when one realizes that even the simple act of seeing something becomes significant only when it is integrated into the context and syntax of the higher order cognitive process, which (along with other things) support our sense of morality, one starts to appreciate the rather extreme view put forward by Haugeland. Intentionality (cf. Brentano 1973) and qualia (cf. Chalmers 1996) are widely considered to be the two major hallmarks of our conscious mental activities, and it is interesting that somehow ideas expressed by a neuroscientist (Rolls) and a philosopher (Haugeland) as well as others converge to the thesis that lower-order properties (such as feels of emotion and visual qualia) are dependent on the higher-order properties of our mental activities (such as morality and language).

When you think about it, emotions are not such lower-order basic mental activities after all. Emotions are in fact tightly coupled with what we call intelligence, including our verbal abilities to communicate, and facilitate cooperation in human society (Goleman 1996; 1998). In this sense, emotional responses provide the syntax and context for sensory stimuli and motor response, and are in a continuous spectrum with such intelligent mental activities as language. One aspect of our emotional ability, namely, the ability to read others' emotion, is closely related to the theory of mind (e.g., Baron-Cohen et al. 1997), which is now considered to be a key element in human intelligence as we know it.

As Rolls points out in Chapter 2, in brain areas devoted to specific sensory information processing, the reinforcement association of the sensory stimuli in general do not alter the response to the sensory stimuli. For example, the independence from reward association seems to be characteristic of neurons right through the temporal visual areas (e.g., Kobatake & Tanaka 1994). There is increasing evidence that the simple act of "seeing" is actually regulated by the top-down feedback to the visual cortex from the prefrontal area (e.g., Lumer et al. 1998), integrating the visual information into the more general cognitive context. However, at the visual cortex, the syntax and context when they are relevant are of such a kind that they naturally extend to language (as is evident from the fact that pattern recognition is a necessary condition for the linguistic processing of alphabets), but not necessarily to the survival value of the stimuli. Such a nature of the information processing in higher sensory areas might have contributed to the view that emotion and intelligence are relatively independent.

It is likely that it is in the frontal lobe that the integration of the sensory information and their survival value occurs, assisting the individual in integrating perception and action in such a way that he or she can survive better, not only in the natural environment, but also in the sociobiological sense (as Rolls discusses in sect. 10.5). The excellent chapter on the orbitofrontal cortex demonstrates its importance in the flexible association of reward value with stimuli. The integration of sensory and motor information processing is a very exciting topic in the neurophysiology of the frontal lobe today. In the same area (dorsal premotor cortex of monkeys) that the "mirror neurons" (Gallese & Goldman 1998) are found, neurons which possibly code for the "affordance" of objects are recorded, indicating that here sensory information is integrated into the processing of the motor information or motor repertoire. The frontal lobe is clearly the area where the sensory, motor, and "value" information are integrated, and it is very likely that exciting findings await us just around the corner.

Human intelligence arose from the biological need for survival. Emotions exist because they assist the survival of the individual, and it is only natural that emotion and intelligence are very tightly coupled. In the near future, much experimental and theoretical work will be done on the relation between emotion and high-order cognitive process, especially in the battlefield of the frontal lobe. Rolls's book on the brain and emotion is an excellent starter for this very important issue. This would be only an "Episode I." I hope very much that we will be able to read a sequel some time soon.

Is what you feel what you don't know?

Simon C. Moore and Mike Oaksford

School of Psychology, Cardiff University, Cardiff, CF10 3YG, Wales, United Kingdom. **{mooresc; oaksford}@cardiff.ac.uk**
www.cf.ac.uk/uwcc/psych/mooresc

Abstract: Rolls defines emotion as innate reward and punishment. This could not explain our results showing that people learn faster in a negative mood. We argue that what people know about their world affects their emotional state. Negative emotion signals a failure to predict negative reward and hence prompts learning to resolve the ignorance. Thus what you don't know affects how you feel.

Rolls argues that emotion can be viewed as the innate product of reward and punishment reinforcement. In this view, emotion is a product of the learning process. However, we have evidence that human emotion, irrespective of its cause, provides an impetus to learn from the environment in its own right. Consequently, Rolls's account of emotion may need to be modified.

We have investigated how emotion modulates the rate at which people acquire procedural skills and learn probabilistic discriminations. We first induced an emotional state using standard laboratory procedures. In one experiment, participants then completed a visual discrimination task (Moore & Oaksford 1999) and in other experiments they completed a probabilistic classification task (adapted from Gluck & Bower 1988). In both cases we measured the rate of learning, either across a single session or over a series of days. According to Rolls's definition of emotion one could not predict that an emotional state would influence the rate of learning for an emotionally unrelated task. However, we found that an induced negative emotion enhanced the consolidation of visual information across a two-week period, whereas positive emotion does not (Moore & Oaksford 1999). Moreover, within a single trial block, negative emotion enhanced the rate at which people learn a probabilistic classification task.

We have sought to understand the causes of these effects at the functional level, that is, why would it make sense for an organism to learn more rapidly about its world when in a negative mood? According to Rolls, emotional response consists of the experience of primary or secondary reinforcement. Fundamental to this definition is the idea that organisms seek to maximize their likelihood of survival. Maximizing survival is achieved by the approach and avoidance behavior associated with rewarding and punishing stimuli. Thus there is an inferred relationship between the behavior and the goal. Approach behavior is initiated when an organism infers a high likelihood of reward, and avoidance behavior is initiated when an organism infers a high likelihood of punishment. That is, to survive, organisms have to make decisions about whether to approach or avoid stimuli in their environments based on the likelihood of experiencing positive or negative reward. This suggests adopting a decision theoretic approach to emotion in which a major factor in determining the appropriate emotion is provided by our prior knowledge of the world.

In any given context, prior knowledge will provide information about the levels of utility associated with events and actions occurring within that context. It will also provide detailed knowledge about the probability of events and the likely consequences of our actions. If the overall goal is to maximize survival, then an organism must seek to perform actions that will minimize their chances of negative reinforcement. If the organism is very familiar with a particular context, then it will be able to avoid negative reinforcement. This suggests that even though two contexts both contain the possibility of the same level of negative reinforcement, the context where you known this can be avoided is less likely to engender a negative emotion. For example, smelling smoke in your own office, where you know where the fire exits are, is likely to induce less fear than smelling smoke in a colleague's office in an unfamiliar building. An important consequence of this decision-theoretic approach is that, given we wish to minimize the expectation of experiencing negative reinforcement, actually encoun-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

tering a punishing stimulus will usually occur unpredictably. To avoid such a stimulus in the future the most appropriate behavior must be to learn about the context in which it occurred, that is, to make it more predictable. For example, suppose you try to rush through a door that has always been unlocked only to find it is locked. From the bump on the head and the ensuing anger, you will learn to approach the door more cautiously in the future and perhaps attempt to discover when and why the door is locked or unlocked. We can now explain why our participants learned faster when in a negative mood. The experience of negative emotion is a cue to learn more about the environment so as to minimize the chances of further negative reinforcement.

Prior knowledge for the causal structure of a context mediates the emotional response. If you can accurately predict the causal structure then you can minimize the likelihood of receiving negative reinforcement. For example, at home, one may not be hungry but still feel happy in the knowledge that there is food available if it is needed. Conversely, if prior knowledge does not allow you to predict the causal structure of the context then your ability to avoid negative reinforcement decreases. The likelihood of finding food if it is required diminishes and the likelihood of running into danger increases. More generally these considerations suggest that simply the levels of uncertainty someone has about a context may sometimes be related to emotion. Consistent with this view, in some further experiments we have found that manipulating the level of uncertainty in a probabilistic classification task alters people's emotional state. The more uncertain they are, the more negative their mood state.

In sum, in this commentary we have been concerned with the etiology of emotion. In Rolls's book, emotion is the experience of reinforcing stimuli. We suggest that emotion may also be modulated by knowledge of the environment. The more you know the inherently safer it is.

Neural behaviorism: From brain evolution to human emotion at the speed of an action potential

Jaak Panksepp

Department of Psychology, Bowling Green State University, Bowling Green, OH 43403. **jpankse@bgnet.bgsu.edu**

Abstract: Rolls shares important data on hunger, thirst, sexuality, and learned behaviors, but is it pertinent to understanding the fundamental nature of emotionality? Important as such work is for understanding the motivated behaviors of animals, Rolls builds a constructivist theory of emotions and primary-process affective consciousness without considering past evidence on specific types of emotional tendencies and their diverse neural substrates.

Rolls provides a useful summary of his many important contributions to understanding motivated behaviors, and discusses their potential implications for our understanding of emotions and consciousness. Although I substantially disagree with his theoretical view, the sharing and contrasting of different perspectives is essential for refining ideas and thereby promoting substantive progress in the field. I shall focus on the relevance of his work and theoretical perspectives to understanding the basic emotions and primary-process consciousness.

All emotions and motivations are assisted in their evolutionarily appointed affairs by generalized learning systems, but does a study of such learning systems address the *fundamental* nature of emotions and consciousness? The relevance of many intrinsic subcortical sensory-motor integrative systems in the generation of emotions is underestimated in Rolls's analysis. General brain theories of emotion should be based on a thorough analysis of the basic emotional processes (e.g., fear, anger, separation-distress, and playful joy). An adequate framework cannot be extrapolated simply from a study of self-stimulation circuitry and motivations such as feeding and drinking.

On the one hand, Rolls accepts a reasonable multiplicity of functions that emotions subserve within the cerebral economy (p. 67–70), but he provides no adequate empirical or conceptual analysis of how many of these functions may actually be achieved within the brain. Evolution surely provided more guidance and organization to animal emotions than the mere influence of simple "good-GO" and "bad-DON'T GO" indicator systems, and a mysterious reinforcement process that connects these values to world events. Rolls does not address vast stretches of data regarding the natural behavioral and affective inclinations of animals and humans and their neural foundations, which do not fit well into his neo-behavioristic point of view (e.g., MacLean 1990; Panksepp 1998a). Although Rolls toys with the possibility of multiple reward and punishment systems, he does not acknowledge the existence of multiple emotion-affect programs within the brain.

As Skinner repeatedly asserted in his twilight years, the two vast gaps in a complete behavioral analysis were between "stimulus and response, and between reinforcement and a resulting change in behavior" – gaps which could only be filled with "the instruments and techniques of neurology" (see Panksepp 1998a, p. 12). Of course, there was also a third chasm – the one between evolutionary processes and the spontaneous psychobehavioral repertoire of the animals – the psycho-ethological gap that Skinner and most behaviorists have assiduously avoided, in both research and discussion. Although Rolls is now filling the first gap with impressive studies, the second remains almost as mysterious as ever, and behaviorists generally continue to ignore the third. Except for a detailed discussion of self-stimulation circuitry, Rolls does not deal with a host of other emotion-relevant circuitries, including the many neuropeptides that appear to be able to provide coordination for a variety of distinct emotional and motivational tendencies (Panksepp 1993; 1998a). Also, within the research program he details, he has yet to study animals extensively in profoundly emotional circumstances. Because of the restricted data base he has chosen to cover, we must wonder how well his observations and admittedly parsimonious theorizing can illuminate the diversity of spontaneous emotional inclinations that ethologists and psychologists have long recognized as the natural emotional repertoire of animals and humans (see Darwin 1872/1998).

Rolls is at his best when he is documenting the way discriminative reward and secondary reinforcement processes may emerge within the brain. He provides compelling data on those issues. However, his analysis of the "lateral-hypothalamic variety" of self-stimulation seems unconvincing considering what we know these days. Rolls still regards the fundamental nature of self-stimulation along the trajectory of the A-10 dopamine systems to reflect the presence of a reasonably straightforward reward/reinforcement process (Chs. 5 and 6), but much recent evidence (as alluded to by Rolls, p. 176), suggests that this system is better conceptualized ethologically as one mediating generalized anticipatory-investigatory phases of appetitive behavior rather than the consummatory-reward components (Ikemoto & Panksepp 2000; Panksepp 1981; 1986; Robinson & Berridge 1993; Schultz 1998). How such concepts as reinforcement and pleasure link up to this fundamental emotional system remains an open question, but most likely, it shall be a smaller part of the overall story than Rolls and most undergraduate texts make it out to be. The psychological effect may be more akin to an appetitive mood of eager anticipation that is clearly expressed in the spontaneous interest and investigatory behaviors of animals. Indeed, contrary to his distinction between emotions and moods (p. 62), it seems likely, from my point of view, that both emerge from essentially the same brain substrates (Panksepp 1994).

Although life events and the resulting appraisals obviously promote emotional episodes, how accurate is the claim that emotions emerge from the reinforcement contingencies of the environment? Although such events surely *trigger* and sculpt emotional

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

responses, there is no robust line of evidence that such contingencies causally integrate emotions in any deeper sense. Might it not be wiser to turn the issue around, and to seek an understanding of reinforcement mechanisms through the operations of the diverse emotional operating systems that exist within the brain (Panksepp 1990)? Even in Rolls's refined behaviorist view, the reinforcement process remains the phlogiston of learning theory. From a developmental systems perspective, the intrinsic emotional-affective processes of young animals appear to be essential mechanisms via which their worldviews are created via interactions with a diversity of world events (Buck 1999), but I am not aware of convincing data that a substantive "reinforcement" *process* is the main mechanism that elaborates such change. Behavioral change may well emerge from a variety of other plasticities within the nervous system. Thus, it seems unrealistic to restrict the term emotion to the cognitive decoding of the reinforcing characteristics of environmental stimuli (p. 62). It must also be related to careful analyses of "instinctive" psychobehavioral states. In short, Rolls's conceptual and evidential bases are not broad enough to handle the richness of emotional phenomena that is evident in the animate world. Rolls does not discuss the possibility that emotional systems operate as *global* state variables within organisms: In their fundamental form the basic emotions are generalized alerting and behavior organizing processes, capable of arousing several distinct forms of affective experience with no propositional contents. We now know from modern proto-oncogene studies that massive swaths of the brain are aroused during practically all emotional states (Beck & Fibiger 1995; Campeau et al. 1997; Kollack-Walker et al. 1997). These global states shift the whole intentional and attentional demeanor (psychobehavioral set) of organisms and thereby prepare them to deal with major types of life challenges.

Rolls does not emphasize that our emotional lives, and those of other animals, are developed more in relation to social situations than other dimensions of existence. A monkey in a restraining chair is not in much of a position to pursue such activities. Although, Rolls mentions in passing the diversity of social processes (e.g., Table 10.1), he ignores such essential topics as playfulness, social-bonding, separation-distress, and maternal nurturance – processes which appear to have dedicated brain systems/processes for achieving those social ends. Rolls does discuss sexuality at some length, but focuses more on evolutionary speculations concerning sperm warfare and genital dynamics than core emotional issues – mate selection, sexual bonding, jealousy, and the passionate nature of lust.

And then there is Rolls's top-down discussion of consciousness. He casts his lot with those who believe that consciousness is constructed from the higher symbolic capacities of the brain/mind. I admire Rolls for taking up the issue of emotional feelings, so commonly ignored by neuroscientists, but I was disappointed by his solution, which is not impressively consistent with a great deal of available evidence. One of many troublesome lines of work for his view is the rewarding effects of drugs (e.g., opioids, psychostimulants, and now others) as evaluated by intracranial self-injection (McBride et al. 1999) and conditioned place preference procedures (e.g., Olmstead & Franklin 1997). Deep subcortical structures are necessary, and perhaps sufficient, for both effects. To envision these influences as information, like any other type, within the higher cognitive reaches of the brain seems categorically incorrect. Emotions have very big brain effects, and all external signs that we have (except for propositional language), including abundant neural evidence, suggest that remarkable homologies and predictive validities exist in the subcortical affective substrates of the basic human and animal emotions (Panksepp 1998a).

There are ways to conceptualize the nature of affective consciousness in the lower reaches of the brain, as long as we respect the complexity and power of genetically ingrained emotional circuits and the body's motor systems in establishing a coherent "center of gravity" for the organism (Panksepp 1998b). To simply have a top-down cognitive view of consciousness which denies simple felt emotional and motivational states to other mammals strikes me not only as an improbable, but an ethically troublesome, proposition. It reflects a sensory-focussed worldview where the body, its various brain representations, and instinctual motor action tendencies are relegated to an undeserved secondary status in emotion and consciousness studies. Surely all would agree with Rolls that much of the higher brain was designed to guide the body sensorially and cognitively in space and time toward important internal and external goals, but the evidence suggests that the basic means to achieve those goals, in the form of affectively tinged action tendencies, were firmed up in primitive integrative systems of the brain long before organisms like us existed at the top of the food chain with aspirations to be closer to angels than the other animals.

Who would deny that our higher thoughts are important determinants for the subtle texture of our sophisticated cognitive consciousness? Also, reinforcement contingencies obviously mold behavioral tendencies. But Rolls's attempt to embed effective experience within those higher networks, even as he takes pains to clarify that his *language* "does not necessarily imply verbal language" (p. 262), strikes me as unrealistic. Were he simply focussing on cognitive issues, I would have found his overall approach congenial, but to envision global affective experience as reflections of cognitive-associative-reinforcement processes, experientially elaborated largely within higher sensory-associative regions of the brain, seems inconsistent with an extensive corpus of evidence (Panksepp 1998a; 1999). Just consider one of many difficulties: Children, who are much less able than adults to "think linguistically about" their "own linguistic thoughts" (p. 263), typically have stronger and more frequent (but mercifully shorter) emotional episodes than adults. Either Rolls should claim that our kids really do not have the strong feelings they seem to have, or he needs to deal with the paradox (at least for his viewpoint) in some more humane way.

In the long run, I anticipate that core emotional processes will be considerably more difficult to model computationally than cognitive processes. I doubt if primary-process affective consciousness can be instantiated in the ways Rolls envisions. If it could, major psychiatric/emotional disorders should have been more prone to alleviation through the application of learning principles than has yet been possible. I am glad Rolls emphasized that for any machine to be propositionally conscious, it must partake of "linguistic thoughts with symbols grounded in the real world" (p. 262). However, I suspect such "organismic" creatures would be zombies unless one were also able to give them substantive and well-crafted central motor and emotional processes of the kind found subcortically in living organisms (Panksepp 1998b). That is a task no one has yet come close to accomplishing (Picard 1997). Indeed, as Freud (1923) suspected (see Solms & Nersessian 1999 for an updating), an understanding of how the higher reaches of the brain achieved subtle propositional types of consciousness, may require a much clearer understanding of how the lower, affective forms are created from "embodied" brain processes that we share with the other animals.

The amygdala – responsible for memories of reward as well as punishment?

Amanda Parker

School of Psychology, University of Nottingham, Nottingham NG7 2RD, United Kingdom. **aep@psychology.nottingham.ac.uk**

Abstract: Rolls's proposal that the amygdala is critical for the association of visual objects with reward is not consistent with recent ablation evidence. Stimulus-reward association learning is more likely to depend on basal forebrain efferents to the inferior temporal cortex, some of which pass through the amygdala. It is more likely that the amygdala is involved in rapid modulation of stimulus reward value.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Rolls reviews the substantial body of evidence, much of it from his own laboratory, on the neuroanatomy of emotion. Although the coverage of this area in the book is somewhat idiosyncratic (e.g., much of it might better be described as dealing with motivation), many important issues in the field are discussed in an accessible manner. A particular strength of the book is its message that we can define emotions in terms of the reaction of the animal to rewards and punishments. This is a convincing argument for monkeys and rats, and is a useful starting point for extending our understanding of emotion generally. In this commentary, I will concentrate on one aspect of the book, the role of the amygdala in emotion. More specifically, I will argue that Rolls's proposal that the amygdala is critical for the association of visual objects with reward is not consistent with recent experimental evidence from ablation studies.

Rolls's proposal is that the ventral visual system projects via the inferior temporal visual cortex to the amygdala and orbitofrontal cortex, which together determine the reward or punishment value of the object, as part of the process of selecting which goal is appropriate. He states that "Lesions of the macaque amygdala impair the learning of both stimulus-reward and stimulus-punisher associations." (Précis, sect. 5.2.2). This statement is not supported by current evidence from lesion studies. A series of elegant and thorough experiments in rat by LeDoux and colleagues have convincingly established the importance of the amygdala in stimulus-punisher learning (for review, see LeDoux 1996), in agreement with Rolls's proposal that "the crucial site of the stimulus-reinforcement associative learning that underlies the responses of amygdala neurons to learned reinforcing stimuli is probably within the amygdala itself" (Précis, sect. 5.2.1). However, the same cannot be said for the importance of the amygdala in stimulus-reward association learning (SRAL). Whilst early evidence from aspiration ablation studies suggested a role for the amygdala in SRAL (see Gaffan 1992, for a review), a very different consensus has now been reached on the basis of two recent sets of findings.

First, the recent work of Murray and colleagues has established that many of the functions previously ascribed to the amygdala are in fact functions performed by the adjacent rhinal cortex (for reviews see Baxter & Murray, in press; Murray 1992), and that the amygdala itself has little or no role in stimulus-reward memory. However, the amygdala does have a role in reinforcer devaluation. In this experimental paradigm the monkey has previously learned that one set of positive objects are rewarded with one type of food and the remaining objects rewarded with a second, equally palatable food. After a selective satiation with one of the foods, normal monkeys will choose far more of the other food than is their normal preference. This effect is abolished in monkeys with excitoxic amygdala lesions (Malkova et al. 1997). The conclusion that can be taken from this is that the amygdala itself is important to reward association memory only on occasions when an abrupt change in the current value of a reinforcer that has an already-established value is required, in order to produce an adaptive response. This conclusion can also be applied to the functional relationship between the amygdala and the orbital prefrontal cortex (Parker et al. 1999).

Second, recent work of Gaffan and colleagues has established that bilateral transection of the amygdala and anterior temporal stem disconnects inferior temporal cortex from its afferent connections in the basal forebrain (Gaffan et al. 1998; submitted). This disconnection severely impairs SRAL, an effect that was previously ascribed to amygdala damage (see Gaffan 1992; Easton & Gaffan 2000, for reviews), and now can be ascribed to interruption of fibres of passage through the amygdala. This conclusion has been strengthened by experiments which use crossed unilateral lesions, in which a heat lesion in the basal forebrain in one hemisphere is combined with a lesion of contralateral inferior temporal cortex. A very similar deficit in SRAL is observed in monkeys with this pattern of lesions to the animals with bilateral amygdala plus anterior temporal stem transection (Easton & Gaffan 1997; submitted), and a similar equivalence between surgical groups in recognition memory deficits can also be seen (Gaffan et al. 1998; submitted; Easton et al., submitted). We can therefore conclude that the basal forebrain, rather than the amygdala, is the crucial structure which modulates memory storage in inferior temporal cortex in SRAL.

Novelty assessment is often proposed to be an important precursor of memory formation, and Rolls and his colleagues have found cells in several key areas of the brain that respond preferentially to novelty. As with the discussion of SRAL above, however, evidence from single cell recording studies alone do not provide a convincing explanation of the relationship of novelty-related processing to memory formation. Rolls, citing the research of Wilson and Rolls (2000), proposes that the amygdala is filtering information about whether visual stimuli are either novel or reward-related, and making decisions about output based on these evaluations (Rolls 1999a, p. 105). In support of this, he states that lesions of the amygdala in macaque negatively affect this process (specific experiments unspecified). This lesion effect is more likely to be due to the effect of damage to the rhinal cortex than to damage to the amygdala itself. Using a series of memory tasks which manipulate the relative novelty of visual stimuli, we have found that aspiration lesions of the amygdala which do not damage rhinal cortex do not abolish the advantage for novel visual stimuli seen in normal monkeys (Parker et al. 1998). In contrast, crossed unilateral lesions of perirhinal cortex and prefrontal cortex do abolish this novelty advantage (Parker et al. 1998), as do bilateral ablations of the perirhinal cortex (Buckley et al. 1999). It therefore seems unlikely that the amygdala is crucial for the novelty-related processing that precedes visual object-reward associations.

Awareness may be existence as well as (higher-order) thought

Jordan B. Peterson

Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 3G3. **peterson@psych.utoronto.ca**
psych.utoronto.ca/~peterson/welcome.htm

Abstract: Rolls attributes to consciousness the functions of reflection, planning, and error-correction. Neuropsychologically grounded cybernetic theory provides an analogous, broader conceptualization: consciousness constructs goals (and plans), alters the valence of goal-related phenomena, registers error-signals, and explores unexpected circumstances (reconfiguring goals and plans as necessary). Consciousness plays a fundamental unrecognized ontological role, as well, conferring the status of "discriminable object" on select aspects of otherwise indeterminate "being."

Rolls identifies consciousness particularly with higher-order thought, describing it as "the state which arises in a system that can think about . . . thoughts" (p. 248). He believes consciousness has two primary functions. The first, potential for reflection on past events, has a corollary, planning for future events. The second, integrally related to the first, is correction of error made by lower-order processes.

Rolls suggests that higher-order thoughts intervene when lower-order-thought-predicated plans fail (p. 250). Higher-order thoughts perform this operation (linguistically) by analyzing the structure of lower-order plans, identifying specific weak links, and replacing those with potentially better alternatives. Rolls's theoretical model can therefore be assimilated to the cybernetic viewpoint (Wiener 1948) (and psychological elaborations thereof). Such assimilation is advantageous because it unites Rolls's work with a well-developed alternative body of theory, and allows his ideas regarding consciousness further integration with current theories of emotion (not excluding his own).

The plans that organisms formulate include goals, from the cy-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

bernetic perspective (Wiener 1948). Goals are hypothetical future world-states similar to those currently obtaining, but modified in a manner both attainable and beneficial. The desired future state may be viewed as an (imaginal) template, against which current insufficient states of the world are compared, as behavior is undertaken (Peterson 1999). A "plan," in this scheme, is that sequence of behavioral steps posited as necessary to produce the desired future condition (Carver & Scheier 1998). Plan-formulation may be the responsibility of the higher-order thought system described by Rolls – but so (explicitly) must be abstract goal-formulation. Higher-order thought systems performing this latter function may modulate emotional valences directly (may affect the "reinforcing value" of stimuli) as well as formulating or altering plans (Peterson 1999). This valence-modulating capacity constitutes a major extension of the role of consciousness, even when defined in a manner otherwise similar in all regards to that proposed by Rolls.

Rolls (p. 61) presumes that emotions are states produced by "instrumental reinforcing stimuli," and suggests that this presumption allows for "operational definition of what causes an emotion." However, what might be instrumentally reinforcing in one situation (whether "subjective" or "objective") may not be in another. This implies that Rolls's definition is less than optimally "operational." All sophisticated behaviorists recognize that the current status of an animal in part determines the reinforcing nature of a given "stimulus" (Rolls in fact admirably details the mechanisms by which such determination occurs, when he describes brain control of feeding and reward). But there is a cognitive component of "current status," analogous to higher-order control and modification of current plans. Formulation of an abstract goal (a future "desired state") instantaneously transforms that goal into something equivalent in all essential features to a consummatory reward (Peterson 1999). It is in this manner that higher-order cognition and emotion meet. A cognitively constructed "consummatory reward" may be something as abstracted away from instinctive significance as a goal scored during a soccer game (an occurrence for which individuals will work, and one whose possibility also colors all other game-events: opposing players and their maneuvers become "threats" or even "punishments," assuming they interfere with scoring chances or score themselves; scoring opportunities are "incentive rewards," indicating progress towards a consummatory goal – and all this in the absence of conditioning!) This means that emotions may be more operationally considered "states produced by stimuli whose reinforcing properties derive from their relationship to some explicitly formulated goal"; and means further that the notion of stimulus-valence is not meaningful, unless current organismal goals are explicitly specified (although sometimes valence evaluation may indeed take place in the absence of higher-order thought; see LeDoux 1996, pp. 161–69).

Consideration of the role played by higher-order thought in goal-specification and consequent valence-determination also allows for a more comprehensive view of error detection and correction. Sokolov (1969) suggested that the orienting reflex (a complex psychophysiological response, associated with initiation of exploratory behavior) emerges automatically when the goal-directed nervous system detects an unexpected "stimulus." "Unexpected" in this context means "deviant from plan (the behavior manifested did not produce the result desired)." Orienting and more complex exploratory behavior garners new information, used either to modify the ongoing plan or, if necessary (and much more problematically), to eradicate and reconstruct the current goal (Peterson 1999). Vinogradova (1975) suggested that the comparator designed to detect such deviation/novelty resided in the hippocampus. Gray (1982; Gray & McNaughton 1996) associated the function of this novelty detection system with the emergence of behavioral inhibition and anxiety. LeDoux (1996) and others, including Rolls, associate anxiety more particularly with amygdalic function, so the relationship between hippocampal novelty-detection and behavioral inhibition/anxiety appears as of yet unclear. However, the notion that novelty is a primary source of anxiety (and knowledge) should take center stage in any discussion of the relationship between cognition, emotion and consciousness (Peterson 1999), as initial caution (and then exploration) is so clearly appropriate, when unexpected relations emerge between desire and world.

Consciousness therefore appears as higher-order (linguistically mediated) correction of lower-order plans, as Rolls suggests, and more. Consciousness also establishes explicit goals (while informed by other brain processes indicating biological necessity); formulates plans designed to attain those goals (or arranges already automatized plans to the same end); re-evaluates the significance of ongoing events (as a consequence of establishing goals and plans); registers signals indicating emergence of unexpected events (feeling not only anxiety but "hope" and "curiosity" while doing so; Peterson 1999); and explores, gathering new information and reconstructing goals and plans in the face of failure.

One more radical attribute of consciousness may also be posited, as a consequence of operation within this expanded conceptual framework. Rolls notes that inputs from sensory systems must be registered within consciousness, for the purposes of planning (p. 251). However, the sensory plenum as such is too complicated to constitute an object (or even many objects) of consciousness (Medin & Aguilar 1999). This means that sensory awareness is selective: the world is necessarily parsed up into the limited set of "objects" functionally relevant to the operations of current goal-directed operations, and the necessarily co-existent category of "all things presently irrelevant" (and therefore "ignored" or "unconscious").

Only functionally relevant objects "exist" at any given moment – constituting figure, so to speak, instead of ground. So the very fact of discriminable things appears as something dependent upon consciousness. This makes consciousness something far more fundamental than generally supposed (makes it something far from epiphenomenal) (Peterson 1999), and helps account at least by inference for its continuing incomprehensibility.

The causal status of emotions in consciousness

Jason T. Ramsay and Marc D. Lewis

Human Development and Applied Psychology, University of Toronto, Toronto, Canada M5S 1V6. **{jramsay; mlewis}@oise.utoronto.ca**

Abstract: Rolls demonstrates how reward/punishment systems are key mediators of cognitive appraisal, and this suggests a fundamental, causal role for emotion in thought and behaviour. However, this causal role for emotion seems to drop out of Rolls's model of consciousness, to be replaced by the old idea that emotion is essentially epiphenomenal. We suggest a modification to Rolls's model in which cognition and emotion activate each other reciprocally, both in appraisal and consciousness, thus allowing emotion to maintain its causal status where it matters most.

Although Rolls covers a wide area of neuroscience research, there are two themes that come to the fore. First is the review and integration of a large amount of basic research on the structure and organization of the brain systems that are central to the production of emotional behaviour. The second theme is the thesis that all emotions are caused by reward/punishment contingencies and that this formulation is consonant with appraisal theories of emotion.

We will argue that Rolls's version of appraisal theory places the two themes, one integrative, and one reductive, at odds. This becomes most evident in Rolls's consideration of the nature of consciousness, where the causal role of emotions is blurred. Finally, we will sketch a solution to the dilemma that is predicated on the idea that emotions inform appraisals even while they are initiated and constrained by those appraisals.

Rolls situates his definition of emotion within the existing frame-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

work of appraisal theory (Smith & Ellsworth 1985). He argues that emotions are the result of cognitions about specific events. He simplifies this proposition further by arguing that it is the evaluation of whether an event is rewarding or punishing that elicits an emotion. Rolls emphasizes a causal role for emotion by enumerating its functions. Emotion instigates autonomic changes in the body, is a prime motivator, and allows for behavioural flexibility. Other functions include the triggering of memory retrieval processes and the consistency of motivation and behaviour across time. Thus, in Rolls's scheme, emotion is the end result of cognitions about reward, but it still has a causal role in determining the contents of thought, memory, and behaviour. So far, Rolls's theory of emotion is consonant with traditional cognitive theories of appraisal. Rolls musters a detailed review of neuroanatomical and neurophysiological research in support of this argument. His main conclusion from this review is that the amygdala and the orbitofrontal cortex are especially important in the production of emotion. This is where Rolls could consolidate his model of cognition-emotion relations. However, whereas Rolls maps the anatomical connections in detail, he is significantly less elaborate on the dynamic interplay between them. For example, lesions or injuries to either the amygdala or the orbitofrontal cortex can produce strikingly similar behavioural sequelae. But there is little discussion of the interconnections between these structures that might account for this parallel. As a result of this underdeveloped account of brain structure and function, Rolls seems to reduce his appraisal theory to a strictly sequential model which in turn limits brain evidence he is willing to consider. At one point, Rolls dismisses evidence of direct connections from the thalamus to the amygdala (LeDoux 1992) on the a priori notion that such connections are meaningless because no emotion can be generated without cortical intervention.

After proposing his model of appraisal-emotion relations, Rolls turns his attention to developing a model of consciousness. Up to this point, emotion and cognition have both been key players in goal-related thoughts and behaviours. In Rolls's model of consciousness, however, emotion falls by the wayside. Cognitions about cognitions (second-order thoughts) are viewed as the basis for consciousness, with emotions relegated to qualia or "raw feels." Consciousness seems to arise from higher-order reflections on a variety of cognitive events. These events would have to include not only initial appraisals but also ongoing goal-directed thoughts such as retrieved memories. Earlier in the book, emotions were argued to be causal antecedents to a variety of mental and behavioural events, including retrieved memories. Yet no functional role is allotted to emotion in consciousness, and its causal status now verges on epiphenomenalism. Rolls asserts that consciousness is the result of lower-order constituent components and processes, such as basic conditioned learning and appraisal processes. Why is emotion not among these constituent processes? It does not seem surprising that consciousness affects emotions, but it does seem surprising that emotions do not affect consciousness. One means of getting around this dilemma would be to renovate Rolls's model of appraisal according to the dynamics of the neural structures that he describes in such detail. Rolls ends his summation by highlighting the importance of the outputs from emotional systems, such as those from the amygdala to the orbitofrontal cortex. Thus, Rolls's appraisal theory already allows for a dynamic, reciprocal interplay between cognition and emotion. Whereas traditional appraisal theories favour a lock-step sequence from cognitive appraisal to emotion, a new generation of models allows emotions both to cause and result from appraisal (Frijda 1993; Lewis 1996). Reciprocal causation between emotion and appraisal permits appraisals to grow – literally to self-organize – in a brain which is highly susceptible to self-amplifying interactions across cortical and limbic subsystems (Freeman 1995). It is no great leap to view consciousness in the same light, as an emergent process whose causal constituents include emotion as well as cognition. The neural evidence points to dynamic, rapidly fluctuating, highly interconnected pathways among all subsystems involved in cognition, perception, and emotion. Most important, feedback is the hallmark of this activity, with each brain structure returning the majority of its outputs to the structure from which it receives most of its inputs (Freeman, in press). In such a system it is entirely possible for emotion to maintain its causal status, not only with respect to appraisal but with respect to consciousness as well, inhibiting certain cognitions while initiating and perpetuating others. This take on emotional causation is not inconsistent with Rolls's punishment/reward model, yet it captures what is known about the reciprocal interconnectedness of the limbic and prefrontal systems.

In conclusion, Rolls's model of the role of emotion in cognition and behaviour does not fully jibe with his model of consciousness nor does it sit well with the neurodynamics of brain function. Modifying Rolls's view of appraisal to include mutual causation between cognitions and emotions could develop a more integrated theory.

Affect systems and neural systems

Eric A. Salzen

Department of Psychology, Aberdeen University, King's College, Aberdeen AB24 2VUB, Scotland. **e.salzen@abdn.ac.uk**

Abstract: The "reward" systems described by Rolls are systems for drive-reinforced associations of contact and distant stimuli and not for emotional behaviours. The neural systems delineated may be associated with distinct categories of "affect," namely "hedonic feelings," "moods," and "emotions." Awareness of these affects requires external perceptual as well as internal feedback. Levels of feedback in evolution and development suggest sensory qualia may not require language.

Rolls's chapters on feeding and drinking are the core of his book and they show the neuronal basis of the association of learned secondary reinforcing stimuli with innately effective primary reinforcing stimuli. In ethological terms primary reinforcers are proximal stimuli acting on contact receptors, touch and taste, and elicit innately organised consummatory behaviour, while secondary reinforcer/incentive stimuli are distant stimuli acting on distance receptors, sight, sound, and smell, to elicit variable responses of approach-withdrawal adjustments or appetitive behaviour that achieves or removes contact stimulation. This association takes place in neurones in the amygdala and in the orbitofrontal cortex but is operative only according to the relevant drive state which, for both feeding and drinking, is the homeostatic balance/imbalance relayed by hypothalamic neurones. Homeostatic balances are the ultimate reinforcers as is clear from the older physiological experiments on motivational systems so that, as ethologists once debated, consummatory stimuli and/or consummatory responses may be immediate goals and reinforcers that terminate behaviour but they are short-term. Rolls comes to the same conclusion. His reviews of brain-stimulation and pharmacology of reward are consistent with this since the effects of "artificial" direct stimulation of the reinforcing neural sites are still drive dependant.

Consequently Rolls's critical neurones in the amygdala and orbitofrontal cortex are association neurones rather than reinforcing. But their association is weighted by drive state so that the output is a call for action which goes to action systems, motor and autonomic. Rolls gives an excellent account of the basal ganglia where the different drive action-cells compete for appetitive actions which can interact with cortical motor and somatosensory inputs representing learned appetitive action-patterns so that both stimuli and responses can become learned incentives to action. But what happens when these action calls are in balance, that is, conflict? This is the core of ethological analyses of agonistic, courtship, parental, and other social displays that are comparable with human behaviours that we would all label as emotional (and

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

which, apart from sociobiological speculations on sexual behavior, are hardly featured by Rolls). These conflict displays are the basis of my thwarted action state signalling (TASS) theory of emotion (Salzen 1991), which gives a restricted (and literal) meaning to the term "emotion."

When considering the neurology of emotion, William James (1890) argued that "either separate and special centres, affected to them alone, are their brain-seat, or else they correspond to processes occurring in the motor and sensory centres already assigned, or in others like them not yet known." His own theory of emotion led him to conclude that there should be "no special brain-centres for emotion." Most studies assume that there are special centres and describe mid-brain and hypothalamic autonomic systems, the limbic system, and related neocortical cognitive systems (cf. LeDoux 1987; 1996). Since these structures are also involved in non-emotional behaviours the result is uncomfortably like James's view. Much depends on what is meant by the term "emotion" and, as Smith and DeVito (1984) have noted, the concept of emotion has to be properly specified before its neural basis can be elucidated. TASS theory makes clear distinctions between "emotions" and other "affective states" and in 1993 (Salzen 1993) I delineated neural systems that correspond with these affective states as follows:

1. Hedonic feelings are the sensations and associations of appetitive and aversive consummatory behaviours (pleasure and pain) involving the inferotemporal-amygdala-hypothalamic-midbrain axis, with a lateral appetitive and a mid-line aversive division (Gray 1987; Olds 1976).

2. Moods are the states of endocrine and hypothalamic drive systems which share a general dopaminergic appetitive motor activation system, the nigro-striatal-accumbens pathway, giving a general level of "get-up-and-go" (mania and depression).

3. Emotions are aroused motivated action states with increased appetitive orientation behaviour and incipient consummatory postures and movements that are blocked or in conflict (unpleasant emotions) and so form displays which induce behaviours in social partners that may end the thwarting state (pleasant emotions). The septo-hippocampal system with serotonergic "stop" and noradrenergic arousal described by Gray (1982) fits this pattern of arrest of specific actions with sustained perceptual scanning for cortical detection, learning, and recall of the thwarting circumstances.

4. Sentiments are affectively biased attitudes and beliefs (cognitive states). They might be expected to involve the posterior association and prefrontal cortices for perception and emission of affective signals as knowledge and language, and for the downward control (cf. Szentagothai 1984), that is, self-control of the affective action systems.

Using these definitions it is apparent that Rolls's book is really about hedonic feelings and their association neurone correlates. His amygdala- and orbitofrontal- hypothalamic- system corresponds with "feelings" (pleasure and pain) rather than emotions. His treatment of the nigro-striatal-accumbens-, orbitofrontal and amygdala dopamine system corresponds with "moods" and the "go" system. His excellent account of the basal ganglia provides a neural correlate that is needed by TASS theory to account for the conflicts of incipient motivational actions that form emotional displays. Gray's septo-hippocampal "stop" system may still be involved when behaviours are blocked by inappropriate or inadequate external stimuli.

I conclude, therefore, that Rolls is dealing with feelings rather than emotions. He is aware that the feeling experience "qualia" of these drive-reinforced (i.e., reward) associations requires "back-projections" for further neocortical processing and specifically by a language system. TASS theory proposed additional feedback by external perceptions of the individual's own bodily behavioural actions and stimulus interactions as crucial at all levels of this interaction (sensorimotor, feeling, emotional, and cognitive) for the experience of affects and suggested corresponding levels of awareness in evolution and in development. In developing this view (Salzen 1998) I have indicated how language can have arisen from emotional vocalisations that are perceived just like those of another person so that one can talk to oneself *as if to another person*. Thus, unlike Rolls, I think that language may be necessary only at the cognitive level of consciousness.

Emotions and reward – but no arousal?

Holger Ursin

Department of Biological and Medical Psychology, Division of Physiological Psychology, University of Bergen, N-5009 Bergen, Norway.
holger.ursin@psych.uib.no

Abstract: This commentary argues for the inclusion of the neurophysiological arousal concept to help understanding the brain mechanisms of emotions and reward and the cognitive mechanisms involved.

In the excellent review of emotions and reward systems by Rolls, I miss a discussion of a few items crucial to my own understanding, at least, of how the brain works. For me, it is hard to conceptualise relationships between needs, drives, emotions, and the brain without discussion or referring to any form of arousal-concept and arousal-theory. Rolls has a few references to brain "arousal," but this concept never attains any meaningful role in his theoretical framework.

In particular, I miss a discussion of the relationship between affective and instrumental aspects of emotional behaviour (Ursin 1985). Two factor-theory is discussed, but within a very limited learning-theory scope (Gray 1975). This is discussed as a transition from classical to instrumental conditioning, but without any apparent consequences for the internal state of the animal, in particular the level of arousal, evident as overt behaviour, or brain arousal, or observable psychophysiological, psychoendocrine, and psychoimmune changes. There is no mention or discussion – as far as I can see – of the dramatic shift from a highly aroused subject in the early phases of avoidance-learning, to the relaxed and nonchalant behaviour of the late phases during this learning. Likewise, I see very little emphasis on the distinction between affective and instrumental aspects of emotional behaviour – in animals or in humans. There seems to be a considerable consensus for this distinction for aggressive behaviour in humans since Festinger (1957), and it has also been used for fear behaviour and in stress theory (Levine & Ursin 1991). Aggressive or fear motivated behaviour may be executed at high levels of perfection and intensity, but with a low level of arousal.

Another cognitive dimension I miss, related to the changes from affective to instrumental behaviour – is the ability of all brains to store and consider the expected results of behaviour. The expectancy concept has been a part of cognitive formulations of learning theory since Tolman (Bolles 1972). In my opinion, this makes learning theory more useful for those of us who want to predict somatic changes due to psychological factors, both physiological and pathophysiological. For me, it is difficult or impossible to handle stress and stress related illness, complaints, and disease without referring to an expectancy theory. Rolls, on the other hand, does not help in making these relationships, as far as I can see. He sticks to an almost extreme S-R tradition. I am reminded of the famous cartoon found in many learning laboratories, also our own, demonstrating a rat contemplating a learning theory psychologist: "S-R, S-R, don't these guys ever think?"

All the impressive, albeit sometimes incomprehensible, network models seem to be S-R models with little or no emphasis on feedback, or with any openings for a cost-benefit evaluation of possible action, based on acquired expectancy of results. The model must at least consider changes from one motivational system to another. A hungry rat or a thirsty rat does not necessarily remain in a state of high arousal. The arousal level depends not on the level of thirst, but the probability of obtaining water. Uncer-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

tainty is associated with high levels of arousal. Certainty of water, or no water, is associated with low levels of arousal (Coover et al. 1984).

Finally, as an old amygdala hand, I appreciate the survival of Rolls and the Oxford group through the hardships of the molecular period of neuroscience. However, when more and more scientists realise – and get funded for – radical notions of the brain actually being involved in behaviour, rather than just pushing sodium and calcium around, we may end up with excessively molar concepts. The amygdala complex remains a complex, and the level of resolution of modern imaging techniques is too low to contribute to localisation studies. Passage of one synapse is a meaningful event, and the concept of systems that jump from one large population of nerve cells to another may be awfully wrong. In an old study with Jim Olds, Reidun Ursin and I demonstrated that the reward properties of neurones in the dentate area differed from those of hippocampal pyramid cells (R. Ursin et al. 1966).

To me, the really exciting thing in contemporary neuroscience is the confirmation of the fruitfulness of a conceptual nervous system. The brain may not only store consequences of behaviour; it may also consider them before further action is to be taken. I am extremely impressed by the reward system of the Rolls brain, but I still think there are good reasons for a rat to be "lost in thought" at any choice point in life. There are not only a lot of stimuli, but also a lot of consequences of behaviour to be considered before further action.

Is the higher order of linguistic thought model of feeling adequate?

Robert Van Gulick

Department of Philosophy, Syracuse University, Syracuse, NY 13244.
rnvangul@syr.edu

Abstract: Despite its explanatory value, the "higher order linguistic thought" model comes up short as an account of the felt aspect of motivational states.

Rolls offers a tentative explanation of the felt experiential aspects of conscious states (Ch. 9). On his model, consciousness arises from higher order linguistic thoughts (HOLTs). A mental state M is conscious just if it is accompanied by a roughly simultaneous higher order thought to the effect that one is in M. Moreover, the qualia or feeling of anything is "the state which is present when linguistic processing that involves second- or higher-order thoughts is being performed." (p. 249) He admits that his model is preliminary, and adds that the criteria for assessing theories of qualia are less clear than for those for the other topics he discusses.

Rolls's HOLT model has important explanatory value, but it comes up short as a general model of consciousness. Fuzzy criteria notwithstanding, it clearly fails to answer some key questions about consciousness. The model may be empirically adequate in so far as higher order linguistic processing may in fact accompany conscious awareness in humans, but evidence of regular correlation in itself fails to explain *why* it does so. In particular it does not suffice for understanding *why* and *how* the neural and organizational substrates give rise to experiential or phenomenal awareness. What is it about such processes that makes it feel like something to undergo them?

Rolls admits that some residual mystery surrounds the step from the sort of processing model he invokes and the phenomenal consciousness he aims to explain. In response he says the following, "if a human with second-order thoughts is thinking about its own first-order thoughts, surely it is very difficult for us to conceive that this would NOT feel like something?" (p. 249). This is not as obvious as Rolls supposes; for example, a Freudian process of repression might requires HOLTs but not feel like anything at all in so far as it is an unconscious process. Moreover even if true, it is beside the point with regard to the residual mystery. As a matter of how humans work, human HOLTs may involve phenomenal awareness, but the mystery is *why* that should be, and *how* the substrate for the former produces the latter. It may be equally difficult to imagine ourselves retrieving some episodic memories without there being any feel to the process. But that too would offer only correlational evidence about the sorts of processes that *in us humans* typically involve phenomenal feel. One must not equate empirical adequacy in the merely correlational sense with explanatory adequacy.

An adequate model of phenomenal consciousness need not explain everything about qualia, just as an adequate biochemical model of genetics need not explain every aspect of growth and reproduction. But it must provide us with an intelligible and mystery-dispelling basic account of how the phenomenal or qualitative supervenes upon the underlying neural and functional organization. We have that sort of insight in the DNA case; we can "see" in some satisfying way how the levels fit together. We cannot as yet do so with consciousness, and the HOLT model fails in that regard as well.

Elsewhere in Chapter 9, Rolls specifically addresses the issue of why qualia accompany emotional and motivational states, for example, "why food deprivation makes one *feel* hungry." (p. 251) However his answer seems to conflate something's being conscious in the phenomenal sense with its being conscious in the sense of being available to higher level processing. Rolls notes that the inputs from sensory and motivational systems must be available to the linguistically based planning system in order for it to do its job. He then argues thus: "I suggest that it would be a very special-purpose system that would allow such sensory inputs, and emotional and motivational states, to be part of (linguistically based) planning, and yet remain unconscious." (p. 251) Perhaps so, but this implies only that they should be conscious in the sense that we have cognitive access to them, not that they should be conscious in the phenomenal sense that there is some way it feels to be aware of them. The issue of qualia and raw feels floats through unexplained.

Rolls distinguishes his account from other higher order thought (HOT) models by adding the condition that the relevant higher order processing be linguistic in nature, though he construes "linguistic" broadly to cover pretty much any generative syntactic mode of representation. This creates both advantages and potential problems. On the plus side, it allows him to exclude simple cases of higher order mentality that do not seem intuitively conscious; for example as described in Chapters 2, 7, and 8, the control systems regulating action must be sensitive to multiple reward parameters of competing motivations. This would seem to require some sort of higher order representation of mental features, but it would not seem sufficient in itself for consciousness. Rolls can exclude these in so far as the higher order representation is not linguistic. On the other hand the requirement threatens to deny consciousness to many nonprimate animals that we intuitively regard as conscious in the phenomenal sense. Even given a liberal reading of "linguistic," how low can we go phylogenetically and still hope to find linguistic higher order thoughts? What of mice, sparrows, frogs, or squid? Yet unless we are willing to attribute HOLTs to all such creatures, Rolls's account would seem to entail that they do not feel any pain or experience any pleasures. They have pains and pleasures in the functional sense vis-à-vis their reward and action control systems, but since Rolls states that the feeling of anything is the state that is present when HOLT processing takes place, it would seem to follow that if there are not any HOLTs then there is not any feeling. At least that is how it seems, if I read Rolls correctly.

A final question. If, as I have argued, the HOLT model is not adequate in key aspects as an account of consciousness, what implications if any does that have for the rest of Rolls's theory? He makes some effort to screen off the material of Chapter 9 as intriguing but less based on scientific fact than the rest. Thus perhaps it could be set to one side with little overall effect. But I am

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

not sure that it can. Though established facts about pathways, informational relations and computational organization would remain intact, questions might arise about the completeness of the functional story. If we assume that the felt aspects of motivational states make a difference to the processes into which they enter (i.e., if motivational qualia are not epiphenomenal) then any functional account will need to explain that distinctive role. Kicking the qualia "upstairs" to the HOLT level has the advantage of confining their effect to that elevated domain of sophisticated syntactic processing. But if they penetrate further down, then any adequate functional account will need to explain what difference they make at those lower levels as well.

Innate psychology and open-ended processes: Finding the middle ground

David Sloan Wilson

Department of Biological Sciences, Binghamton University, Binghamton, NY 13902-6000. **dwilson@binghamton.edu**

Abstract: Rolls's mechanistic account of emotion can help to bridge a rift within the field of evolutionary psychology. One side of the rift emphasizes the importance of innate psychological mechanisms that evolved to solve specific problems encountered in the ancestral environment. The other side emphasizes learning, development, and culture as open-ended evolutionary processes in their own right. Rolls shows how these two views can be reconciled, allowing a productive middle ground to be explored.

Long term progress in biology requires equal attention to function and mechanism. Knowing what a biological system has evolved to do is essential for understanding its properties, but so also is knowledge about the particular pathways whereby adaptation is partially realized and partially constrained. As an evolutionary biologist who thinks mostly in terms of function, I have profited from Rolls's mechanistic account of the neurobiology of emotion and the central role that emotion plays as a midstation in cognitive processing that begins with perception and ends with action.

I would like to use this essay to comment on a rift that exists within the field of evolutionary psychology, which can be attributed in part to a lack of mechanistic understanding. One side of the rift emphasizes the importance of innate psychological mechanisms that evolved to solve specific problems encountered in the ancestral environment (Barkow et al. 1992; Buss 1999). At the extreme, the mind is envisioned as a jukebox of pre-evolved mechanisms that are played in response to environmental stimuli. The other side of the rift emphasizes learning and cultural transmission (Boyd & Richerson 1985; Campbell 1960). At the extreme, these are envisioned as unconstrained evolutionary processes in their own right that rely on blind variation and selective retention (to use Campbell's 1960 felicitous phrase) to create new solutions to modern problems. The truth presumably lies somewhere in between but lack of mechanistic understanding has prevented productive exploration of the middle ground. The mechanistic details provided by Rolls reveals a psychology that is elaborately innate but which does not exclude, and indeed has evolved to enable, open-ended processes such as learning and cultural transmission.

The neurobiology of hunger and thirst provide model examples of innate special-purpose psychological mechanisms that evolved to solve specific problems in ancestral environments. According to the jukebox view, the mind can be fully understood simply by expanding the number of such systems that have evolved by natural selection; a cheater detection system, a sexual jealousy system, a habitat selection system, and so on. Rolls embraces this possibility with enthusiasm in the second and more speculative half of his book, culminating in a long list of primary reinforcers in Table 10.1. It would be easy to criticize the details of this table but the general point is clear enough; there are many more records in the jukebox than the hunger and thirst records. One point that Rolls does not address is whether there is a theoretical limit to the number of primary reinforcers. Is it possible to imagine the list in Table 10.1 running into the hundreds or thousands, as some evolutionary psychologists have speculated? It seems to me that there must be a limit, beyond which adaptive behaviors must be orchestrated via secondary reinforcers. I hope that Rolls will address this important theoretical question in his response or in future publications.

Although Rolls appreciates the importance of special-purpose mechanisms and is willing to consider a large number of them, he equally appreciates the importance of open-ended learning and cultural processes, on at least two levels. First, learning exists at the level of each special-purpose mechanism. Indeed, the whole purpose of emotion is to provide the rewards and punishments required for a learning process to occur. This is in contrast to the image of a jukebox in which pre-evolved "records" completely specify behavior and the environment is reduced to the role of the button-pusher determining which record is played. If each "record" is a guided system for learning, then the environment becomes much more than a button pusher, even at the level of a single special-purpose mechanism. Second, the symbolic processing system discussed in Chapter 9 provides a new level of relatively domain-general reasoning that may be uniquely human (see also Deacon 1998). Both levels provide a mechanistic basis for learning and culture as evolutionary processes in their own right, capable of providing new behavioral adaptations to modern environmental problems.

To summarize, *The brain and emotion* is a welcome fusion of functional and mechanistic approaches to mind and behavior, which may help to bridge a rift within evolutionary psychology that has developed in the absence of mechanistic understanding.

Author's Response

On *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: There are many advantages to defining emotions as states elicited by reinforcers, with the states having a set of different functions. This approach leads towards an understanding of the nature of emotion, of its evolutionary adaptive value, and of many principles of brain design. It also leads towards a foundation for many of the processes that underlie evolutionary psychology and behavioral ecology. It is shown that recent as well as previous evidence implicates the amygdala and orbitofrontal cortex in positive as well as negative emotions. The issue of why emotional states feel like something is part of the much larger problem of phenomenal consciousness. It is argued that thinking about one's own thoughts would have adaptive value by enabling first order linguistic thoughts to be corrected. It is suggested that reflecting on and correcting one's own thoughts and plans would feel like something, and that phenomenal consciousness may occur when this type of monitoring process is taking place.

Introduction

I thank the commentators for their constructive and interesting points. The following responses are part of the process of moving our understanding forward.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

R1. What are emotions? The nature and functions of emotion

R1.1. The definition of emotions. In *The brain and emotion,* emotions are defined as states elicited by instrumental reinforcers, that is, by rewards and punishers. These states have particular functions.[1] The functions include (pp. 67–70; note all page, section, Figure, and Table numbers refer to the book) eliciting autonomic responses, providing the goals for instrumental actions and thus allowing flexibility in which action is performed, motivating actions to achieve these goals, communication, social bonding, providing a way for genes to specify some of the goals for actions, influencing cognitive evaluations of events and memories, facilitating the storage of memories,[2] producing persistent and consistent behavior by lasting for minutes or hours (cf. **Ben-Ze'ev**), and triggering particular memories. Although some of the functions of emotion are social, as outlined on pp. 67–70 and in Table 10.1, and as emphasized by **Adophs** and **Grafman,** many are not, for example, fear induced by the sight of a painful stimulus and many others noted in Table 10.1.[3]

A number of commentators had the reaction that this definition was too simple (e.g., **Ben-Ze'ev, Adolphs, Buck**). This is a natural reaction: can emotion really be conceptualised this simply? Well, not quite, although using this as the key part of the definition provides the foundation for understanding what events produce emotions, and for understanding its value in the design of a brain designed by natural selection of genes that influence behavior. A very wide range of complex emotions can be accounted for by adding to this fundamental reward/punishment starting concept that the resultant emotion depends on what behavioral (or "coping") strategies are possible (e.g., one difference in the emotions elicited by frustrative non-reward is that anger, frustration, etc., may occur if active behavioral responses can be made, whereas sadness or grief may occur if only passive behavior is appropriate, cf. **Izard** and **Grafman**); that different emotions may arise for every case where the primary and secondary reinforcers are different; that many emotions may involve multiple associations of a stimulus or event with a reinforcer, producing, for example, conflict or guilt (**Fentress**); and that the intensity of the reinforcer can produce differences in the degree of an emotional state.

(A point relevant to **Katz**'s, **Mac Aogáin**'s and **Frijda**'s commentaries is that although reinforcers are usually conceptualised as stimuli or events, remembered stimuli or events have many of the same effects, and in fact the process of recall activates the same higher cortical sensory processing areas and representations in those areas as are activated by the original stimulus or event (p. 65 and Rolls & Treves 1998). **Izard** reminds us of the related point that in modeling, just seeing a fearful facial expression may indicate that the signified object is aversive.

I did not give exhaustive examples of the ways in which many emotions can then be defined with this starting point. However, the partial list of primary reinforcers provided in Table 10.1 should give readers a foundation for starting to understand the rich classification scheme that can arise.[5] To take two examples, jealousy might be an emotional state that arises in the context of needing to protect a sexual investment made with a partner; and guilt may arise when there is a conflict between an available reward and a rule or law of society (**Izard**). Many similar examples can be surmised from the area of evolutionary psychology (see, e.g., Buss 1999; Ridley 1993).

R1.2. Advantages of this approach to emotion. What are the advantages of and justifications for starting with the concept that emotions are states elicited by instrumental reinforcers, even though one proposes that a full definition requires the points just summarized, including a statement of the functions elicited by those states?[6] One advantage is that this definition in terms of rewards and punishers may provide a concise operational definition of the environmental stimuli or events that actually lead to emotions. If we can agree that the environmental conditions that lead to emotions are those that can be described as rewarding or punishing, and that those that are not rewards or punishers do not lead to states that are described as emotional, then we are a long way forward in producing a conceptualisation of what emotions may be.[7] No commentator actually produced clear exceptions to this correspondence. (However I note in both the Précis and the book that one may wish to delimit states produced by reinforcers such as the taste of food that are relevant to internal homeostatic variables as not necessarily producing emotional states, even though they do result in pleasure. We may wish to reserve the concept of emotion for states elicited by reinforcers that are not relevant to internal homeostatic needs, for example, fear produced by the sight of a stimulus associated with a painful stimulus.) If we accept this operational definition, it provides us with a powerful way to start examining emotions (because we accept that they are states elicited by rewards or punishers, and have a useful delimitation of what events produce emotion). This leads directly to an analysis of the brain mechanisms that implement emotions as those that decode environmental stimuli as primary reinforcers, and those that implement stimulus-reinforcer association learning.

A second advantage of this definition is that it enables us to see emotions in the context of what I propose is their most important function, namely, as a mechanism for the genes to influence behavior in a brain that evolves by gene selection. The genes do this by specifying the stimuli or events that the animal is built to find rewarding or punishing, so that the genes specify the goals for action, not the actions themselves. It would be very uneconomical genetically and inflexible behaviorally if genes were to specify large numbers of (typically species-specific) fixed action patterns.

A third advantage is that the definition offers a principled approach to emotion. Different emotions can be classified and understood in terms of different reinforcement contingencies, and hence directly in terms of their functions. This is more advantageous than categorising emotions based on clusters of variables or factors which result from multidimensional analysis of questionnaires or by correlation with autonomic or face expression measures, which do not lead directly to an understanding of the different functions of different emotions (and run the risk of producing seven plus or minus two categories, cf. Miller 1956). Moreover, this principled way of understanding emotions provides a systematic and fundamental way to approach the brain mechanisms involved in emotion, in that brain regions decoding primary reinforcers, and brain regions learning associations of events to primary reinforcers can be seen to have a clear information-processing role in emotion.

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

A fruitful approach to neural computation (Rolls & Treves 1998) is to analyse the information processed at each connected stage in the brain. In the context of emotion, this approach is more principled and systematic than identifying categories of behavior (sometimes described ethologically) as playfulness and aggression, and looking for brain centers specialised for each category (**Panksepp**). The specification of actions such as fixed action patterns (in contrast to goals) by genes is not only genetically expensive, but having brain regions specialised for actions (such as playfulness and rage, cf. Panksepp) would lead to a multitude of specialised brain action/emotion systems – potentially one for every possible type of emotional response. In contrast, specifying emotions as states elicited by reinforcers leaves open and flexible the particular action that may be taken in particular circumstances, and has the great advantage of economy of genetic specification (the genes need only specify what is rewarding and punishing). (Of course, the available type of coping actions may influence the emotional state, as in the case of sadness vs. anger.) Specifying emotions in terms of the types of rewarding and punishing stimuli that elicit the emotion may also lead to spatially separated brain systems especially involved in different types of emotion, for the primary reinforcers (such as taste, touch, pain, the failure to receive an expected reward, or a face expression, and learning about these reinforcers) may be decoded and represented in different brain regions, leading to some specialisation of different brain regions and systems in different types of emotion.

A fourth advantage of conceptualising emotions as states elicited by reinforcers is that this provides an immediate way to understand the relation between emotion and personality (pp. 73–74). In particular, if emotions are states elicited by rewards and punishers or alterations of reward and punishment contingencies, then it would be very likely that sensitivity to changes in different reinforcement contingencies would be different. For example, some individuals might, partly because genes specify primary reinforcers, be differently sensitive to punishment, or non-reward, than other individuals. And it is exactly this type of difference in sensitivity to reinforcers that may account causally (at least in part) for differences in personality, producing in the case of an individual not very sensitive to punishment what is described as extraversion (pp. 73–74; Eysenck & Eysenck 1968; Gray 1970).

R1.3. Cognition and emotion. It may be noted that while the definition of emotions as states elicited by reinforcers (with particular functions) is operational, it should not be criticised as behaviorist (**Katz**). For example, the definition has nothing to do with stimulus-response (habit) associations (cf. **Ursin**), but instead with a two-stage type of learning, in which a first stage is learning which environmental stimuli or events are associated with reinforcers, which is potentially a very rapid and flexible process,[8] and a second stage is in producing appropriate instrumental and arbitrary actions performed in order to achieve the goal (which might be to obtain a reward or avoid a punisher). In the instrumental stage, animals learn about the outcomes of their responses (see Dickinson 1994; Pearce 1997; cf. **Ursin**). To determine what is a goal for an action, every type of cognitive operation may be involved. The proposal is this: Whatever cognitive operations are involved, if the outcome is that a certain event, stimulus, thought (or any one of these remembered) leads to the evaluation that the event is rewarding or punishing, then an emotion will be produced. So cognition is far from excluded. Indeed, cognitive operations may produce emotions when operating at three levels of the architecture.

The first is the implicit level (Fig. 9.4), where a primary reinforcer, or a stimulus or event associated with a primary reinforcer, may lead to emotions. The second level is where a first order syntactic symbol processing system performing "what if" computations to implement planning results in the identification of a rewarding or punishing outcome. The third is the higher order linguistic thought (HOLT) level, where thinking about and evaluating the operations of a first order linguistic processor may result in a reinforcing outcome such as "I should not spend further time thinking about that set of plans, as it would be better now to devote my linguistic resources (which are limited and serial) to this other set of plans." One of the effects of mood on cognitive processing is to promote continuity of behavior (see p. 70 of the book). A mechanism described (in sect. 4.8) utilises backprojections to cortical areas from the amygdala and orbitofrontal cortex, so that reciprocal interactions between cognition and emotion are made possible (**Ramsay & Lewis, Mogi**).

This definition of emotion also leads to an operational and thus clearly specified approach to emotions, whereas approaches such as appraisal theory (referred to by **Frijda** and **Dalgleish**) may suffer from the disadvantage that they quickly become somewhat underspecified and intractable. In appraisal theory (see e.g., Frijda 1986; 1993), primary appraisal of the situation leads to an immediate sense of affect, and secondary appraisal is concerned with coping potential, for example, whether a plan can be constructed, and how successful it is likely to be. However, I do note that appraisal theory is in many ways quite close to the theory that I outline and I do not see them as rivals. Instead, I hope that those who have an appraisal theory of emotion will consider whether much of what is encompassed by primary appraisal is not actually rather close to assessing whether stimuli or events are reinforcers; and whether much of what is encompassed by secondary appraisal is rather close to taking into account the actions that are possible in particular circumstances. An aspect of appraisal theory with which I do not agree is that one of the functions of emotions is to release particular actions, which seems to make a link with species-specific action tendencies or responses. Rarely are responses programmed by genes (see Table 10.1); instead genes optimise their effects on behavior if they specify the goals for (flexible) actions. The difference is quite considerable, in that specifying goals is much more economical in terms of the information that must be encoded in the genome; specifying goals for actions allows much more flexibility to the actual actions produced. Of course I acknowledge that there is some preparedness to learn (**Fentress**), and see this just as an economy of sensory-sensory convergence in the brain, whereby, for example, it does not convey much advantage to be able to learn that flashing lights (as contrasted with the taste of a food just eaten) are followed by sickness.

A related issue concerns where the boundaries for emotional states should be set. Should our definition result in emotions in invertebrates such as Aplysia, as suggested by **Kupfermann**? My own answer to this is to set off from emotions those behaviors that are performed with fixed re-

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

sponses, that is, without the possibility for selecting arbitrary types of behavior as the goals for actions (see Ch. 10). Such behaviors included taxes. (An operant is demonstrated most precisely by the bidirectional criterion that either a response, or its opposite, may be performed as an action to obtain a goal.) One reason why types of behavior with fixed responses are excluded from emotion (though they may be forerunners to it) is that the behavior does not occur by elicitation of a persistent or continuing state to a reinforcing stimulus which provides the motivation for (arbitrary) instrumental responses to obtain the goal. This intervening state elicited by reinforcing stimuli is a mechanism by which stimuli may be interfaced to arbitrary responses; this is one of the prime functions of emotion and is not part of the functional architecture of organisms such as Aplysia as described by Kupfermann.

R1.4. Emotion, motivation, reward, and mood. It is useful to be clear about the difference between motivation, emotion, reward, and mood (cf. **Ben-Ze'ev, Laming, Kralik & Hauser, Buck, Frijda**). Motivation makes one work for a reward. One example of motivation is hunger, another thirst, which in these cases are states set largely by internal homeostatically related variables such as plasma glucose concentration and plasma osmolality.[9] A reward is a stimulus or event that one works to obtain, such as food,[10] and a punisher is what one works to escape from or avoid, such as a painful stimulus or the sight of an object associated with a painful stimulus. Obtaining the reward or avoiding the punisher is the goal for the action. An emotion is a state elicited by an instrumental reinforcer (i.e., a reward or punisher, or omission or termination of a reward or punisher), for example, fear produced by the sight of the object associated with pain. This makes it clear that emotions are not rewards or punishers (cf. **Adolphs**), but states elicited by rewards or punishers that have particular functions. Of course, one of the functions of emotions is that they are motivating, as exemplified by the case of the fear produced by the sight of the object that can produce pain, which motivates one to avoid receiving the painful stimulus, which is the goal for the action. In that emotion-provoking stimuli or events produce motivation, arousal is likely to occur, especially for reinforcers that lead to the active initiation of actions. However, arousal alone is not sufficient to define motivation or emotion (**Ursin, Buck**), in that the motivational state must specify the particular type of goal that is the object of the motivational state, such as water if we are thirsty, food if we are hungry, and avoidance of the painful unconditioned stimulus signalled by a fear-inducing conditioned stimulus (**Frijda**). A mood is a continuing state normally elicited by a reinforcer, and is thus part of what an emotion is. (The other part of an emotion is the decoding of the stimulus in terms of reward and punishment, that is, what causes the emotion, or in philosophical terminology, what the emotion is about, the object of the emotion.) Mood states help to implement some of the persistence-related functions of emotion (**Frijda**), can continue when the originating stimulus may be forgotten (by the explicit system), and may occur spontaneously, not because such spontaneous mood swings may have been selected for, but because of the difficulty of maintaining stability of the neuronal firing which implements mood (or affective) state (pp. 62, 66). Mood states are thus not necessarily about an object.

R1.5. Emotional states and their underlying mechanisms. The question is raised by **Kralik & Hauser** whether it is helpful to maintain the concept of an emotional state when one starts to understand the mechanisms of reward and punishment decoding, the selection of actions, etc. My view is that emotion is a helpful concept, for a number of reasons. First, the state is produced by clearly defined stimuli (see above). Second, the state has many different functions, summarized in the first paragraph above (and pp. 67–72 and in Ch. 10), so that a model in which a stimulus is connected to a single output is inappropriate. In these circumstances, an intervening state which implements many functions is useful. Third, one of the functions of emotion is to support the selection of any appropriate action to a rewarding or punishing stimulus, or its omission or termination, as in two-process learning. In the first stage, an emotional state is produced, and in the second stage, any action is selected that is appropriate given the emotional state. For example, if fear is the emotional state produced by a pain-associated stimulus, an action will be selected to escape from or avoid the emotion-provoking stimulus. In that emotion is a state which facilitates the elicitation of an action to a stimulus, the emotional state is not itself a behavioral response.[11] Fourth, other functions of emotional states include the biassing of cognitive function to influence the interpretation of future events, which is clearly not a response. Fifth, emotional states have the important property of persisting for times in the order of minutes or hours, thus maintaining persistence of behavior and consistency of action even after the emotion-provoking stimulus has disappeared. Sixth, the concept of emotional states just described maps neatly onto folk-psychological concepts of emotions, and provides a convenient conceptual level which bridges to the low-level description of exactly how the stimuli are decoded to elicit the state, how the state is maintained, and how it performs its many functions. The concept of an emotional state is thus clearly defined in terms of how stimuli elicit the state, and of the many functions of the state, including the selection of action. Emotional states are not the stimuli themselves, nor the stimulus decoding, nor the responses finally selected, but consist of ongoing states elicited by stimuli in the way described, and performing the functions described. We are starting to understand exactly how the different types of processing involved are implemented in the brain, and indeed this is one of the types of advance described in *The brain and emotion.* But understanding the implementation of the processes involved in emotion does not mean that emotion itself as a useful concept at its own level will disappear.

R1.6. The functions of emotions. One of the key points about the functions of emotion (made in Ch. 10) is that emotions provide a way for genes to produce appropriate behavior in the animals they specify by identifying goals (such as the taste of food when hungry), and making these goals the targets of arbitrary actions. This produces much more flexibility in the final behavior than would programming-in behavioral responses to particular stimuli (such as fixed action patterns and "innate emotion-specific and species-specific behavioral programs and 'expressive' behaviors" to use the terminology of **Frijda,** and which do occur, but not in great numbers), and requires much less information to be encoded in the genome, and moreover provides a way via a common reward/punishment currency

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

for different goals to compete. Hence this is an argument about the type of brain that natural selection acting on genetic variation would produce through a thoroughly (neo-) Darwinian evolutionary process. Commentators seemed to accept this fundamental role for emotion in brain design, and even encouraged further developments in evolutionary thinking about emotion and brain design (e.g., **Kralik & Hauser, Wilson**). Many of the other points I made about the adaptive value of different types of emotional behavior were rooted in studies described by Ridley (1993), Betzig (1997), and Baker and Bellis (1995), and I certainly agree with Kralik & Hauser that it is very important in that research area to attempt to provide more than conjectures, by for example the use of comparative studies (cf. Buss 1999).

R2. The functions of the amygdala and other brain regions in emotion

R2.1. The amygdala. The evidence supports the hypothesis that at least a part of the amygdala's role in emotion is in learning associations between stimuli and primary reinforcers. Some of the evidence for this is the following: The amygdala is a brain region where potential learned reinforcers (the sight of objects) are brought together anatomically with inputs from primary reinforcers, such as taste, touch, and pain. Neurophysiologically, it can be shown that some neurons respond to primary reinforcers such as taste or touch, and others to visual or auditory stimuli. This convergence can be demonstrated onto single neurons. Some neurons respond, for example, to visual stimuli associated with a primary reinforcer such as a taste (although such neurons are not found in the preceding visual stage of processing, the inferior temporal visual cortex in primates), or to auditory stimuli associated with a primary reinforcer such as shock in rats. The learning of these neuronal responses can be demonstrated by neuronal recordings in rats. This learning can be blocked in rats by intra-amygdaloid administration of blockers of synaptic modification such as NMDA-receptor blockers. And lesions of the amygdala can impair the learning of such associations between previously neutral stimuli and primary reinforcers.

The brain and emotion, in the Précis, and Rolls (2000a) all point out that one potential problem is that after aspiration or coagulative lesions of the amygdala, fibres of passage are also likely to be damaged, and this could account for some of the deficits. For this reason, it is important to assess whether neurotoxic lesions which damage amygdala neurons but leave intact fibres of passage produce emotional changes and affect the learning of associations between stimuli and primary reinforcers.

Parker reviews recent evidence on the effects of neurotoxic lesions of the macaque amygdala. I agree with her that recognition memory effects at one time ascribed to the amygdala are due to damage to rhinal cortical areas, as suggested by the findings of Zola-Morgan et al. (1989; see also Baxter & Murray 2000). Recognition memory is not a function that is related to emotion. However, what about associations between stimuli and primary reinforcers, and emotional behavior? Are these affected by neurotoxic amygdala lesions in primates? Using such lesions (made with ibotenic acid) in monkeys, Malkova et al. (1997) showed that amygdala lesions did not impair visual discrimination learning when the reinforcer was an auditory secondary reinforcer learned as being positively reinforcing preoperatively. This was in contrast to an earlier study by Gaffan and Harrison (1987; see also Gaffan 1992) using aspiration lesions. In the study by Malkova et al. (1997), the animals with amygdala lesions were somewhat slower to learn a visual discrimination task for food reward, and made more errors, but with the small numbers of animals (the numbers in the groups were 3 and 4), the difference did not reach statistical significance.

However, it would be interesting to test non-human primates with neurotoxic amygdala lesions when the association to be learned is directly between a visual stimulus and a primary reinforcer such as taste. It is this type of association learning, between a previously neutral visual or auditory stimulus and a primary (unlearned) reinforcer such as rewarding or punishing taste or touch, that the amygdala and orbitofrontal cortex are hypothesized to implement (see Rolls 1990a; 1992b; 2000a; 2000b; Rolls & Treves 1998). Part of the basis for this hypothesis is that the amygdala and orbitofrontal cortex are brain regions where pathways from high order visual and auditory cortical areas converge anatomically. Neurophysiologically, single neurons in these regions can be activated by primary reinforcers, or by potential secondary reinforcers (visual and auditory stimuli), or show convergence between both; they also alter their responses during visual-to-primary reinforcer association learning (see Rolls 1996; 1999a; 1999b; 2000a; 2000b).

In most non-human primate studies, the reward being given is solid food (typically a pellet of laboratory chow), which is seen before it is tasted, and for which the food delivery mechanism makes a noise. These factors mean that the reward for which the animal is working includes secondary reinforcing components, the sight and sound. When the association to be learned is a purely sensory-sensory (i.e., visual-to-visual or visual-to-auditory) association where neither is a primary reinforcer, cortical areas where these particular sensory signals converge, such as the rhinal cortex, may be able to learn these associations. However, the hypothesis that it would be particularly useful to see tested in non-human primate lesion studies is that associations of sensory stimuli to primary reinforcers depend on the amygdala. (For the orbitofrontal cortex, there is already evidence showing this, see Baylis & Gaffan 1991.) Many of the studies that interfere with amygdala neuronal activity in rats are indeed consistent with this hypothesis, that associations of sensory stimuli to primary reinforcers including both painful stimuli and rewarding stimuli depend on the amygdala (see Everitt & Robbins 1992; Everitt et al. 1999; 2000; LeDoux 1992; 1995).

Consistent with this hypothesis that the amygdala is involved in learning associations to primary reinforcers and therefore in emotion (see book and below), Malkova et al. (1997) showed in macaques that amygdala lesions made with ibotenic acid impair the processing of reward-related stimuli, in that when the reward value of one set of foods was reduced by feeding it to satiety (i.e., sensory-specific satiety, see Rolls 1997), the monkeys still chose the visual stimuli associated with the foods with which they had been satiated. Further evidence that neurotoxic lesions of the amygdala in primates affect behavior to stimuli learned as being reward-related as well as punishment-related is that monkeys with neurotoxic lesions of the amygdala showed abnormal patterns of food choice, picking up and eating

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

foods not normally eaten such as meat, and picking up and placing in their mouths inedible objects (Murray et al. 1996). Further, Meunier et al. (1996; see Baxter & Murray 2000) showed that macaques with neurotoxic amygdala lesions showed altered emotional behavior, including reduced fear and aggressiveness, increased submission, and excessive manual and tactile exploration. These symptoms produced by selective amygdala lesions are classical Kluver-Bucy symptoms. None of these effects is ascribable to rhinal cortex damage (see Baxter & Murray 2000). Thus, in primates, there is evidence that selective amygdala lesions impair some types of behavior to learned reward-related stimuli as well as to learned punishment-related stimuli, thus including stimuli that normally elicit emotional behavior. However, we should not conclude that the amygdala is the only brain structure involved in this type of learning, for especially when rapid stimulus-reinforcement association learning is performed in primates, the orbitofrontal cortex is involved, as discussed below and by Rolls (1999a; 1999b).

Although **Parker** and Easton and Gaffan (2000) revise the previous view of Gaffan (1992) that the amygdala is involved in stimulus-reinforcement association learning, the evidence they discuss (part of it based on the neurotoxic amygdala lesion studies described above) is also concerned with stimulus-stimulus (e.g., visual-to-visual and visual-to-auditory) rather than stimulus-to-primary-reinforcer association learning. Their view that basal forebrain neurons are important in stimulus-reinforcement association learning is consistent with the neurophysiological evidence that basal forebrain neurons can be activated by primary reinforcers such as tastes; by secondary reinforcers such as the sight of food; by both; reverse their responses during stimulus-to-primary reinforcer learning; and reflect reward and even sensory-specific satiety in that their responses are modulated by hunger and sensory-specific satiety (Burton et al. 1976; Mora et al. 1976; Rolls et al. 1976; 1979; 1980; 1986; Wilson & Rolls 1990a; 1990b; 1990c; see Rolls 1997 and Ch. 2). The signals that activate these basal forebrain neurons are actually received from the orbitofrontal cortex and amygdala. However, although these neurons probably project to many cortical areas including the inferior temporal cortex, the suggestion is not that they provide the reward or unconditioned stimulus that enables cortical neurons to respond to visual stimuli associated with rewards (as implied by Easton & Gaffan 2000), because inferior temporal cortex neurons do not respond to visual stimuli based on the association of visual stimuli with reward or punishment (Rolls et al. 1977). Instead, Rolls (1999a; 1992b; see also Rolls & Treves 1998) has suggested that one function of the basal forebrain neurons is, via their cortical terminals, to facilitate learning in the cerebral cortex, enabling whatever type of learning is implemented in a cortical area to take place better at times when basal forebrain neurons are active, consistent with the evidence that acetyl choline is involved in cortical long term potentiation (Bear & Singer 1989).

In summary, in nonhuman primates emotional changes are produced by neurotoxic lesions of the amygdala. There is some evidence for a deficit in one type of stimulus-to-primary reinforcer association learning, when a primary reinforcer is devalued. There is a need for further studies of visual-to-primary reinforcer association learning. Associations between stimuli that do not include primary reinforcers (e.g., visual-to-visual) may indeed not depend on the amygdala, but may be performed in temporal lobe cortical areas. In rats there is evidence that associations to primary reinforcers are impaired by neurotoxic amygdala lesions. This is the case when the primary reinforcer is a punisher (see Davis, 1994; LeDoux 1995). **Killcross** implies that the situation is not so clear in rats for associations to rewards, but it is not clear whether he is referring to associations to *primary* positive reinforcers and lesions of the whole of the amygdala. The evidence for the view is not cited, and he seems to be ignoring a large body of evidence that the rat amygdala is involved in learning associations between stimuli and rewards (Everitt et al. 1999; 2000). Indeed, Everitt et al. (2000) review evidence that neurotoxic lesions of the rat basolateral amygdala impair learning between events and appetitive stimuli when such learning affects instrumental behavior, and evidence that neurotoxic lesions of the central amygdala impair learning between events and appetitive stimuli when such learning affects some responses that are produced by classical conditioning (including autoshaping; see Burns et al. 1994; 1999b; Everitt et al. 1999; 2000). Another type of evidence is that blockade of NMDA receptors in the rat amygdala impairs the learning but not the later performance (as expected, given the role of NMDA receptors in associative learning) of appetitive learning tasks (Burns et al. 1994).

Much additional evidence that the amygdala is involved in stimulus-reinforcement (including stimulus-reward) association learning is described, including evidence that primate amygdala neurons are activated by (1) rewarding primary reinforcers (e.g., the taste of glucose, cf. **Killcross**), (2) aversive primary reinforcers (e.g., the taste of saline), (3) visual stimuli associated by learning with these (see also Rolls 2000a), and, in humans, that the primate amygdala is just as well activated (as shown by fMRI) by (4) the positive primary reinforcer sweet taste as by (5) the primary negative reinforcer salty taste (O'Doherty et al. 2000); it is also activated by (6) the sight of the positive reinforcer food (LaBar et al. 1999). The orbitofrontal cortex is implicated in a similar type of learning, especially in primates (see Ch. 2 and Ch. 4), so that the relative importance of the amygdala in primates may be less than in rodents.

Although for very simple stimuli such as pure tones a subcortical route to the amygdala to produce emotion may be used in performing these functions as investigated by LeDoux (1995), normally the stimuli that elicit emotions are objects and faces. It is for this reason, and to produce the required invariant representations described in Chapter 4, that the normal route to the amygdala and orbitofrontal cortex is via high order cortical processing areas (cf. **Ramsay & Lewis**).

R2.2. The orbitofrontal cortex. The issue of the functions of different parts of the prefrontal cortex in emotion is raised by **Grafman.** The evidence described in Chapters 2 and 4 indicates that the orbitofrontal cortex of primates (and, in the case of humans where the exact limits of brain damage are not usually confined to a circumscribed brain area, what is more generally referred to as the ventral prefrontal cortex), is especially involved in emotion, in that it decodes and represents some primary reinforcers (e.g., taste and touch), and is involved in the rapid learning of associations between visual and olfactory stimuli and primary reinforcers. In addition, it has neuronal representations of

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

faces, and in humans not only is it involved in the effects of rewards and punishers on behavior (Damasio 1994; Rolls et al. 1994), but also damage to it impairs the identification of face expressions (Hornak et al. 1996), which are primary social reinforcers. Disruption of these functions may underlie the deficits that Grafman refers to as impairments in "domain specific social knowledge." In contrast, other parts of the prefrontal cortex are involved in different functions, including spatial and object short term memory, which are foundations for the ability to plan, which requires several steps to be held together in an ordered sequence (see sect. 9.3; Goldman-Rakic 1996; Rolls & Treves 1998; Petrides 1996; Shallice & Burgess 1996). Now the ability to perform such multistep thinking ahead to the consequences of events, planning with "what if" statements, enables animals (instead of having emotions and performing actions to immediately decoded or remembered reinforcing events) to have emotions because they can see the likely consequences of events decoded in this multistep way, or for their emotions to be influenced by what types of action or "coping strategy" become evident as a result of such multistep syntactic planning. Thus the cognitive functions of other parts of the frontal lobe are involved in emotion insofar as they may provide the cognitive apparatus for determining the reinforcing value of possible future events, and for providing coping strategies to deal with reinforcing events (see Ch. 4). This may provide a structural foundation in understanding some of the findings described by Grafman and conceptualising some of the views of appraisal theorists such as **Dalgleish.** As pointed out in Chapter 10, it may be that these types of cognitive process which enable one to think many steps ahead contribute to the fact that emotions can become apparently nonadaptive in humans because they can be so strong.

R3. The initiation of action

It is right for **Ursin** to emphasise that a cost-benefit evaluation should be performed before actions are initiated. It is suggested that evaluation of stimuli and events (including remembered stimuli and events) in terms of their (predicted) reward and punishment outcomes provides a common currency in which the net outcomes of different actions can be compared and conflicts between possible actions can be resolved (see also **Salzen**). It is suggested that the actual evaluation is performed for implicit actions in structures such as the basal ganglia, and for explicit actions involving "what if" multiple step plans in areas of the brain that support linguistic (in particular, syntactic) processing (see below). I also made a link to investigations of optimality in behavioral ecology (see **Houston & McNamara**), and made the point that one way in which the "optimal" behavior heuristic operates may involve the specification by genes of the relative reward and punishment values of different classes of stimuli and events, of how the values are modulated by internal states such as hunger and thirst, and of factors such as incentive motivation, sensory-specific satiety, and novelty.[12] I also pointed out that this may provide a useful route into the neural and physiological mechanisms that control what animals treat as"optimal," which Houston & McNamara agree is an underdeveloped area of optimality theory. A real issue for optimality theory (and behavioral ecology) is what constitutes "optimal" behavior in a given situation, where optimality is justified in the general sense of increasing fitness, but fitness is difficult to define in a particular situation or point in time. The idea of linking the concept of optimal behavior to the behavioral heuristics that result from the operation of the reward and punishment systems built by natural selection to enable genes to influence behavior may be a useful way forward here, and helps to prevent the risk of "optimality" being a post-hoc construct.[13]

Panksepp suggests that the A-10 dopamine system is involved in anticipatory-investigatory phases of ingestive behavior, rather than in reward. I actually went further than this in *The brain and emotion*, and suggested that dopamine neuron firing is related to "Go" or "prepare for action" events, and not to reward. On the other hand, I described evidence that one way in which subcortical events can produce reward is by dopamine influencing transmission in the ventral striatum of reward signals received from the amygdala and orbitofrontal cortex. I also showed that neurons in subcortical structures such as the lateral hypothalamus are implicated in reward in that they are activated by brain-stimulation reward and by natural rewards such as food for a hungry animal. Also, I described evidence that activity in subcortical structures such as the midbrain periacqueductal gray is involved in behavioral responses to painful stimuli. However, I showed that at least in primates (and this may be less true in rodents), cortical processing is involved in decoding very many rewards, including taste, the sight of visual stimuli associated with primary reinforcers, and stimuli important in social behavior such as face expression, and in this sense, cortical systems in primates are very important for emotional behavior.

R4. Emotional feelings and consciousness

In Chapters 4, 6, and 9, evidence that there are two main types of route to action for emotional events is described. One is for instrumental behavior to primary reinforcers or to stimuli associated with them, and involves working for immediate goals (including goals expected as a result of stimulus-reinforcement association learning). There is evidence that we (and split brain subjects and other patients, see Ch. 9) can perform many of these actions automatically, without conscious awareness, and this is therefore sometimes described as an implicit system.[14] The second type of route uses a "what if" type of reasoning involving flexible ("on-line") syntactic operations on symbols and allows long-term planning for strategies to obtain goals, and immediate goals to be deferred. It is argued that there is a credit assignment problem in such a first order syntactic system in that if a plan does not result in the desired outcome, then which of the multiple steps in the plan leads to the error is not clear. To solve this problem, it is suggested that it would be useful to have a higher order thought system to enable each step of the plan to be thought about (evaluating for example the premises for each step of the plan), so that the plan could be corrected. It is then suggested that it is difficult to imagine such a system thinking about its own thoughts grounded in the world without it feeling like something. That is, it is suggested that phenomenal experience ("what it feels like") arises as a property of the operation of such a higher order linguistic thought (HOLT) system. (This type of system is sometimes described as the

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

explicit system.) Having provided a computational advantage for a system to have thoughts about thoughts, and suggested that phenomenal experience arises by virtue of this system, I note that sometimes this system must include sensory processes, emotional states, and so on, in its operations, and suggest that such sensory events and emotional and motivational processes feel like something by virtue of participating in this system.

A number of commentators wished to discuss these issues further. There are of course deep philosophical issues here, and what I and the commentators have to say are points for discussion, rather than any conclusion that should be taken to have practical implications. **DeLancey, Aydede,** and **Korb & Nicholson** ask about the symbols in the HOLT system. The symbols (or symbolic representations) are symbols in the sense that they can take part in syntactic processing. The symbolic representations are grounded in the world in that they refer to events in the world. The symbolic representations must have a great deal of information about what is referred to in the world, including the quality and intensity of sensory events, emotional states, and so on. The need for this is that the reasoning in the symbolic system must be about stimuli, events, and states, and remembered stimuli, events, and states and for the reasoning to be correct, all the information that can affect the reasoning must be represented in the symbolic system, including, for example, just how light or strong the touch was.

Indeed, it is pointed out (pp. 252–53) that it is no accident that the shape of the multidimensional phenomenal (sensory, etc.) space does map so clearly onto the space defined by neuronal activity in sensory systems, for if this were not the case, reasoning about the state of affairs in the world would not map onto the world, and would not be useful. Good examples of this close correspondence are found in the taste system, in which subjective space maps simply onto the multidimensional space represented by neuronal firing in primate cortical taste areas. In particular, if a three-dimensional space reflecting the distances between the representations of different tastes provided by macaque neurons in the cortical taste areas is constructed, then the distances between the subjective ratings by humans of different tastes is very similar (Plata-Salaman et al. 1996; Smith-Swintowsky et al. 1991; Yaxley et al. 1990). [See also Palmer: "Color, Consciousness, and the Isomorphism Constraint" *BBS* 22(6) 1990.] Similarly, the changes in human subjective ratings of the pleasantness of the taste, smell, and sight of food parallel very closely the responses of neurons in the macaque orbitofrontal cortex (Ch. 2). The representations in the first order linguistic processor that the HOLTs process include beliefs (for example "food is available," or at least representations of this), and the HOLT system (as pointed out by **Aydede**) would then have available to it the concept of a thought (so that it could represent "I believe [or there is a belief] that food is available").

As summarised in the first paragraph of this section (R4), however, representations of sensory processes and emotional states must be processed by the first order linguistic system, and HOLTs may be about these representations of sensory processes and emotional states capable of taking part in the syntactic operations of the first order linguistic processor. Such sensory and emotional information may reach the first order linguistic system from many parts of the brain, including those such as the orbitofrontal cortex and amygdala implicated in emotional states (Fig. 9.3 and p. 253). When the sensory information is about the identity of the taste, the inputs to the first order linguistic system must come from the primary taste cortex, in that the identity of taste, independent of its pleasantness (in that the representation is independent of hunger) must come from the primary taste cortex. In contrast, when the information that reaches the first order linguistic system is about the pleasantness of taste, it must come from the secondary taste cortex, in that there the representation of taste depends on hunger (see Fig. 9.3). Further, to address an issue raised by **Van Gulick,** the higher order linguistic thought system, with its role in correcting planning, must be able to influence behavior, as indicated in Figure 9.3.

Another issue is that of the type of syntax that is required (see **Korb & Nicholson,** and **Katz**). What is required for the first order linguistic symbol processing system is the ability to link together representations in multiple "if . . . then" steps, to form a flexible plan. The plan involves flexible linking in that one plan might be formulated now, and another one, using many of the same symbols or representations, might be formed two minutes later. (Such a system formally *requires* syntax to bind the symbols; cf. Korb & Nicholson). Thus no claim is made about human verbal language being required, and a number of nonhuman animals may be able to form this type of plan. The higher order thought system needs to be able to understand and correct the plans of the first order syntactic system, and for this reason itself needs to be able to process syntax, and in this sense is termed a higher order linguistic thought (HOLT) system. Given that the suggestion is that phenomenal experience is associated with the higher order thought system, these definitions would thus exclude thermostats from having phenomenal experience.[15]

On the issue of the architectural features and functionality associated with phenomenology, mine is a correlative theory, which suggests that there is a correlation between processing in the higher order linguistic thought processing system and phenomenology. The relevant features are first order linguistic processing as described above, and the ability to think about those thoughts to correct them. According to this view, the information processing that is relevant to phenomenal consciousness is the type of processing implemented by HOLTs, which, we should note, have a defined and important function in correcting plans (cf. **Aydede**).[16,17] The phenomenal aspects arise as a property of performing this type of information processing on representations of evidence grounded in the world of the animal, but cannot be said to have a separate, additional function. However, in the context of emotion, we should note that when such a system reasons about its plans and goals, and the goals feel like something in this system, it is consistent that they have an affective component (emotional feelings), as these are states that the animal is planning to obtain or avoid, and it is appropriate that the feelings should be consistent with them being the goals that the system wants to perform actions to obtain (cf. **Ramsay & Lewis,** who raise the issue of the functional role of emotion in consciousness). Systems that have less than this extent of architecture *and grounding in the world* (see pp. 251, 276–78; cf. **DeLancey; Van Gulick; Korb & Nicholson**) may be expected to have less in the way of the properties (including the phenomenal properties) of the system, but where important boundaries are remain to be determined.

The theory of consciousness outlined in *The brain and*

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

emotion is thus a correlative theory of consciousness, which may be contrasted with an explanatory theory, which was what was wished for by **Van Gulick** in his interesting commentary. I doubt that we will ever produce a fully explanatory theory of why a certain type of information processing *has* to feel like something, and believe that the best at present is to try to define what type of information processing is *correlated* with phenomenal consciousness. **Izard** really begs the question when he writes "obvious adaptive advantages accrue from the generation of emotion experience as a direct function of sense perception and lower order cognition." The problem is that it is not at all obvious what classes of machine would have the property of phenomenal perception – could we not just build a machine with the same functionality but without it feeling like anything to be operating as that machine? And what are the functions that would be implemented by its feeling like something to be that machine? These are difficult questions, to which the answers are not very obvious.

Bermúdez raises a related issue when he writes that "primary reinforcers have qualitative aspects. It is impossible to divorce pain's being a negative reinforcer from its feeling the way it does." One has to be very careful with one's terminology here. Primitive animals with no cortex, and pain pathways implemented by C fibres, may perform operant learning to obtain that reinforcer, but is the feel, the conscious awareness or phenomenology, the quale, the *same* as it is in humans? We should not let consciousness slip in through the back door by using loaded terms without very clear definitions. To answer Bermúdez's interesting question about whether higher order thoughts (HOTs) are involved when we are conscious about secondary reinforcers, my hypothesis is that the HOLT system for explicit linguistic planning should usually be monitoring our behavior when it is being performed implicitly and automatically, in case the different types of computation made possible by the syntactic planning system would result in a better outcome by deferring an immediate goal and acting for a goal achievable only by multistep planning (see Ch. 9 and 10); and that it is by virtue of the operation of the HOLT system that we are conscious of the secondary reinforcer. That is, some behavioral responses to the secondary reinforcer may be learned about in an implicit system (one to which there is no conscious access), but there may nevertheless be explicit access to the stimuli involved because they reach a HOLT linguistic system that is continually monitoring. This may lead to the explicit system confabulating sometimes about causes or reasons for actions, as described in Chapter 9.

For those who wondered whether it was proposed that human verbal language was necessary for qualia and feelings (**Aydede, Ben-Ze'ev, Izard**) or even implicit emotional behavior (**Izard**), it should be clear that this was not implied by the proposal. In any case, theories of consciousness are not sufficiently developed that they should be taken to have practical implications. However, the complex issue of consciousness must be addressed if we are to address the issue of emotional *feelings,* and as findings in neuroscience are relevant, these issues are the subject of Chapter 9.

R5. Conclusion

In the course of discussing the many interesting and valuable points raised by commentators, it is shown that there are many advantages to defining emotions as states elicited by reinforcers, with the states having a set of different functions. This approach leads towards an understanding of the nature of emotion, of its evolutionary adaptive value, and of many principles of brain design. It also leads towards a foundation for many of the processes that underlie evolutionary psychology and behavioral ecology.

It shows that recent evidence on the functions of the amygdala and orbitofrontal cortex in emotion supports the hypothesis that they are important in emotion in primates because not only are they involved in stimulus-reinforcement association learning, but also they play an important role in decoding and representing a number of primary reinforcers.

The issue of why emotional states feel like something is part of the much larger problem of phenomenal consciousness. It is argued that thinking about one's own thoughts would have adaptive value by enabling first order linguistic thoughts to be corrected. It is suggested that reflecting on and correcting one's own thoughts and plans would feel like something, and that phenomenal consciousness may occur when this type of monitoring process is taking place.

ACKNOWLEDGMENTS

I wish to thank Martin Davies, Joe Lau (Hong Kong University), John O'Doherty, and Morten Kringelbach for extremely helpful and interesting discussions.

NOTES

1. Commentator **Frijda** asked about the consequences of emotional states. They are its functions, described on pp. 67–70, in Table 10.1, etc.

2. Experiments are described by **Moore & Oaksford** in which learning and memory can be modulated by emotional state. It is generally adaptive when one has received a primary or secondary reinforcer to store information, as any information stored at that time may be useful when later being faced with similar situations (p. 70). Moreover, some of the possible brain mechanisms for this are described on pp. 140–43.

3. It is suggested by **Grafman** that "stored social knowledge dominates our behavior and controls emotional states, thereby reducing emotions to a subservient role in behavior." This statement cannot be correct because many powerful emotions (such as the sight of the pain-inducing stimulus) may have nothing to do with social behavior. Many social factors which are indeed primary reinforcers include facial expression, parental bonding, and mother-infant attachment, and in that they are reinforcing, these social factors lead to emotional states. The emotions that occur can also be influenced by the actions or "coping strategies" that are available, as in the example of frustrative non-reward and the difference between sadness and anger. In just the same way, stored social knowledge can lead to a cognitive evaluation of what strategies may be possible in the emotion-provoking situation, and the outcome of this cognitive evaluation can influence the emotional state in just the same way, by specifying what coping action may or may not be possible.

4. A point relevant to **Ben-Ze'ev**'s commentary is that, given that each emotion may be categorised using a *combination* of the factors just described, the intensity of that particular emotion can be influenced by the magnitude of the reinforcers.

5. If **Panksepp** accepts the conceptual point that different genes code for all the different types of putative primary reward and punisher made in Table 10.1, and that they build reward and punishment systems that are each sensitive to the appropriate reward and punisher, he probably sees that here is a way to understand how the brain produces many types of "natural behavioral and affective inclinations." Where I differ from Panksepp (1998a) is that I take an information-processing view of brain organisation,

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

and analyse where in information processing in the brain different rewards and punishers are decoded, where learning of stimulus-reinforcer associations occurs, and how the representations of reinforcers are altered by ongoing internal and external stimuli (cf. **Fentress**), whereas Panksepp takes a more ethological approach and attempts to map whole behaviors onto separate brain systems.

6. Katz notes that not all effects elicited by reinforcers are emotional states. I agree, making it clear that the states elicited by reinforcers that have certain functions are emotional states. For example, a primary reinforcer such as the taste of water for a thirsty animal leads to the ingestion of water (the goal object), which in turn has physiological consequences, such as fluid replenishment. The relation between the goal which produces the emotional state and the replenishment in this example is that natural selection has resulted in depletion's setting the taste of water to be rewarding, i.e., making it the goal object. But the rehydration of the body tissues (resulting from the consummatory behavior of drinking the water) is not itself an emotional state, and consistent with this, plasma dilution is a very poor reinforcer, as described in Chapter 7.

7. Commentators **Kralik & Hauser** raise the issue of the best way to define reinforcers. I use the terminology of Gray (1975), and refer to Mackintosh (1983), Dickinson (1980), and Pearce (1997) for further discussion. The definitions used are based on the properties of the stimulus, which is what is decoded by the brain. Thus, a punisher, aversive stimulus, or negative reinforcer is said to decrease the probability of a response on which it is made contingent. Then one specifies separately what happens if that stimulus is omitted or terminated. In particular, omitting or terminating the punisher increases the probability of the responses. This approach becomes useful when considering rewards, appetitive stimuli or positive reinforcers, which increase the probability of responses on which they are made contingent. Omission or termination of a reward can decrease (or increase in the short term as in frustrative non-reward, cf. **Houston & McNamara**) the probability of actions. Thus it is useful to keep separate in the definitions the valence of the stimulus as defined by whether the animal will work to obtain or avoid it, and what happens instrumentally when the stimulus is omitted or terminated. In an alternative definition (see Mazur 1998), the "positive" qualifier refers to presentation of the reinforcer, and negative to termination or omission of the reinforcer, rather than to the valence. It is helpful in the context of emotion to place the emphasis on the valence of the stimulus, that is, on whether it is rewarding or punishing, and on whether a rewarding or punishing stimulus has been omitted or terminated, because this is the type of stimulus and event decoding that must first be performed and is relevant to emotional states, before instrumental actions are selected depending on what possibilities are available in this two-stage process (cf. Gray 1975).

8. The process of stimulus-reinforcer association learning uses, to answer **Katz,** the heuristic of associative Hebbian learning (see Rolls & Treves 1998) and related processes to reflect contingency (see Pearce 1997) to identify which environmental stimuli are (causally) associated with primary reinforcers in the implicit system. To answer **Ursin,** this is how expectancies are learned.

9. There are representations of external stimuli, notes **Mac Aogáin,** that can be used for a number of functions, including, in some parts of the brain, evaluation of the reward or punishment value of the stimuli or events. However, I also believe that motivational states such as hunger are based on representations, in this case of sensed physiological variables such as gastric distension and plasma glucose concentration.

10. In relation to the ethological approach of **Salzen,** I note that primary reinforcers need not be proximal stimuli acting on contact receptors. There are a number of examples in Table 10.1, including warning calls and facial expression. I also note that when Salzen refers to consummatory stimuli there are actually two types of stimuli involved, those that produce the reward and are the goals for action, and others that produce satiety and may be very poor reinforcers. In the case of feeding, the orosensory stimuli such as taste provide the reward and little satiety, while gastric distension and the absorption of food produce satiety but are poor reinforcers (see Ch. 2). This distinction is crucial for understanding the brain mechanisms involved.

11. Kralik & Hauser are correct to point out the misprint on p. 125, where emotional response was printed instead of emotional state, with the latter intended and used elsewhere.

12. The issue of novelty and emotion is raised by **Izard.** I actually argued that novel stimuli produce reward by activating the same neurons in the amygdala that are activated by reward; and that the function of this was to encourage animals to explore the high-dimensional space within which their genes operate.

13. Another issue for optimality theory is the need to separate reward from satiety processes. **Houston & McNamara** wish to question the evidence on p. 66 that behavioral contrast and the regression of reward sensitivity to a long-term resettable average may not be adaptive, by discussing a particular case of foraging. They say that when reward availability becomes greater in foraging, then the reward value may not increase. But here there is an influence of satiety. When animals are satiated (or perhaps see that they can easily be satiated), the reward value of the relevant sensory input decreases. The modulation of reward value by satiety signals is an important theme of *The brain and emotion.* Similar behavior can also be seen in drug self-administration behavior, where an increase in the amount of reward obtained by each self-administration decreases the rate of working. The reason in this case is that the larger reward takes longer to metabolize, and so the need for another self-administration to maintain a given drug level is delayed. In relation to another point Houston & McNamara make, of course animals may communicate their emotional states for many functions, and indeed it may be adaptive in some cases to use the emotional communication channel "dishonestly" to miscommunicate the emotional state.

14. Note that, contrary to a thought of **Izard,** we cannot recall the details of implicit skilled actions performed automatically, such as the balancing required on a bicycle.

15. Contrary to **DeLancey**'s remark, Chalmers (1996) does seem to think that thermostats have at least some phenomenal experience: he writes (p. 295) of "a thermostat having phenomenology" and "experience." Indeed, he writes (p. 297) that "there is 'experience' whenever there is causal interaction," and he does define phenomenology and experience (conventionally) as being concerned with what it "feels" like (p. 11).

16. Higher order linguistic thoughts need not themselves be conscious (following Rosenthal [1993] and to avoid an infinite regress) (cf. **Izard**).

17. Commentator **Peterson** outlines a cybernetic approach to information processing rather than a theory of phenomenal consciousness. I agree that cognitive processing of the "what if" type enables goals to be identified, which may be evaluated as being more desirable than the goals identified by immediate decoding of the valence of sensory stimuli, which may then be deferred.

References

Letters "a" and "r" before authors' initials refer to target article and response, respectively.

Abercrombie, E. D., Keefe, K. A., DiFrischia, D. S. & Zigmond, M. J. (1989) Differential effect of stress on *in vivo* dopamine release in striatum, nucleus accumbens, and medial frontal cortex. *Journal of Neurochemistry* 52:1655–58. [aETR]

Abrams, P. A. (1991) Life history and the relationship between food availability and foraging effort. *Ecology* 72:1242–52. [AIH]

Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372:669–72. [aETR]

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

(1995) Fear and the human amygdala. *The Journal of Neuroscience* 15:5879–92. [RA]
Allman, J. M. (1999) *Evolving brains.* Scientific American Library. [JDK]
Amaral, D. G., Price, J. L., Pitkanen, A. & Carmichael, S. T. (1992) Anatomical organization of the primate amygdaloid complex. In: *The amygdala,* ed. J. P. Aggleton. Wiley-Liss. [aETR]
Aristotle. (1984) *De anima.* In: *The complete works of Aristotle: The revised Oxford translation,* ed. J. Barnes. Princeton. [LDK]
(1984) *Eudemian ethics. Op. cit.* [LDK]
(1984) *Rhetoric. Op. cit.* [LDK]
Aydede, M. (1998) The language of thought: State of the art. http://humanities.uchicago.edu/faculty/aydede/LOTH.SEP.html. Shorter version in the electronic *Stanford Encyclopedia of Philosophy,* ed. E. Zalta. http://plato.stanford.edu/entries/thought-language/ [MA]
(in preparation) Naturalism, qualia and pain. The University of Chicago. http://humanities.uchicago.edu/faculty/aydede/pain.pdf [MA]
Baker, R. & Bellis, M. (1995) *Human sperm competition: Copulation, competition and infidelity.* Chapman and Hall. [rETR]
Balleine, B. W. & Dickinson, A. (1998) Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37(4–5):407–19. [SK]
Barbas, H. (1995) Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. *Neuroscience Biobehavioral Review* 19(3):499–510. [JG]
Barkow, J. H., Cosmides, L. & Tooby, J., eds. (1992) *The adapted mind: Evolutionary psychology and the generation of culture.* Oxford University Press. [DSW]
Baron-Cohen, S., Tooby, J. & Cosmides, L. (1997) *Mindblindness: An essay on autism and theory of mind.* Bradford. [EM]
Bates, J. (1994) The role of emotion in believable agents. *Communications of the ACM, Special Issue on Agents* 37(7):122–25. [KBK]
Batson, C. D. (1990) How social is an animal? The human capacity for caring. *American Psychologist* 45(3):336–46. [CI]
Baxter, M. G. & Murray, E. A. (2000) Reinterpreting the behavioural effects of amygdala lesions in non-human primates. In: *The amygdala, a functional analysis,* ed. J. P. Aggleton. Oxford University Press. [AP, rETR]
Baylis, L. L. & Gaffan, D. (1991) Amygdalectomy and ventromedial prefrontal ablation produce similar deficits in food choice and in simple object discrimination learning for an unseen reward. *Experimental Brain Research* 86:617–22. [rETR]
Bear, M. F. & Singer, W. (1986) Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature* 320:172–76. [rETR]
Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. W. (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50:7–15. [aETR]
Bechara, A., Damasio, H., Tranel, D. & Anderson, S. W. (1998) Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience* 18:428–37. [aETR]
Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–95. [aETR]
Bechara, A., Tranel, D., Damasio, H. & Damasio, A. R. (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex* 6:215–25. [aETR]
Beck, C. H. M. & Fibiger, H. (1995) Conditioned fear-induced changes in behavior and the expression of the immediate early gene *c-fos:* With and without diazepam treatment. *Journal of Neuroscience* 15:709–20. [JP]
Ben-Ze'ev, A. (2000) *The subtlety of emotions.* MIT Press. [AB-Z]
Bermúdez, J. L. (1998a) *The paradox of self-consciousness.* MIT Press. [JLB]
(1998b) Philosophical psychopathology. *Mind and Language* 13:287–307. [JLB]
Berridge, K. C. (1999) Pleasure, pain, desire, and dread: Hidden core processes of emotion. In: *Foundations of hedonic psychology: Scientific perspectives on enjoyment and suffering,* ed. D. Kahneman, E. Diener & N. Schwarz. Sage. [NF]
Betzig, L., ed. (1997) *Human nature: A critical reader.* Oxford University Press. [rETR]
Block, N. (1980) What is functionalism? In: *Readings in philosophy of psychology, vol. 1,* ed. N. Block. Harvard University Press. [MA]
Bolles, R. C. (1972) Reinforcement, expectancy, and learning. *Psychological Review* 79:391–409. [HU]
Boyd, R. & Richerson, P. J. (1985) *Culture and the evolutionary process.* University of Chicago Press. [DSW]
Breland, K. & Breland, M. (1961) The misbehaviour of organisms. *American Psychologist* 16:681–84. [DRJL]
Brentano, F. C. (1874/1973/1995) *Psychologie vom empirischen Standpunkt.* Duncker & Humbolt. English translation, *Psychology from an empirical standpoint,* trans. O. Kraus, 1973 edition, Routledge & Kegan Paul. [EM]
2nd edition translation of 3rd German edition (1995) ed. L. L. McAlister. Routledge. [LDK]
Bretherton, I., McNew, S. & Beeghly-Smith, M. (1981) Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind"? In: *Infant social cognition,* ed. M. Lamb & L. R. Sherrod. Erlbaum. [CI]
Bromhall, C. (1994) *The sexual imperative: 1. The importance of sex.* Channel Four, 14th May. [DRJL]
Brothers, L. (1997) *Friday's footprint.* Oxford University Press. [RA]
Brothers, L. & Ring, B. (1993) Mesial temporal neurons in the macaque monkey with responses selective for aspects of social stimuli. *Behavioural Brain Research* 57:53–61. [aETR]
Buck, R. (1985) Prime theory: An integrated view of motivation and emotion. *Psychological Review* 92:389–413. [RB]
(1999) The biological affects: A typology. *Psychological Review* 106:301–36. [RB, JP]
Buck, R. & Ginsburg, B. (1991) Emotional communication and altruism: The communicative gene hypothesis. In: *Altruism. Review of personality and social psychology, vol. 12,* ed. M. Clark. Sage. [RB]
(1997) Communicative genes and the evolution of empathy. In: *Empathic accuracy,* ed. W. Ickes. Guilford. [RB]
Buckley, M. J., Easton, A., Parker, K., Wilding, E. L. & Parker, A. (1999) Novelty, memory and the perirhinal cortex in monkeys: Associative object and scene learning. *Society for Neuroscience Abstracts* 25:93. [AP]
Burns, L. H., Everitt, B. J. & Robbins, T. W. (1994) Intra-amygdala infusion of the N-methyl-D-aspartate receptor antagonist AP5 impairs acquisition but not performance of discriminated approach to an appetitive CS. *Behavioral and Neural Biology* 61:242–50. [rETR]
Burns, L. H., Everitt, B. J. & Robbins, T. W. (1999a) Differential effects of excitotoxic lesions of the basolateral amygdala, ventral subiculum and medial prefrontal cortex on responding with conditioned reinforcement and locomotor activity potentiated by intra-accumbens infusions of D-amphetamine. *Behavioural Brain Research* 55:167–83. [rETR]
(1999b) Effects of excitotoxic lesions of the basolateral amygdala on conditioned discrimination learning with primary and conditioned reinforcement. *Behavioural Brain Research* 100:123–33. [rETR]
Burton, M. J., Rolls, E. T. & Mora, F. (1976) Effects of hunger on the responses of neurones in the lateral hypothalamus to the sight and taste of food. *Experimental Neurology* 51:668–77. [rETR]
Buss, D. M. (1999) *Evolutionary psychology.* Allyn and Bacon. [rETR, DSW]
Campbell, T. D. (1960) Blind variation and selective retention in creative thought and other knowledge processes. *Psychological Review* 67:380–400. [DSW]
Campeau, S., Falls, W. A., Cullinan, W. E., Helmreich, D. L., Davis, M. & Watson, S. J. (1997) Elicitation and reduction of fear: Behavioural and neuroendocrine indices and brain induction of the immediate-early gene *c-fos. Neuroscience* 78:1087–104. [JP]
Carver, C. S. & Scheier, M. F. (1998) *On the self-regulation of behavior.* Cambridge University Press. [JBP]
Caryl, P. G. (1979) Communication by agonistic displays: What can game theory contribute to ethology? *Behaviour* 68:136–69. [AIH]
Chalmers, D. J. (1996) *The conscious mind: In search of a fundamental theory.* Oxford University Press. [CI, EM, rETR]
Clutton-Brock, T. H. & Harvey, P. H. (1980) Primates, brains and ecology. *Journal of Zoology* 207:151–69. [JDK]
Cooper, J. R., Bloom, F. E. & Roth, R. H. (1996) *The biochemical basis of neuropharmacology, 7th edition.* Oxford University Press. [aETR]
Coover, G. D., Murison, R., Sundberg, H., Jellestad, F. & Ursin, H. (1984) Plasma corticosterone and meal expectancy in rats: Effects of low probability cues. *Physiology and Behavior* 33:179–84. [HU]
Crick, F. (1994) *The astonishing hypothesis.* Macmillan. [JCF]
Damasio, A. R. (1994) *Descartes' error: Emotion, reason, and the human brain.* Grosset/Putnam. [RA, CI, KBK, arETR]
Darwin, C. (1872/1998) *The expression of the emotions in man and animals, 3rd edition.* University of Chicago Press. [aETR] John Murray edition. [JCF] Oxford University Press, 1998 edition. [JP]
Davenport, P. and staff reporters (1985) Passengers panic as fire races through fuselage. *The Times,* 23rd August, pp.1–2. [DRJL]
Davidson, R. J. & Irwin, W. (1999) The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Science* 3:11–21. [aETR]
Davis, M. (1992) The role of the amygdala in conditioned fear. In: *The amygdala,* ed. J. P. Aggleton. Wiley-Liss. [aETR]
(1994) The role of the amygdala in emotional learning. *International Review of Neurobiology* 36:225–66. [rETR]
Dawkins, M. S. (1998) *Through our eyes only.* Oxford University Press. [JLB]
Dawkins, R. (1986) *The blind watchmaker.* Longman. [aETR]
Deacon, T. W. (1997) *The symbolic species: The co-evolution of language and the brain.* W. W. Norton. [JDK, DSW]
De Sousa, R. D. (1987) *The rationality of emotion.* MIT Press. [KBK]

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Dickinson, A. (1980) *Contemporary animal learning theory.* Cambridge University Press. [JLB, rETR]
(1994) Instrumental conditioning. In: *Animal learning and cognition,* ed. N. J. Mackintosh. Academic Press. [JDK, rETR]
Dimitrov, M., Grafman, J. & Hollnagel, C. (1996) The effects of frontal lobe damage on everyday problem solving. *Cortex* 32(2):357–66. [JG]
Dretske, F. (1995) *Naturalizing the mind.* MIT Press. [MA, JLB]
Easton, A. & Gaffan, D. (1997) Hypothalamic-cortical disconnection impairs visual reward-association learning in monkeys. *Society for Neuroscience Abstracts* 23:11. [AP]
(2000) Amygdala and the memory of reward: The importance of fibres of passage from the basal forebrain. In: *The amygdala, a functional analysis,* ed. J. P. Aggleton. Oxford University Press. [AP, rETR]
(2000) Hypothalamic-cortical disconnection impairs object-reward association learning in monkeys. *Neuropsychologia.* [AP]
Easton, A., Parker, A. & Gaffan, D. (submitted) Hypothalamic-cortical disconnection impairs object recognition memory in monkeys. (submitted to *Journal of Neuroscience*). [AP]
Ekman, P. (1982) *Emotion in the human face, 2nd edition.* Cambridge University Press. [aETR]
(1992) An argument for basic emotions. *Cognition and Emotion* 6:169–200. [RA]
(1993) Facial expression and emotion. *American Psychologist* 48:384–92. [RA, aETR]
Ekman, P., Friesen, W. V. & Ellsworth, P. C. (1972) *Emotion in the human face: Guidelines for research and integration of findings.* Pergamon Press. [CI]
Enquist, M. (1985) Communication during aggressive interactions with particular reference to variation in choice of behaviour. *Animal Behaviour* 33:1152–61. [AIH]
Everitt, B. J., Cardinal, R. N., Hall, J., Parkinson, J. A. & Robbins, T. W. (2000) Differential involvement of amygdala subsystems in appetitive conditioning and drug addiction. In: *The amygdala, a functional analysis,* ed. J. P. Aggleton. Oxford University Press. [rETR]
Everitt, B. J., Parkinson, J. A., Olmstead, M. C., Arroyo, M., Robledo, P. & Robbins, T. W. (1999) Associative processes in addiction and reward: The role of amygdala- ventral striatal subsystems. *Annals of the New York Academy of Sciences* 877:412–38. [rETR]
Everitt, B. J. & Robbins, T. W. (1992) Amygdala-ventral striatal interactions and reward- related processes. In: *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction,* ed. J. P. Aggleton. Wiley-Liss. [arETR]
Eysenck, H. J. & Eysenck, S. B. G. (1968) *Personality structure and measurement.* R. R. Knapp. [rETR]
Fentress, J. C. (1990) Organizational patterns in action: Local and global issues in action pattern formation. In: *Signal and senses: Local and global order in perceptual maps,* ed. G. M. Edelman, W. E. Gall & W. M. Cowan. Wiley. [JCF]
(1991) Analytical ethology and synthetic neuroscience. In: *The development and integration of behaviour,* ed. P. Bateson. Cambridge University Press. [JCF]
(1999) The organization of behaviour revisited. *Canadian Journal of Experimental Psychology* 53:8–19. [JCF]
Festinger, L. (1957) *A theory of cognitive dissonance.* Harper Row. [HU]
Francis, S., Rolls, E. T., Bowtell, R., McGlone, F., O'Doherty, J., Browning, A., Clare, S. & Smith, E. (1999) The representation of the pleasantness of touch in the human brain, and its relation to taste and olfactory areas. *NeuroReport* 10:453–60. [aETR]
Frank, R. H. (1988) *Passions within reason: The strategic role of the emotions.* W. W. Norton. [CD, JG]
(1989) Honesty as an evolutionarily stable strategy. *Behavioral and Brain Sciences* 12:705–706. [JG]
Freeman, W. J. (1995) *Societies of brains.* Erlbaum. [JTR]
(in press) Emotion is essential to all intentional behaviors. In: *Emotion, development, and self-organization,* ed. M. D. Lewis & I. Granic. Cambridge University Press. [JTR]
Freud, S. (1923/1981) *The ego and the id.* In: *The standard edition of the complete psychological works of Sigmund Freud,* ed. J. Strachey. The Hogarth Press. [JP]
Fridlund, A. J. (1994) *Human facial expression: An evolutionary view.* Academic Press. [aETR]
Frijda, N. H. (1986) *The emotions.* Cambridge University Press. [TD, IK, arETR]
(1993) The place of appraisal in emotion. *Cognition and Emotion* 7:357–87. [JTR, rETR]
Gaffan, D. (1992) Amygdala and the memory of reward. In: *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction,* ed. J. P. Aggleton. Wiley-Liss. [rETR]
Gaffan, D. & Harrison, S. (1987) Amygdalectomy and disconnection in visual learning for auditory secondary reinforcement by monkeys. *Journal of Neuroscience* 7:2285–92. [rETR]
Gaffan, D., Parker, A. & Easton, A. (1998) Dense amnesia in macaques after section of anterior temporal stem, amygdala and fornix. *Society for Neuroscience Abstracts* 24:18. [AP]
(submitted) Dense amnesia in macaques after transaction of anterior temporal stem, amygdala, and fornix. (submitted to *Neuropsychologia*). [AP]
Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2:493–500. [EM]
Gilbert, C. D. (1995) Dynamic properties of adult visual cortex. In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. MIT Press. [JCF]
Gluck, M. A. & Bower, G. H. (1988) From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General* 117:227–47. [SCM]
Goel, V., Grafman, J., Tajik, J., Gana, S. & Danto, D. (1997) A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain* 120(Pt 10):1805–22. [JG]
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* 351:1445–53. [arETR]
Goleman, D. (1996) *Emotional intelligence.* Bloomsbury. [EM]
(1998) What makes a leader? *Harvard Business Review* 76(6):93–102. [EM]
Grafen, A. (1990) Biological signals as handicaps. *Journal of Theoretical Biology* 144:517–46. [JG]
Grafman, J. (1995) Similarities and distinctions among current models of prefrontal cortical functions. *Annals of the New York Academy of Sciences* 769:337–68. [JG]
Grafman, J., Holyoak, K. J. & Boller, F., eds. (1995) Structure and functions of the human prefrontal cortex. *Annals of the New York Academy of Sciences* 769:337–68. [JG]
Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R. & Salazar, A. M. (1996) Frontal lobe injuries, violence, and aggression: A report of the Vietnam Head Injury Study. *Neurology* 46(5):1231–38. [JG]
Gray, J. A. (1970) The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy* 8:249–66. [rETR]
(1975) *Elements of a two-process theory of learning.* Academic Press. [arETR, HU]
(1982) *The neuropsychology of anxiety: An enquiry into the functions of the septal-hippocampal system.* Oxford University Press. [JBP, EAS]
(1987) *The psychology of fear and stress, 2nd edition.* Cambridge University Press. [NF, aETR]
Gray, J. A. & McNaughton, N. (1996) The neuropsychology of anxiety: Reprise. *Nebraska Symposium on Motivation* 43:61–134. [JBP]
Gray, J. A., Young, A. M. J. & Joseph, M. H. (1997) Dopamine's role. *Science* 278:1548–49. [aETR]
Griffiths, P. E. (1997) *What emotions really are.* University of Chicago Press. [RA, KBK]
Halgren, E. (1992) Emotional neurophysiology of the amygdala within the context of human cognition. In: *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction,* ed. J. P. Aggleton. Wiley-Liss. [aETR]
Hall, B. K. (1992) *Evolutionary developmental biology.* Chapman and Hall. [JCF]
Haugeland, J. (1999) Authentic intentionality. *Proceedings of the New Trends in Cognitive Science Conference, Vienna, May 1999,* 60–69. Austrian Cognitive Society. [EM]
Hauser, M. D. (1996) *The evolution of communication.* MIT Press. [LDK]
(2000) *Wild minds: What animals really think.* Henry Holt. [JDK]
Hauser, M. D. & Nelson, D. A. (1991) Intentionality signaling in animal communication. *Trends in Ecology and Evolution* 6:186–89. [AIH]
Hebb, D. O. (1949) *The organization of behavior.* Wiley. [JCF]
Hinde, R. A. (1970) *Animal behavior: A synthesis of ethology and comparative psychology, 2nd edition.* McGraw-Hill. [JCF]
Hoebel, B. G. (1997) Neuroscience and appetitive behavior research: 25 years. *Appetite* 29:119–33. [aETR]
Hoffman, M. L. (1975) Developmental synthesis of affect and cognition and its implications for altruistic motivation. *Developmental Psychology* 11:607–22. [CI]
Hornak, J., Rolls, E. T. & Wade, D. (1996) Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia* 34:247–61. [arETR]
Houk, J. C., Adams, J. L. & 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,* ed. J. C. Houk, J. L. Davies & D. G. Beiser. MIT Press. [aETR]
Houston, A. I. & McNamara, J. M. (1989) The value of food: Effects of open and closed economies. *Animal Behaviour* 37:546–62. [AIH]
Houston, A. I. & McNamara, J. M. (1999) *Models of adaptive behaviour.* Cambridge University Press. [AIH]

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Houston, A. I. & Sumida, B. (1985) A positive feedback model for switching between two activities. *Animal Behaviour* 33:315–25. [AIH]
Hume, D. (1738/1911) *A treatise of human nature, vol. 1.* Dent. [EMA]
Ikemoto, S. & Panksepp, J. (2000) The role of nucleus accumbens dopamine in behavior: A unifying interpretation with special reference to reward seeking. *Brain Research Reviews.* (in press). [JP]
Izard, C. E. (1971) *The face of emotion.* Appleton-Century Crofts. [CI]
(1991) *The psychology of emotions.* Plenum. [aETR]
(1993) Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review* 100:68–90. [TD]
(1994) Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin* 115:288–99. [CI]
Izard, C. E., Fantauzzo, C. A., Castle, J. M., Haynes, O. M., Rayias, M. F. & Putnam, P. H. (1995) The ontogeny and significance of infants' facial expressions in the first nine months of life. *Developmental Psychology* 31:997–1013. [CI]
Izard, C. E., Hembree, E. A. & Huebner, R. R. (1987) Infants' emotion expressions to acute pain: Developmental change and stability of individual differences. *Developmental Psychology* 23:105–13. [CI]
James, W. (1884) What is an emotion? *Mind* 9:188–205. [aETR]
(1890) *The principles of psychology.* Reprinted 1950. Dover. [EAS]
Johnson, M. K. & Multhaup, K. S. (1992) Emotion and MEM. In: *The handbook of emotion and memory: Current research and theory,* ed. S. A. Christensen. Erlbaum. [TD]
Johnstone, R. A. (1997) The evolution of animal signals. In: *Behavioural ecology, 4th edition,* ed. J. R. Krebs & N. B. Davies. Blackwell. [AIH]
Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. & Sagi, D. (1994) Dependence on REM sleep of overnight improvement of a perceptual skill. *Science* 265:679–82. [SCM]
Kleinginna, P. R. & Kleinginna, A. M. (1981) A categorized list of motivation definitions, with suggestions for a consensual definition. *Motivation and Emotion* 5:263–91. [RB]
Kobatake, E. & Tanaka, K. (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology* 71(3):856–67. [EM]
Koechlin, E., Basso, G., Pietrini, P., Panzer, S. & Grafman, J. (1999) The role of the anterior prefrontal cortex in human cognition. *Nature* 399(6732):148–51. [JG]
Kollack-Walker, S., Watson, S. J. & Akil, H. (1997) Social stress in hamsters: Defeat activates specific neurocircuits within the brain. *Journal of Neuroscience* 17:8842–55. [JP]
Krebs, J. R. & Kacelnik, A. (1991) Decision making. In: *Behavioural ecology,* ed. J. R. Krebs & N. B. Davies. Blackwell. [aETR]
Kupferman, I. (1974) Feeding behavior in *Aplysia:* A simple system for the study of motivation. *Behavioral Biology* 10:1–26. [IK]
Kupferman, I., Rosen, S. C., Teyke, T., Cropper, E. C., Miller, M., Vilim, F. S. & Weiss, K. R. (1989) Neurobiology of behavioral states in *Aplysia:* Non-associative forms of plasticity of feeding responses. Dynamics and plasticity in neuronal systems. *Proceedings of the 17th Göttingen Neurobiology Conference,* ed. N. Elsner & W. Singer. Georg Thieme Verlag. [IK]
Kupferman, I., Brezina, V., Cropper, E. C., Deodhar, D., Probst, W. C., Rosen, S. C., Vilim, F. S. & Weiss, K. R. (1997) Reconfiguration of the peripheral plant during various forms of feeding behaviors in the mollusc *Aplysia.* In: *Neurons, networks, and motor behavior,* ed. P. S. G. Stein, S. Grillner, A. I. Selvertson & D. G. Stuart. MIT Press. [IK]
LaBar, K. S., Gitelman, D. R., Parrish, T. B., Kim, Y.-H., Nobre, A. C. & Mesulam, M.-M. (1999) Motivational state selectively modulates amygdala activation to appetitive visual stimuli. *NeuroImage,* June Supplement, Fifth International Conference on Functional Mapping of the Human Brain. [rETR]
Lackie, J. M. (1986) *Cell movement and cell behaviour.* Allen & Unwin. [RB]
Lange, C. (1885) The emotions. In: *The emotions,* ed. E. Dunlap. Williams and Wilkins. [aETR]
Larkin, S. & McFarland, D. J. (1978) The cost of changing from one activity to another. *Animal Behaviour* 26:1237–46. [AIH]
Lazarus, R. S. (1991) *Emotion and adaptation.* Oxford University Press. [aETR]
LeDoux, J. E. (1987) Emotion. In: *Handbook of physiology, vol. V. Section 1: The nervous system. Part 1: Higher functions of the brain,* ed. F. Plum. American Physiological Society. [EAS]
(1992) Emotion and the amygdala. In: *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction,* ed. J. P. Aggleton. Wiley-Liss. [JTR, arETR]
(1994) Emotion, memory and the brain. *Scientific American* 270:32–39. [aETR]
(1995) Emotion: Clues from the brain. *Annual Review of Psychology* 46:209–35. [TD, rETR]
(1996) *The emotional brain: The mysterious underpinnings of emotional life.* Simon and Schuster. [RA, RB, CI, EMA, AP, JBP, aETR, EAS]
Leibowitz, S. F. & Hoebel, B. G. (1998) Behavioral neuroscience and obesity. In: *The handbook of obesity,* ed. G. A. Bray, C. Bouchard & P. T. James. Dekker. [aETR]
Leonard, C. M., Rolls, E. T., Wilson, F. A. W. & Baylis, G. C. (1985) Neurons in the amygdala of the monkey with responses selective for faces. *Behavioural Brain Research* 15:159–76. [aETR]
Leventhal, H. & Scherer, K. (1987) The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion* 1:3–28. [TD]
Levine, S. & Ursin, H. (1991) What is stress? In: *Stress, neurobiology and neuroendocrinology,* ed. M. R. Brown, C. Rivier & G. Koob. Marcel Decker. [HU]
Lewis, M. D. (1996) Self-organising cognitive appraisals. *Cognition and Emotion* 10:1–25. [JTR]
Losick, R. & Kaiser, D. (1997) Why and how bacteria communicate. *Scientific American* 276(2):68–73. [RB]
Lumer, E. D., Friston, K. J. & Rees, G. (1998) Neural correlates of perceptual rivalry in the human brain. *Science* 280(5371):1930–34. [EM]
Lycan, W. (1996) *Consciousness and experience.* MIT Press. [MA]
Mac Aogáin, E. (1999) Information and appearance. *Behavioral and Brain Sciences* 22:159–60. [EMA]
Mackintosh, N. J. (1983) *Conditioning and associative learning.* Oxford University Press. [arETR]
MacLean, P. D. (1990) *The triune brain in evolution: Role in paleocerebral functions.* Plenum Press. [JP]
Maes, P. (1995) Artificial life meets entertainment: Interacting with lifelike autonomous agents. *Communications of the ACM, Special Issue on New Horizons of Commercial and Industrial AI* 38(11):108–14. [KBK]
Malkova, L., Gaffan, D. & 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* 17:6011–20. [AP, arETR]
Mazur, J. E. (1997/98) *Learning and behavior.* (4th edition, 1998). Prentice Hall. [JDK, rETR]
McBride, W. J., Murphy, J. M. & Ikemoto, S. (1999) Localization of brain reinforcement mechanisms: Intracranial self-administration and intracranial place-conditioning studies. *Behavioural Brain Research* 101:129–232. [JP]
McCulloch, G. (1994) *An analytical introduction to early Sartrean themes.* Routledge. [CI]
McFarland, D. J. & McFarland, F. J. (1968) Dynamic analysis of an avian drinking response. *Medical and Biological Engineering* 6:659–68. [AIH]
McNamara, J. M. (1996) Risk-prone behaviour under rules which have evolved in a changing environment. *American Zoologist* 36:484–95. [AIH]
McNamara, J. M. & Houston, A. I. (1980) The application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology* 85:673–90. AIH]
McNamara, J. M. & Houston, A. I. (1994) The effect of a change in foraging options on intake rate and predation rate. *American Naturalist* 144:978–1000. [AIH]
Medin, D. L. & Aguilar, C. M. (1999) Categorization. In: *MIT encyclopedia of cognitive sciences,* ed. R. A. Wilson & F. Keil. MIT Press. [JBP]
Meunier, M., Bachevalier, J., Murray, E. A., Malkova, L. & Mishkin, M. (1996) Effects of aspiration vs. neurotoxic lesions of the amygdala on emotional reactivity in rhesus monkeys. *Society for Neuroscience Abstracts* 22:1867. [rETR]
Millenson, J. R. (1967) *Principles of behavioral analysis.* Macmillan. [NF, aETR]
Miller, G. A. (1956) The magic number seven, plus or minus two: Some limits on our capacity for the processing of information. *Psychological Review* 63:81–93. [rETR]
Milner, A. D. & Goodale, M. A. (1995) *The visual brain in action.* Oxford University Press. [aETR]
Mineka, S., Davidson, M., Cook, M. & Keir, R. (1984) Observational conditioning of snake fear in rhesus monkeys. *Journal of Abnormal Psychology* 93(4):355–72. [CI]
Mirenowicz, J. & Schultz, W. (1996) Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 279(6564):449–51. [SK, aETR]
Moore, S. C. & Oaksford, M. (1999) Feeling low but learning faster: The long term effects of emotion on human cognition. In: *Proceedings of the 21st Annual Conference of the Cognitive Science Society, Simon Fraser University, Vancouver, Canada, August 19–21, 1999.* Erlbaum. [SCM]
Mora, F., Rolls, E. T. & Burton, M. J. (1976) Modulation during learning of the responses of neurones in the lateral hypothalamus to the sight of food. *Experimental Neurology* 53:508–19. [rETR]
Morris, J. S., Frith, C. D., Perrett, D. L., Rowland, D., Young, A. W., Calder, A. J. & Dolan, R. J. (1996) A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383:812–15. [aETR]
Mowrer, O. H. (1960) *Learning theory and behavior.* Wiley. [NF]
Murray, E. A. (1982) Medial temporal lobe structures contributing to recognition memory: The amygdala complex versus the rhinal cortex. In: *The amygdala:*

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

Neurobiological aspects of emotion, memory and mental dysfunction, ed. J. P. Aggleton. Wiley-Liss. [AP]
Murray, E. A., Gaffan, E. A. & Flint, R. W. (1996) Anterior rhinal cortex and amygdala: Dissociation of their contributions to memory and food preference in rhesus monkeys. *Behavioral Neuroscience* 110:30–42. [rETR]
Nagel, T. (1974) What is it like to be a bat? *Philosophical Review* 83(4):435–50. [MA]
Nichelli, P., Grafman, J., Pietrini, P., Clark, K., Lee, K. Y. & Miletich, R. (1995) Where the brain appreciates the moral of a story. *NeuroReport* 6(17):2309–13. [JG]
Oatley, K. & Jenkins, J. M. (1996) *Understanding emotions.* Backwell. [aETR]
Oatley, K. & Johnson-Laird, P. N. (1987) Towards a cognitive theory of emotions. *Cognition and Emotion* 1:29–50. [TD]
O'Doherty, J., Rolls, E. T., Francis, S., McGlone, F. & Bowtell, R. (2000) The representation of pleasant and aversive taste in the human brain. *Journal of Neurophysiology.* [rETR]
Olds, J. (1976) Behavioral studies of hypothalamic functions: Drives and reinforcements. In: *Biological foundations of psychiatry, vol. 1,* ed. R. G. Grenell & S. Gabay. Raven Press. [EAS]
Olmstead, M. C. & Franklin, K. B. (1997) The development of a conditioned place preference to morphine: Effects of microinjections into various CNS sites. *Behavioral Neuroscience* 111:1324–34. [JP]
Ono, T. & Nishijo, H. (1992) Neurophysiological basis of the Kluver-Bucy syndrome: Responses of monkey amygdaloid neurons to biologically significant objects. In: *The amydgala,* ed. J. P. Aggleton. Wiley-Liss. [aETR]
Panksepp, J. (1981) Hypothalamic integration of behavior: Rewards, punishments, and related psychobiological process. In: *Handbook of the hypothalamus, vol. 3. Part A: Behavioral studies of the hypothalamus,* ed. P. J. Morgane & J. Panksepp. Marcel Dekker. [JP]
(1986) The anatomy of emotions. In: *Emotion: Theory, research and experience: Vol. 3: Biological foundations of emotions,* ed. R. Plutchik & H. Kellerman. Academic Press. [JP]
(1990) Gray zones at the emotion-cognition interface: A commentary. *Cognition and Emotion* 4:289–302. [JP]
(1993) Neurochemical control of moods and emotions: Amino acids to neuropeptides. In: *The handbook of emotions,* ed. M. Lewis & J. Haviland. Guilford. [JP]
(1994) Basic emotions ramify widely in the brain, yielding many concepts that cannot be distinguished unambiguously . . . yet. In: *Questions about emotions,* ed. R. Davidson & P. Ekman. Oxford University Press. [JP]
(1998a) *Affective neuroscience: The foundations of human and animal emotions.* Oxford University Press. [RA, RB, CD, NF, JP, rETR]
(1998b) The periconscious substrates of consciousness: Affective states and the evolutionary origins of the self. *Journal of Consciousness Studies* 5:566–82. [JP]
(1999) Emotions as viewed by psychoanalysis and neuroscience: An exercise in consilience. *NeuroPsychoanalysis* 1:15–38. [JP]
Parker, A., Wilding, E. & Akerman, C. (1998) The von Restorff effect in visual object recognition memory in humans and monkeys: The role of frontal/perirhinal interaction. *Journal of Cognitive Neuroscience* 10:691–703. [AP]
Parker, A., Baxter, M. G., Lindner, C., Izquierdo, A. D. & Murray, E. A. (1999) Interaction of the amygdala with orbital prefrontal cortex in reinforcer devaluation in rhesus monkeys. *Society for Neuroscience Abstracts* 25:790. [AP]
Partiot, A., Grafman, J., Sadato, N., Wachs, J. & Hallett, M. (1995) Brain activation during the generation of non-emotional and emotional plans. *NeuroReport* 6(10):1397–1400. [JG]
Pearce, J. M. (1997) *Animal learning and cognition, 2nd edition.* Psychology Press. [rETR]
Peterson, J. B. (1999) *Maps of meaning: The architecture of belief.* Routledge. http://psych.utoronto.ca/~peterson/mom.htm [JBP]
Petrides, M. (1996) Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philosophical Transactions of the Royal Society B* 351:1455–62. [arETR]
Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A. & David, A. S. (1997) A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389:495–98. [RA]
Picard, R. (1997) *Affective computing.* MIT Press. [KBK, JP]
Pinker, S. (1997) *How the mind works.* W. W. Norton. [JDK]
Plata-Salaman, C. R., Smith-Swintosky, V. L. & Scott, T. R. (1996) Gustatory neural coding in the monkey cortex: Mixtures. *Journal of Neurophysiology* 75:2369–79. [rETR]
Power, M. J. & Dalgleish, T. (1997) *Cognition and emotion: From order to disorder.* Psychology Press. [TD]
Pribram, K. H. (1981) Emotions. In: *Handbook of clinical neuropsychology,* ed. S. B. Filskov & T. J. Boll. Wiley. [NF]
Pylyshyn, Z. (1999) Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences* 22(3):341–423. [EMA]
Rada, P., Mark, G. P. & Hoebel, B. G. (1998) Dopamine in the nucleus accumbens released by hypothalmic stimulation-escape behavior. *Brain Research* 782:228–34. [aETR]
Reisenzein, R. (1983) The Schachter theory of emotion: Two decades later. *Psychological Bulletin* 94:239–64. [aETR]
Renniger, K. A. & Wozniak, R. H. (1985) Effect of interest on attentional shift, recognition, and recall in young children. *Developmental Psychology* 21(4):624–32. [CI]
Rey, G. (1997) *Contemporary philosophy of mind.* Blackwell. [MA]
Ridley, M. (1993) *The red queen: Sex and the evolution of human nature.* Penguin. [rETR]
Robbins, T. W., Cador, M., Taylor, J. R. & Everitt, B. J. (1989) Limbic-striatal interactions in reward-related processes. *Neuroscience and Biobehavioral Reviews* 13:155–62. [aETR]
Robinson, T. & Berridge, K. (1993) The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews* 18:247–91. [JP]
Rolls, B. J. & Rolls, E. T. (1982) *Thirst.* Cambridge University Press. [aETR]
Rolls, E. T. (1975) *The brain and reward.* Pergamon. [aETR]
(1986a) Neural systems involved in emotion in primates. In: *Emotion: Theory, research, and experience. Vol. 3. Biological foundations of emotion,* ed. R. Plutchik & H. Kellerman. Academic Press. [aETR]
(1986b) A theory of emotion, and its application to understanding the neural basis of emotion. In: *Emotion. Neural and chemical control,* ed. Y. Oomura. Karger. [aETR]
(1990) A theory of emotion, and its application to understanding the neural basis of emotion. *Cognition and Emotion* 4:161–90. [aETR]
(1992a) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society* 335:11–21. [aETR]
(1992b) Neurophysiology and functions of the primate amygdala. In: *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction,* ed. J. P. Aggleton. Wiley-Liss. [arETR]
(1996) The orbitofrontal cortex. *Philosophical Transactions of the Royal Society B* 351:1433–44. [arETR]
(1997) Taste and olfactory processing in the brain and its relation to the control of eating. *Critical Reviews in Neurobiology* 11:263–87. [arETR]
(1999a) *The brain and emotion.* Oxford University Press. [arETR]
(1999b) The functions of the orbitofrontal cortex. *Neurocase* 5:301–12. [rETR]
(2000a) Neurophysiology and functions of the primate amygdala, and the neural basis of emotion. In: *The amygdala, a functional analysis,* ed. J. P. Aggleton. Oxford University Press. [rETR]
(2000b) The orbitofrontal cortex and reward. *Cerebral Cortex* 10:284–94. [rETR]
Rolls, E. T., Burton, M. J. & Mora, F. (1976) Hypothalamic neuronal responses associated with the sight of food. *Brain Research* 111:53–66. [rETR]
(1980) Neurophysiological analysis of brain-stimulation reward in the monkey. *Brain Research* 194:339–57. [arETR]
Rolls, E. T., Hornak, J., Wade, D. & McGrath, J. (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry* 57:1518–24. [arETR]
Rolls, E. T. & Johnstone, S. (1992) Neurophysiological analysis of striatal function. In: *Neuropsychological disorders associated with subcortical lesions,* ed. G. Vallar, S F. Cappa & C. W. Wallesch. Oxford University Press. [aETR]
Rolls, E. T., Judge, S. J. & Sanghera, M. (1977) Activity of neurones in the inferotemporal cortex of the alert monkey. *Brain Research* 130:229–38. [rETR]
Rolls, E. T., Murzi, E., Yaxley, S., Thorpe, S. J. & Simpson, S. J. (1986) Sensory-specific satiety: Food-specific reduction in responsiveness of ventral forebrain neurons after feeding in the monkey. *Brain Research* 368:79–86. [rETR]
Rolls, E. T., Sanghera, M. K. & Roper-Hall, A. (1979) The latency of activation of neurons in the lateral hypothalamus and substantia innominata during feeding in the monkey. *Brain Research* 164:121–35. [rETR]
Rolls, E. T., Thorpe, S. J. & Maddison, S. P. (1983) Responses of striatal neurons in the behaving monkey: I. Head of the caudate nucleus. *Behavioural Brain Research* 7:179–210. [aETR]
Rolls, E. T. & Treves, A. (1998) *Neural networks and brain function.* Oxford University Press. [JCF, arETR]
Rolls, E. T., Critchley, H. D., Browning, A. S., Hernadi, A. & Lenard, L. (1999) Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex. *Journal of Neuroscience* 19:1532–40. [aETR]
Rosenthal, D. M. (1990) A theory of consciousness. Technical Report ZIF No.40, Zentrum für Interdisciplinaire Forschung. [KBK]
(1991) Two concepts of consciousness. In: *The nature of mind,* ed. D. M. Rosenthal. Oxford University Press. [JLB]

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press

(1993) Thinking that one thinks. In: *Consciousness,* ed. M. Davies & G. W. Humphreys. Blackwell. [rETR]
Rozeboom, W. W. (1960) Do stimuli elicit behavior? A study in the logical foundations of behavioristics. *Philosophy of Science* 27:159–70. [EMA]
Rueckert, L. & Grafman, J. (1996) Sustained attention deficits in patients with right frontal lesions. *Neuropsychologia* 34(10):953–63. [JG]
Rumelhart, D. E., McClelland, J. L. & the PDP Research Group (1986) *Parallel distributed processing: Explorations in the microstructure of cognition.* MIT Press. [KBK]
Russell, S. & Norvig, P. (1995) *Artificial intelligence: A modern approach.* Prentice-Hall. [KBK]
Sackett, G. (1966) Monkeys reared in isolation with pictures as visual input: Evidence for an innate releasing mechanism. *Science* 154:1468–73. [CI]
Salzen, E. A. (1991) On the nature of emotion. *International Journal of Comparative Psychology* 5:47–88. [EAS]
(1993) The neural systems of emotion. *European Journal of Neuroscience, Suppl.* 6:155. [EAS]
(1998) Emotion and self-awareness. *Applied Animal Behaviour Science* 57:299–313. [EAS]
Sanghera, M. K., Rolls, E. T. & Rooper-Hall, A. (1979) Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Experimental Neurology* 63:610–26. [aETR]
Schacter, S. & Singer, J. E. (1962) Cognitive, social and physiological determinants of emotional state. *Psychological Review* 69:379–99. [RA]
Scherer, K. S. (1999) Appraisal theory. In: *Handbook of cognition and emotion,* ed. T. Dalgleish & M. J. Power. Wiley. [TD]
Schore, A. N. (1994) *Affect regulation and the origin of the self: The neurobiology of emotional development.* Erlbaum. [RA]
Schultz, W. (1998) Predictive reward signal of dopamine neurons. *Journal of Neurophysiology* 80:1–27. [JP, aETR]
Schultz, W., Romo, R., Ljunberg, T., Mirenowicz, J., Hollerman, J. R. & Dickinson, A. (1995) Reward-related signals carried by dopamine neurons. In: *Models of information processing in the basal ganglia,* ed. J. C. Houk, J. L. Davis & D. G. Beiser. MIT Press. [aETR]
Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P. & Johnson, M. (1997) Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385:254–57. [aETR]
Scott, T. R., Yan, J. & Rolls, E. T. (1995) Brain mechanisms of satiety and taste in macaques. *Neurobiology* 3:281–92. [aETR]
Searle, J. (1980) Minds, brains, and programs. *Behavioral and Brain Sciences* 3:417–24. [KBK]
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. [aETR]
(1976) Electrical stimulation and self-stimulation in man with chronic implanted electrodes. Interpretation and pitfalls of results. In: *Brain-stimulation reward,* ed. A. Wauquier & E. T. Rolls. North-Holland. [aETR]
Shallice, T. & Burgess, P. (1996) The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society B* 351:1405–11. [arETR]
Sherry, D. F. (1997) Cross-species comparisons. In: *Characterizing human psychological adaptations,* ed. M. Daly. Wiley. [JDK]
Sirigu, A., Cohen, L., Zalla, T., Pradat-Diehl, P., Van Eeckhout, P., Grafman, J. & Agid, Y. (1998) Distinct frontal regions for processing sentence syntax and story grammar. *Cortex* 34(5):771–78. [JG]
Sirigu, A., Zalla, T., Pillon, B., Grafman, J., Agid, Y. & Dubois, B. (1996) Encoding of sequence and boundaries of scripts following prefrontal lesions. *Cortex* 32(2):297–310. [JG]
Sirigu, A., Zalla, T., Pillon, B., Grafman, J., Dubois, B. & Agid, Y. (1995) Planning and script analysis following prefrontal lobe lesions. *Annals of the New York Academy Sciences* 769:277–88. [JG]
Smith, C. A. & Ellsworth, P. C. (1985) Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology* 48:813–38. [JTR]
Smith, O. A. & DeVito, J. L. (1984) Cental neural integration for the control of autonomic responses associated with emotion. *Annual Review of Neuroscience* 7:43–65. [EAS]
Smith-Swintosky, V. L., Plata-Salaman, C. R. & Scott, T. R. (1991) Gustatory neural coding in the monkey cortex: Stimulus quality. *Journal of Neurophysiology* 66:1156–65. [rETR]
Sokolov, E. N. (1969) The modeling properties of the nervous system. In: *Handbook of contemporary Soviet psychology,* ed. I. Maltzman & K. Coles. Basic Books. [JBP]
Solms, M. & Nersessian, E. (1999) Freud's theory of affect. *NeuroPsychoanalysis* 1:5–14. [JP]
Sorce, J. F., Emde, R. N., Campos, J. J. & Klinnert, M. D. (1985) Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology* 21(1):195–200. [CI]
Spinoza, B. (1677/1985) *Ethics.* In: *The collected works of Spinoza,* ed. E. Curley. Princeton University Press. [AB-Z]
Strongman, K. T. (1996) *The psychology of emotion, 4th edition.* Wiley. [aETR]
Szentagothai, J. (1984) Downward causation? *Annual Review of Neuroscience* 7:1–11. [EAS]
Teasdale, J. & Barnard, P. (1993) *Affect, cognition and change.* Erlbaum. [TD]
Termine, N. T. & Izard, C. E. (1988) Infants' responses to their mothers' expressions of joy and sadness. *Developmental Psychology* 24:223–29. [CI]
Thierry, A. M., Tassin, J. P., Blanc, G. & Glowinski, J. (1976) Selective activation of mesocortical DA system by stress. *Nature* 263:242–44. [aETR]
Thistlethwaite, D. (1951) A critical review of latent learning and related experiments. *Psychological Bulletin* 48:97–129. [DRJL]
Thorpe, S. J., Rolls, E. T. & Maddison, S. (1983) Neuronal activity in the orbitofrontal cortex of the behaving monkey. *Experimental Brain Research* 49:93–115. [aETR]
Tinbergen, N. (1951) *The study of instinct.* Oxford University Press. [JCF, aETR]
Tolman, E. C. (1959) Principles of purposive behavior. In: *Psychology: A study of a science, vol. 2.* McGraw-Hill. [SK]
Trivers, R. L. (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57. [AIH]
Tye, M. (1996) *Ten problems of consciousness.* MIT Press. [MA]
Ursin, H. (1985) The instrumental effects of emotional behavior. In: *Perspectives in ethology, vol. 6,* ed. P. P. G. Bateson & P.-H. Klopfer. Plenum Press. [HU]
Ursin, R., Olds, J. & Ursin, H. (1966) Self-stimulation of hippocampus in rats. *Journal of Comparative Physiological Psychology* 61:353–59. [HU]
Van Gulick, R. (1994) Deficit studies and the function of phenomenal consciousness. In: *Philosophical psychopathology,* ed. G. Graham & L. S. Stephens. MIT Press. [JLB]
Vinogradova, O. (1975) Functional organization of the limbic system in the process of registration of information: Facts and hypotheses. In: *The hippocampus, neurophysiology and behavior, vol. 2,* ed. R. Isaacson & K. Pribram. Plenum Press. [JBP]
Wallis, G. & Rolls, E. T. (1997) Invariant face and object recognition in the visual system. *Progress in Neurobiology* 51:167–94. [aETR]
Weiskrantz, L. (1968) Emotion. In: *Analysis of behavioural change,* ed. L. Weiskrantz. Harper and Row. [aETR]
Weiss, K. R., Brezina, V., Cropper, E. C., Heirhorst, J., Hooper, S. L., Probst, W. C., Vilim, F. S. & Kupferman, I. (1993) Physiology and biochemistry of peptidergic cotransmission in *Aplysia. Journal of Physiology (Paris)* 87:141–51. [IK]
Wiener, N. (1948) *Cybernetics, or control and communication in the animal and the machine.* Cambridge University Press. [JBP]
Williams, G. V., Rolls, E. T., Leonard, C. M. & Stern, C. (1993) Neuronal responses in the ventral striatum of the behaving macaque. *Behavioural Brain Research* 55:243–52. [aETR]
Wilson, F. A. W. & Rolls, E. T. (1990a) Neuronal responses related to reinforcement in the primate basal forebrain. *Brain Research* 502:213–31. [rETR]
(1990b) Neuronal responses related to the novelty and familiarity of visual stimuli in the substantia innominata, diagonal band of Broca and periventricular region of the primate. *Experimental Brain Research* 80:104–20. [rETR]
(1990c) Learning and memory are reflected in the responses of reinforcement-related neurons in the primate basal forebrain. *Journal of Neuroscience* 10:1254–67. [rETR]
(1993) The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. *Experimental Brain Research* 93:367–82. [aETR]
(2000) The primate amygdala and reinforcement: A dissociation between rule-based and associatively-mediated memory revealed in amygdala neuronal activity.
Yaxley, S, Rolls, E. T. & Sienkiewicz, Z. J. (1990) Gustatory responses of single neurons in the insula of the macaque monkey. *Journal of Neurophysiology* 63:689–700. [rETR]
Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P. & Hanley, J. R. (1995) Face processing impairments after amygdalotomy. *Brain* 118:15–24. [aETR]
Young, A. W., Hellawell, D. J., Van de Wal, C. & Johnson, M. (1996) Facial expression processing after amygdalotomy. *Neuropsychologia* 34:31–39. [aETR]
Zahn-Waxler, C., Radke-Yarrow, M. & King, R. A. (1979) Child rearing and children's prosocial initiations toward victims of distress. *Child Development* 50(2):310–30. [CI]
Zola-Morgan, S., Squire, L. R. & Amaral, D. G. (1989) Lesions of the amygdala that spare adjacent cortical regions do not impair memory or exacerbate the impairment following lesions of the hippocampal formation. *Journal of Neuroscience* 9:1922–36. [rETR]

https://doi.org/10.1017/S0140525X00512424 Published online by Cambridge University Press