



Emotion, motivation, decision-making, the orbitofrontal cortex, anterior cingulate cortex, and the amygdala

Edmund T. Rolls^{1,2}

Received: 27 January 2023 / Accepted: 12 April 2023 / Published online: 13 May 2023
© Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The orbitofrontal cortex and amygdala are involved in emotion and in motivation, but the relationship between these functions performed by these brain structures is not clear. To address this, a unified theory of emotion and motivation is described in which motivational states are states in which instrumental goal-directed actions are performed to obtain rewards or avoid punishers, and emotional states are states that are elicited when the reward or punisher is or is not received. This greatly simplifies our understanding of emotion and motivation, for the same set of genes and associated brain systems can define the primary or unlearned rewards and punishers such as sweet taste or pain. Recent evidence on the connectivity of human brain systems involved in emotion and motivation indicates that the orbitofrontal cortex is involved in reward value and experienced emotion with outputs to cortical regions including those involved in language, and is a key brain region involved in depression and the associated changes in motivation. The amygdala has weak effective connectivity back to the cortex in humans, and is implicated in brainstem-mediated responses to stimuli such as freezing and autonomic activity, rather than in declarative emotion. The anterior cingulate cortex is involved in learning actions to obtain rewards, and with the orbitofrontal cortex and ventromedial prefrontal cortex in providing the goals for navigation and in reward-related effects on memory consolidation mediated partly via the cholinergic system.

Keywords Emotion · Motivation · Reward · Human orbitofrontal cortex · Cingulate cortex · Amygdala · Ventromedial prefrontal cortex · Memory · Depression · Consciousness · Mind-brain problem · Welfare

Introduction and aims

There have been considerable advances recently in understanding the connectivity and connections of the human orbitofrontal cortex and amygdala, and how they relate to emotion (Rolls et al. 2023a, d), but how these systems and processes are related to motivation has been much less explored. This paper shows how the brain systems involved in motivation are similar to those involved in emotion, and provides a framework for understanding how emotion and motivation are related to each other, and how similar brain systems are involved in both. This paper aims to make key advances in our understanding of how the orbitofrontal

cortex and amygdala structure (anatomy and connectivity) is related to the two key functions performed by these brain regions, emotion and motivation.

To understand the neuroscience of both emotion and motivation, it is important to have a framework for understanding the relation between emotion and motivation. This paper first sets out a theory of emotion, and a framework for understanding the relation between emotion and motivation, and then considers how brain regions involved in emotion and motivation, the orbitofrontal cortex, anterior cingulate cortex, and amygdala, are involved in emotion and motivation. Special reference is made to these brain regions in primates including humans, to ensure that what is described is relevant to understanding brain systems involved in emotion and motivation in humans, and their disorders. Recent evidence about the connectivity of these systems in humans makes this paper very timely (Rolls et al. 2023a, d). A second aim is to show how emotion and its brain systems are highly adaptive from an evolutionary and gene specification perspective. The third aim is to consider where and how

✉ Edmund T. Rolls
Edmund.Rolls@oxcns.org
https://www.oxcns.org

¹ Oxford Centre for Computational Neuroscience, Oxford, UK

² Department of Computer Science, University of Warwick, Coventry, UK

decisions are made about reward and emotional value, and separately about where and how decisions are made about the actions to obtain the rewards. The fourth aim is to consider some of the implications of this research for understanding brain function in health and disease; evolution to select for brain systems that respond to stimuli that encode rewards and punishers; memory and memory consolidation; and personality.

The approach taken here is new, in that it produces a unified approach to understanding emotion and motivation and their underlying brain mechanisms; in that it updates our understanding of the brain mechanisms of emotion (Rolls 2014b, 2018) by incorporating new evidence on the effective connectivity as well as the functional connectivity and the tractography of the brain systems involved in humans (Rolls et al. 2023a, d); in that it emphasises how evolution operates in part by selecting for brain reward systems that increase reproductive fitness; and in that it considers implications for understanding brain function in neurological and psychiatric states, how reward and emotional systems relate to episodic and semantic memory and memory consolidation, and welfare. The new results and understanding from taking this approach, including the advances related to new investigations of effective connectivity of the human brain, are summarised in "[Conclusions and highlights](#)".

A theory of emotion relevant to brain systems involved in reward value and emotion

First a definition and theory of emotion and its functions are provided, and then key brain regions involved in emotion are considered, including the orbitofrontal cortex, anterior cingulate cortex, amygdala, striatum, the dopamine system, and the insula.

A definition of emotion

A clear working definition of emotion is helpful before we consider its brain mechanisms. Emotions can usefully be defined (operationally) as states elicited by the presentation, termination or omission of rewards and punishers which have particular functions (Rolls 1999, 2000a, 2013b, 2014b, 2018). A reward is anything for which an animal (which includes humans) will work. A punisher is anything that an animal will escape from or avoid. As shown in Fig. 1, different reward/punishment contingencies are associated with different types of emotion. An example of an emotion associated with a reward might be the happiness produced by being given a particular reward, such as a pleasant touch, praise, or winning a large sum of money. An example of an emotion produced by a punisher 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 punishing stimuli. An example of an emotion produced by the omission or termination or loss of a reward is frustration or anger (if some action can be taken), or sadness (if no action can be taken). An example of an emotion produced by the omission or termination of a punisher (such as the removal of a painful stimulus, or sailing out of danger) would be relief. These examples indicate how emotions can be produced by the delivery, omission, or termination of rewarding or punishing stimuli, and go some way to indicate how different emotions could be produced and classified in terms of the rewards and punishers received, omitted, or terminated. Figure 1 summarizes some of the emotions associated with the delivery of a reward or punisher or a stimulus associated with them, or with the omission of a reward or punisher.

The subjective feelings of emotions are part of the much larger problem of consciousness (Rolls 2020). The brain bases of subjective experience are a topic of considerable current interest, not only with higher order thought (HOT) theories (Rosenthal 2004; Brown et al. 2019), but also with the higher order syntactic thought (HOST) theory of consciousness (Rolls 2007a, 2012b, 2014b, 2016c, 2018, 2020) which is more computationally specific and addresses the adaptive value of the type of processing related to consciousness; and a point made here is that the orbitofrontal cortex is at least on the route to human subjective experience of emotion and affective value (see below).

I consider elsewhere a slightly more formal definition than rewards or punishers, in which the concept of reinforcers is introduced, and it is shown that emotions can be usefully seen as states produced by instrumental reinforcing stimuli (Rolls 2014b). Instrumental reinforcers are stimuli which, if their occurrence, termination, or omission is made contingent upon the making of a response, alter the probability of the future emission of that response (Cardinal et al. 2002).

Some stimuli are unlearned (innate), "primary", reinforcers (e.g., the taste of food if the animal is hungry, or pain). Some examples of primary reinforcers are shown in Table 1 (Rolls 2014b). There may be in the order of 100 such primary reinforcers, each specified by different genes (Rolls 2014b). Each primary reinforcer can produce a different type of affective state, for example the taste of a pleasant sweet or sweet/fat texture food such as ice cream is very different from the feel of a pleasant touch vs pain; which are all in turn very different from attraction to or love for someone. Thus different types of affective state are produced by each different primary reinforcer, and the reinforcement contingencies shown in Fig. 1 apply to each of these primary reinforcers. For example, not receiving ice cream is very different emotionally from not receiving pleasant touch.

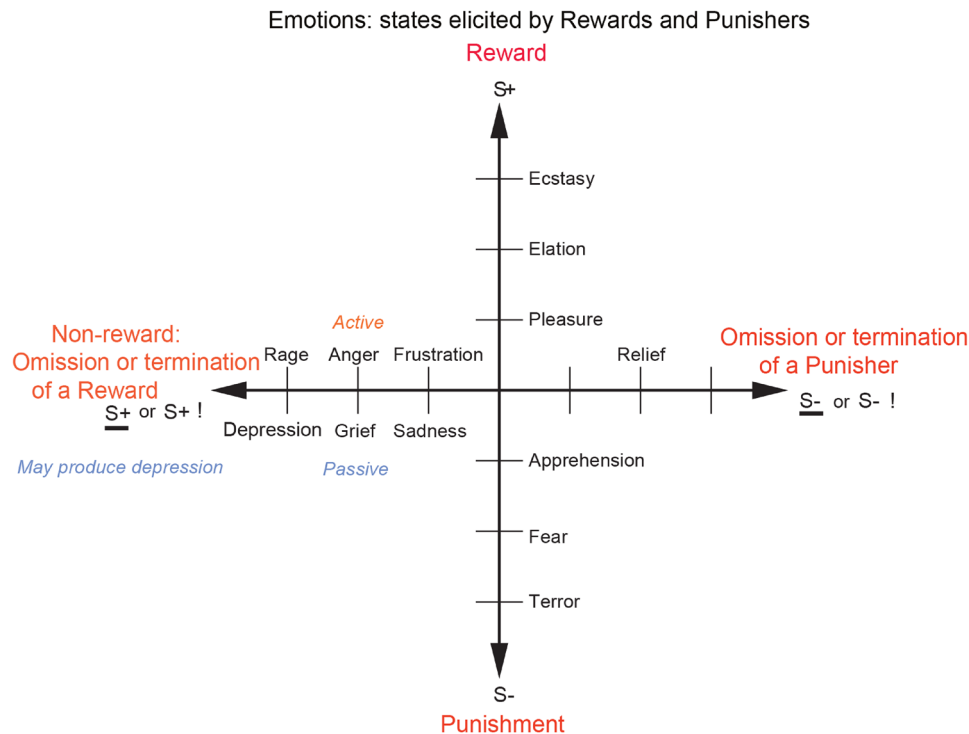


Fig. 1 Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the centre of the diagram, on a continuous scale. The classification scheme created by the different reinforcement contingencies consists with respect to the action of (1) the delivery of a reward (S+), (2) the delivery of a punisher (S−), (3) the omission of a reward (S±) (extinction) or the termination of a reward (S+!) (time out), and (4) the omission of a punisher (S−) (avoidance) or the termination of a punisher (S−!) (escape). Note that the vertical axis describes emotions associated with the delivery of a reward (up) or punisher (down). The horizontal axis describes emotions associated with the non-delivery of an expected reward (left) or the non-delivery of an expected punisher (right). For the contingency of non-reward (hori-

zontal axis, left) different emotions can arise depending on whether an active action is possible to respond to the non-reward, or whether no action is possible, which is labelled as the passive condition. In the passive condition, non-reward may produce depression. Frustration could include disappointment. The diagram summarizes emotions that might result for one reinforcer as a result of different contingencies. Every separate reinforcer has the potential to operate according to contingencies such as these. This diagram does not imply a dimensional theory of emotion, but shows the types of emotional state that might be produced by a specific reinforcer. Each different reinforcer will produce different emotional states, but the contingencies will operate as shown to produce different specific emotional states for each different reinforcer

Other stimuli may become reinforcing by associative learning, because of their association with such primary reinforcers, thereby becoming "secondary reinforcers". An example might be the sight of a painful stimulus. Brain systems that learn and unlearn these associations between stimuli or events in the environment and reinforcers are important in understanding the neuroscience and neurology of emotions, as we will see below.

This foundation has been developed (Rolls 2014b) 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 reward and of a punisher, allowing states such as conflict and guilt to arise.)
4. Emotions elicited by stimuli associated with *different primary reinforcers* will be different, as described above, with some primary reinforcers each of which will produce different affective states shown in Table 1.
5. Emotions elicited by *different secondary reinforcing stimuli* will be different from each other (even if the primary reinforcer is similar). For example, the same touch to the arm but by different people might give rise to very different emotions. Cognitive states and semantic knowledge can contribute to emotion in these ways, as well as in other ways that might arise because for example of reasoning in the rational brain system.

Table 1 Some primary reinforcers and the dimensions of the environment to which they are tuned

Taste	
Salt taste	Reward in salt deficiency
Sweet	Reward in energy deficiency
Bitter	Punisher, indicator of possible poison
Sour	Punisher
Umami	Reward, indicator of protein; Produced by monosodium glutamate and inosine monophosphate
Tannic acid	Punisher: it prevents absorption of protein; found in old leaves; probably somatosensory not gustatory (Critchley and Rolls 1996)
Odor	
Putrefying odor	Punisher; hazard to health
Pheromones	Reward (depending on hormonal state)
Somatosensory	
Pain	Punisher
Touch	Reward
Grooming	Reward; to give grooming may also be a primary reinforcer
Washing	Reward
Temperature	Reward if tends to help maintain normal body temperature; otherwise punisher
Visual	
Snakes, etc.	Punisher for, e.g., primates
Youthfulness	Reward, associated with mate choice
Beauty, e.g., symmetry	Reward
Secondary sexual characteristics	Rewards
Face expression	Reward (e.g., smile) or punisher (e.g., threat)
Blue sky, cover, open space	Reward, indicator of safety
Flowers	Reward (indicator of fruit later in the season?)
Auditory	
Warning call	Punisher
Aggressive vocalization	Punisher
Soothing vocalization	Reward (part of the evolutionary history of music, which at least in its origins taps into the channels used for the communication of emotions)
Reproduction	
Courtship	Reward
Sexual behavior	Reward (different reinforcers, including a low waist-to-hip ratio, and attractiveness influenced by symmetry and being found attractive by members of the opposite sex)
Mate guarding	Reward for a male to protect his parental investment Jealousy results if his mate is courted by another male, because this may ruin his parental investment
Nest building	Reward (when expecting young)
Parental attachment (love)	Reward (good for the parent's genes both when the attachment is to the other parent or an infant)
Infant attachment to parents (love)	Reward (good for the infant's genes)
Crying of infant	Punisher to parents; produced to promote successful development
Power, status, wealth, resources	Attractive to females, who may benefit from resources for their offspring Attractive to males as they make males attractive to females
Body size	Large in males may be attractive to females as a signal for the provision of protection and of the ability of her male offspring to compete for a mate Small in females may be attractive to males as a neotenus sign of youth, and therefore fertility
Other	
Novel stimuli	Rewards (encourage animals to investigate the full possibilities of the multidimensional space in which their genes are operating)

Table 1 (continued)

Sleep	Reward; minimizes nutritional requirements and protects from danger
Altruism to genetic kin	Reward (kin altruism)
Altruism to other individuals	Reward while the altruism is reciprocated in a ‘tit-for-tat’ reciprocation (reciprocal altruism). Forgiveness, honesty, and altruistic punishment are some associated heuristics. May provide underpinning for some aspects of what is felt to be moral
Altruism to other individuals	Punisher when the altruism is not reciprocated
Group acceptance, reputation	Reward (social greeting might indicate this) These goals can account for why some cultural goals are pursued
Control over actions	Reward
Play	Reward
Danger, stimulation, excitement	Reward if not too extreme (adaptive because of practice?)
Exercise	Reward (keeps the body fit for action)
Mindreading	Reward; practice in reading others’ minds, which might be adaptive
Solving an intellectual problem	Reward (practice in which might be adaptive)
Storing, collecting	Reward (e.g., Food)
Habitat preference, home, territory	Reward
Some responses	Reward (e.g., pecking in chickens, pigeons; adaptive because it is a simple way in which eating grain can be programmed for a relatively fixed type of environmental stimulus)
Breathing	Reward

6. The emotion elicited can depend on whether an *active* or *passive behavioural response* is possible. (For example, if an active behavioural response can occur to the omission of a positive reinforcer, then anger might be produced, but if only passive behaviour is possible, then sadness, depression or grief might occur: see Fig. 1.)

By combining these six factors, it is possible to account for a very wide range of emotions, as described by Rolls (2014b). This is important: the range of emotions that can be accounted for in this way is enormous (Rolls 2014b), and is not limited (Adolphs and Anderson 2018). It is also worth noting that emotions can be produced just as much by the recall of reinforcing events as by external reinforcing stimuli; that cognitive processing (whether conscious or not) is important in many emotions, for very complex cognitive processing may be required to determine whether or not environmental events are reinforcing. Indeed, emotions normally consist of cognitive processing that analyses the stimulus, and then determines its reinforcing valence; and then an elicited affective (emotional) state or longer term mood change if the valence is positive or negative. I note that a mood or affective state may occur in the absence of an external stimulus, as in some types of depression, but that normally the mood or affective state is produced by an external stimulus, with the whole process of stimulus representation, evaluation in terms of reward or punishment, and the resulting mood or affect being referred to as emotion (Rolls 2014b).

The functions of emotions

The most important function of emotion is as part of the processes of learning goal-directed actions to obtain rewards or avoid punishers. The first process is stimulus-reinforcer association learning; emotional states are produced as a result (Rolls 2014b). An example might be learning that the sight of a person is associated with rewards, which might produce the emotion of happiness. This process is implemented in structures such as the orbitofrontal cortex and amygdala (Figs. 2, 3, 4) (Rolls and Grabenhorst 2008; Grabenhorst and Rolls 2011; Rolls 2014b).

The second process is instrumental learning of an action made to approach and obtain the reward (an outcome of the action) or to avoid or escape from the punisher (an outcome). This is action–outcome learning, and involves brain regions such as the anterior cingulate cortex when the actions are being guided by the goals (Rushworth et al. 2011, 2012; Rolls 2014b, 2018, 2019a, 2021b, 2023d). 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 (Rolls 2014b). The behaviour is under control of the reward value of the goal, in that if the reward is devalued, for example by feeding a food until satiety is reached, then on the very next occasion that the stimulus (the food) is offered, no action will be performed to try to obtain it (Rolls 2014b).

The striatum, rest of the basal ganglia, and dopamine system can become involved when the behaviour becomes automatic, and habit-based, that is, uses stimulus–response connections (Figs. 2, 3). In this situation, very little emotion

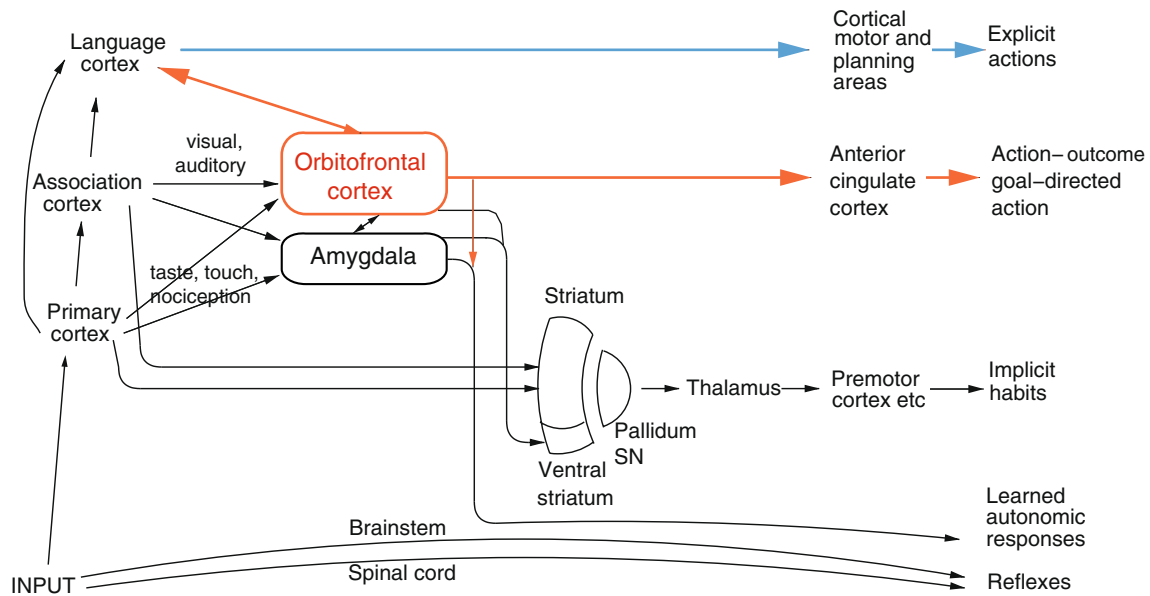


Fig. 2 Multiple routes to the initiation of actions and responses to rewarding and punishing stimuli in primates including humans. The lowest (spinal cord and brainstem) levels in the hierarchy are involved in reflexes, including for example reflex withdrawal of a limb to a nociceptive stimulus, and unlearned autonomic responses. The second level in the hierarchy involves associative learning in the amygdala and orbitofrontal cortex between primary reinforcers such as taste, touch and nociceptive stimuli and neutral stimuli such as visual and auditory stimuli from association cortex (e.g. inferior temporal visual cortex) to produce learned autonomic and some other behavioural responses such as approach. The anteroventral viscer-autonomic insula may be one link from the orbitofrontal cortex to autonomic output. A third level in the hierarchy is the route from the orbitofrontal cortex and amygdala via the basal ganglia especially

the ventral striatum to produce implicit stimulus–response habits. A fourth level in the hierarchy important in emotion is from the orbitofrontal cortex to the anterior cingulate cortex for actions that depend on the value of the goal in action–outcome learning. For this route, the orbitofrontal cortex implements stimulus–reinforcer association learning, and the anterior cingulate cortex action–outcome learning (where the outcome refers to receiving or not receiving a reward or punisher). A fifth level in the hierarchy is from the orbitofrontal cortex [and much less the amygdala (Rolls et al. 2023a)] via multiple step reasoning systems involving syntax and language. Processing at this fifth level may be related to explicit conscious states. The fifth level may also allow some top-down control of emotion-related states in the orbitofrontal cortex by the explicit processing system. Pallidum/SN—the globus pallidus and substantia nigra

may be elicited by the stimulus, as the behaviour has now become automated as a stimulus–response habit. For this type of learning, if the reward is devalued outside the situation, then the very next time that the stimulus is offered, the automated response is likely to be performed, providing evidence that the behaviour is no longer being guided by the reward value of the stimulus. The dopamine system is involved in this type of rather slow habit-based learning, it is thought by providing an error signal to the striatum which implements this type of habit learning (Schultz 2016c, b, 2017). The dopamine system probably receives its inputs from the orbitofrontal cortex (Rolls 2017; Rolls et al. 2023d). These brain systems are considered further below.

Other functions of emotion include the elicitation of autonomic responses, via pathways for example from the orbitofrontal cortex to the anteroventral visceral/autonomic insula and to the subgenual cingulate cortex (Critchley and Harrison 2013; Rolls 2013b, 2014b, 2019b, a; Quadt et al. 2022).

Stimuli can elicit behaviours in a number of ways via different routes to action in primates including humans, as

shown in Fig. 2. An important point made by Fig. 2 is that there are multiple routes to output including to action that can be produced by stimuli that produce emotional states. Here emotional states are the states elicited by reward and punishing/non-reward stimuli, as illustrated in Fig. 1. The multiple routes are organized in a set of hierarchies, with each level in the system added later in evolution, but with all levels left in operation over the course of evolution (Rolls 2016c). The result of this is that a response such as an autonomic response to a stimulus that happens also to be rewarding might be produced by only the lower levels of the system operating, without necessarily the highest e.g. explicit levels being involved. The lowest levels in the hierarchy illustrated in Fig. 2 are involved in reflexes, including for example reflex withdrawal of a limb to a nociceptive stimulus, and autonomic responses. The second level in the hierarchy can produce learned autonomic and some other behavioural responses to for example a previously neutral visual or auditory stimulus after it has been paired with a nociceptive stimulus or with a good taste stimulus. This route involves stimulus–reinforcer learning in the amygdala

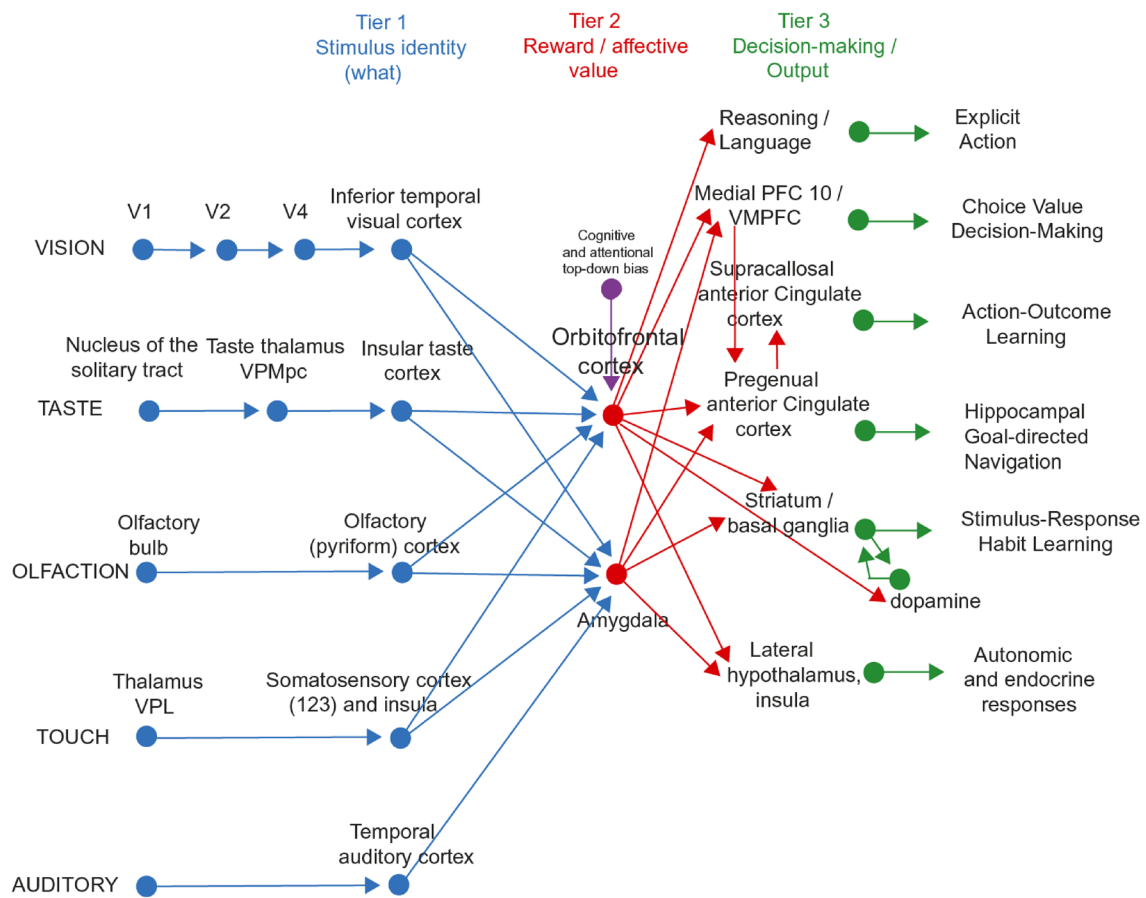


Fig. 3 The systems level organization of the brain for emotion in primates including humans. In Tier 1, representations are built of visual, taste, olfactory and tactile stimuli that are independent of reward value and therefore of emotion. In Tier 2, reward value and emotion are represented. A pathway for top-down attentional and cognitive modulation of emotion is shown in purple. In Tier 3 actions are learned in the supracallosal (or dorsal) anterior cingulate cortex to obtain the reward values signaled by the orbitofrontal cortex and amygdala that are relayed in part via the pregenual anterior cingulate cortex and vmPFC. Decisions between stimuli of different reward value can be taken in the ventromedial prefrontal cortex, vmPFC. In Tier 3, orbitofrontal cortex inputs to the reasoning/language sys-

tems enable affective value to be incorporated and reported. In Tier 3, stimulus–response habits can also be produced using reinforcement learning. In Tier 3 autonomic responses can also be produced to emotion-provoking stimuli. Auditory inputs also reach the amygdala. V1—primary visual (striate) cortex; V2 and V4—further cortical visual areas. PFC—prefrontal cortex. The Medial PFC area 10 is part of the ventromedial prefrontal cortex (vmPFC). VPL—ventro-postero-lateral nucleus of the thalamus, which conveys somatosensory information to the primary somatosensory cortex (areas 1, 2 and 3). VPMpc—ventro-postero-medial nucleus pars parvocellularis of the thalamus, which conveys taste information to the primary taste cortex

and orbitofrontal cortex. A third level in the hierarchy shown in Fig. 2 is the route from the orbitofrontal cortex and amygdala via the basal ganglia especially the ventral striatum to produce implicit stimulus–response habits. A fourth level in the hierarchy that is important in emotion is from especially

the orbitofrontal cortex to the anterior cingulate cortex for goal-directed action. The emotional states implemented at this level may not necessarily be conscious. A fifth level in the hierarchy shown in Fig. 2 is from the orbitofrontal cortex [and much less the amygdala (Rolls et al. 2023a)] via

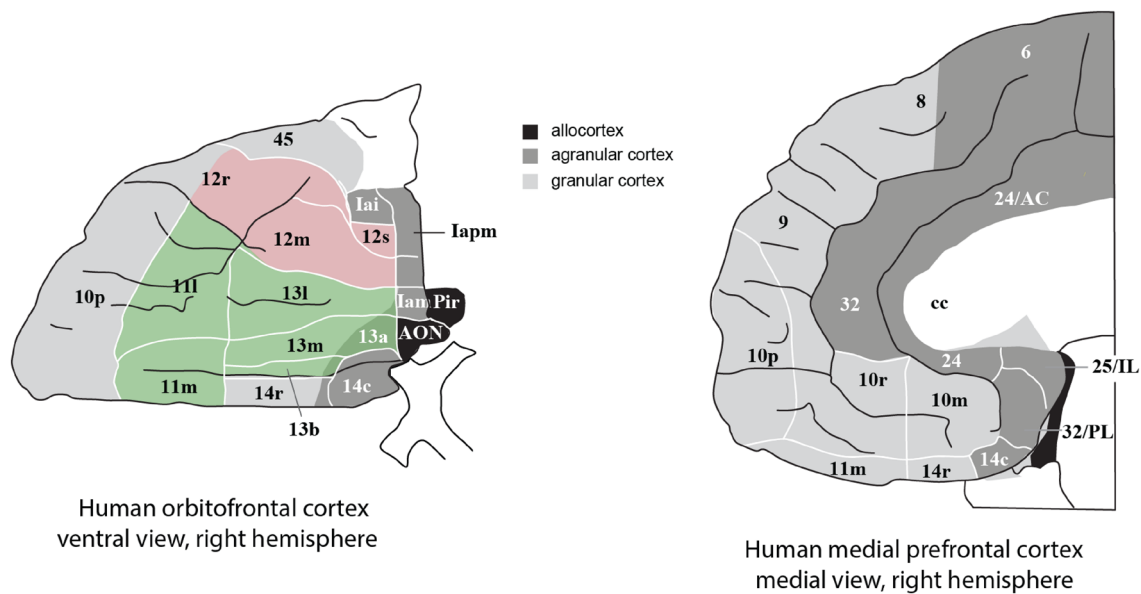


Fig. 4 Maps of architectonic areas in the orbitofrontal cortex (left, ventral view of the brain) and medial prefrontal cortex (right, medial view of the brain) of humans. Left: the medial orbitofrontal cortex includes areas 13 and 11 (green). The lateral orbitofrontal cortex includes area 12 (red). (Area 12 is sometimes termed area 12/47 in humans. The figure shows three architectonic subdivisions of area 12.) Almost all of the human orbitofrontal cortex except area 13a is granular. Agranular cortex is shown in dark grey. The part of area 45

shown is the orbital part of the inferior frontal gyrus pars triangularis. Right: the anterior cingulate cortex includes the parts shown of areas 32, 25 (subgenual cingulate), and 24. The ventromedial prefrontal cortex includes areas 14 (gyrus rectus) 10m and 10r. AON—anterior olfactory nucleus; Iai, Ial, Iam, Iapm—subdivisions of the agranular insular cortex. (After Öngür et al. (2003) *Journal of Comparative Neurology* with permission of John Wiley & Sons, Inc., modified from a redrawn version by Passingham and Wise (2012).)

multiple step reasoning systems involving syntax and language, which can be associated with explicit conscious states (especially I argue if a higher order syntactic thought system for correcting lower order thoughts is involved (Rolls 2008, 2014b, 2020, 2023d), see "A reasoning, rational, route to action"). It is emphasized that each of these types of output have adaptive value in preparing individuals to deal physiologically and behaviourally with what may generally be described as emotion-provoking events.

The neuroscience of emotion in humans and other primates

A framework for understanding the neuroscience of emotion in humans and other primates

A framework is shown in Fig. 3, and is built on evidence from neuronal recordings, the effects of brain damage, and fMRI in humans and macaques some of which is summarized below (Rolls 2014b, 2018, 2019a, 2021b, 2023d; Rolls et al. 2020b). Part of the evidence for what is shown in Fig. 3 comes from reward devaluation, in which when the reward value is changed, for example by feeding to satiety, neural

responses to stimuli are little affected in Tier 1, but decrease to zero in Tier 2. Part of the evidence comes from the learning of associations between stimuli and reward value, which occurs mainly in Tier 2. Part of the evidence comes from the effects of brain damage on emotion, which occur primarily after damage to the orbitofrontal cortex and amygdala in Tier 2, and the cingulate cortex in Tier 3 (Rolls 2021c). The organization of reward value processing and therefore emotion in the rodent brain is very different (Rolls 2019b, 2021b, 2023d), and a brief summary about this is provided in "Brain systems for emotion and motivation in primates including humans compared to those in rodents".

In the context of what is shown in Fig. 3, the focus next is on key brain areas involved in emotion in humans and other primates, the orbitofrontal cortex, anterior cingulate cortex, and amygdala.

The orbitofrontal cortex

The connections and connectivity of the orbitofrontal cortex

The orbitofrontal cortex cytoarchitectonic areas of the human brain are shown in Fig. 4 (left). The medial

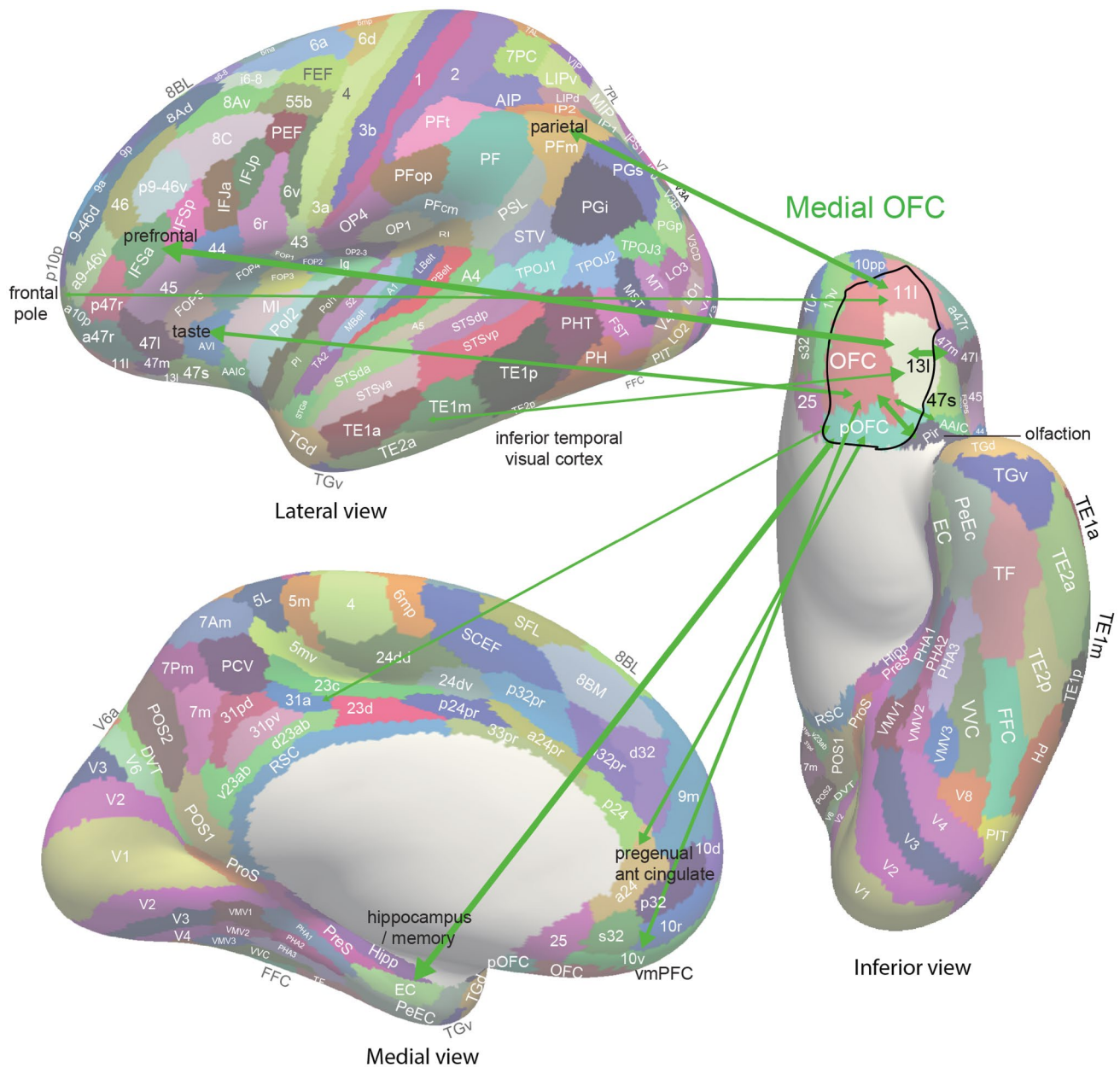


Fig. 5 Summary of the effective connectivity of the human medial orbitofrontal cortex. The medial orbitofrontal cortex has taste, olfactory and inferior temporal visual cortex inputs, and connectivity with the hippocampus, pregenual anterior cingulate cortex, ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (e.g. 31), parietal cortex, inferior prefrontal cortex, and frontal pole. The main

regions with which the medial OFC has connectivity are indicated by names with the words in black font. The width of the arrows and the size of the arrow heads in each direction reflects the strength of the effective connectivity. The abbreviations are listed in Rolls et al. (2023d)

orbitofrontal cortex includes areas 13 and 11 (Öngür et al. 2003). The lateral orbitofrontal cortex includes area 12 (sometimes in humans termed 12/47) (Öngür et al. 2003). The anterior cingulate cortex includes the parts shown in Fig. 4 (right) of areas 32, 25 (subgenual cingulate), and 24 (see also Figs. 5 and 6). The ventromedial prefrontal cortex includes areas 14 (gyrus rectus), 10m and 10r.

Some of the main connections of the orbitofrontal cortex in primates are shown schematically in Fig. 3 (Carmichael and Price 1994, 1995; Barbas 1995, 2007; Petrides and Pandya 1995; Pandya and Yeterian 1996; Öngür and Price 2000; Price 2006, 2007; Saleem et al. 2008; Mackey and Petrides 2010; Petrides et al. 2012; Saleem et al. 2014; Henssen et al. 2016; Rolls 2017, 2019d, b, Rolls et al. 2020b). The orbitofrontal cortex receives inputs from the ends of every

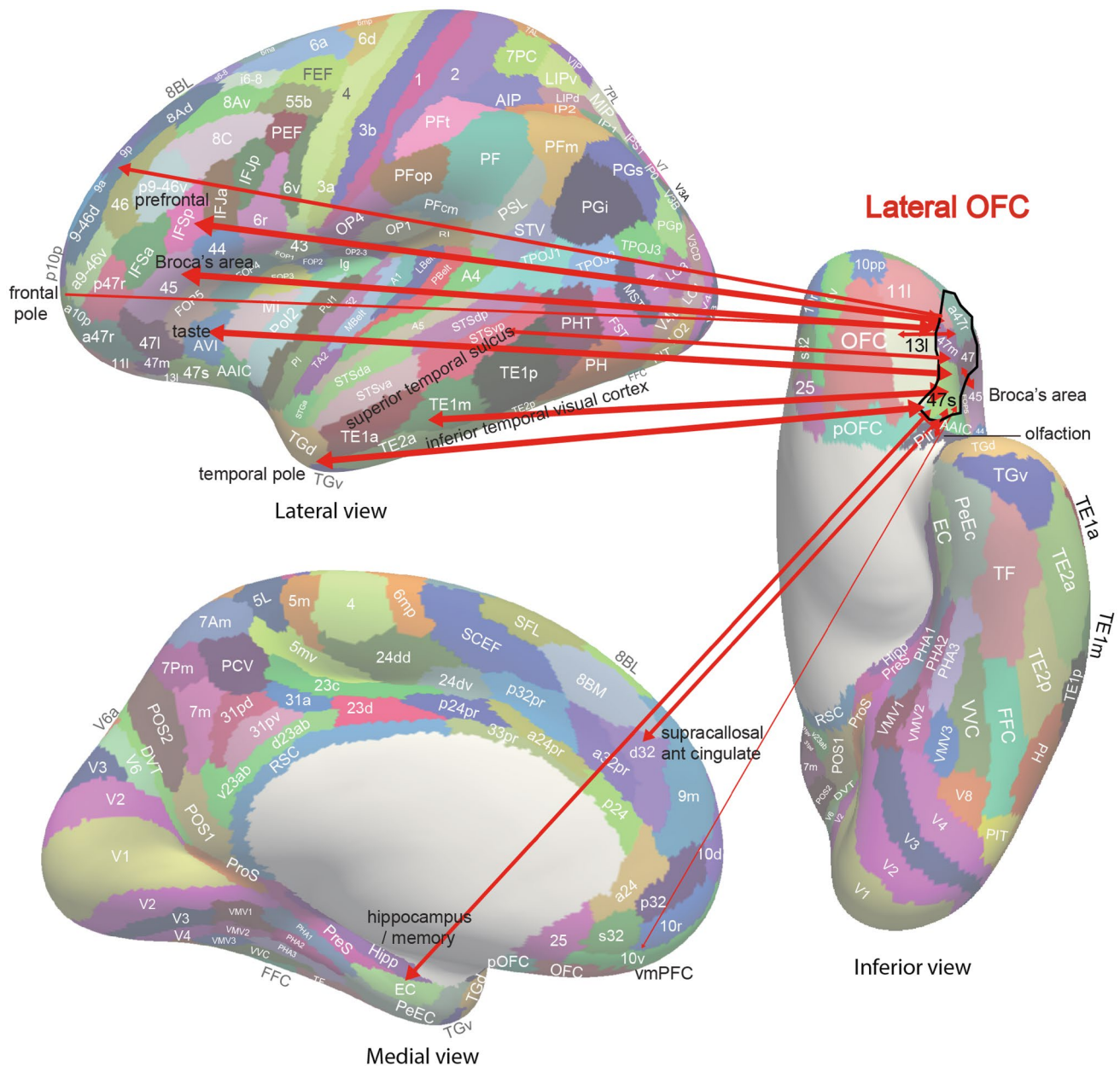


Fig. 6 Summary of the effective connectivity of the human lateral orbitofrontal cortex. The lateral orbitofrontal cortex has taste, olfactory and inferior temporal visual cortex inputs, and connectivity with the hippocampus, supracallosal (dorsal) anterior cingulate cortex, inferior and dorsolateral prefrontal cortex, and frontal pole. However, the lateral OFC also has connectivity with language regions (the

cortex in the superior temporal sulcus and Broca's area). The main regions with which the lateral OFC has connectivity are indicated by names with the words in black font. The width of the arrows and the size of the arrow heads in each direction reflects the strength of the effective connectivity. The abbreviations are listed in Rolls et al. (2023d)

ventral cortical stream that processes the identity of visual, taste, olfactory, somatosensory, and auditory stimuli (Rolls 2019b, 2023d). At the ends of each of these cortical processing streams, the identity of the stimulus is represented independently of its reward value (Rolls 2023d). This is shown by neuronal recordings in primates (Rolls 2019b). For example, the inferior temporal cortex represents objects and faces independently of their reward value as shown by visual

discrimination reversals, and by devaluation of reward tests by feeding to satiety (Rolls et al. 1977; Rolls 2012c, 2016c, 2019b). Similarly, the insular primary taste cortex represents what the taste is independently of its reward value (Yaxley et al. 1988; Rolls 2015, 2016d, 2019b, 2023d).

Outputs of the orbitofrontal cortex reach the anterior cingulate cortex, the striatum, the insula, and the inferior frontal gyrus (Rolls 2019a, 2023d; Rolls et al. 2023d),

and enable the reward value representations in the orbitofrontal cortex to influence behaviour (Fig. 3, green). The orbitofrontal cortex projects reward value outcome information (e.g. the taste of food) to the anterior cingulate cortex, where it is used to provide the reward outcomes for action–outcome learning (Rushworth et al. 2012; Rolls

a, 2019b, 2023d). The orbitofrontal cortex also projects expected reward value information (e.g. the sight of food) to the anterior cingulate cortex where previously learned actions for that goal can be selected. The orbitofrontal cortex projects reward-related information to the ventral striatum (Williams et al. 1993), and this provides a route, in part

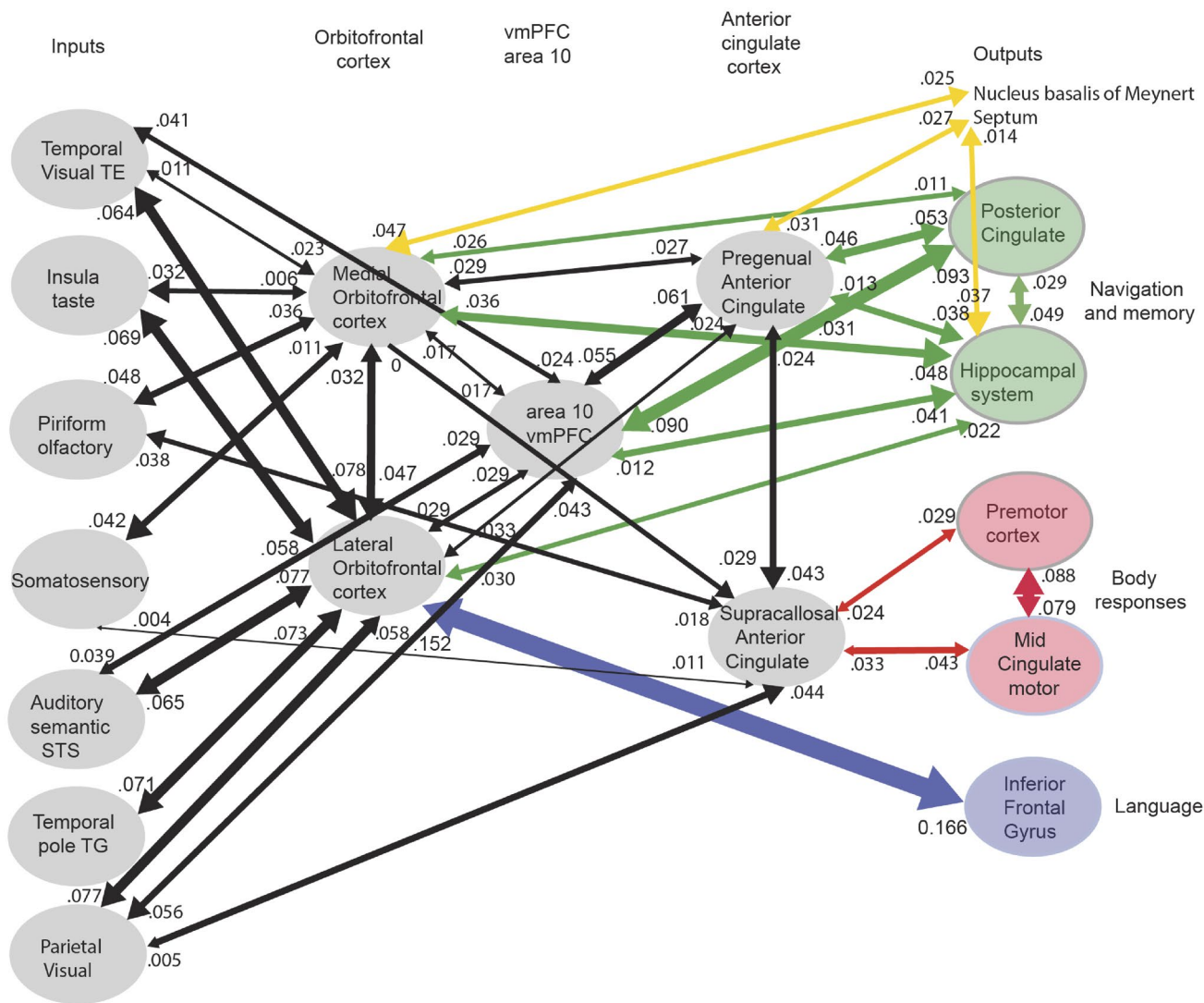


Fig. 7 Effective connectivity of the human orbitofrontal cortex, vmPFC, and anterior cingulate cortex shown in the middle, with inputs on the left and outputs on the right. The effective connectivity was measured in 171 participants imaged at 7 T by the Human Connectome Project, and was measured between the 360 cortical regions in the HCP-multimodal parcellation atlas (Glasser et al. 2016a), with subcortical regions using the HCPex atlas (Huang et al. 2022). The effective connectivity measures the effect in each direction between every pair of cortical regions, uses time delays to assess the directionality using a Hopf computational model which integrates the dynamics of Stuart–Landau oscillators in each cortical region, has a maximal value of 0.2, and is described in detail elsewhere (Rolls et al. 2022a; b, 2023d). The width of the arrows is proportional to the effective connectivity in the highest direction, and the size of the

arrowheads reflects the strength of the effective connectivity in each direction. The effective connectivities shown by the numbers are for the strongest link where more than one link between regions applies for a group of brain regions. Effective connectivities with hippocampal memory system regions are shown in green; with premotor/mid-cingulate regions in red; with the inferior prefrontal language system in blue; and in yellow to the basal forebrain nuclei of Meynert which contains cholinergic neurons that project to the neocortex and to the septal nuclei which contain cholinergic neurons that project to the hippocampus. The Somatosensory regions include 5 and parietal PF and PFop, which also connect to the pregenual anterior cingulate but are not shown for clarity; the Parietal regions include visual parietal regions 7, PGI and PFm. (From Rolls et al (2023d))

via the habenula, for reward-related information to reach the dopamine neurons (Rolls 2017), which respond inter alia to positive reward prediction error (Bromberg-Martin et al. 2010; Schultz 2016b). The striatal/basal ganglia route is used for stimulus–response, habit, learning (Everitt and Robbins 2013; Rolls 2014b, 2023d), with dopamine used to provide reward prediction error in reinforcement learning (Schultz 2016c; Cox and Witten 2019). As that system uses dopamine in reinforcement learning of stimulus–response habits, it is much less fast to learn than the orbitofrontal cortex (outcome) with anterior cingulate cortex (action) system for action–outcome goal-based learning, and for emotion (Rolls 2021b). The orbitofrontal cortex projects to the insula as an output pathway and includes a projection to the visceromotor cortex in the antero-ventral insula (Hasanpour et al. 2018; Quadt et al. 2022) that helps to account for why the insula is activated in some tasks in which the orbitofrontal cortex is involved (Rolls 2016d, 2019b, 2023d). This antero-ventral part of the insula (Quadt et al. 2022) is just ventral to the primary taste cortex, and has very strong connections in primates to (and probably from) the orbitofrontal cortex (Baylis et al. 1995). The orbitofrontal cortex also projects to the inferior frontal gyrus, a region that on the right is implicated in stopping behaviour (Aron et al. 2014).

New evidence on the connectivity of the orbitofrontal cortex in humans is shown in Figs. 5, 6, 7, based on measurements of effective connectivity between 360 cortical regions and 24 subcortical regions measured in 171 humans from the Human Connectome Project, and complemented with functional connectivity and diffusion tractography (Rolls et al. 2023d). Effective connectivity measures ‘causal’ effects (in that they take into account time delays) in each direction between every pair of brain regions. (Although time delays are a signature of causality, further evidence is needed to prove causality, such as interventions (Rolls 2021f, e).) The effective connectivities of the orbitofrontal cortex with other brain regions are summarised in Figs. 5, 6, 7 (Rolls et al. 2023a; d). The medial and lateral orbitofrontal cortex between them (and they have effective connectivity with each other) receive taste, somatosensory, olfactory, visual, and auditory inputs that are needed to build the reward and punishment value representations that are found in these regions but much less in the preceding cortical areas that provide these inputs (Rolls 2019d, 2019b, 2021a, 2023d). Taste and somatosensory inputs provide information about primary reinforcers or outcome value, and the orbitofrontal cortex contains visual and olfactory neurons that can learn and reverse in one trial the associations with primary reinforcers and so represent expected value (Thorpe et al. 1983). This is consistent with the schematic diagram in Fig. 3.

In more detail (Fig. 5) (Rolls et al. 2023a; d), parts of the medial orbitofrontal cortex (11l, 13l, OFC and pOFC, which are interconnected) have effective connectivity with

the taste/olfactory/visceral anterior agranular insular complex (AAIC); the piriform (olfactory) cortex; the entorhinal cortex (EC); the inferior temporal visual cortex (TE1p, TE2a, TE2p); superior medial parietal 7Pm; inferior parietal PF which is somatosensory (Rolls et al. 2023e, f); with parts of the posterior cingulate cortex (31pv, 7m, d23ab) related to memory (Rolls et al. 2023i); with the pregenual anterior cingulate cortex (s32, a24, p24, p32, d32) and much less with the supracallosal anterior cingulate cortex (only 33pr); with ventromedial prefrontal 10r, 10d and 9m; with the frontal pole (10pp, p10p, a10p); with lateral orbitofrontal cortex (47m, 47s, a47r); and dorsolateral prefrontal cortex (46 and a9-46v) (Rolls et al. 2023e). Medial orbitofrontal cortex regions also have effective connectivity directed towards the caudate nucleus and nucleus accumbens (Rolls et al. 2023d).

Also with some detail, the lateral orbitofrontal cortex areas a47r, p47r and 47m share generally similar effective connectivities (Fig. 6) (Rolls et al. 2023a; d) from the visual inferior temporal cortex (TE areas); from parts of the parietal cortex [PFm which receives visual and auditory object-level information and IP2 which is visuomotor (Rolls et al. 2023f)]; from the medial orbitofrontal cortex (11l, 13l, pOFC); from the inferior frontal gyrus regions including IFJ, IFS and BA45; from the dorsolateral prefrontal cortex (8Av, 8BL, a9-46v and p9-46v) implicated in short-term memory (Rolls 2023d; Rolls et al. 2023e); and from the frontal pole (a10p, p10p, 10pp) (Rolls et al. 2023a; d). 47m (which is relatively medial in this group) also has effective connectivity with the hippocampal system (Hipp, EC, perirhinal, and TF), and with ventromedial prefrontal region 10r; and with the frontal pole [10d, and 9m (Rolls et al. 2023c)]. The diffusion tractography provides in addition evidence for connections of these parts of the lateral orbitofrontal cortex with the anterior ventral insular region (AVI) and the frontal opercular areas FOP4 and FOP5 which include the insular primary taste cortex (Rolls 2015, 2016d; Rolls et al. 2023a, d); with the anterior agranular insular complex (AAIC) which may be visceral (Rolls 2016d) and also has taste-olfactory convergence (De Araujo et al. 2003a); with the middle insular region (MI) which is somatosensory (Rolls et al. 2023e); and with the piriform (olfactory) cortex.

The human orbitofrontal cortex has connectivity to the hippocampal memory/navigation system that is both direct, and via the ventromedial area 10 regions (10r, 10d, 10v and 9m), pregenual anterior cingulate cortex, and the memory-related parts of the posterior cingulate cortex (Fig. 7). It is proposed that this connectivity provides a key input about reward/punishment value for the hippocampal episodic memory system, adding to the ‘what’, ‘where’, and ‘when’ information that are also key components of episodic memory (Rolls 2022b; Rolls et al. 2023d). Damage to the vmPFC/anterior cingulate cortex system is likely to contribute to episodic memory impairments by impairing

a key component of episodic memory, the reward/punishment/emotional value component (Rolls 2022b; Rolls et al. 2023d). Moreover, the medial orbitofrontal cortex connects to the nucleus basalis of Meynert and the pregenual cingulate to the septum, and damage to these cortical regions may contribute to memory impairments by disrupting cholinergic influences on the neocortex and hippocampus (Rolls 2022b; Rolls et al. 2023d). Navigation is generally towards goals, usually rewards, and it is proposed that this connectivity provides the goals for navigation to the hippocampal system to enable the hippocampus to be involved in navigation towards goals (Rolls 2022b, 2023c; Rolls et al. 2023d).

Two regions of the lateral orbitofrontal cortex, 47l and 47s, are especially connected with language systems in the temporal pole, cortex in the superior temporal sulcus (STS), and inferior frontal gyrus including Broca's area 45 and 44 (Rolls et al. 2022a). This provides a route for subjective reports to be made about the pleasantness or unpleasantness of stimuli and events (Rolls 2023d).

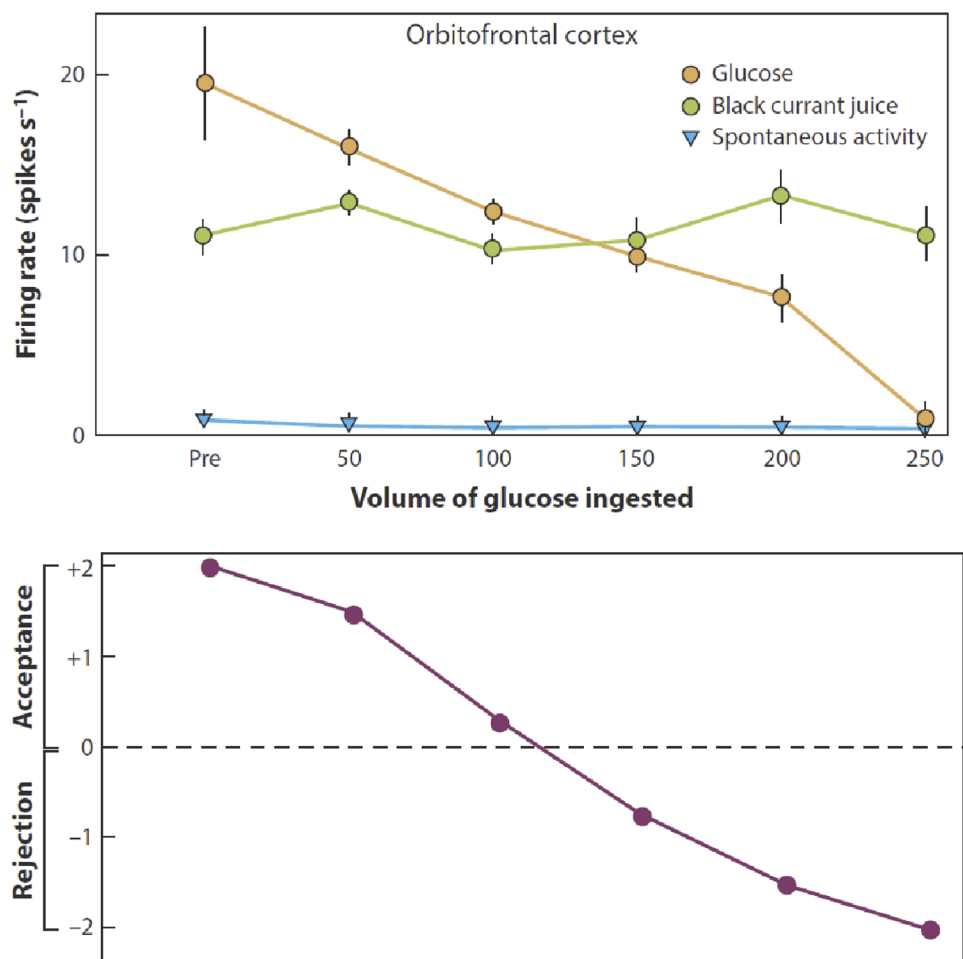
In the context that the anterior cingulate cortex is implicated in learning associations between actions and the

rewards or punishers associated with the actions (Noonan et al. 2011; Rushworth et al. 2012; Rolls 2019a, 2023d), the part of the anterior cingulate cortex that is most likely to be involved in action–outcome learning is the supracallosal (or dorsal) anterior cingulate cortex. That part has effective connectivity with somato-motor areas involved in actions, but which as shown in Fig. 7 receives inputs from the medial orbitofrontal cortex and pregenual anterior cingulate cortex that it is proposed provide the reward/punishment ‘outcome’ signals necessary for action–outcome learning (Rolls 2023d; Rolls et al. 2023d).

The human medial orbitofrontal cortex represents reward value

The primate including human orbitofrontal cortex is the first stage of cortical processing that represents reward value (red in Fig. 3) (Rolls 2019b, d, 2021b). For example, in devaluation experiments, taste, olfactory, visual, and oral texture neurons in the macaque orbitofrontal respond to food when hunger is present, and not after feeding to satiety when the food is no longer rewarding (Rolls et al. 1989; Critchley

Fig. 8 The effect of feeding to satiety with glucose solution on the responses (firing rate \pm s.e.m.) of a neuron in the orbitofrontal (secondary taste) cortex to the taste of glucose (open circles) and of blackcurrant juice (BJ). The spontaneous firing rate is also indicated (SA). Below the neuronal response data, the behavioural measure of the acceptance or rejection of the solution on a scale from +2 (strong acceptance) to -2 (strong rejection) is shown. The solution used to feed to satiety was 20% glucose. The monkey was fed 50 ml of the solution at each stage of the experiment as indicated along the abscissa, until he was satiated as shown by whether he accepted or rejected the solution. Pre is the firing rate of the neuron before the satiety experiment started. (Reproduced from Rolls et al. 1989, Copyright 1989 Society for Neuroscience.)



and Rolls 1996). An example of a devaluation experiment is shown in Fig. 8, which shows that as the value of the taste of glucose is reduced by feeding glucose to satiety, a typical orbitofrontal cortex neuron responding to the taste of food when it is rewarding at the start of the experiment gradually reduces its response to zero as the reward value reaches zero because glucose had been consumed. In fact, the experiment shows more than this, for the effect is relatively specific to the food eaten to satiety: there was little reduction of the firing rate to the flavour of fruit (black currant) juice after glucose had been fed to satiety. Correspondingly, the black currant juice was still rewarding after feeding to satiety with glucose (Fig. 8). Thus satiety is somewhat specific to the reward that has been received, and this is termed sensory-specific satiety. In fact, sensory-specific satiety was discovered when we were recording from lateral hypothalamic neurons responding to the taste and/or sight of food (Rolls et al. 1986). We traced back the computation to the orbitofrontal cortex, in which neurons show sensory-specific satiety to a primary reinforcer, the taste of food (Rolls et al. 1989), and to a secondary reinforcer, the sight and smell of food (Critchley and Rolls 1996). Devaluation effects are not found in the stages that provide taste information to the orbitofrontal cortex, the insular/opercular primary taste cortex (Rolls et al. 1988; Yaxley et al. 1988), nor in the brain region that provides visual inputs to the orbitofrontal cortex, the inferior temporal visual cortex (Rolls et al. 1977). This is some of the evidence on which Fig. 3 is based. The devaluation procedure has been adopted by others (Rudebeck et al. 2017; Murray and Rudebeck 2018; Murray and Fellows 2022).

This discovery of sensory-specific satiety has enormous implications, for it is proposed to apply to all rewards and to no punishers (Rolls 2014b, 2018, 2022a), and has the evolutionary adaptive value that behaviour switches from one reward to another. This ensures for example that a wide range of nutrients will be ingested [as we showed in experiments we performed with Oxford undergraduates after the neurophysiological discovery (Rolls et al. 1981a, b, c)] (though obesity is a resulting risk if a wide range of nutrients becomes easily available for humans) (Rolls 2016a); and more generally tends to promote reproductive success for the genes, in that a wide range of possible rewards will be explored (Rolls 2014b, 2018) (see "[Some implications and extensions of the understanding of emotion, motivation, and their brain mechanisms](#)"). Sensory-specific satiety is thus a key factor in emotion.

Further evidence that reward value is represented in the orbitofrontal cortex is that in visual discrimination reversal experiments, neurons in the macaque orbitofrontal cortex reverse the visual stimulus to which they respond in as little as one trial when the reward vs punishment taste received as an outcome for the choice reverses (Thorpe et al. 1983; Rolls et al. 1996). This is rule-based reversal, in that after a

previously rewarded visual stimulus is no longer rewarded, the macaques choose the other stimulus on the very next trial, even though its previous reward association was with punishment, as illustrated in Fig. 10c which also illustrates a non-reward neuron active at the time of the reversal (Thorpe et al. 1983). (Non-reward refers here to not obtaining an expected reward.) This capability requires a rule to be held in memory and reversed by non-reward (Deco and Rolls 2005c; Rolls and Deco 2016) (which is described as model-based), is very appropriate for primates including humans who in social situations may benefit from being very responsive to non-reward vs reward signals, and may not occur in rodents (Rolls 2019b, 2021b; Hervig et al. 2020). The macaque orbitofrontal cortex also contains visual neurons that reflect face expression and face identity (both necessary to decode the reward/punishment value of an individual) (Thorpe et al. 1983; Rolls et al. 2006), and also social categories such as young faces (Barat et al. 2018). Information about face expression and movements important in social communication probably reaches the orbitofrontal cortex from neurons we discovered in the cortex in the macaque superior temporal sulcus that respond to these stimuli (Hasselmo et al. 1989a, b), in what is a region now accepted as important for decoding visual stimuli relevant to social behaviour (Pitcher et al. 2019; Pitcher and Ungerleider 2021). Economic value is represented in the orbitofrontal cortex, in that for example single neurons reflect the trade-off between the quality of a reward and the amount that is available (Padoa-Schioppa and Cai 2011; Padoa-Schioppa and Conen 2017). These investigations show that some orbitofrontal cortex neurons respond to outcome value (e.g. the taste of food), and others to expected value (of future rewards). The expected value neurons are not positive reward prediction error neurons, for they keep responding to the expected reward even when there is no prediction error (Rolls 2021b). Consistent with this, lesions of the macaque medial orbitofrontal cortex areas 13 and 11 make the animals less sensitive to reward value, as tested in devaluation experiments in which the animal is fed to satiety (Rudebeck et al. 2017). Neurotoxic lesions of the macaque orbitofrontal cortex produce effects that are difficult to interpret (Murray and Rudebeck 2018; Sallet et al. 2020), perhaps because these lesions have not always been based on knowledge of where neurons and activations related to reversal learning are found, and the difficulty of disabling all such orbitofrontal cortex neurons. Further, the tasks used in these studies are sometimes complicated, whereas a prototypical task is deterministic one-trial Go-NoGo rule-based visual discrimination reversal between the a visual stimulus and taste (Thorpe et al. 1983; Rolls et al. 1996), or in humans between a visual stimulus and winning or losing points or money (Rolls et al. 2020c). Rodents appear not to be able to perform this one-trial rule-based visual-reward reversal task (Hervig et al. 2020).

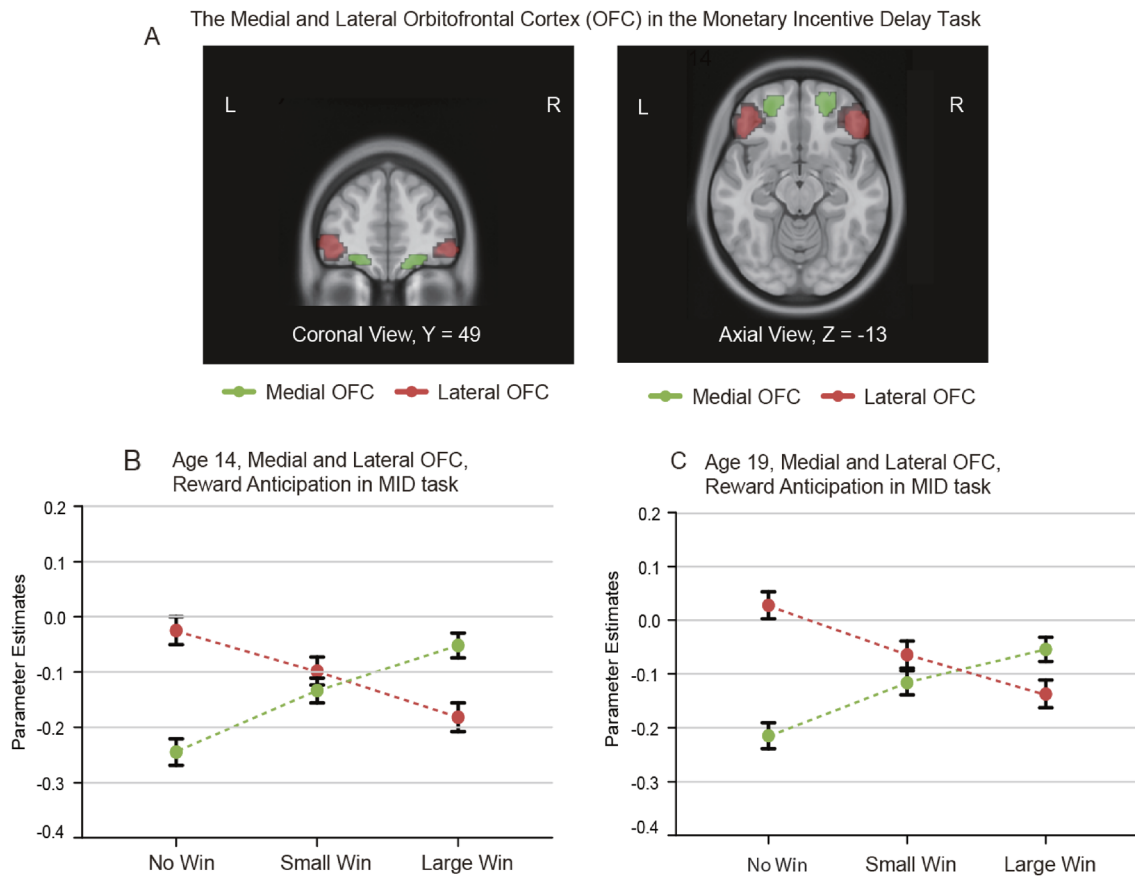


Fig. 9 The lateral orbitofrontal cortex is activated by not winning, and the medial orbitofrontal cortex by winning, in the monetary incentive delay task. The lateral orbitofrontal cortex region in which activations increased towards no reward (No Win) in the monetary incentive delay task are shown in red in 1140 participants at age 19 and in 1877 overlapping participants at age 14. The conditions were Large win (10 points) to Small Win (2 points) to No Win (0 points) (at 19; sweets were used at 14). The medial orbitofrontal cortex region in which activations increased with increasing reward from No Win to Small Win to High Win) is shown in green. The parameter estimates are shown from the activations for the par-

ticipants (mean \pm sem) with the lateral orbitofrontal in red and medial orbitofrontal cortex in green. The interaction term showing the sensitivity of the medial orbitofrontal cortex to reward and the lateral orbitofrontal cortex to non-reward was significant at $p=10^{-50}$ at age 19 and $p < 10^{-72}$ at age 14. In a subgroup with depressive symptoms as shown by the Adolescent Depression Rating Scale, it was further found that there was a greater activation to the No Win condition in the lateral orbitofrontal cortex; and the medial orbitofrontal cortex was less sensitive to the differences in reward value. (Modified from Xie et al. 2021.)

Neuroimaging experiments in humans produce consistent evidence (De Araujo et al. 2003a; Kringelbach et al. 2003; Kringelbach and Rolls 2003; Grabenhorst and Rolls 2008; Grabenhorst et al. 2008a), and allow the types of reward to be extended to include monetary reward (O'Doherty et al. 2001; Xie et al. 2021), face expressions (Kringelbach and Rolls 2003), and face beauty (O'Doherty et al. 2003). Further, in humans activations of the medial orbitofrontal cortex are linearly related to the subjective (conscious) pleasantness of stimuli (Grabenhorst and Rolls 2011; Rolls 2019b). These reward-related effects are found for odors (Rolls et al. 2003b), flavor (De Araujo et al. 2003a; Kringelbach et al. 2003), pleasant touch (Rolls et al. 2003c; McCabe et al. 2008), monetary reward (O'Doherty et al. 2001; Xie et al.

2021), and amphetamine (Völlm et al. 2004). A recent study with 1140 participants emphasizes these points, by showing that the medial orbitofrontal cortex is activated by reward [such as winning money or candies], and that the lateral orbitofrontal cortex is activated by not winning (Fig. 9 (Xie et al. 2021)).

Further, humans with orbitofrontal cortex lesions may also be less sensitive to reward, as shown by their reduced subjective emotional feelings (Hornak et al. 2003), and their difficulty in identifying face and voice emotion-related expressions, which are important for emotional and social behaviour (Hornak et al. 1996, 2003).

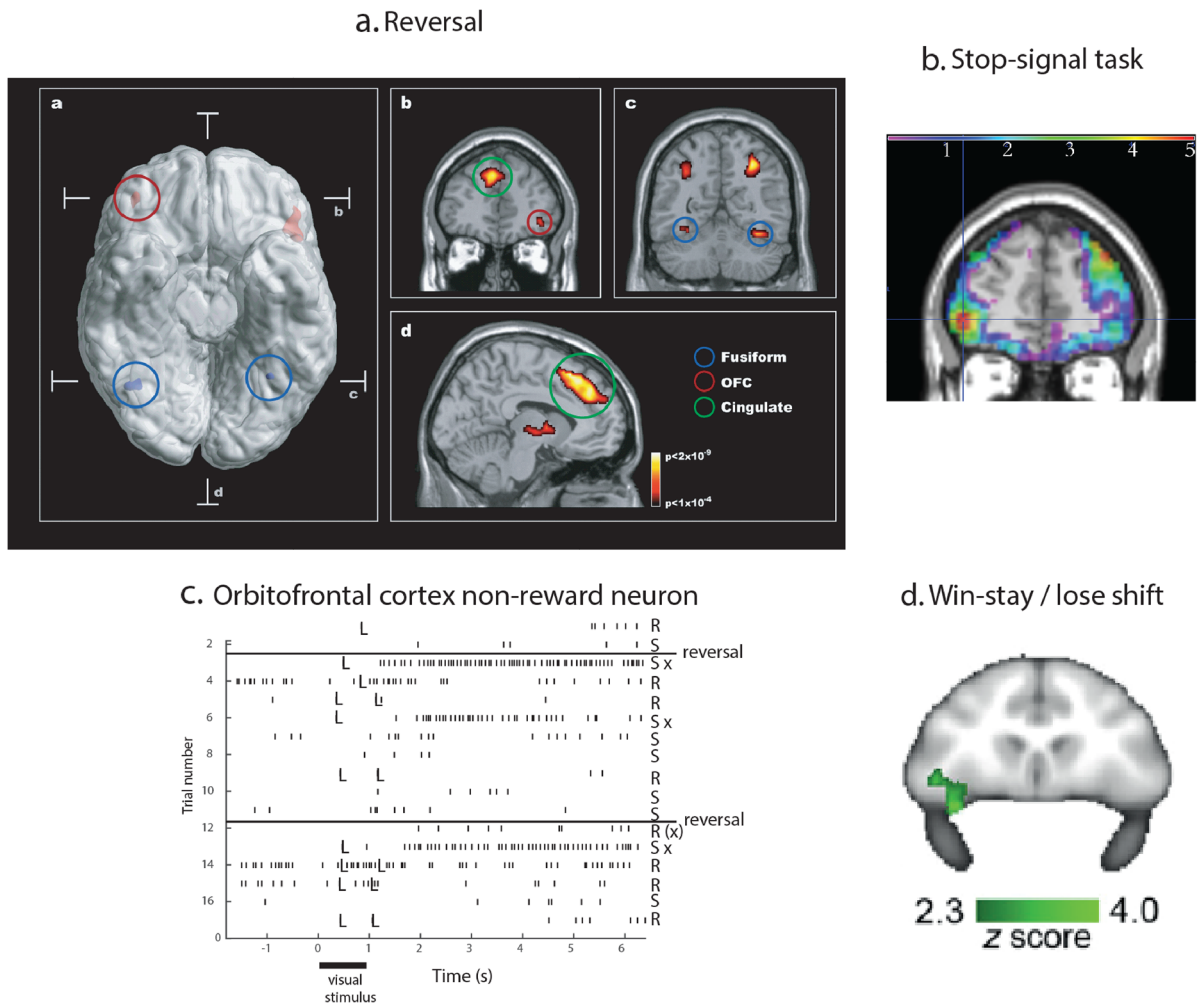


Fig. 10 **a** Evidence that the human lateral orbitofrontal cortex is activated by non-reward. Activation of the lateral orbitofrontal cortex in a visual discrimination reversal task on reversal trials, when a face was selected but the expected reward was not obtained, indicating that the subject should select the other face in future to obtain the reward. **a** A ventral view of the human brain with indication of the location of the two coronal slices (**b**, **c**) and the transverse slice (**d**). The activations with the red circle in the lateral orbitofrontal cortex (OFC, peaks at [42 42 – 8] and [– 46 30 – 8]) show the activation on reversal trials compared to the non-reversal trials. For comparison, the activations with the blue circle show the fusiform face area produced just by face expressions, not by reversal, which are also indicated in the coronal slice in **c**. **b** A coronal slice showing the activation in the right orbitofrontal cortex on reversal trials. Activation is also shown in the supracallosal anterior cingulate region (Cingulate, green circle) that is also known to be activated by many punishing, unpleasant, stimuli (see Grabenhorst and Rolls (2011)). (From NeuroImage 20 (2), Morten L. Kringelbach and Edmund T. Rolls, Neural correlates of rapid reversal learning in a simple model of human social interaction, pp. 1371–83, Copyright, 2003, with permission from Elsevier.). **b** Activations in the human lateral orbitofrontal cortex are related to a signal to change behaviour in the stop-signal task. In the task, a left or right arrow on a screen indicates which button to touch. However on some trials, an up-arrow then appears, and the participant must change the behaviour, and stop the response. There is a larger response on trials on which the participant successfully changes the behaviour and stops the response, as shown by the contrast stop-success–stop-failure, in the ventrolateral prefrontal cortex in a region including the lateral orbitofrontal cortex, with peak at [– 42 50

– 2] indicated by the cross-hairs, measured in 1709 participants. There were corresponding effects in the right lateral orbitofrontal cortex [42 52 – 4]. Some activation in the dorsolateral prefrontal cortex in an area implicated in attention is also shown. (After Deng, Rolls et al. 2016). **c** Non-reward error-related neurons maintain their firing after non-reward is obtained. Responses of an orbitofrontal cortex neuron that responded only when the macaque licked to a visual stimulus during reversal, expecting to obtain fruit juice reward, but actually obtained the taste of aversive saline because it was the first trial of reversal (trials 3, 6, and 13). Each vertical line represents an action potential; each L indicates a lick response in the Go-NoGo visual discrimination task. The visual stimulus was shown at time 0 for 1 s. The neuron did not respond on most reward (R) or saline (S) trials, but did respond on the trials marked S x, which were the first or second trials after a reversal of the visual discrimination on which the monkey licked to obtain reward, but actually obtained saline because the task had been reversed. The two times at which the reward contingencies were reversed are indicated. After responding to non-reward, when the expected reward was not obtained, the neuron fired for many seconds, and was sometimes still firing at the start of the next trial. It is notable that after an expected reward was not obtained due to a reversal contingency being applied, on the very next trial the macaque selected the previously non-rewarded stimulus. This shows that rapid reversal can be performed by a non-associative process, and must be rule-based. (After Thorpe et al. 1983.) **d** Bold signal in the macaque lateral orbitofrontal related to win-stay/lose-shift performance, that is, to reward reversal performance. (After Chau et al 2015)

The human lateral orbitofrontal cortex represents punishers and non-reward, and is involved in changing emotional behaviour

The macaque orbitofrontal cortex has neurons that respond when an expected reward is not received (Thorpe et al. 1983), and these have been termed *non-reward neurons* (Rolls 2014b, 2019b, d, 2021b) (see example in Fig. 10c). They can be described as negative reward prediction error neurons, in that they respond when a reward outcome is less than was expected (Rolls 2019b). These neurons do not respond to expected punishers [e.g. the discriminative stimulus for saline in Fig. 10c (Thorpe et al. 1983)], but other neurons do respond to expected punishers (Rolls et al. 1996), showing that non-reward and punishment are represented by different neurons in the orbitofrontal cortex. The finding of non-reward neurons is robust, in that 18/494 (3.6%) of the neurons in the original study responded to non-reward (Thorpe et al. 1983), consistent results were found in different tasks in a complementary study (10/140 non-reward neurons in the orbitofrontal cortex or 7.1%) (Rosenkilde et al. 1981), and an fMRI study has shown that the macaque lateral orbitofrontal cortex is activated when an expected reward is not obtained during reversal (Chau et al. 2015) (Fig. 10d). The hypothesis is that the non-reward responsiveness of these neurons is computed in the orbitofrontal cortex, because this is the first brain region in primates at which expected value and outcome value are represented, as summarized in Fig. 3 and with the evidence set out fully by Rolls (2019b, 2021b, 2023d), and these two signals are those required to compute non-reward, that is, that reward outcome is less than the expected value.

Corresponding to this, the human lateral orbitofrontal cortex is activated when a reward is not obtained in a visual discrimination reversal task (Kringelbach and Rolls 2003) (Fig. 10a), and when money is not received in a monetary reward task (O'Doherty et al. 2001; Xie et al. 2021), and in a one-trial reward reversal task (Rolls et al. 2020c). Further, the human lateral orbitofrontal cortex is also activated by punishing, subjectively unpleasant, stimuli (Grabenhorst and Rolls 2011; Rolls 2019b, d, 2021b). Examples include unpleasant odors (Rolls et al. 2003b), pain (Rolls et al. 2003c), losing money (O'Doherty et al. 2001), and receiving an angry face expression indicating that behaviour should change in a reversal (Kringelbach and Rolls 2003). The human right lateral orbitofrontal cortex/inferior frontal gyrus is also activated when behavioural correction is required in the stop-signal task (Fig. 10b) (Aron et al. 2014; Deng et al. 2017). These discoveries show that one way in which the orbitofrontal cortex is involved in decision-making and emotion is by representing rewards, punishers, and errors made during decision-making. This is supported by

the problems that orbitofrontal cortex damage produces in decision-making, which including failing to respond correctly to non-reward, as described next.

Consistent with this neurophysiological and neuroimaging evidence, lesions of the orbitofrontal cortex can impair reward reversal learning during decision-making in humans (Rolls et al. 1994; Hornak et al. 2004; Fellows 2011), who continue responding to the previously rewarded, now non-rewarded, stimulus. The change in contingency between the stimulus and reward vs non-reward is not processed correctly. In macaques, damage to the lateral orbitofrontal cortex impairs reversal and extinction (Butter 1969; Iversen and Mishkin 1970), and damage of the lateral orbitofrontal cortex area 12 extending around the inferior convexity impaired the ability to make choices based on whether reward vs non-reward had been received (Rudebeck et al. 2017; Murray and Rudebeck 2018). Further evidence that the lateral orbitofrontal cortex is involved in learning contingencies between stimuli and reward vs non-reward is that in humans, lateral orbitofrontal cortex damage impaired this type of 'credit assignment' (Noonan et al. 2017). This type of flexibility of behaviour is important in primate including human social interactions, and indeed many of the effects of damage to the human orbitofrontal cortex, including the difficulty in responding appropriately to the changed circumstances of the patient, and the changed personality including impulsivity, can be related to these impairments in responding to non-reward and punishers (Rolls et al. 1994; Berlin and Rolls 2004; Berlin et al. 2004; Hornak et al. 2004; Rolls 2018, 2019b, d, 2021c, b; Rolls et al. 2020b).

The ventromedial prefrontal cortex and reward-related decision-making

The ventromedial prefrontal cortex (vmPFC, which can be taken to include the gyrus rectus area 14 and parts of 10m and 10r, Fig. 4) receives inputs from the orbitofrontal cortex, and has distinct connectivity (with strong functional connectivity with the superior medial prefrontal cortex, cingulate cortex, and angular gyrus Du et al. 2020; Rolls et al. 2023d). The vmPFC has long been implicated in reward-related decision-making (Bechara et al. 1997, 2005; Glascher et al. 2012), this region is activated during decision-making contrasted with reward valuation (Grabenhorst et al. 2008b; Rolls and Grabenhorst 2008), and it has the signature of a decision-making region of increasing its activation in proportion to the difference in the decision variables, which correlates with decision confidence (Rolls et al. 2010a, b; Rolls 2019b, 2021b). Consistently, in macaques single neurons in the ventromedial prefrontal cortex rapidly come to signal the value of the chosen offer, suggesting that this vmPFC system serves to produce a choice (Strait et al.

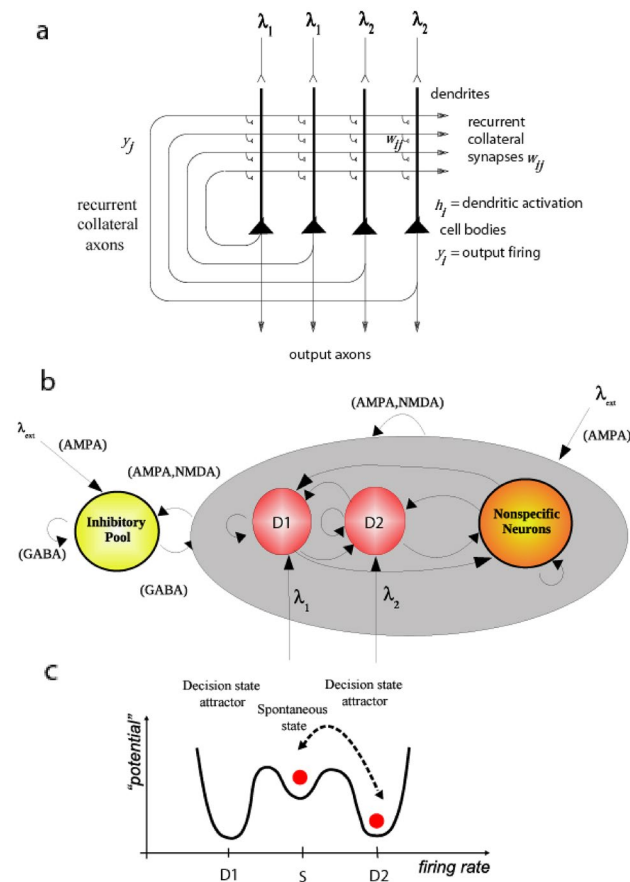


Fig. 11 **a** Attractor or autoassociation single network architecture for decision-making. The evidence for decision 1 is applied via the λ_1 inputs, and for decision 2 via the λ_2 inputs. The synaptic weights have been associatively modified during training in the presence of λ_1 and at a different time of λ_2 . When λ_1 and λ_2 are applied, each attractor competes through the inhibitory interneurons (not shown), until one wins the competition, and the network falls into one of the high firing rate attractors that represents the decision. The noise in the network caused by the random spiking of the neurons means that on some trials, for given inputs, the neurons in the decision 1 (D1) attractor are more likely to win, and on other trials the neurons in the decision 2 (D2) attractor are more likely to win. This makes the decision-making probabilistic, for, as shown in **c**, the noise influences when the system will jump out of the spontaneous firing stable (low energy) state S, and whether it jumps into the high firing state for decision 1 (D1) or decision 2 (D2). **b** The architecture of the integrate-and-fire network used to model decision-making (see text). **c** A multistable “effective energy landscape” for decision-making with stable states shown as low “potential” basins. Even when the inputs are being applied to the network, the spontaneous firing rate state is stable, and noise provokes transitions into the high firing state decision attractor state D1 or D2 (see Rolls and Deco 2010; Rolls et al. 2010a, b; Rolls 2021b)

2014), also consistent with the attractor model of decision-making (Rolls and Deco 2010; Rolls et al. 2010a, b; Rolls 2014b, 2016c, 2021b).

The attractor model of decision-making is a neuronal network with associatively modifiable recurrent collateral

synapses between the neurons of the type prototypical of the cerebral cortex (Wang 2002; Rolls and Deco 2010; Rolls 2021b) (see Fig. 11). The decision variables (the inputs between which a decision needs to be made) are applied simultaneously, and the network, after previous training with these decision variables, reaches a state where the population of neurons representing one of the decision variables has a high firing rate (Rolls and Deco 2010; Deco et al. 2013; Rolls 2016c, 2021b). There is noise or randomness in this model of decision-making that is related to the approximately Poisson distributed firing times of neurons for a given mean firing rate. This approach to decision-making (see also Rolls et al. 2010a, b), illustrated in Fig. 11, provides a much more biologically well-founded model with integrate-and-fire neurons coupled in an attractor network than accumulator models of decision-making in which noise is added to two variables to see which one wins (Deco et al. 2013; Shadlen and Kiani 2013).

A key conceptual point can be made here about reward-related decision-making, which will typically be between two or more rewards. The inputs (decision variables, λ_1 and λ_2 in Fig. 11) that drive each of the reward attractor neuronal populations in Fig. 11, need to produce as output the identity of the reward signal, so that behaviour can be directed towards making these goal neurons fire. Effectively, the two sets of output neurons in Fig. 11, each driven by λ_1 and λ_2 in Fig. 11, are the reward neurons, competing with each other through the inhibitory interneurons. This results in the output of the decision-making network being the identity of the reward that won, and that can be used as the goal for behaviour. It is not useful to have a common currency for reward, if common currency means some general reward representation (Cabanac 1992). Instead, the output of the decision-making needs to be the specific reward that won in the computation, and the fact that this is an attractor network provides a way to maintain the firing of the winning neurons so that they can continue firing to act as the goal for the motivated behaviour (Rolls 2014b, 2021b). To place this in the context of emotion: each pleasure associated with each type of reward (with examples in Table 1) must be different, and feel different, so that we know that we have been successful in obtaining the correct reward that was being sought. Of course, having different rewards on the same scale of magnitude is useful, so that the decision-making network weights the two inputs on the same reward value scale (Grabenhorst et al. 2010a).

The amygdala

The amygdala in rodents, in which the orbitofrontal cortex is much less developed than in primates (Passingham and Wise 2012; Passingham 2021), has been implicated in emotion-related responses such as a conditioned autonomic responses,

conditioned freezing behavior, cortical arousal, and learned incentive effects in fear conditioning in which an auditory tone is associated with foot shock (LeDoux 1995, 1996; Quirk et al. 1996). Synaptic modification in the amygdala is implicated in the learning of these types of response (Davis 1992, 1994; Davis et al. 1995; Rogan et al. 1997; LeDoux 2000a, b; Davis 2011). In macaques, bilateral lesions of the amygdala impair the learning of fear-potentiated startle to a visual cue (Antoniadis et al. 2009). In macaques, connections reach the lateral and basal amygdala 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 (Van Hoesen 1981; Amaral et al. 1992; Ghashghaei and Barbas 2002; Freese and Amaral 2009). The visual and auditory inputs from these cortical regions may be associated in the primate amygdala with primary reinforcers such as taste from the anterior insular primary taste cortex, and with touch and nociceptive input from the insular somatosensory cortex (Leonard et al. 1985; Rolls 2000c, 2014b; Kadohisa et al. 2005a, b; Wilson and Rolls 2005; Rolls et al. 2018). The outputs of the primate amygdala include connections to the hypothalamus, autonomic centres in the medulla oblongata, and ventral striatum (Heimer et al. 1982; Amaral et al. 1992; Freese and Amaral 2009; Rolls 2014b). In addition, the monkey amygdala has direct projections back to many areas of the temporal, orbitofrontal, and insular cortices from which it receives inputs (Amaral et al. 1992), including even V1 (Freese and Amaral 2005), and to the hippocampal system (Stefanacci et al. 1996). In addition, different fMRI responses of the macaque inferior temporal cortex to different face expressions were reduced after amygdala lesions (Hadj-Bouziane et al. 2012).

Although the primate amygdala thus has some of the same connections as the orbitofrontal cortex in monkeys (see Fig. 3) (Rolls 2014b, 2023d), in humans it has much less connectivity with the neocortex than the orbitofrontal cortex (Fig. 12) (Rolls et al. 2023a). In humans, the amygdala receives primarily from auditory cortex A5, and semantic regions in the superior temporal gyrus and temporal pole regions; the piriform (olfactory) cortex; the lateral orbitofrontal cortex 47m; somatosensory cortex; the memory-related hippocampus, entorhinal cortex, perirhinal cortex, and parahippocampal cortex; and from the cholinergic nucleus basalis (Rolls et al. 2023a). The amygdala has effective connectivity to the hippocampus, entorhinal and perirhinal cortex; to the temporal pole; and to the lateral orbitofrontal cortex (Fig. 12) (Rolls et al. 2023a). Given the paucity of amygdalo-neocortical effective connectivity in humans, and the richness of its subcortical outputs in rodents (Quirk et al. 1996) and in humans (Klein-Flugge et al. 2022), it is proposed that the human amygdala is involved primarily in autonomic and conditioned responses via brainstem

connectivity, rather than in reported (declarative) emotion (Rolls et al. 2023a).

This new evidence about the connectivity of the human amygdala is consistent with the evidence that the amygdala is an evolutionarily old brain region, and appears to be overshadowed by the orbitofrontal cortex in humans (Rolls 2014b, 2019b, 2021b, c, 2023d; Rolls et al. 2020b). For example, the effects of damage to the human amygdala on emotion and emotional experience are much more subtle (Adolphs et al. 2005; Whalen and Phelps 2009; Delgado et al. 2011; Feinstein et al. 2011; Kennedy and Adolphs 2011; Damasio et al. 2013; LeDoux and Pine 2016; LeDoux et al. 2018; Rolls et al. 2023a) than of damage to the orbitofrontal cortex (Rolls et al. 1994; Hornak et al. 1996, 2003, 2004; Camille et al. 2011; Fellows 2011; Rolls 2019b). Indeed, LeDoux and colleagues have emphasized the evidence that the human amygdala is rather little involved in subjective emotional experience (LeDoux 2012; LeDoux and Pine 2016; LeDoux and Brown 2017; LeDoux et al. 2018; LeDoux 2020; Taschereau-Dumouchel et al. 2022). That is in strong contrast to the orbitofrontal cortex, which is involved in subjective emotional experience, as shown by the evidence just cited. LeDoux's conundrum is: if not the amygdala for subjective emotional experience, then what (LeDoux 2020)? My answer is: the human orbitofrontal cortex is the key brain region involved in subjective emotion (Rolls 2014b, 2019b, 2023d; Rolls et al. 2023a). Further, consistent with the poor rapid reversal learning found by amygdala neurons (Sanghera et al. 1979; Rolls 2014b, 2021b) compared to orbitofrontal cortex neurons, it has been found that neuronal responses to reinforcement predictive cues in classical conditioning update more rapidly in the macaque orbitofrontal cortex than amygdala, and activity in the orbitofrontal cortex but not the amygdala was modulated by recent reward history (Saez et al. 2017).

The problem of over-interpreting the role of the amygdala in emotion was that rodent studies showed that some responses such as classically conditioned autonomic responses and freezing are elicited by the amygdala with its outputs to brainstem systems, and it was inferred that therefore the amygdala is involved in emotion in the way that it is experienced by humans (LeDoux 1995, 1996, 2000a; Quirk et al. 1996). It turned out later that humans with amygdala damage had similar response-related changes, but little impairment in subjectively experienced and reported emotions (Whalen and Phelps 2009; Delgado et al. 2011; Damasio et al. 2013; LeDoux and Pine 2016; LeDoux et al. 2018; Rolls et al. 2023a). It is important therefore it is argued not to infer subjective reported emotional states in humans from responses such as conditioned autonomic and freezing responses (Rolls et al. 2023a). This dissociation of autonomic response systems from subjectively felt and reported emotions in humans is further evidence against

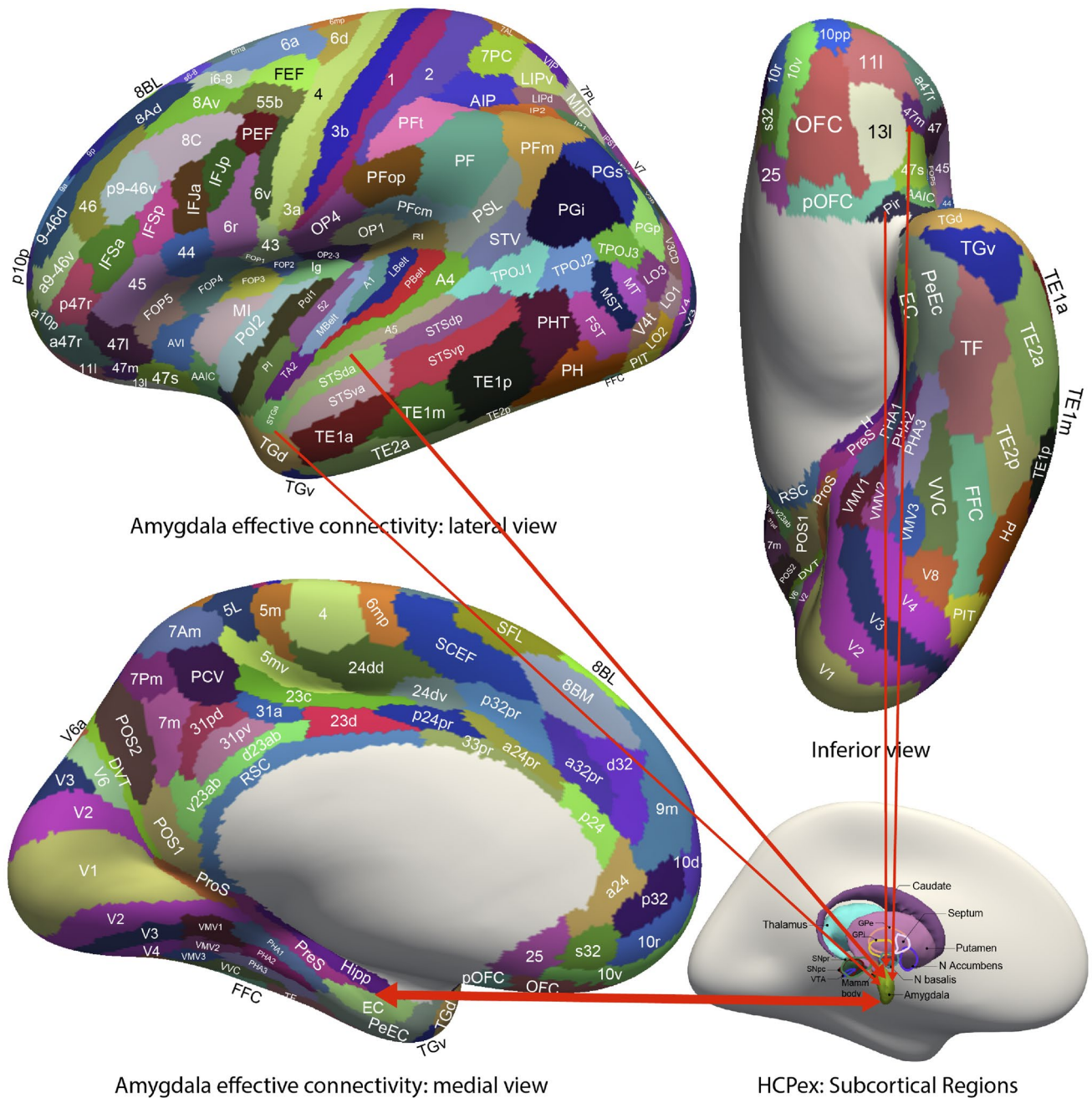


Fig. 12 Effective connectivity of the human amygdala: schematic diagram. The width of the arrows reflects the effective connectivity with the size of the arrowheads reflecting the connectivity in each direction. The connectivity from most cortical areas (anterior temporal lobe STGa and TGd, STSda and A5, and pyriform olfactory cortex) is only towards the amygdala. The connectivity with the hippocampal

system (Hipp, entorhinal cortex EC, and perirhinal cortex PeEc) is in both directions. The sulci have been opened sufficiently to show the cortical regions in the sulci. The cortical regions are defined in the Human Connectome Project Multi-Modal Parcellation atlas (Glasser et al. 2016a; Huang et al. 2022). The abbreviations are provided elsewhere (Huang et al. 2022; Rolls et al. 2023a)

the James-Lange theory of emotion and the related somatic marker hypothesis (Damasio 1994, 1996) (see Rolls (2014b) and the Appendix).

Although as described further below the amygdala may be overshadowed in humans by the orbitofrontal cortex, which has connectivity with the amygdala and that could influence amygdala neuronal responses, it is of interest that

in macaques, some amygdala neurons not only respond to faces (Leonard et al. 1985), but also respond to socially relevant stimuli when macaques interact socially (Grabenhorst et al. 2019; Grabenhorst and Schultz 2021).

The anterior cingulate cortex

Based on cytoarchitecture, connectivity and function, the anterior cingulate cortex can be divided into a pregenual part (regions s32, a24, p24, p32, and d32 in Fig. 5) that is activated by rewards, and a supracallosal or dorsal part (regions a32pr, a24pr, 33pr, p32pr and p23pr in Fig. 6) activated by punishers and non-reward (Grabenhorst and Rolls 2011; Rolls et al. 2023d), with further background provided in Vogt (2009, 2019).

The human pregenual cingulate cortex is activated by many of the same rewards as the medial orbitofrontal cortex; and the supracallosal anterior cingulate cortex is activated by many of the same punishers, and by non-reward during reward reversal, as the lateral orbitofrontal cortex (Grabenhorst and Rolls 2011; Rolls 2019a, 2021b; Rolls et al. 2020c) (see e.g. Fig. 10a). Thus value representations reach the anterior cingulate cortex (ACC). To provide examples, pain activates an area typically 10–30 mm posterior to and above the most anterior (i.e. pregenual) part of the ACC, in what can be described as the supracallosal (or dorsal) anterior cingulate cortex (Vogt et al. 1996; Vogt and Sikes 2000; Rolls et al. 2003c). Pleasant touch activated the pregenual cingulate cortex (Rolls et al. 2003c; McCabe et al. 2008). Pleasant temperature applied to the hand also produces a linear activation proportional to its subjective pleasantness in the pregenual cingulate cortex (Rolls et al. 2008b). Somatosensory oral stimuli including viscosity and the pleasantness of the texture of fat in the mouth also activate the pregenual cingulate cortex (De Araujo and Rolls 2004; Grabenhorst et al. 2010b). Pleasant (sweet) taste also activates the pregenual cingulate cortex (de Araujo et al. 2003b; De Araujo and Rolls 2004) where attention to pleasantness (Grabenhorst and Rolls 2008) and cognition (Grabenhorst et al. 2008a) also enhances activations. Pleasant odours also activate the pregenual cingulate cortex (Rolls et al. 2003b), and these activations are modulated by word-level top-down cognitive inputs that influence the pleasantness of odours (De Araujo et al. 2005), and also by top-down inputs that produce selective attention to odour pleasantness (Rolls et al. 2008a). Unpleasant odours activate the supracallosal ACC (Rolls et al. 2003b). The pregenual cingulate cortex is also activated by the ‘taste’ of water when it is rewarding because of thirst (de Araujo et al. 2003c), by the flavour of food (Kringelbach et al. 2003), and by monetary reward (O’Doherty et al. 2001). Moreover, the outcome value and the expected value of monetary reward activate the pregenual cingulate cortex (Rolls et al. 2008c).

Grabenhorst and Rolls (2011) show the brain sites of some of these activations.

In these investigations, the anterior cingulate activations were linearly related to the subjective pleasantness or unpleasantness of the stimuli, providing evidence that the anterior cingulate cortex represents value on a continuous scale, which is characteristic of what is found in the sending region, the orbitofrontal cortex (Rolls 2019a, b, d, 2021b). Moreover, evidence was found that there is a common scale of value in the pregenual cingulate cortex, with the affective pleasantness of taste stimuli and of thermal stimuli delivered to the hand producing identically scaled BOLD activations (Grabenhorst et al. 2010a).

We now consider how these value representations are used in the anterior cingulate cortex (ACC). We start with the evidence that primate orbitofrontal cortex neurons represent value, but not actions or behavioural responses (Thorpe et al. 1983; Padoa-Schioppa and Assad 2006; Grattan and Glimcher 2014; Rolls 2019b, d, 2023d), and therefore project value-related information but not action information to the anterior cingulate cortex. In contrast, there is evidence that the anterior cingulate cortex is involved in associating potential actions with the value of their outcomes, in order to select an action that will lead to the desired goal (Walton et al. 2003; Rushworth et al. 2007, 2011; Grabenhorst and Rolls 2011; Kolling et al. 2016; Morris et al. 2022). Indeed, consistent with its strong connections to motor areas (Morecraft and Tanji 2009), lesions of the ACC impair reward-guided action selection (Kennerley et al. 2006; Rudebeck et al. 2008), in humans the ACC is activated when information about outcomes guides choices (Walton et al. 2004; Morris et al. 2022), and neurons in the ACC encode information about actions, outcomes, and prediction errors for actions (Matsumoto et al. 2007; Luk and Wallis 2009; Kolling et al. 2016). For example, if information about three possible outcomes (different juice rewards) had to be associated with two different actions, information about both specific actions and specific outcomes was encoded by neurons in the ACC (Luk and Wallis 2009).

Given the evidence described above, and the connectivity shown in Fig. 7 (Rolls et al. 2023d), it is now proposed that the part of the anterior cingulate cortex involved in action–outcome learning is the supracallosal (dorsal) part, because this has effective connectivity to premotor cortical areas involved in actions with the body, including the mid-cingulate cortex. The route for value input to reach the supracallosal anterior cingulate cortex appears to be from the pregenual anterior cingulate cortex and medial orbitofrontal cortex (Fig. 7 (Rolls et al. 2023d)). The findings that aversive stimuli including pain activate the supracallosal anterior cingulate cortex may relate to the fact that actions to escape from or avoid aversive, unpleasant, stimuli often involve actions of the body, such as those involved in fight,

flight or limb withdrawal (Rolls et al. 2023d). The supracallosal anterior cingulate cortex was also implicated in human action–outcome learning in a learning theory-based analysis (Morris et al. 2022).

Further, given the evidence described above, and the connectivity shown in Fig. 7 (Rolls et al. 2023d), it is now proposed that the pregenual anterior cingulate cortex, which receives from the medial orbitofrontal cortex and ventromedial prefrontal cortex (vmPFC), and connects to the hippocampal system (Rolls et al. 2023d), in part via the memory-related parts of the posterior cingulate cortex (Rolls et al. 2023i), provides a route for affective value to be incorporated into hippocampal system episodic memory (Rolls 2022b, 2023a, c), and also to provide the information about goals that is required for navigation (Rolls 2022b, 2023c; Rolls et al. 2023d). Indeed, it has been pointed out that navigation typically involves multistep routes to reach a goal (Rolls 2021d, 2023c; Rolls et al. 2023d).

Further, the pregenual anterior cingulate cortex has connectivity to the septal region which has cholinergic neurons that project to the hippocampus (Fig. 7) (Rolls et al. 2023d), and this may contribute (Rolls 2022b) to the memory problems that can be present in humans with damage to the anterior cingulate cortex and vmPFC region (Bonnici and Maguire 2018; McCormick et al. 2018; Ciaramelli et al. 2019).

Putting all this evidence together, it appears that the anterior cingulate cortex is a key brain region in emotion, for it provides part of the route via which actions can be learned and guided by the reward/punishment outcomes that are received via the orbitofrontal cortex after a goal-directed action is performed. When instrumental learning under control of the goal was referred to previously, this is the brain region that appears to be involved in this aspect of emotion. This capability requires a computational ability to remember previous actions, which could be implemented by attractor networks in the anterior cingulate cortex, and then to associate the remembered actions with the reward or punishment outcome. However, an interesting new extension to this concept is that while the supracallosal anterior cingulate cortex is implicated in body responses such as an action with the limbs to obtain a goal or avoid a punisher, or, fight, or flight, the pregenual anterior cingulate cortex has outputs to the hippocampal system to enable actions such as navigation to obtain goals, as well as to allow reward information to be incorporated into hippocampal episodic memory, for possible future use in finding goals again (Rolls 2021d, 2022b, 2023c; Rolls et al. 2023d).

Consistent with these concepts that the anterior cingulate cortex is involved in emotion, damage to the human anterior cingulate cortex can produce emotional changes and problems in identifying face and voice emotional expressions (Hornak et al. 2003). Also consistent with inputs from reward systems in the orbitofrontal cortex driving actions

via the anterior cingulate cortex, the functional connectivity between the medial orbitofrontal cortex and the anterior cingulate cortex is higher in sensation-seekers (Wan et al. 2020), in risk-takers (Rolls et al. 2022c), and in those with a high BMI that may relate to being over-stimulated by the sensory properties of food (Rolls et al. 2023g).

The subgenual cingulate cortex (area 25), and also the orbitofrontal cortex, may link rewards and punishers to autonomic output (Critchley and Harrison 2013; Rolls 2016d, 2019a; Quadt et al. 2018, 2022). Although it has been argued by Rolls (2014b) that the autonomic system is not normally in a circuit through which behavioural responses are produced (i.e., against the James-Lange and related somatic theories of emotion (Damasio 1996), see Appendix), there may be some influence from effects produced through the endocrine system (and possibly the autonomic system, through which some endocrine responses are controlled) on behaviour (Quadt et al. 2018, 2022), or on the dual emotional and rational systems ("A reasoning, rational, route to action") that control behaviour.

A comparison of my theory of emotion with other theories of emotion is provided in the Appendix, but to maintain the continuity of the argument presented in this paper, I now move to relate my theory of emotion to my theory of motivation.

A theory of motivation, and brain systems that implement motivation

I now describe Rolls' theory of motivation, which complements and utilises many of the same brain systems, as Rolls' theory of emotion described above.

The outline of a theory of motivation

First, the essence of Rolls' approach to motivation is described. My definition of motivation is that motivational states are states that are present when rewards and punishers, that is, instrumental reinforcers, are the goals for action (Rolls 2014b, 2016f). A reward is anything for which an animal (and this includes humans) will work. A punisher is anything that an animal will work to escape or avoid, or that will suppress actions on which it is contingent (Rolls 2014b). The force of 'instrumental' in this definition is that the motivational states are seen as defining the goals for arbitrary behavioural actions, made to obtain the instrumental reinforcer. This is very different from classical conditioning, in which a response, typically autonomic, may be elicited to a stimulus without any need for an intervening state (Rolls 2014b). The motivational states modulate the reward value of instrumental reinforcers that have particular functions (Rolls 2014b, 2016f). It is important in this definition that

the reward values of potential goals are controlled appropriately, with for example factors such as plasma glucose, gastric distension and absorbed food acting to control the reward value of food (Rolls 2016a), and cellular and extracellular dehydration modulating the reward value of water (Rolls et al. 1980a, b; Rolls and Rolls 1981).

An example of a motivational state might thus be a hunger state in which the animal will perform goal-directed actions to obtain the reinforcer or goal. Another example is that the omission or termination of a reward ('extinction' and 'time out' respectively) can produce a motivational state of frustration, in which the probability of the action may become reduced if no action is possible to regain the reward, or may increase if further motivated attempts are likely to lead to the reward (Rolls 2014b, 2016f).

These examples show that the reinforcement contingency as well as the particular reinforcer or goal object (e.g. food,

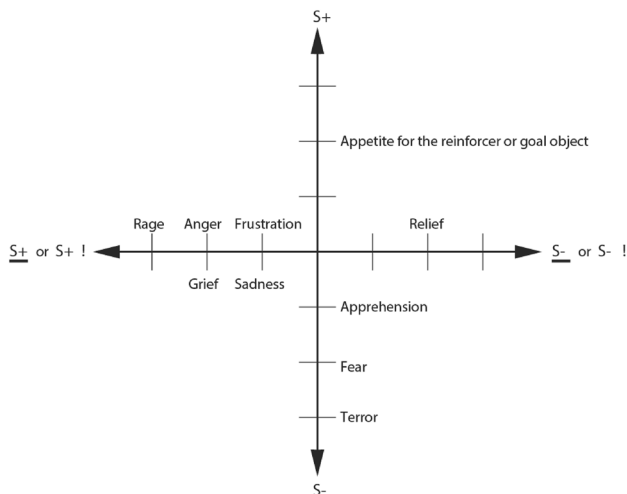


Fig. 13 Some of the motivational states associated with different reinforcement contingencies are indicated. Intensity increases away from the centre of the diagram, on a continuous scale. The classification scheme shows how different motivational states created by the different reinforcement contingencies consist of being in a state that leads to (1) performing an action to obtain a reward (S+), (2) performing an action to escape from or avoid a punisher (S-), (3) performing an action, or not, because of the omission of a reward (S±) (extinction) or the termination of a reward (S+!) (time out), and (4) performing an action, or not, because of the omission of a punisher (S-±) (avoidance) or the termination of a punisher (S-!) (escape). Note that the vertical axis describes motivational states associated with the delivery of a reward (up) or punisher (down). The horizontal axis describes motivational states associated with the non-delivery of an expected reward (left) or the non-delivery of an expected punisher (right). The diagram summarizes motivational states that might arise for one reinforcer as a result of different contingencies. Every separate reinforcer has the potential to operate according to contingencies such as these. Each different reinforcer will produce different motivational states, but the contingencies will operate as shown to produce different specific motivational states for each different reinforcer

water, aversive stimulation) lead to particular motivational states. The types of motivational state produced by different reinforcement contingencies are illustrated in Fig. 13. The diagram summarizes motivational states that might arise for one reinforcer as a result of different contingencies. Every separate reinforcer has the potential to operate according to contingencies such as these. Each different reinforcer will produce different motivational states, but the contingencies will operate as shown to produce different specific motivational states for each different reinforcer. Thus hunger might be present when the appetite is for the goal object of food, and thirst when the appetite is for the goal object of water. Definitions of reinforcers, and of the contingencies with which they operate, are elaborated by Rolls (2014b).

We must be clear about the difference between motivation and emotion. According to Rolls' theory of emotion described above, emotion is the state that results from having received, or not having received, the instrumental reinforcer, the goal object (Rolls 2014b, 2018). In contrast, motivation is the state when the instrumental reinforcer is being worked for, before the outcome stage, where the outcome is the delivery or not of the reinforcer. An important attribute of this theory of motivation and emotion is that the goal objects can be the same for motivation and emotion, simplifying the biological specification, with the difference being that motivation is the phase before the outcome, and emotion is the phase after the outcome. An additional property is that emotions, states occurring after the delivery or not of the reinforcer, can be motivating (Rolls 2014b). A good example is that if an expected reward is not obtained, then the frustrative non-reward can be motivating, and make the animal (including humans) work harder to obtain the goal object (Rolls 2014b).

As described above for emotion, reinforcers, that is rewards or punishers, may be unlearned or *primary reinforcers*, or learned, that is secondary reinforcers. An example of a primary reinforcer is pain, which is innately a punisher. The first time a painful stimulus is ever delivered, it will be escaped from, and no learning that it is aversive is needed. Similarly, the first time a sweet taste is delivered, it acts as a positive reinforcer, so it is a primary positive reinforcer or reward. Other stimuli become reinforcing by learning, because of their association with primary reinforcers, thereby becoming '*secondary reinforcers*'. For example, the sight of a food that regularly precedes the flavour of the food can rapidly become a secondary reinforcer.

Some examples of primary reinforcers are provided in Table 1. All of the primary reinforcers or goal objects can elicit different, specific, motivational states. As these are primary reinforcers, they are likely to be gene-specified.

Motivational states are states that modulate the reward or punishment value of instrumental reinforcers, and are different from the mechanisms involved in taxes, approach responses, classical conditioning, and stimulus–response habits

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 that 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 toward 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 response mechanisms in such a way that the stimuli are approached or escaped from.

Habit or stimulus–response learning

In this second level of complexity, involving reinforcers, learning may occur. If an organism performs trial-and-error responses, and as the result of performing one particular response is more likely to obtain a reward, then the response may become linked by a learning process to that stimulus as a result of the reinforcement received. The reward is said to reinforce the response to that stimulus, and we have what is described as stimulus–response or habit learning. The reward acts as a positive reinforcer in that it increases the probability of a response on which it is made contingent. A punisher reduces the probability of a response on which it is made contingent. It should be noted that this is an operational definition, and that there is no implication that the punisher feels like anything in particular—the punisher just has, in the learning mechanism, to reduce the probability of responses followed by the punisher. Stimulus–response or habit learning is typically evident after over-training, and once habits are being executed, the behaviour becomes somewhat independent of the reward value of the goal, as shown in experiments in which the reward is devalued. This is described in more detail in the "[Wanting vs liking and goal-directed motivational behaviour](#)". (Stimulus–response habit learning is quite different from action–outcome learning, in which actions are performed and learned to obtain a

goal, and in which the value of the goal therefore influences the actions Cardinal et al. 2002; Rolls 2014b).

Once a behaviour becomes automated as a habit, aversive stimuli can be avoided with very little sign of an emotional state.

The dopamine system is implicated in providing the training signal for this type of habit learning implemented in the striatum (Rolls 2014b, 2021b, 2023d; Schultz 2016a, c).

Rewards and punishers: instrumental goals for action towards which motivation is directed

As soon as we have approach to 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 an animal’s response consists of a fixed response to obtain the stimulus (e.g. locomotion up a chemical gradient), we shall call this a taxis not a reward. If a fixed behavioural response or action pattern such as skeletomotor freezing and autonomic responses are elicited by a stimulus, they may be adaptive, but are essentially stimulus–response reflexes, with no need for an intervening motivational state, such as the representation of a goal to be reached. On the other hand, if an arbitrary operant action (an instrumental action) can be performed by the animal in order to approach the stimulus or goal, then we will call this rewarded behaviour, and the stimulus that the animal works to obtain a reward, the goal for the action, and the state of wanting and being willing to work for the goal a motivational state. The arbitrary operant response can be thought of as any arbitrary action that 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 its tail, or lower its tail, in order to obtain a piece of food, then we can be sure that there is no fixed relationship between the stimulus (e.g. the sight of food) and the response, as there is in a taxis. I, and a number of other authors (Teitelbaum 1974), reserve the term ‘motivated behaviour’ for that in which an arbitrary operant action, an instrumental action, will be performed to obtain a reward or to escape from or avoid a punisher. This is the action–outcome learning described above in which the anterior cingulate cortex is implicated. If this criterion is not met, and only a fixed response can be performed, then the term ‘drive’ can be used to describe the state of the animal when it will work to obtain or escape from the stimulus.

We can thus distinguish a first level of approach/avoidance mechanism complexity in a taxis, with a fixed response available for the stimulus, from a second level of complexity

in which any arbitrary response (or action) can be performed, in which case we use the term reward when a stimulus is being approached, and punisher when the action is to escape from or avoid the stimulus. The motivational, intervening, state is one in which the animal will perform an arbitrary action to obtain a goal. Again, we should distinguish habit-related stimulus–response implemented instrumental behaviour, from action–outcome instrumental behaviour that is under the control of the goal.

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 of the animals can lead to better survival to enable genes to be passed on to the next generation, which is what we mean by fitness. Fitness refers to the fitness of genes, but this must be measured by the effects that the genes have on the organism. The animals must be built by such natural selection to perform actions 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 perform actions that will enable them to escape from, or avoid (when learning mechanisms are introduced), stimuli that will reduce their fitness. There are likely to be many dimensions of environmental stimuli along which actions of the animal can alter fitness. Each of these dimensions may be a separate reward–punisher 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 behavioural responses to obtain such reward stimuli when in that need or motivational state. Similarly, another dimension is water reward, in which the taste of water becomes rewarding when there is body-fluid depletion (Rolls 2005).

One aspect of the operation of these reward–punisher systems that these examples illustrate is that with very many reward–punisher dimensions for which actions may be performed, there is a need for a selection mechanism for actions performed to these different dimensions. In this sense, each specific reward and punisher is on a common scale (Grabenhorst et al. 2010a) to facilitate the operation of action selection mechanisms. Evolution must set the magnitudes of each of the different reward systems so that each will be chosen for action in such a way as to maximize overall fitness. Food reward must be chosen as the aim for action if some nutrient depletion is present, but water reward as a target for action must be selected if current water depletion poses a greater threat to fitness than does the current degree of food depletion. This indicates that for a competitive selection process for rewards, each reward must be carefully calibrated in evolution to have the right value on a common scale for the selection process (but not converted into a common currency) (Rolls 2014b). Other types of behaviour, such as sexual behaviour, must be performed sometimes, but probably

less frequently, in order to maximize fitness (as measured by gene transmission into the next generation).

There are many processes that 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 that reduce the rewards within a dimension (such as hunger signals described below), but also sensory-specific satiety mechanisms, which facilitate switching to another reward stimulus (sometimes within and sometimes outside of the same main dimension), and attraction to novel stimuli. Attraction to novel stimuli, i.e. finding novel stimuli rewarding, is one way that organisms are encouraged to explore the multidimensional space within which their genes are operating. The suggestion is that animals should be built to find somewhat novel stimuli rewarding, for this encourages them to explore new parts of the environment in which their genes might do better than others' genes. Unless animals are built to find novelty somewhat rewarding, the multidimensional genetic space being explored by genes in the course of evolution might not find the appropriate environment in which they might do better than others' genes (Rolls 2014b). The primate orbitofrontal cortex contains neurons that respond to novel stimuli (Rolls et al. 2005).

Motivation, and instrumental, action–outcome, goal-directed, learning

When behaviour is under control of the goal, the reward or punisher, then we call this motivated behaviour. A test of whether the behaviour is under the control of the goal is reward devaluation. For example, if humans and other animals are fed to satiety with a food, they show sensory-specific satiety for the food, rate its subjective pleasantness as zero, and are no longer motivated to obtain and ingest it. The motivation for other foods not eaten to satiety usually remains. The hallmark of a devaluation experiment showing that a behaviour is under the control of the goal and therefore qualifies for being described as 'motivated' is that when the goal is devalued, the human or other animal will not perform an instrumental action to obtain it the first time that the stimulus is presented (see "[Wanting vs liking and goal-directed motivational behaviour](#)").

Two-stages of learning may be involved in such motivational goal-controlled instrumental learning. Rewards and punishers provide the basis for guiding behaviour within a dimension, and for selecting the dimension towards which action should be directed.

The first stage of the learning is stimulus–reinforcer association learning, in which the reinforcing value of a previously neutral, e.g. visual or auditory, stimulus is learned because of its association with a primary reinforcer, such as a sweet or salt taste (Kehoe and Blass 1985) or a painful

touch. This learning is of an association between one stimulus, the conditioned or secondary reinforcer, and the primary reinforcer, and is thus stimulus-stimulus association learning. This stimulus-reinforcer learning can be very fast, in as little as one trial. For example, if a new visual stimulus is seen and then placed in the mouth and a sweet taste is obtained, an instrumental action such as reaching for the object will be made on the next trial. Moreover, this stimulus-reinforcer association learning can be reversed very rapidly, at least in primates including humans though not in rodents. For example, if subsequently the object is made to taste of salt, then the visual stimulus is no longer reached for, and the stimulus is even likely to be actively pushed away. This stimulus-reinforcer association learning is implemented in the primate including human orbitofrontal cortex, and leads to representations of expected value (Rolls 2014b, 2018, 2019b, 2021b, 2023d).

The second process or stage in this type of learning is instrumental learning of an action (or ‘operant response’) made in order to obtain the stimulus now associated with reward (or avoid the stimulus associated by learning with the punisher). This is action–outcome learning (implemented in brain regions such as the anterior cingulate cortex as described above Grabenhorst and Rolls 2011; Rushworth et al. 2011; Rolls 2014b, 2019a, 2023d). The outcome could be a primary reinforcer, but often involves a secondary reinforcer learned by stimulus-reinforcer association learning. The action–outcome learning may be much slower than the stimulus-reinforcer learning, for action–outcome learning may involve trial-and-error learning of which action is successful in enabling the animal to obtain the stimulus now associated with reward or avoid the stimulus now associated with a punisher. However, this second stage may be greatly speeded if an operant response or strategy that has been learned previously to obtain a different type of reward (or avoid a different punisher) can be used to obtain (or avoid) the new stimulus now known to be associated with reinforcement. It is in this flexibility of the action that two-factor learning has a great advantage over stimulus–response learning. The advantage is that any action (even, at its simplest, approach or withdrawal) can be performed once an association has been learned between a stimulus and a primary reinforcer. This flexibility in the action is much more adaptive (and could provide the difference between survival or not) than no learning, as in taxes, or stimulus–response habit learning. The different processes that are involved in instrumental learning are described in more detail by Rolls (2014b).

Another key advantage of this type of two-stage learning is that after the first stage the different rewards and punishers available in an environment can be compared in a selection mechanism, using the common scale of different rewards and punishers for the comparison and selection

process (Grabenhorst et al. 2010a; Rolls 2014b). In this type of system, the many dimensions of rewards and punishers are again the basis on which the selection of an action to perform is made (Rolls 2014b).

Gene-specified rewards and the mechanisms of evolution

Part of the process of evolution can be seen as identifying the factors or dimensions that affect the (reproductive) fitness of an animal, and providing the animal with sensors that lead to rewards and punishers that are tuned to the environmental dimensions that influence fitness. The example of sweet or salt taste receptors being set up by evolution to provide reward when physiological nutrient need is present (Kehoe and Blass 1985) has been given above, and shows how genes are involved in specifying motivational states.

We can ask whether there would need to be a separate sensing mechanism tuned to provide primary (unlearned) reinforcers for every dimension of the environment to which it may be important to direct motivational behaviour. (The motivated behaviour has to be directed to climb up the reward gradient to obtain the best reward, or to climb a gradient up and away from punishers). It appears that there may not be. For example, in the case of the so-called specific appetites, for perhaps a particular vitamin lacking in the diet, it appears that a type of stimulus-reinforcer association learning may actually be involved, rather than having every possible flavour set up to be a primary reward or punisher. The way that this happens is by a form of association learning. If an animal deficient in one nutrient is fed a food with that nutrient, it turns out that the animal’s physiological state is ‘better’ some time after ingesting the new food, and the animal associates this better physiological state with the taste of that particular food. Later, that food will be chosen. The point here is that the first time the animal is in the deficient state and tastes the new food, that food may not be chosen instead of other foods. It is only after the post-ingestive conditioning that, later, that particular food will be selected (Rozin and Kalat 1971; Berthoud et al. 2021; Rolls 2023e). Thus in addition to a number of specific primary (unlearned) reward systems (e.g. sweet taste for nutrient need, salt taste for salt deficiency (Kehoe and Blass 1985), pain for potentially damaging somatosensory stimulation), there may be great opportunity for other arbitrary sensory stimuli to become conditioned rewards or punishers by association with some quite general change in physiological state. The implication here is that a number of bodily signals can influence a general bodily state, and we learn to improve the general state, rather than to treat the signal as a specific reinforcer that directs us to a particular goal. Another example might be social reinforcers. It would be difficult to build-in a primary reinforcer system for every possible type of social reinforcer. Instead, there may be a number of

rather general primary social reinforcers, such as acceptance within a group, approbation, greeting, face expression, and pleasant touch, which are among the primary rewards; and by association with these primary rewards, other stimuli can become secondary social reinforcers.

To help specify the way in which stimulus-reinforcer association learning operates, a list of what may be in at least some species primary reinforcers is provided in Table 1. The reader will doubtless be able to add to this list, and it may be that some of the reinforcers in the list are actually secondary reinforcers. The reinforcers are categorized where possible by modality, to help the list be systematic. Possible dimensions to which each reinforcer is tuned are suggested.

In Rolls' theories of motivation and emotion, there may be a set of approximately 100 gene-specified primary reinforcers of the type described in Table 1. Each primary reinforcer accounts for a motivational state in which the reward is the target of an instrumental action, and for the emotional state that is produced when the reward or punisher is or is not received. These motivational and emotional states must all be specific; for example hunger must increase food reward but not water reward. These reward value systems must be modulated by the correct selective signals; for example, sensors of metabolic state that relate to hunger must increase the reward value of food but not of water. In so doing, there must be mechanisms to lead animals, when in a motivational state, to navigate and perform appropriate actions to find a specific reward (Deutsch 1960). The reward is produced by the sensory input produced by taste, smell, flavour, touch, sight, and sound, etc., and not by a reduction in the motivational signal. Some of the evidence for this is that very small sensory inputs, such as a drop of food, act as powerful rewards, but reducing hunger by placing food into the stomach produces little reward (Rolls 2014b, 2023e). Consistent with this, reducing the firing of hunger neurons has only a minor rewarding effect (Sternson 2013), so reducing hunger or more generally motivation does not normally act as the reward for instrumental behaviour.

In the reward-based motivational system that I describe, each reward must be scaled to a similar range, so that the different rewards are selected at least sometimes by competing in a decision-making process, so that each reward can contribute to survival and reproductive success (Rolls 2014b). Motivational behaviour can be seen from this approach as an animal operating with a set of initially gene-specified goals for actions (though subject to learned re-evaluation) which compete in a high-dimensional space of rewards for a decision to be taken about which is most rewarding at the time, depending on modulators such as hunger signals, sensory-specific satiety, etc. (Rolls 2014b). The decision taken will also reflect the costs of the actions required to obtain the different rewards (Rolls 2014b). Evidence about how the underlying mechanisms operate are described in

Emotion and Decision-Making Explained (Rolls 2014b) and elsewhere (Rolls 2018, 2021b, 2023d).

Biological economy in the specification of rewards and punishers, for they can be used to implement both motivation and emotion

We now come to the heart of the adaptive value of my approach to motivation and emotion.

My proposal is that the same gene-specified rewards and punishers can be used for both motivation and emotion. This produces great simplification in the genetic specification of motivation and emotion, for the genes have to specify just one set of primary rewards and punishers. The reward has to be motivating, in that animals need to be built to want to perform actions to obtain rewards. Each gene-specified reward then needs to be modulated by the appropriate motivational state. For example, the motivational state of hunger, which modulates the reward value of the taste, smell and sight of food, is signalled by many factors including plasma and gut nutrients and metabolic hormones, as described in detail elsewhere (Rolls 2014b, 2016a, f, 2018). The motivational state of thirst, which modulates the reward value of the taste and sight of water, is signalled by cellular and extracellular fluid volume (Rolls et al. 1980a, b; Rolls and Rolls 1982; de Araujo et al. 2003c; Rolls 2005). Factors that influence the reward value of stimuli involved in sexual behaviour are numerous, and typically adaptive for the genes (Buss 1989, 2015; Rolls 2014b, 2018). For example, in males, the reward value of sexual behaviour typically decreases soon after ejaculation, as a further ejaculate in the same female soon would be unlikely to increase markedly the probability of reproductive success, and it may be adaptive to conserve some sperm for a possible opportunity for reproductive success with another female, with sensory-specific satiety here being referred to as the Coolidge effect (Buss 1989, 2015; Rolls 2014b, 2018). The reward value of sexual behaviour in females is also subject to modulation by many factors that influence reproductive success (Buss 1989, 2015; Rolls 2014b, 2018). The key point here is that the value of each type of reward must be modulated by appropriate motivational signals, such as gut and plasma nutrient signals for food reward, cellular and extracellular volume reductions for water reward, and factors such as the probability of reproductive success in passing on genes to the next generation for sex rewards (Buss 1989, 2015; Rolls 2005, 2014b, 2018).

The same set of rewards, and punishers, when received after for example an instrumental action, lead to emotional states, as described above.

The great utility of both emotional and motivational states relating to rewards and punishers is that this is a highly efficient way for behaviour to be organised, in that the genes specify stimuli that are rewards and punishers, and leave

it open to the animal to perform any instrumental action to obtain the reward or avoid the punisher. This is very much more efficient than having genes specify a fixed response to stimuli, such as pecking at small grains as they may be food. The latter type of mechanism of gene-specified responses can have utility for a few responses to a few stimuli, as in the case of chickens pecking at grains of corn. But the genetic specification of many such stimulus–response pairs would be genetically expensive, and would have the great disadvantage that there would be no or little flexibility of the response. Instead, when genes are used to specify rewards and punishers, of the type set out in Table 1, then an almost unlimited set of actions can be learned to obtain the rewards or avoid the punishers. For this reason, I argue that the specification of rewards and punishers by genes, rather than fixed behavioural responses, is a major factor in the design of brains for evolutionary success.

These concepts (including that an important way for genes to influence behaviour is by specifying the reward and punishment value of stimuli) were developed and made clear by Rolls (2005, 2014b, 2016f, 2018), but were not featured in *The Selfish Gene* and subsequent books (Dawkins 1976, 1982, 1986). These concepts are key to understanding how in practice genes frequently increase their (selfish) success by specifying stimuli that are rewards and punishers. Operating in this way, so that the same genes specify rewards and punishers appropriate for both motivation and emotion, and do not specify actions, leads to great adaptiveness and elegance in brain design (Rolls 2016c, f, 2021b, 2023d).

Wanting vs liking and goal-directed motivational behaviour

Rolls' theory of motivation holds that each gene-specified reward is a goal for action, that is, accounts for motivation (Rolls 2016f); and also, when the reward is received, it generates emotion (Rolls 2014b, 2018). An important attribute of these theories of motivation and emotion is that the same specification of a goal object, a reward, perhaps genetically or by stimulus-reward learning, accounts for both the motivation, which has to be produced if the animal is ever to seek the reward, and the emotion, which is associated with the reward when it is received. This makes for great economy in evolution, for genes are needed to specify goal objects, and in doing this, have to produce both working to obtain those goal objects ('wanting') and the emotional state when the goal object is received or not received ('liking') (Rolls 2014b).

It is useful in this context to discuss an apparent dissociation between 'wanting' and 'liking' (or 'desire' vs 'pleasure') that has been raised (Berridge 1996; Berridge and Robinson 1998; Berridge et al. 2009). 'Wanting' or conditioned 'incentive salience' effects are used to describe classically

conditioned approach behaviour to rewards (Berridge and Robinson 1998, 2003), and this learning is implemented via the amygdala and ventral striatum, is under control of dopamine (Cardinal et al. 2002), and contributes to addiction (Robinson and Berridge 2003). Conditioned 'incentive salience' effects can influence instrumental responses made, for example, to obtain food.

A first point is that Berridge and Robinson (1998) suggest that 'liking' can be measured by orofacial reflexes such as ingesting sweet solutions or rejecting bitter solutions. There is evidence that brain opioid systems are involved in influencing the palatability of and hedonic reactions to foods, in that humans report a reduction in the pleasantness of sucrose solution following administration of naltrexone which blocks opiate receptors, but can still discriminate between sucrose solutions (Gosnell and Levine 2009; Stice et al. 2013). One problem here is that orofacial reflexes may reflect brainstem mechanisms that are not at all closely related to the reward value of food as reflected in instrumental actions performed to obtain food (see Fig. 2). Some of the evidence for this is that these responses occur after decerebration, in which the brainstem is all that remains to control behaviour (Grill and Norgren 1978) [with consistent evidence from anencephalic humans (Steiner et al. 2001)]. Care must be taken about such inferences as there are many routes to behavioural responses (Rolls 2014b, 2021b, 2023d; Balleine 2019) (Fig. 2).

A second point is that normally the rated reward value or pleasantness given in humans to food is closely related to instrumental actions performed to obtain food, as shown by the close relation between pleasantness ratings ('liking') by humans given to a food in a sensory-specific satiety experiment, and whether that food is subsequently eaten in a meal ('wanting') (Rolls et al. 1981c).

Third, a confusion may arise when a stimulus–response habit is formed by overlearning, and persists even when the reward is devalued by, for example, feeding to satiety. This persistence of stimulus–response habits after reward devaluation should not necessarily be interpreted as 'wanting' when not 'liking', for it may just reflect the operation of a stimulus–response habit system that produces responses after overlearning without any guidance from reward, pleasantness, and liking (Cardinal et al. 2002; Rolls 2014b; Balleine 2019). Indeed, I emphasize that after overtraining, responses may become inflexibly linked to stimuli, and the goals, and the reward value of the goals, may no longer be directly influencing behaviour in an ongoing way. If behaviour becomes overlearned and a habit or stimulus–response connection is built up by another brain system (such as the basal ganglia), then animals may make automatic responses that are not goal directed (Cardinal et al. 2002; Rolls 2014b; Balleine 2019). There has been considerable confusion in the literature caused by overlooking this point (Berridge and Robinson 1998; Berridge et al. 2009; Berridge and Dayan

2021; Nguyen et al. 2021; Warlow and Berridge 2021). Indeed, just as in the research on the amygdala described above in which LeDoux inferred full emotions from conditioned responses, it is unwarranted and potentially misleading to use subjective emotion-laden words such as ‘wanting’ and ‘liking’ that describe emotional feelings (Berridge 1996; Berridge and Robinson 2003; Robinson and Berridge 2003; Berridge et al. 2009; Berridge and Dayan 2021; Nguyen et al. 2021; Warlow and Berridge 2021), when classically conditioned responses such as Pavlovian-Instrumental Transfer, and orofacial reflexes and stimulus–response habits are what has been measured (Cardinal et al. 2002; Rolls 2014b; Balleine 2019). The fact that behaviour can become stimulus–response and no longer under the control of the goal need not surprise us. Normally, and certainly during learning before habits set in, we want a goal, and when we get the goal we like it: goal stimuli normally specify what is wanted, and what is liked. Indeed, my theory is that normally we want because we like. This is inherent in my theory, for the genes that make a stimulus (such as a sweet taste) rewarding (i.e. wanted, a goal for action) also make the stimulus liked (i.e. accepted, with a subjective correlate of pleasure, pleasantness, and affective liking).

My approach is that I believe that liking, defined by pleasantness ratings of stimuli, is normally very closely related to wanting, that is being willing to perform behaviour (instrumental actions) to obtain a reward of the pleasant stimulus (Rolls 2014b, 2016f). Thus motivational behaviour is normally controlled by reward stimuli or goals (unless the behaviour is overlearned), and motivational state (e.g. hunger) modulates the reward value of unconditioned and conditioned stimuli such as the taste and sight of food. Thus normally, liking a goal object and wanting it are different aspects of how reward systems control instrumental behaviour, and this follows from the approach to gene-specified goal or value representations which in a unifying way account for wanting a goal, and liking the goal object when it is obtained (Rolls 2014b, 2016f, 2018, 2021b, 2023d). For further clarification, consider a probabilistic decision-making task in which the probability P of obtaining a reward outcome (such as the taste of food) is 0.5, and the reward outcome value is 1 if the reward is delivered (e.g. fruit juice), and is 0 when the reward is not delivered. Then the Expected (reward) Value when the offer is made is 0.5 (Expected reward Value = $P \times$ Outcome Value (Rolls 2014b)), and the value of the motivational state at that time (which is *before* the outcome is known) is 0.5. Then later in the trial, the affective/emotional state (which is *after* the outcome is delivered) is 1 for the fruit juice reward, and 0 if nothing is obtained as the outcome.

Thus, it is possible to identify the brain mechanisms involved in ‘wanting’ and ‘liking’ experimentally, and to distinguish them from the classically conditioned ‘incentive

salience’ stimuli that influence approach and instrumental actions and which influence ‘appetitive’ behaviour, implemented in part separately from the reward systems that are activated by a primary reinforcer such as the taste of food during ‘consummatory’ behaviour (Cardinal et al. 2002; Rolls 2014b).

Some implications and extensions of the understanding of emotion, motivation, and their brain mechanisms

Top-down cognitive effects on reward value and affective responses, for example on the reward value and pleasantness of taste, olfactory, and flavor stimuli

To what extent does cognition influence the reward value of stimuli, and how far down into the sensory system does the cognitive influence reach? Alternatively, is the reward value in brain regions such as the orbitofrontal cortex independent of cognitive factors, with reward value being interfaced to cognition in other, perhaps language-related, brain regions?

We discovered that word-level cognitive effects have top-down modulatory effects on reward value processing in the orbitofrontal cortex and anterior cingulate cortex. This was shown for olfactory (De Araujo et al. 2005), taste (Grabenhorst et al. 2008a), and touch and the sight of touch (McCabe et al. 2008) reward value. For example, a standard test odor (isovaleric acid combined with cheddar cheese odor, presented orthonasally using an olfactometer) was paired with a descriptor word on a screen, which on different trials was “Cheddar cheese” or “Body odor”. Participants rated the affective value of the standard test odor, isovaleric acid, as significantly more pleasant when labelled “Cheddar Cheese” than when labeled “Body odor”, and these effects reflected activations in the medial orbitofrontal cortex and pregenual anterior cingulate cortex (De Araujo et al. 2005). The implication is that cognitive factors can have profound effects on our responses to the reward value and subjective reported pleasantness of olfactory stimuli, in that these effects are manifest quite far down into reward value processing (in the orbitofrontal cortex), so that hedonic representations of odors are affected (De Araujo et al. 2005).

Similar cognitive effects and mechanisms have also been found for the taste and flavor of food, where the cognitive word level descriptor was, for example, ‘rich delicious flavor’ and activations to flavor were increased in the orbitofrontal cortex and regions to which it projects including the pregenual anterior cingulate cortex and ventral striatum, but were not influenced in the insular primary taste cortex where activations reflected the intensity (concentration) of the stimuli (Grabenhorst et al. 2008a).

For the sight of touch, the cognitive modulation was produced by word labels, ‘Rich moisturizing cream’ or ‘Basic cream’, while cream was being applied to the forearm, or was seen being applied to a forearm. The cognitive labels influenced the activations to the sight of touch and also the correlations with pleasantness in the pregenual anterior cingulate/orbitofrontal cortex and ventral striatum (McCabe et al. 2008).

The wider implication of these discoveries is that our cognitive processes can actually modulate the representation of reward value and subjective pleasantness in brain regions involved in reward value representations such as the orbitofrontal and pregenual anterior cingulate cortex, and this can potentially provide important ways in which the appreciation of other rewards such as music, art and aesthetics can be influenced by cognitive factors acting on the reward value representations in parts of the brain that represent reward value and subjective pleasantness. In this way, the appropriate top-down cognitive bias could enhance the pleasure being experienced.

The mechanisms of top-down cognitive modulation are understood as biased activation being applied to the orbitofrontal cortex from brain regions such as the prefrontal cortex which maintain the biasing information in short-term memory (Deco and Rolls 2003, 2004, 2005a, 2005b; Deco et al. 2005; Rolls and Deco 2006, 2010; Rolls 2013a, 2016c, 2021b, 2023d).

Effects of top-down selective attention to affective value versus intensity on representations of stimuli including those involved in taste, olfactory, and flavour processing

We have found that with taste, flavor, and olfactory food-related stimuli, selective attention to pleasantness modulates representations in the orbitofrontal cortex, whereas selective attention to intensity modulates activations in areas such as the primary taste cortex (Grabenhorst and Rolls 2008, 2010; Rolls et al. 2008a; Ge et al. 2012; Luo et al. 2013; Rolls 2013a).

This differential biasing of brain regions engaged in processing a sensory stimulus depending on whether the cognitive or attentional demand is for affect-related vs more sensory-related processing may be an important aspect of cognition and attention which has implications for how strongly the reward system is driven by food, and thus for eating and the control of appetite, but also for other types of reward (Grabenhorst and Rolls 2008, 2011; Rolls et al. 2008a; Rolls 2012a, 2014b).

The wider implication is that top-down attention directed to the reward value and subjective pleasantness of stimuli can enhance activations to these stimuli in reward-related brain regions, and this has potential applications to enhance

the subjective pleasantness of many types of reward, including aesthetic types of reward (e.g. music and art). Attention applied in this way may divert brain systems from maintaining unpleasant ruminating events in memory, and this is of potential use in for example the treatment of depression and other unpleasant states including pain (Rolls 2016b, 2018, 2021b; Rolls et al. 2020b).

The mechanisms of top-down attentional modulation are understood as biased competition and biased activation being applied to the orbitofrontal cortex from brain regions such as the prefrontal cortex which maintain the biasing information in short-term memory (Deco and Rolls 2003, 2004, 2005a, b; Deco et al. 2005; Rolls and Deco 2006, 2010; Luo et al. 2013; Rolls 2013a, 2016c, 2021b, 2023d).

Individual differences in the reward systems, evolution, and personality

An important hypothesis is that different humans may have reward systems that differ in how strongly each of their reward systems are activated, driven by the sensory and cognitive factors that make stimuli attractive. In a test of this, we showed that activations to the sight and flavor of chocolate in the orbitofrontal and pregenual cingulate cortex were much higher in chocolate cravers than non-cravers (Rolls and McCabe 2007), though there were no differences at the level of the insular taste cortex where taste perception is represented. This provides evidence that differences in specific reward systems, and not necessarily in earlier sensory processing, can lead to individual differences in behaviour to different types of rewarding stimuli. (Rolls 2014b, 2018). This concept that individual differences in responsiveness to food reward are reflected in brain activations in regions related to the control of food intake (Beaver et al. 2006; Rolls and McCabe 2007) may provide a way for understanding and helping to control food intake and obesity (Rolls 2012a, 2014b, 2016a, 2018).

But the concept is much more general than this. The wider implication is that part of the way in which evolution operates is by utilizing natural variation in each of the specific reward systems (examples of which are shown in Table 1), and selecting for reward systems with sensitivities that lead to reproductive success. This results in each individual having a different set of sensitivities of perhaps 100 different gene-specified reward systems of the type shown in Table 1 (Rolls 2014b, 2018).

The sensitivity of an individual to different rewards and punishers, and the ability to learn and be influenced by rewards and punishers, provide a basis for different personalities (Rolls 2014b, 2018). Part of the basis of personality may be differential sensitivity to different rewards and punishers, and omission and termination of different rewards and punishers (the reinforcement contingencies shown in

Fig. 1), and this could give rise to many types of personality when we take into account that there are many different types of reinforcer. This can be related to brain function, in that for example the medial orbitofrontal cortex involved in different rewards and the lateral orbitofrontal cortex involved in different punishers and non-rewards could have different sensitivities of these systems for different types of reward. An extreme example might be that if humans were insensitive to social punishers following orbitofrontal cortex damage, we might expect social problems and impulsive behaviour, and indeed Tranel et al (2002) have used the term ‘acquired sociopathy’ to describe some of these patients. Indeed, we might expect sensitivity to different types of reinforcer (including social reinforcers) to vary between individuals both as a result of gene variation and as a result of learning, and this, operating over a large number of different social reinforcers, might produce many different variations of personality based on the sensitivity to a large number of different reinforcers.

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

Jeffrey Gray (1970) reinterpreted the findings, suggesting that introverts are more sensitive to punishment and frustrative non-reward than are extraverts; and that neuroticism reflects the extent of sensitivity to both reward and punishment. However, in addition extraverts may perform less well at vigilance tasks (in which the subject must detect stimuli that occur with low probability); may tend to be more impulsive; and perform better when arousal is high (e.g. later in the day), and when rapid responses rather than reflective thought is needed (Matthews and Gilliland 1999).

In functional neuroimaging studies, it has been found that happy face expressions are more likely to activate the human amygdala in extraverts than in introverts (Canli et al. 2002). This supports the conceptually important point made above that part of the basis of personality may be differential sensitivity to different rewards and punishers, and omission and termination of different rewards and punishers.

In one update of this approach, links have been made to behavioural economics by relating loss aversion to greater negative valuation sensitivity compared to positive valuation sensitivity; by suggesting that tendencies to approach or avoid have distinct sensitivities to those of the valuation systems; that approach-avoidance conflict is a distinct process from the basic approach and avoidance systems; and linking these to a reinforcer sensitivity theory of personality (Corr and McNaughton 2012).

Another example is the impulsive behaviour that is a part of Borderline Personality Disorder (BPD), which could reflect factors such as less sensitivity to the punishers associated with waiting for rational processing to lead to a satisfactory solution, or changes in internal timing processes that lead to a faster perception of time (Berlin and Rolls 2004; Berlin et al. 2004). It was of considerable interest that the BPD group (mainly self-harming patients), as well as a group of patients with damage to the orbitofrontal cortex, scored highly on a Frontal Behaviour Questionnaire that assessed inappropriate behaviours typical of orbitofrontal cortex patients including disinhibition, social inappropriateness, perseveration, and uncooperativeness. In terms of measures of personality, using the Big Five personality measure (which identifies five major components of personality (Trull and Widiger 2013), both groups were also less open to experience (i.e. less open-minded). In terms of other personality measures and characteristics, the orbitofrontal and BPD patients performed differently: BPD patients were less extraverted and conscientious and more neurotic and emotional than the orbitofrontal group (Berlin and Rolls 2004; Berlin et al. 2004, 2005). Thus some aspects of personality, such as impulsiveness and being less open to experience, but not other aspects, such as extraversion, neuroticism and conscientiousness, were especially related to orbitofrontal cortex function.

However, in terms of detailed understanding of how the computations in different brain regions relate to personality, research is still in its early stages (DeYoung et al. 2022).

A reasoning, rational, route to action

Routes to action that are related to the emotional system are described in "[The neuroscience of emotion in humans and other primates](#)", and are indicated in Fig. 2.

Another main route to action in humans, and perhaps some other species, involves a computation with many "if ... then" statements, to implement a multi-step plan to obtain a reward. In this case, the reward may actually be *deferred* as part of the plan, which might involve working first to obtain one reward, and only then to work for a second more highly valued reward, if this was thought to be overall an optimal strategy in terms of resource usage (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 involvement of language systems in the brain is required for this type of planning (Rolls 2014b, 2018, 2021b, 2023d). Thus the explicit language system in humans may allow working for deferred rewards by enabling use of a one-off, individual, 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. This short-term memory may be for example in non-human primates 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 in place 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 in order to form a multiple step plan. Such short-term memories are implemented in the (dorsolateral and inferior convexity) prefrontal cortex of non-human primates and humans (Goldman-Rakic 1996; Fuster 2014; Fuster 2015; Rolls 2016c; Miller et al. 2018; Kolb and Wishaw 2021; Rolls 2021b, 2023d; Rolls et al. 2023e), and may be part of the reason why prefrontal cortex damage impairs planning (Shallice and Cipolotti 2018).

Decisions between the emotional and reasoning systems

The question then arises of how decisions are made in animals such as humans that have both the implicit, direct reward-based, and the explicit, rational, planning systems (Rolls 2011, 2014b, 2018, 2019c, 2021b, 2023b). One particular situation in which the first, implicit, system may be especially important is when rapid reactions to stimuli with reward or punishment value must be made, for then the direct connections from structures such as the orbitofrontal cortex to the basal ganglia may allow rapid actions (Rolls 2014b). Another is when there may be too many factors to be taken into account easily by the explicit, rational, planning, system, when the implicit system may be used to guide action. In contrast, when the implicit system continually makes errors, it would then be beneficial for the organism to switch from direct, action based on obtaining what the orbitofrontal cortex system decodes as being the most positively rewarding choice currently available, to the explicit conscious control system which can evaluate with its long-term planning algorithms what action should be performed next. Indeed, it would be adaptive for the explicit system to regularly be assessing performance by the goal-based and habit-based systems, and to switch itself in to control

behaviour quite frequently, as otherwise the adaptive value of having the explicit system would be less than optimal.

There may also be a flow of influence from the explicit, verbal system to the implicit system, in that the explicit system may decide on a plan of action or strategy, and exert an influence on the implicit system that will alter the reinforcement evaluations made by and the signals produced by the implicit system (Rolls 2014b, 2018).

It may be expected that there is often a conflict between these systems, in that the first, implicit, emotional, system is able to guide behaviour particularly to obtain the greatest immediate reinforcement, whereas the explicit system can potentially enable immediate rewards to be deferred, and longer-term, multi-step, plans to be formed (Rolls 2019c, 2023b). This type of conflict will occur in animals with a syntactic planning ability, that is in humans and any other animals that have the ability to process a series of "if ... then" stages of planning. This is a property of the human language system, and the extent to which it is a property of non-human primates is not yet fully clear. In any case, such conflict may be an important aspect of the operation of at least the human mind, because it is so essential for humans to correctly decide, at every moment, whether to invest in a relationship or a group that may offer long-term benefits, or whether to directly pursue immediate benefits (Rolls 2014b, 2018, 2019c, 2023b).

The thrust of the argument (Rolls 2014b, 2018, 2019c, 2020) thus is that much complex animal including human behaviour can take place using the implicit, emotional, often unconscious, route to action. We should be very careful not to postulate intentional states (i.e., states with intentions, beliefs and desires) unless the evidence for them is strong, and it seems to me that a flexible, one-off, linguistic processing system that can handle propositions is needed for intentional states. What the explicit, linguistic, system does allow is exactly this flexible, one-off, multi-step planning ahead type of computation, which allows us to defer immediate rewards based on such a plan.

The selfish gene vs the selfish phene, and evolution

I have provided evidence above that there are two main routes to decision-making and goal-directed action. The first route selects actions by gene-defined goals for action, and is closely associated with emotion. The second route involves multistep planning and reasoning which requires syntactic processing to keep the symbols involved at each step separate from the symbols in different steps (Rolls 2019c, 2020, 2023b, d). (This second route is used by humans and perhaps by closely related animals.) Now the 'interests' of the first and second routes to decision-making and action are different. As argued very convincingly by Richard Dawkins in *The Selfish Gene* (Dawkins 1989), and by others (Hamilton 1964,

1996; Ridley 1993), many behaviours occur in the interests of the survival of the genes, not of the individual (nor of the group), and much behaviour can be understood in this way.

I have extended this approach by arguing that an important role for some genes in evolution is to define the goals (i.e. the rewards and punishers) for actions that will lead to better survival of those genes; that emotions are the states associated with these gene-defined goals; and that the defining of goals for actions rather than actions themselves is an efficient way for genes to operate, as it leaves flexibility of choice of action open until the animal is alive (Rolls 2014b, 2018). This provides great simplification of the genotype as action details do not need to be specified, just rewarding and punishing stimuli, and also flexibility of action in the face of changing environments faced by the genes. This is a useful and interesting advance beyond what was considered in *The Selfish Gene* and later books (Dawkins 1976, 1982, 1986, 1989). In any case, the interests that are implied when the first route to action is chosen are those of the “selfish genes”, not those of the individual, the phenotype.

However, the second route to action allows, by reasoning, decisions to be taken that might not be in the interests of the genes, might be longer term decisions, and might be in the interests of the individual, the phenotype (Rolls 2011, 2014b). An example might be a choice not to have children, but instead to devote oneself to science, medicine, music, or literature. The reasoning, rational, system presumably evolved because taking longer-term decisions involving planning rather than choosing a gene-defined goal might be advantageous at least sometimes for genes. But an unforeseen consequence of the evolution of the rational system might be that the decisions would, sometimes, not be to the advantage of any genes in the organism. After all, evolution by natural selection operates utilizing genetic variation like a Blind Watchmaker (Dawkins 1986). In this sense, the interests when the second route to decision-making is used are at least sometimes those of the “selfish phenotype”. (Indeed, we might euphonicly say that the interests are those of the “selfish phene” (where the etymology is Gk phaino, ‘appear’, referring to appearance, hence the thing that one observes, the individual. The term ‘phene’ was coined (Rolls 2011, 2014b) to refer to the individual or phenotype, but to emphasize that here we have an individual who can choose between the goals specified by the genes from earlier stages of evolution, and the goals that may be relevant to the reasoning individual who might make a choice using reasoning that might not be in the interests of those emotion-related genes.) Hence the decision-making referred to above is between a first system where the goals are gene-defined, and a second rational system in which the decisions may be made in the interests of the genes, or in the interests of the phenotype and not in the interests of the genes. Thus we may

speak of the choice as sometimes being between the “Selfish Genes” and the “Selfish Phene”.

Now what keeps the decision-making between the “Selfish Genes” and the “Selfish Phene” more or less under control and in balance? If the second, rational, system chose too often for the interests of the “Selfish Phene”, the genes in that phenotype would not survive over generations. Having these two systems in the same individual will only be stable if their potency is approximately equal, so that sometimes decisions are made with the first route, and sometimes with the second route. If the two types of decision-making, then, compete with approximately equal potency, and sometimes one is chosen, and sometimes the other, then this is exactly the scenario in which stochastic processes in the decision-making mechanism are likely to play an important role in the decision that is taken (Rolls and Deco 2010; Rolls 2014b, 2016c, 2019c, 2023b). The same decision, even with the same evidence, may not be taken each time a decision is made, because of noise in the system.

The system itself may have some properties that help to keep the system operating well. One is that if the second, rational, system tends to dominate the decision-making too much, the first, gene-based emotional system might fight back over generations of selection, and enhance the magnitude of the reward value specified by the genes, so that emotions might actually become stronger as a consequence of them having to compete in the interests of the selfish genes with the rational decision-making processes.

Another property of the system may be that sometimes the rational system cannot gain all the evidence that would be needed to make a rational choice. Under these circumstances the rational system might fail to make a clear decision, and under these circumstances, basing a decision on the gene-specified emotions is an alternative. Indeed, Damasio (1994) argued that under circumstances such as this, emotions might take an important role in decision-making. In this respect, I agree with him, basing my reasons on the arguments above. He called the emotional feelings gut feelings, and, in contrast to me, hypothesized that actual feedback from the gut was involved. His argument seemed to be that if the decision was too complicated for the rational system, then rely on outputs sent to the viscera, and whatever is sensed by what they send back could be used in the decision-making, and would account for the conscious feelings of the emotional states. My reading of the evidence is that the feedback from the periphery is not necessary for the emotional decision-making, or for the feelings, nor would it be computationally efficient to put the viscera in the loop given that the information starts from the brain, but that is a matter considered in the Appendix and elsewhere (Maia and McClelland 2004; Rolls 2014b, 2018).

Another property of the system is that the interests of the second, rational, system, although involving a different form

of computation, should not be too far from those of the gene-defined emotional system, for the arrangement to be stable in evolution by natural selection. One way that this could be facilitated would be if the gene-based goals felt pleasant or unpleasant in the rational system, and in this way contributed to the operation of the second, rational, system. This is something that I propose is the case (Rolls 2011, 2014b, 2018, 2020).

If the multistep syntactic reasoning/planning system fails, there may be a credit assignment problem, in that the faulty step in the series of steps needs to be identified. It has been argued that this computation requires a higher order thought system, which can think about the first order thought (the plan), and can correct it. This higher order thought system needs syntax, as it has to perform computations on the first order syntactic thoughts, the plan. I have proposed that it is a property of this Higher Order Syntactic Thought system that when it operates, it would feel like something, to be thinking about one's own first order thoughts, and that is a basis for my HOST theory of consciousness (Rolls 2007b, 2008, 2011, 2020). It is suggested that anything that is being dealt with by the HOST computational system becomes conscious, and that is also how qualia, raw sensory feels, are produced (Rolls 2020).

This raises the issue of what the relation is between the mind and the brain (Descartes 1644; Dennett 1991; Block 2005; Carruthers 2019). In the neuroscience-based approach that I propose for the relation between the mind and the brain, the proposal is that events at the sub-neuronal, neuronal, and neuronal network levels take place simultaneously to perform a computation that can be described at a high level as a mental state, with content about the world (Rolls 2021e, f). It is argued that as the processes at the different levels of explanation take place at the same time, they are linked by a non-causal relationship: causality can best be described in brains as operating within but not between levels. This mind-brain theory allows mental events to be different in kind from the mechanistic events that underlie them; but does not lead one to argue that mental events cause brain events, or vice versa: they are different levels of explanation of the operation of the computational system (Rolls 2021e, f). This computational neuroscience levels of explanation approach to causality (Rolls 2023d) provides an opportunity to proceed beyond Cartesian dualism (Descartes 1644) and physical reductionism (Carruthers 2019) in considering the relations between the mind and the brain (Rolls 2020, 2021e, f).

The dopamine system in a broader context of brain reward systems and emotion

The primate striatum receives from all cortical regions including motor and premotor cortical regions, and is

implicated in stimulus–response habit learning, probably by having the ability to associate any stimulus with any response (Rolls 2016c, 2021b, 2023d). Dopamine neurons, by responding to reward prediction errors, are thought to strengthen or weaken particular stimulus–response connections depending on whether the reward outcome is better or worse than predicted (Schultz 2013, 2016b, c, 2017). That hypothesis does not specify where the dopamine neurons receive their reward prediction error input from. It has been proposed that the dopamine system receives its inputs from the orbitofrontal cortex, partly directly, and partly via the ventral striatum and habenula (Rolls 2017), and this hypothesis is supported by the effective connectivity recently demonstrated in humans from the orbitofrontal cortex to the dopaminergic substantia nigra pars compacta (Rolls et al. 2023d) (Fig. 7). The proposal is also supported by the fact that the primate ventral striatum and adjoining part of the head of the caudate nucleus receive connections from the orbitofrontal cortex and amygdala (Haber and Knutson 2010; Rolls 2014b). Consistent with this, some neurons in these striatal regions respond to the taste, flavor, and/or sight of food (Rolls et al. 1983b; Rolls and Williams 1987; Williams et al. 1993; Rolls 2014b; Strait et al. 2015).

One comment is that this dopaminergic system is a small part of the circuitry involved in emotion-related responses described here, as is evident in Fig. 3, where it is shown as part of the habit system, which is not the instrumental system involved in action under control of the goal that is related to emotion in this paper. Other brain systems than the basal ganglia/dopamine system appear to be more involved in goal-dependent actions to obtain rewards and avoid punishers such as the orbitofrontal cortex, vmPFC and anterior cingulate cortex, and which are therefore closely involved in emotion. It is true that there are some dopamine inputs to the orbitofrontal and anterior cingulate cortex, but these regions already have the relevant reward/punisher information, so the roles of dopamine inputs to these regions are not clear. There are also some dopamine inputs to temporal cortex regions, but whether they implement some stamping in of connectivity as proposed for the striatum is not known.

A second comment is that although the striatum receives a dopaminergic input that it has been suggested is a positive reward prediction error signal (Schultz 2013), there may be too much diversity in the activity of dopamine neurons for this to apply in a simple way (Bromberg-Martin et al. 2010; Rolls 2014b, 2023d). Moreover, there is no evidence that the dopamine neurons encode a specific reward signal (for example for the taste of food vs. the texture of fat) in the way that is required to account for the control of goal-directed motivated behaviour and that is present in the primate orbitofrontal cortex (Rolls 2014b). Further, the activity of ventral striatal neurons appears to be more influenced by orbitofrontal cortex types of signals rather than by positive

reward prediction error signals (Strait et al. 2015). The role of the striatum and dopamine in the control of behaviour is considered in more detail elsewhere (Rolls 2014b, 2021b, 2023d), but they appear to be much less important than the orbitofrontal cortex and anterior cingulate cortex in emotion, as described here.

Decision-making and noise in the brain

The attractor model of decision-making is a neuronal network with associatively modifiable recurrent collateral synapses between the neurons of the type prototypical of the cerebral cortex as shown in Fig. 11 (Wang 2002; Rolls and Deco 2010; Rolls 2021b). The decision variables are applied simultaneously, and the network, after previous training with these decision variables, reaches a state where the population of neurons representing one of the decision variables has a high firing rate (Rolls and Deco 2010; Deco et al. 2013; Rolls 2016c, 2021b). There is noise or randomness in this model of decision-making that is related to the approximately Poisson distributed firing times of neurons for a given mean firing rate (Rolls and Deco 2010; Deco et al. 2013; Rolls 2016c, 2021b). It is this noise that makes decision-making probabilistic.

An implication is that if the odds are almost equal, it is wise to take any decision at least 3 times, as noise in the brain might have influenced a single decision. Another implication is that variability in behaviour can be produced by the randomness in this type of decision-making, and this is important for understanding the variability of emotional states, for understanding how decisions are made between the emotional and reasoning systems in our brains, and for understanding many related processes including the advantageous unpredictability of some behaviour, and how creativity is facilitated in the brain by this ‘noise in the brain’ (Rolls and Deco 2010).

The neurology of human emotion

Some disorders of human emotion produced by brain damage or disease can be understood using the approach taken here (Rolls 2021c). Humans with damage to the orbitofrontal cortex and anterior cingulate cortex fail to reverse correctly in a stimulus-reward reversal task, revealing that they cannot change their behaviour rapidly when the reinforcement contingencies change (Rolls et al. 1994; Hornak et al. 2004; Fellows 2011). These patients also have an impaired ability to identify facial and voice expressions of emotions, and this is likely to contribute to their changes in social behaviour (Hornak et al. 1996, 2003; Tsuchida and Fellows 2012). For these reasons, these patients are often impulsive and disinhibited, have an altered personality, and have impaired subjective feelings of emotion.

In an fMRI study, we showed that the human orbitofrontal cortex is especially involved in pleasant touch and pain, relative to the somatosensory cortex which is more activated by physically strong somatosensory stimuli (Rolls et al. 2003c; Rolls 2010b, 2016e). This raises the issue of where the emotional aspects of pain are represented in the brain. The orbitofrontal cortex, together with the supracallosal anterior cingulate cortex (Vogt and Sikes 2000; Rolls et al. 2003c, 2023d), are thereby implicated in pain processing, and consistent with this, clinical reports provide evidence that patients with orbitofrontal cortex damage know when a painful stimulus is applied, but have a reduced emotion pain reaction. There is room for this evidence on the important contributions of the orbitofrontal cortex in the emotional, affective, subjective, aspects of pain to be incorporated into investigations of pain systems in the brain (Segerdahl et al. 2015; Tracey 2017).

In humans, amygdala damage has much less effect on emotion than does orbitofrontal cortex damage (Rolls 2021c). For example, the effects of damage to the human amygdala on emotion and emotional experience are much more subtle (Whalen and Phelps 2009; Delgado et al. 2011; LeDoux and Pine 2016; LeDoux et al. 2018) than the effects of damage to the orbitofrontal cortex (Rolls et al. 1994; Hornak et al. 1996, 2003, 2004; Camille et al. 2011; Fellows 2011; Rolls 2019b). Indeed, LeDoux and colleagues have emphasized the evidence that the human amygdala is rather little involved in subjective emotional experience (LeDoux and Pine 2016; LeDoux and Brown 2017; LeDoux et al. 2018). This is in strong contrast to the orbitofrontal cortex, which is involved in subjective emotional experience, as described above. The orbitofrontal cortex provides the answer to LeDoux’s conundrum: if not the amygdala for subjective emotional experience, then what? The role of the amygdala in the processing of emotions may be reduced in humans because of the great evolutionary development of the orbitofrontal cortex, which with its cortical design contains attractor networks that are useful in remembering previous emotion-related inputs, and that are useful in decision-making (Rolls 2021c, 2023d).

A psychiatric disorder of emotion: depression

Not obtaining an expected reward can lead to sadness, and feeling depressed (Rolls 2018; Rolls et al. 2020b) (Fig. 1). The concept is advanced that an important brain region in depression is the orbitofrontal cortex, with depression related to over-responsiveness and over-connectedness of the non-reward-related lateral orbitofrontal cortex, and to under-responsiveness and under-connectivity of the reward-related medial orbitofrontal cortex (Rolls 2016b, 2017, 2019d; Zhang et al. 2023). Evidence from large-scale voxel-level studies and supported by an activation study has been

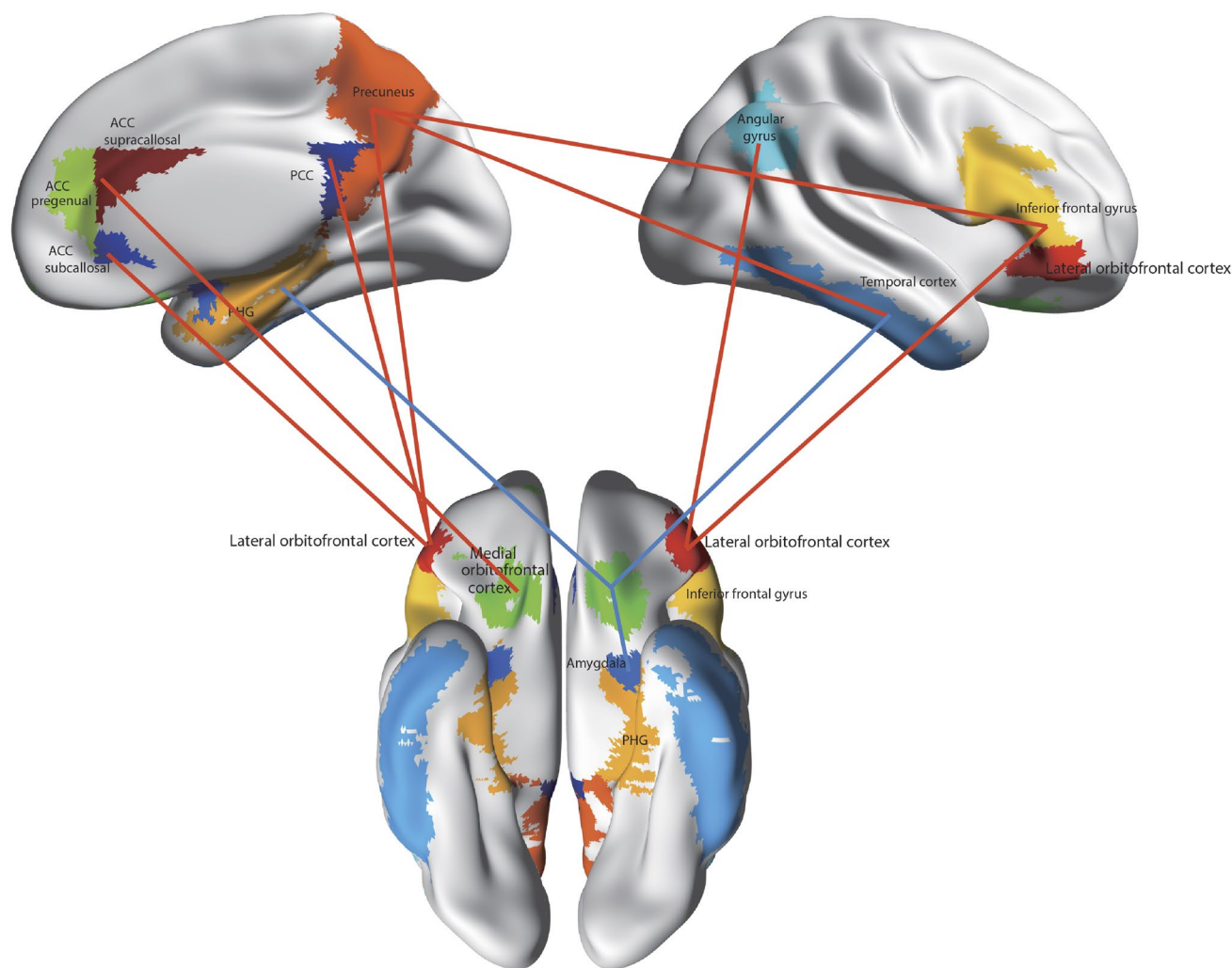


Fig. 14 Functional connectivity (FC) differences of the medial and lateral orbitofrontal cortex in major depressive disorder. Higher functional connectivity in depression is shown by red connecting lines, and includes higher functional connectivity of the non-reward/punishment-related lateral orbitofrontal cortex with the precuneus, posterior cingulate cortex (PCC), subcallosal anterior cingulate cortex (ACC), angular gyrus, and inferior frontal gyrus. Lower functional connectivity in depression is shown by blue connecting lines, and includes lower functional connectivity of the medial orbitofrontal cortex with

the parahippocampal gyrus memory system (PHG), amygdala, temporal cortex and supracallosal anterior cingulate cortex (ACC). The part of the medial orbitofrontal cortex in which voxels were found with lower functional connectivity in depression is indicated in green. The areas apart from the medial orbitofrontal cortex shown are as defined in the automated anatomical labelling atlas (Rolls et al. 2015), although the investigations that form the basis for the summary were at the voxel level

described that provides support for this hypothesis (Rolls 2016b, 2017; Rolls et al. 2020b) (Fig. 14).

Increased functional connectivity of the lateral orbitofrontal cortex with brain areas that include the precuneus and posterior cingulate cortex and angular gyrus is found in patients with depression, and is reduced towards the levels in controls when treated with medication (Cheng et al. 2016, 2018a, 2018b; Rolls et al. 2020a). This is interpreted as related to negative self-esteem and enhanced language-related rumination in depression (Cheng et al. 2016; Rolls et al. 2020b; Zhang et al. 2023).

Decreased functional connectivity of the medial orbitofrontal cortex with medial temporal lobe areas involved in memory is found in patients with depression (Cheng et al. 2016). This is interpreted as being related to less processing or efficacy of systems involved in happy memories in depression (Rolls et al. 2020b; Zhang et al. 2023).

In an activation study with more than 1000 participants, it was found that in individuals with some symptoms of depression, the lateral orbitofrontal cortex has increased sensitivity to not winning, and the medial orbitofrontal cortex has decreased sensitivity to winning in the monetary incentive delay task (Xie et al. 2021). This provides support

for Rolls' theory of depression (Rolls 2016b, 2018; Rolls et al. 2020b).

Some treatments for depression may act by reducing activity or connectivity of the lateral orbitofrontal cortex (Rolls et al. 2020b). New treatments are needed that increase activity or connectivity of the medial orbitofrontal cortex as a possible treatment for depression (Rolls et al. 2020b), and one possibility is that ketamine implements this (Zhang et al. 2023).

These concepts, and that of increased activity in non-reward attractor networks, have potential for advancing our understanding and treatment of depression. Indeed, the hypothesis is developed that the orbitofrontal cortex has a special role in emotion and decision-making and depression in part because as a cortical area and because of its connectivity with other cortical regions it can implement attractor networks useful in maintaining reward and emotional states online including ruminating thoughts, and in decision-making (Rolls et al. 2020b; Rolls 2021b). Maintaining language-related ruminating thoughts because of cortico-cortical feedback loops involving attractor networks may make depressive states particularly severe in humans (Rolls 2016b, 2018).

Role of reward and emotion in episodic and semantic memory

The human orbitofrontal cortex connecting with the vmPFC and anterior cingulate cortex provides a route to the hippocampus for reward and emotional value to be incorporated into episodic memory, enabling memory of where a reward was seen (Rolls 2022b; Rolls et al. 2023b, 2023d) (Figs. 5 and 15). In particular, the green arrows in Fig. 15 show how reward regions of the orbitofrontal cortex, vmPFC (pOFC, 10r, 10v) and pregenual anterior cingulate cortex (a24 and p32), and punishment/non-reward regions of the lateral orbitofrontal cortex (47m) have effective connectivity with the hippocampus, entorhinal cortex, and perirhinal cortex. Consistent with this, some neurons in the primate hippocampus respond to a combination of a spatial view and the reward value that is available at that location in a scene location—reward memory task (Rolls and Xiang 2005). It is argued that reward, punishment, and more generally emotional value are important components of episodic memory (Rolls 2022b).

Beyond this, it is proposed that this reward value component results in primarily episodic memories with some value component to be repeatedly recalled from the hippocampus so that they are more likely to become incorporated into neocortical semantic and autobiographical memories (Rolls 2022b). This is thus a theory of how reward or utility value is important in influencing what information is stored in

semantic long-term memory (Rolls 2022b), which is a key aspect of memory consolidation.

The same orbitofrontal and anterior cingulate regions also connect in humans to the septal and basal forebrain cholinergic nuclei (Rolls et al. 2023d) (Fig. 7), which in turn project to the hippocampus and neocortex respectively, where acetylcholine is known to be important in memory consolidation (Hasselmo and Sarter 2011; Zaborszky et al. 2018). It is therefore proposed that key brain systems in the orbitofrontal cortex, vmPFC, and anterior cingulate cortex involved in reward value decoding and emotion play key roles in consolidation of information into hippocampal episodic and also semantic long-term neocortical memory (Rolls 2022b). This also helps to account (Rolls 2022b) for why damage to the vmPFC and anterior cingulate cortex impairs memory (Bonnici and Maguire 2018; McCormick et al. 2018).

Brain systems for emotion and motivation in primates including humans compared to those in rodents

Emphasis is placed on research in primates and humans, because there is evidence that the rodent taste and food reward systems operate somewhat differently in primates and humans vs rodents (Rolls 2014b, 2015, 2016d, 2023d). In brief, the taste system is different in rodents in that there is a pontine taste area which then projects subcortically, but in primates there is no pontine taste area and cortical processing is performed first (Fig. 16).

Second, in rodents, the taste and olfactory systems are modulated peripherally [in the nucleus of the solitary tract and the olfactory bulb respectively (Pager et al. 1972; Palouzier-Paulignan et al. 2012; Rolls 2015)] by hunger so that reward is represented peripherally and is entangled with sensory processing, whereas in primates and humans food perception is separated from its reward value, as described elsewhere (Rolls 2014b, 2016d, 2019b, 2021b, 2023d) (Fig. 16). A perceptual correlate of this is that when humans feed to satiety, the intensity of the flavor changes very little, whereas the pleasantness of the flavor decreases to zero (Rolls et al. 1983a; Rolls and Rolls 1997), showing that in humans perceptual representations of taste and olfaction are kept separate from hedonic representations. This is adaptive, in that we do not go blind to the sight, taste, and smell of food after eating it to satiety, and can therefore still learn about where food is located in the environment even when we are not hungry (Rolls 2014b).

Third, the orbitofrontal cortex is very little developed in rodents (with only an agranular part) (Wise 2008), yet is one of the major brain areas involved in taste and olfactory processing, and emotion and motivation, in primates, including humans (Rolls 2014b, 2023d). These findings make the rodent taste and olfactory system a poor model of neural reward processing in humans, and for that reason emphasis

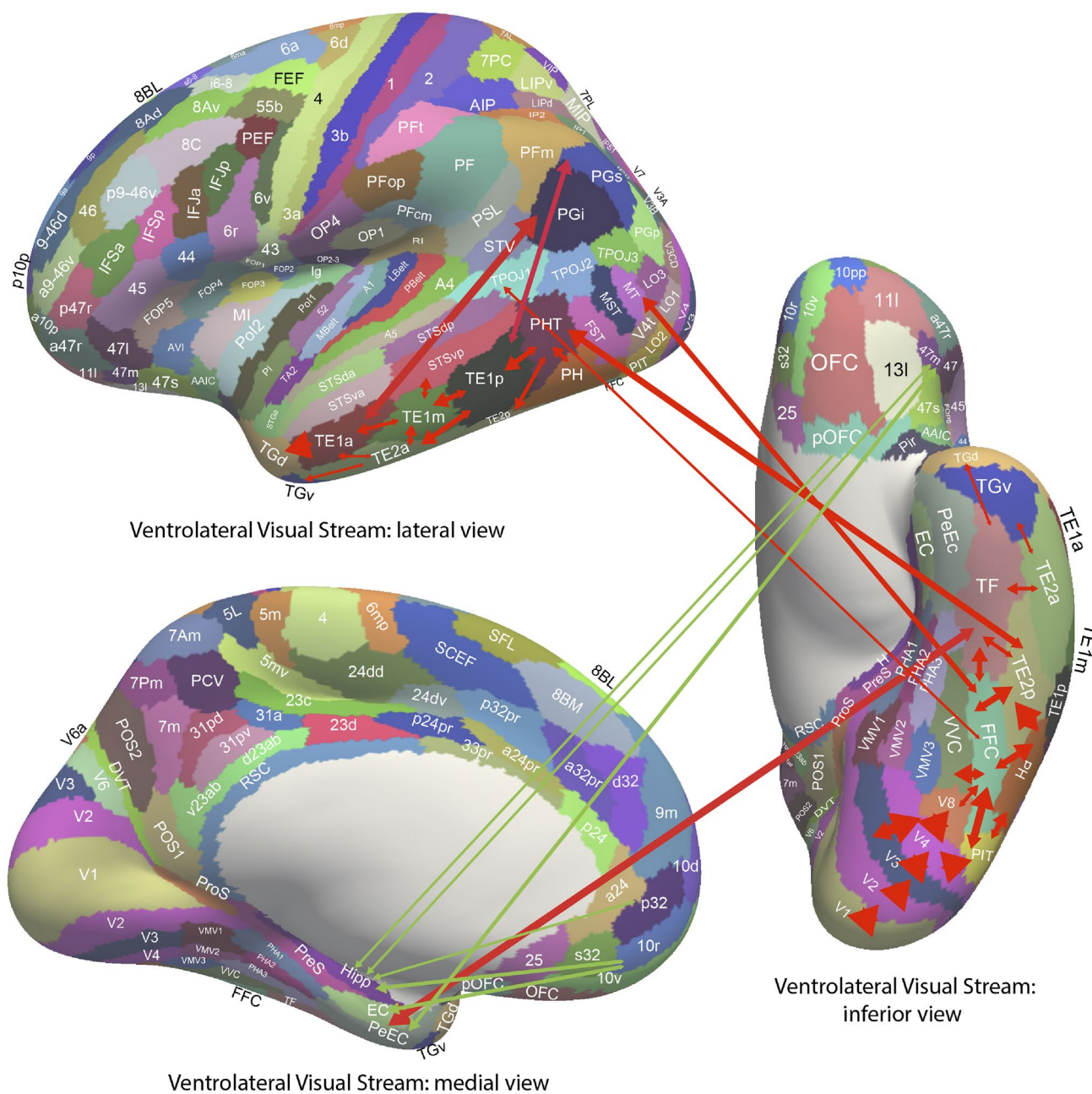


Fig. 15 Effective connectivity of the Ventrolateral Visual Stream which reaches inferior temporal cortex TE regions in which objects and faces are represented (red arrows): schematic overview. One of the red arrows shows how the Ventrolateral Visual Stream provides ‘what’ input to the hippocampal memory system via parahippocampal gyrus TF to perirhinal PeEc connectivity from FFC, PH, TE1p, TE2a and TE2p. The green arrows show how reward regions of the orbitofrontal cortex, vmPFC (pOFC, 10r, 10v) and pregenual anterior cingulate (a24 and p32), and punishment/non-reward regions of the

lateral orbitofrontal cortex (47m) have effective connectivity with the hippocampus (Hipp), entorhinal cortex (EC), and perirhinal cortex (PeEC). The Ventrolateral Visual Stream also provides input to the semantic language system via TGd. The Ventrolateral Visual Stream also has connectivity to the inferior parietal visual area PFm, PGs and PGi as indicated by 2 green arrows. The widths of the lines and the size of the arrowheads indicate the magnitude and direction of the effective connectivity. After Rolls et al (2023i)

is placed here on discoveries in primates and humans (Rolls 2014b, 2015, 2016d, 2019b, 2021b, 2023d). For example, in rodents, movements and actions may be represented in the

‘orbitofrontal cortex’ or what in fact may be the agranular insula (Wilson et al. 2014; Sharpe et al. 2015).

Fourth, the visual system is very different in primates including humans vs rodents, partly because of foveate

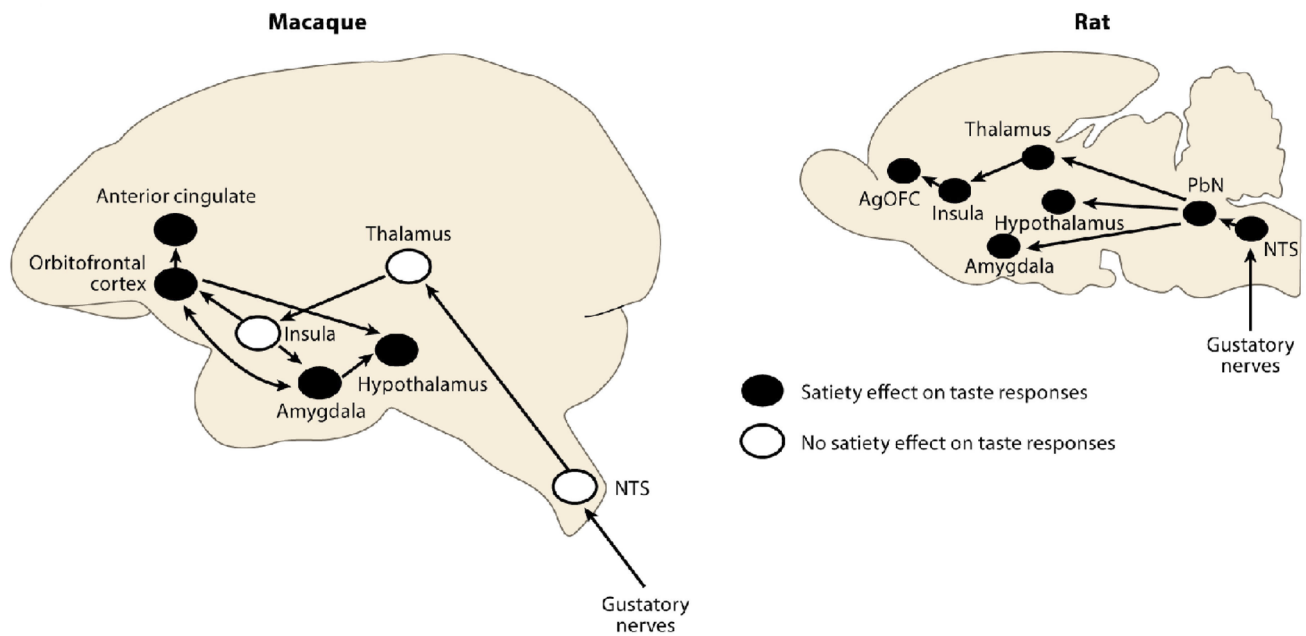


Fig. 16 Taste pathways in the macaque and rat. In the *macaque*, gustatory information reaches the nucleus of the solitary tract (NTS), which projects directly to the taste thalamus (ventral posteromedial nucleus, pars parvocellularis, VPMpc) which then projects to the taste cortex in the anterior insula (Insula). The insular taste cortex then projects to the orbitofrontal cortex and amygdala. The orbitofrontal cortex projects taste information to the anterior cingulate cortex. Both the orbitofrontal cortex and the amygdala project to the hypothalamus (and to the ventral striatum). In macaques, feeding to normal self-induced satiety does not decrease the responses of taste neurons in the NTS or taste insula (and by inference not VPMpc) (see text). In the *rat*, in contrast, the NTS projects to a pontine taste area, the parabrachial nucleus (PbN). The PbN then has projections

directly to a number of subcortical structures, including the hypothalamus, amygdala, and ventral striatum, thus bypassing thalamo-cortical processing. The PbN in the rat also projects to the taste thalamus (VPMpc), which projects to the rat taste insula. The taste insula in the rat then projects to an agranular orbitofrontal cortex (AgOFC), which probably corresponds to the most posterior part of the primate OFC, which is agranular. (In primates, most of the orbitofrontal cortex is granular cortex, and the rat may have no equivalent to this Wise 2008; Small and Scott 2009; Passingham and Wise 2012; Rolls 2014b, 2015). In the rat, satiety signals such as gastric distension and satiety-related hormones decrease neuronal responses in the NTS (see text), and by inference therefore in the other brain areas with taste-related responses, as indicated in the Figure

vision (Rolls 2023a, d). In primates, ventral visual stream object and face recognition become computationally tractable because the fovea provides for fixation in which only one or a very few objects are present, and performs object recognition with a relatively small receptive field in complex natural scenes (Rolls et al. 2003a; Aggelopoulos et al. 2005; Aggelopoulos and Rolls 2005; Rolls 2021b, a, 2023d). In the primate dorsal visual system, there is also great development to provide mechanisms to provide for the fixation of objects in complex natural scenes, and then to perform actions such as reaching and grasping of fixated objects (Itti and Koch 2000; Bisley and Goldberg 2010; Rolls and Webb 2014; Fattori et al. 2017; Galletti and Fattori 2018; Gamberini et al. 2021; Rolls 2023d; Rolls et al. 2023b). Foveate vision has an enormous impact too on other brain systems, including the hippocampal memory system (Rolls 2021b, 2023a, c). There are even differences at a trivial level, in that locomotion greatly increases the firing of rodent visual

cortex (Zatka-Haas et al. 2021), whereas the primate inferior temporal visual cortex is little affected by whether an action is performed or not (Rolls et al. 1977, 2003a; Aggelopoulos et al. 2005; Aggelopoulos and Rolls 2005), and instead elegantly represents the identity of the stimulus that is being shown without the great interference described for the rodent. Further, rodents have no posterior cingulate cortex (Vogt 2009; Rolls 2021b). Further, the prefrontal cortex appears to be very different in primates including humans compared to rodents (Passingham 2021).

Fifth, understanding of the functions of different subregions of the rodent orbitofrontal cortex is still limited (Izquierdo 2017; Barreiros et al. 2021).

For these reasons, emphasis here is placed on systems-level investigations in primates and humans, for there is evidence that many cortical systems operate so differently in rodents (Passingham 2021; Rolls 2021b, 2023d). Some of these differences are set out in Sect. 19.10 of Rolls (2023d).

Implications for welfare

An implication of the above approach to emotion and motivation is that when considering animal welfare, it is likely to be important to take into account what value each species places on different rewards and the avoidance of possible aversive stimuli. This can in principle be measured by measuring the choices that animals make between different rewards or avoiding different potential punishers. The procedures are well known in neuroeconomics, in which it is possible to measure for example how many drops of fruit juice A are chosen equally often as two drops of fruit juice B (Padoa-Schioppa and Assad 2006; Padoa-Schioppa 2007; Padoa-Schioppa and Assad 2008; Padoa-Schioppa 2011; Padoa-Schioppa and Cai 2011; Cai and Padoa-Schioppa 2012; Glimcher and Fehr 2013; Rolls 2014b; Platt et al. 2016; Padoa-Schioppa and Conen 2017; Yamada et al. 2018; Cai and Padoa-Schioppa 2019; Kuwabara et al. 2020; Dawkins 2021). Similar titration procedures could be used to measure what value, measured by choice, a species places on for example food vs bedding vs having other animals nearby vs overcrowding vs being able to take a swim or shower vs being able to sit on a perch vs being able to reach a branch high above the ground vs being able to perform reproductive behaviour, etc. (Dawkins 2023). When measuring these choices, it is important to ensure that the choice is being made by the goal-directed reward system for instrumental action, and not by any system involved in a reflex or fixed action pattern, or a learned habit.

Measuring instrumental goal-directed choices made by particular species may be useful to minimize over-anthropomorphic inferences about the value that a species may place on different ‘goods’ (the term used in neuroeconomics). Further, even the evidence taken from humans may need to be carefully assessed, for humans are able to provide reasons with their declarative system for their choices made with their syntactic learning system, but may confabulate reasons why they chose a good when the choice has been made by the emotional or by the automatic habit system (Gazzaniga and LeDoux 1978; Rolls 2010a, 2011, 2020, 2023d).

Another implication is that the taste, olfactory and food texture systems present in different species may result in adequate nutrition in their natural environment, but care may be needed to ensure in other environments that the nutrition being made available is appropriate. In this context, it must be remembered that animals do not have flavour mechanisms built to ensure that every possible nutrient needed is being selected by specific appetites for different nutrients. Instead, in the natural environment animals condition to new foods that provide useful nutrients by physiological effects that may occur some time after the food is ingested (Berthoud et al. 2021; Rolls 2023d).

When considering emotional states, we should remember that there are a set of hierarchically organised neural systems that connect what might be defined as emotion-provoking inputs to different types of output, as illustrated in Fig. 2. I define emotional states as states elicited by rewards and punishers that guide goal-directed instrumental behaviour and that are under the control of the reward (or punishment) value of the goal. I refer here therefore to action–outcome learning where the behaviour is under the control of the reward value of the goal in that devaluation of the goal (e.g. by feeding to satiety) results in an immediate cessation of the behaviour, which is not the case for habits or stimulus–response behaviour (Fig. 2). Emotional states are therefore the states involved in implementing this type of goal-dependent instrumental behaviour. We must be very careful to distinguish these emotional states from further states that are related to the subjective feelings, that is to states that in humans are declarative and can be reported and are described as conscious (Rolls 2020). Thus the word ‘emotion’ is ambiguous, and it is always essential to make it clear whether the emotional state is one that might link a stimulus input to for example goal-directed goal-dependent behaviour that utilises action–outcome learning, from a state involved in subjective experience, that may involve further brain systems (Fig. 2). The corresponding situation arises for the word motivation, which as used here and by others refers to goal-directed behaviour of the type just described, with the word ‘drive’ used for simpler behaviors such as approach to food that may not require internal functional states of the type involved in action–outcome learning (Teitelbaum 1974). Thus, for the word ‘motivation’, we should always distinguish systems involved in goal-directed actions, from subjective feelings of being motivated, having ‘desires’.

A set of criteria for achieving good welfare in farm animals, known as the Five Freedoms (Farm Animal Welfare Council 2009) consist of: 1. Freedom from hunger and thirst. 2. Freedom from discomfort. 3. Freedom from pain, injury and disease. 4. Freedom to express normal behaviour. 5. Freedom from fear and distress. The present approach suggests that when assessing (4), it will be useful to measure the value of the different types of ‘normal behaviour’ to help assess priorities. The present approach suggests that when assessing (5), farm animals may often be protected from the fears, stressors, and predators that are present in the natural world, but that these provide a scale against which other fear and distress might be calibrated.

Conclusions and highlights

1. A new approach is taken here to produce a unified understanding of emotion and motivation and their underlying brain mechanisms. In this unified theory

of emotion and motivation, motivational states are states in which instrumental goal-directed actions are performed to obtain rewards or avoid punishers, and emotional states are states that are elicited when the reward or punisher is or is not received. This greatly simplifies our understanding of emotion and motivation, for the same set of genes and associated brain systems can define the primary or unlearned rewards and punishers such as sweet taste or pain that can be used for both emotion and motivation.

2. New evidence on the connectivity in humans of brain systems involved in emotion and motivation is described, which measures the effective connectivity between 360 cortical regions in the Human Connectome Project MultiModal Parcellation atlas (HCP-MMP) (Glasser et al. 2016a), and is complemented by the addition of 66 subcortical regions (Huang et al. 2022). The cortical regions in this atlas are defined by anatomical characteristics (cortical myelin content and cortical thickness), functional connectivity, and task-related fMRI, and provide a useful basis for understanding brain regions with different connectivity and potentially different computational functions. Some of the following points reflect advances in our understanding of brain systems involved in emotion by taking into account the effective connectivity of the human brain, complemented by functional connectivity and diffusion tractography (Rolls et al. 2022a, 2023a, b, d, e, f, i).
3. It is shown that the primate including human orbitofrontal cortex represents primary reinforcers such as taste, pain, and pleasant touch, with this information reaching the orbitofrontal cortex from the primary taste cortex in the anterior insula and from somatosensory cortical regions. It is shown that the primate including human orbitofrontal cortex learns associations between these primary reinforcers and secondary reinforcers such as the sight of food or of an aversive stimulus in one trial, and can reverse these associations in one trial using a rule- or model-based computation. These stimulus-stimulus learned representations are of expected value. The representations in the orbitofrontal cortex are value-based, and are appropriate for being the goals for motivated behaviour, and for eliciting emotional states. Actions are not represented in the primate orbitofrontal cortex. Other inputs to the orbitofrontal cortex are about socially relevant stimuli such as face expression and face identity, and relate to inputs from the cortex in the superior temporal sulcus. Rewards tend to be represented in the human medial orbitofrontal cortex, and punishers and non-reward in the lateral orbitofrontal cortex. This evidence is complemented by the effects of damage to the orbitofrontal cortex in humans, which impairs reward-related reversal learning, emotional responses, and subjective emotional feelings. In primates, reward and punisher value is not represented in cortical stages of sensory processing prior to the orbitofrontal cortex, such as the insular primary taste cortex and inferior temporal visual cortex.
4. The ventromedial prefrontal cortex (vmPFC) receives from the orbitofrontal cortex, is activated by rewards, is implicated in reward-related decision-making, and has connectivity to the pregenual and supracallosal anterior cingulate cortex (Rolls et al. 2023d) (Fig. 7).
5. The human medial and lateral orbitofrontal cortex, and the vmPFC, have connectivity to the pregenual anterior cingulate cortex, which is strongly activated by rewards, and which projects to the hippocampal system, both directly, and via the posterior cingulate cortex (Figs. 5, 6, 7) (Rolls et al. 2023d). It is proposed that this provides the route for rewards and emotional states to become part of episodic memory. It is further proposed that the reward/emotional value of recalled episodic memories is important in influencing which memories are further processed and become incorporated into long-term semantic memory. It is further proposed that this route enables goals for navigation to enter the human hippocampal system, and indeed navigation is almost always to obtain goals, and which are reflected in hippocampal neuronal activity (Rolls 2023a, c).
6. The human pregenual anterior cingulate cortex has effective connectivity to the septum, from which cholinergic neurons important in memory consolidation project to the hippocampus (Fig. 7) (Rolls et al. 2023d). The human medial orbitofrontal cortex (region pOFC) has effective connectivity to the basal forebrain magnocellular nucleus of Meynert, from which cholinergic neurons important in memory consolidation project to the hippocampus (Fig. 7) (Rolls et al. 2023d). It is proposed that by these routes the value system can influence memory consolidation. Consistent with this, damage to the vmPFC/anterior cingulate cortex in humans impairs memory. It is argued that the human orbitofrontal cortex/vmPFC/pregenual anterior cingulate cortex is not a memory system, but a value system, and that this value system influences memory and memory consolidation by these connectivities (Rolls 2022b).
7. The orbitofrontal cortex and pregenual anterior cingulate cortex have connectivity in humans to the sup-

racallosal anterior cingulate cortex, which in turn has connectivity to premotor cortical regions including the midcingulate premotor cortex. It is proposed that these routes provide for action–outcome learning in the supracallosal anterior cingulate cortex, where the outcome is the reward or punisher received from the orbitofrontal cortex and pregenual anterior cingulate cortex.

8. With this foundation, it is proposed that the function of the primate orbitofrontal cortex in emotion is to represent rewards and punishers, and to implement stimulus – reward/punisher association learning and reversal (i.e. stimulus-stimulus learning). It is argued that in contrast, the role of the supracallosal anterior cingulate cortex is to learn associations between actions and the rewards/punishers that follow the actions, and with this action–outcome learning to influence the future choice of actions when reward/aversive expected value stimuli are received from the orbitofrontal cortex.
9. It is shown that the human amygdala has effective connectivity from relatively few cortical regions, primarily those in the anterior temporal lobe, and even less effective connectivity back to the neocortex. The outputs of the human amygdala are directed primarily to brainstem regions involved in autonomic responses, cortical arousal, and some behavioural responses. In line with this, there is evidence that the human amygdala is much less involved in reported, experienced, declarative emotion than the orbitofrontal cortex. This is a key re-evaluation of the functions of the human amygdala in human emotion (Rolls, Deco, Huang and Feng 2023a).
10. It is shown that in addition to these emotion-related outputs to behaviour, in humans and perhaps in other animals there is a rational, reasoning, route to action, that may over-ride the genes selected during evolution to specify the rewards and punishers important in the control of goal-directed behaviour. The reasoning route to action may make choices in the interests of the individual, the phenotype, not in the interests of the gene-specified rewards.
11. Damage to the orbitofrontal cortex in humans can produce neurological changes such as reduced ability to respond correctly to emotion-relevant stimuli such as face and voice expression, and to learn and change behaviour in response to reinforcement contingencies. It is shown that altered connectivity of the orbitofrontal cortex with other brain regions, and sensitivity of the medial orbitofrontal cortex to rewards and of the lateral orbitofrontal cortex to punishers is involved in human depression.
12. In relation to welfare, it is proposed that measurement by choice of the value of goal-related options is important to consider, and needs to be distinguished from other routes to responses such as fixed action patterns, reflexes, taxes, and habits that result from over-training.

Appendix

Other theories of emotion

For completeness, I now outline some other theories of emotion, and compare them with the above (Rolls') theory of emotion. Surveys of some of the approaches to emotion that have been taken in the past are provided by Strongman (2003) and Keltner et al. (2018).

The James-Lange and other bodily theories of emotion including Damasio's theory

James (1884) believed that emotional experiences were produced by sensing bodily changes, such as changes in heart rate or in skeletal muscles. Lange (1885) had a similar view, although he emphasized the role of autonomic feedback (for example from the heart) in producing the experience of emotion. The theory, which became known as the James-Lange theory, suggested that there are three steps in producing emotional feelings. The first step is elicitation by the emotion-provoking stimulus of peripheral changes, such as skeleto-muscular activity to produce running away, and autonomic changes, such as alteration of heart rate. But, as pointed out above, the theory leaves unanswered perhaps the most important issue in any theory of emotion: Why do some events make us run away (and then feel emotional), whereas others do not? This is a major weakness of this type of theory. The second step is the sensing of the peripheral responses (e.g. running away, and altered heart rate). The third step is elicitation of the emotional feeling in response to the sensed feedback from the periphery.

The history of research into peripheral theories of emotion starts with the fatal flaw that step one (the question of which stimuli elicit emotion-related responses in the first place) leaves unanswered this most important question. The history continues with the accumulation of empirical evidence that has gradually weakened more and more the hypothesis that peripheral responses made during emotional behaviour have anything to do with producing the emotional behaviour (which has largely already been produced anyway

according to the James-Lange theory), or the emotional feeling. Some of the landmarks in this history are described by Rolls (2014b).

First, the peripheral changes produced during emotion are not sufficiently distinct to be able to carry the information that would enable one to have subtly different emotional feelings to the vast range of different stimuli that can produce different emotions. The evidence suggests that by measuring many peripheral changes in emotion, such as heart rate, skin conductance, breathing rate, and hormones such as adrenaline and noradrenaline (known in the United States by their Greek names epinephrine and norepinephrine), it may be possible to make coarse distinctions between, for example, anger and fear, but not much finer distinctions (Keltner et al. 2018). Brain processing must of course produce the somewhat different autonomic responses in the first place, and there is evidence that the orbitofrontal and anterior cingulate cortices, perhaps acting via an insular visceral cortex region, are involved in producing autonomic responses (Critchley and Harrison 2013; Quadt et al. 2018). Of course there are pathways from the viscera to the brain, and visceral changes can influence the brain (Critchley and Harrison 2013; Quadt et al. 2018; Quadt, Critchley and Nagai 2022), but whether those visceral changes are in the normal causal chain for the elicitation of emotional states is much more difficult to prove.

Second, when emotions are evoked by imagery, then the peripheral responses are much less marked and distinctive than during emotions produced by external stimuli (Ekman et al. 1983; Levenson et al. 1990). This makes sense in that although an emotion evoked by imagery may be strong, there is no need to produce strong peripheral responses, because no behavioural responses are required.

Third, disruption of peripheral responses and feedback from them either surgically (for example in dogs (Cannon 1927, 1931), or as a result of spinal cord injury in humans (Hohmann 1966; Bermond et al. 1991), does not abolish emotional responses. What was found was that in some patients there was apparently some reduction in emotions in some situations (Hohmann 1966), but this could be related to the fact that some of the patients were severely disabled (which could have produced its own consequences for emotionality), and that in many cases the patients were considerably older than before the spinal cord damage, and this could have been a factor. What was common to both studies was that emotions could be felt by all the patients; and that in some cases, emotions resulting from mental events were even reported as being stronger (Hohmann 1966; Bermond et al. 1991).

Fourth, when autonomic changes are elicited by injections of, for example, adrenaline or noradrenaline, particular emotions are not produced. Instead, the emotion that is produced depends on the cognitive decoding of the reinforcers

present in the situation, for example an actor who insults your parents to make you angry, or an actor who plays a game of hula hoop to make you feel happy (Schachter and Singer 1962). In this situation, the hormone adrenaline or noradrenaline can alter the magnitude of the emotion, but not which emotion is felt. This is further evidence that it is the decoded reinforcement value of the input stimulus or events that determines which emotion is felt. The fact that the hormone injections produced some change in the magnitude of an emotion is not very surprising. If you felt your heart pounding for no explicable reason, you might wonder what was happening, and therefore react more or abnormally.

Fifth, if the peripheral changes associated with emotion are blocked with drugs, then this does not block the perception of emotion (Reisenzein 1983).

Sixth, it is found that in normal life, behavioural expressions of emotion (for example smiling when at a bowling alley) do not usually occur when one might be expected to feel happy because of a success, but instead occur when one is looking at one's friends (Kraut and Johnson 1979). These body responses, which can be very brief, thus often serve the needs of communication, or of action, not of producing emotional feelings.

Despite this rather overwhelming evidence against an important role for body responses in producing emotions or emotional feelings, Damasio (1994) effectively tried to resurrect a weakened version of the James-Lange theory of emotion from the nineteenth century, by arguing with his *somatic marker hypothesis* that after reinforcers have been evaluated, a bodily response ('somatic marker') normally occurs, then this leads to a bodily feeling, which in turn is appreciated by the organism to then make a contribution to the decision-making process. [In the James-Lange theory, it was emotional feelings that depend on peripheral feedback; for Damasio, it is the decision of which behavioural response to make that is normally influenced by the peripheral feedback. A quotation from Damasio (1994, p 190) follows: "The squirrel did not really think about his various options and calculate the costs and benefits of each. He saw the cat, was jolted by the body state, and ran." Here it is clear that the pathway to action uses the body state as part of the route. Damasio would also like decisions to be implemented using the peripheral changes elicited by emotional stimuli. Given all the different reinforcers that may influence behaviour, Damasio (1994) even suggests that the net result of them all is reflected in the net peripheral outcome, and then the brain can sense this net peripheral result, and thus know what decision to take.] The James-Lange theory has a number of major weaknesses just outlined that apply also to the somatic marker hypothesis.

The somatic marker hypothesis postulates that emotional decision-making is facilitated by peripheral feedback from

for example the autonomic nervous system. In a direct test of this, emotional decision-making was measured using the Iowa Gambling Task (Bechara, Damasio, Tranel and Damasio 2005) in patients with pure autonomic failure (Heims et al. 2004). In this condition, there is degeneration of the peripheral autonomic system, and thus autonomic responses are severely impaired, and there can be no resulting feedback to the brain. It was found that performance in the Iowa Gambling Task was not impaired, and nor were many other tests of emotion and emotional performance, including face expression identification, theory of mind tasks of social situations, and social cognition tasks. Thus emotional decision-making does not depend on the ongoing feedback from somatic markers related to autonomic function. Damasio might argue that feedback from the autonomic system is not actually important, and that it is feedback from skeleto-motor responses such as arm movements or muscle tension that is important. He might also argue that the autonomic feedback is not usually necessary for emotional decision-making, because it can be ‘simulated’ by the rest of the brain. However, the study by Heims et al (2004) does show that ongoing autonomic feedback is not necessary for normal emotional decision-making, and this leaves the somatic marker hypothesis more precarious.

Part of the evidence for the somatic marker hypothesis was that normal participants in the Iowa Gambling Task were described as deciding advantageously before knowing the advantageous strategy (Bechara, Damasio, Tranel and Damasio 1997). The interpretation was that they had implicit (unconscious) knowledge implemented via a somatic marker process that was used in the task, which was not being solved by explicit (conscious) knowledge. Maia and McClelland (2004; 2005) however showed that with more sensitive questioning, normal participants at least had available to them explicit knowledge about the outcomes of the different decks that was as good as or better than the choices made, weakening the arguments that the task was being solved implicitly and using somatic markers (Bechara, Damasio, Tranel and Damasio 1997; Bechara, Damasio, Tranel and Damasio 2005).

Another argument against the somatic marker hypothesis is that there can be dissociations between autonomic and other indices of emotion, thus providing evidence that behaviour may not follow from autonomic and other effects. For example, lesions of different parts of the amygdala influence autonomic responses and instrumental behaviour differently (Rolls 2014b, 2018).

Another major weakness, which applies to both the James-Lange and to Damasio’s somatic marker hypothesis, and to the roles of feedback from the autonomic system to the brain (Quadt, Critchley and Nagai 2022), is that they do not take account of the fact that once an information processor has determined that a response should be made

or inhibited based on reinforcement association, a function attributed in Rolls’ theory of emotion (Rolls 2014b, 2018) to the orbitofrontal cortex, it would be very inefficient and noisy to place in the execution route a peripheral response, and transducers to attempt to measure that peripheral response, itself a notoriously difficult procedure. Even for the cases when Damasio (1994) might argue that the peripheral somatic marker and its feedback can be by-passed using conditioning of a representation in, e.g., the somatosensory cortex to a command signal (which might originate in the orbitofrontal cortex), he apparently would still wish to argue that the activity in the somatosensory cortex is important for the emotion to be appreciated or to influence behaviour. (Without this, the somatic marker hypothesis would vanish.) The prediction would apparently be that if an emotional response were produced to a visual stimulus, then this would necessarily involve activity in the somatosensory cortex or other brain region in which the ‘somatic marker’ would be represented. This prediction could be tested (for example in patients with somatosensory cortex damage), but it seems most unlikely that an emotion produced by a visual reinforcer would require activity in the somatosensory cortex to feel emotional or to elicit emotional decisions. However, it has been found that the more damage there is to somatosensory cortex, the greater the impairment in the emotional state reported by patients (Adolphs et al. 2000). However, the parts of the somatosensory system that appear to be damaged most frequently in the patients with emotional change are often in the anterior and ventral extensions of the somatosensory cortex in insular and nearby areas, and it would be useful to know whether this damage interrupted some of the connections or functions of the orbitofrontal cortex areas just anterior.

More recently, Damasio has stated the somatic marker hypothesis in a weak form, suggesting that somatic markers do not even reflect the valence of the reinforcer, but just provide a signal that depends on the intensity of the emotion, independently of the type of emotion. On this view, the role of somatic markers in decision-making would be very general, providing, as Damasio says, just a jolt to spur the system on (A.R. Damasio, paper delivered at the 6th Annual Wisconsin Symposium on Emotion, April 2000).

The alternative view proposed here and elsewhere (Rolls 1990, 2000a, b, 2014b, 2018) is that where the reinforcement value of the visual stimulus is decoded, namely in the orbitofrontal cortex and the amygdala, is the appropriate part of the brain for outputs to influence behaviour (via, e.g., the orbitofrontal to cingulate cortex and orbitofrontal-to-striatal connections), and that the orbitofrontal cortex and amygdala, and brain structures that receive connections from them, are the likely places where neuronal activity is directly related to emotional states.

Appraisal theory

Appraisal theory (Frijda 1986; Moors et al. 2013; Keltner et al. 2018) generally holds that two types of appraisal are involved in emotion. Primary appraisal holds that “an emotion is usually caused by a person consciously or unconsciously evaluating an event as relevant to a concern (a goal) that is important; the emotion is felt as positive when a concern is advanced and negative when a concern is impeded” (Oatley and Jenkins 1996). The concept of appraisal presumably involves assessment of whether something is a reward or punisher, that is whether it will be worked for or avoided. The description in terms of rewards and punishers adopted here (and by Rolls (2014a) simply seems much more precisely and operationally specified. If primary appraisal is defined with respect to goals, it might be helpful to note that goals may just be the reinforcers specified in Rolls’ theory of emotion (Rolls 1999, 2000a, 2005, 2014a), and if so the reinforcer/punisher approach provides clear definitions of goals. Secondary appraisal is concerned with coping potential, that is with whether for example a plan can be constructed, and how successful it is likely to be.

Scherer (2009) summarizes his appraisal theory approach as follows. He suggests that there are four major appraisal objectives to adaptively react to a salient event:

- (a) Relevance: How relevant is this event for me? Does it directly affect me or my social reference group?
- (b) Implications: What are the implications or consequences of this event and how do they affect my well-being and my immediate or long-term goals?
- (c) Coping Potential: How well can I cope with or adjust to these consequences?
- (d) Normative Significance: What is the significance of this event for my self-concept and for social norms and values?

To attain these objectives, the organism evaluates the event and its consequences on a number of criteria or stimulus evaluation checks, with the results reflecting the organism’s subjective assessment (which may well be unrealistic or biased) of consequences and implications on a background of personal needs, goals, and values. Scherer (2009) states that an important feature of the model is that it does not include overt instrumental behaviour. Instead he sees emotion as a reaction to significant events that prepares action readiness and different types of alternative, possibly conflicting, action tendencies but not as a sufficient cause for their execution. This is a clear difference from my theory, in that my theory is that emotions are states that have a key role in brain design by providing a way for stimuli to produce states that are the goals for instrumental actions. Of

course stimuli that are instrumental reinforcers, goals for action, can also produce adaptive autonomic and skeletomotor reflexes (such as freezing), but these are responses, and can be classically conditioned, but do not require intervening goal-related representations or states, which are emotional and motivational states.

I note that appraisal theory is in many ways quite close to the theory that I outline here and elsewhere (Rolls 1999, 2000a, 2005, 2014a), 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 instrumental 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 some flavours of appraisal theory with which I do not agree is that emotions have as one of their functions releasing particular actions, which seems to make a link with species-specific action tendencies or responses, or ‘fixed action patterns’ (Panksepp 1998, 2011) or more ‘open motor programs’ (Ekman 2003). I argue that rarely are behavioural responses programmed by genes, but instead genes optimize their effects on behaviour if they specify the goals for (flexible) actions, that is if they specify rewards and punishers (Rolls 2014b). 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; and in that specifying goals for actions allows much more flexibility in the actual actions that are produced. Of course I acknowledge that there is some preparedness to learn associations between particular types of secondary and primary reinforcers (Seligman 1970), 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.

Panksepp’s theory of emotion

Panksepp’s approach to emotion had its origins in neuroethological investigations of brainstem systems that when activated lead to behaviours like fixed action patterns, including escape, flight and fear behaviour (Panksepp 1998, 2011). Using evidence from brain stimulation that elicits behaviours, he has postulated that there are a set of basic emotions, including for example Seeking, Rage, Fear, Lust, Care, Panic/Grief and Play. He argued that these are ‘natural kinds’, things that exist in nature as opposed to being inventions (constructions) of the human mind. My view is that there are not a few basic emotions, that emotions do not involve fixed action patterns as these do not require intervening emotional states to support goal-directed instrumental

actions, and that emotions can be classified based on the specific reinforcer, the specific reinforcement contingency, the actions that are available, etc. as described in Rolls' theory of emotion (Rolls 2014b, 2018).

Dimensional, categorical, and other theories of emotion

Dimensional and categorical theories of emotion suggest that there are a number of fundamental or basic emotions. Charles Darwin for example in his book *The Expression of the Emotions in Man and Animals* (Darwin 1872) showed that some basic expressions of emotion are similar in animals and humans. Some of the examples he gave are shown in Table 1. His focus was on the continuity between animals and humans of how emotion is expressed.

In a development of this approach, Ekman (1992; 2003) has suggested that humans categorize face expressions into a number of basic categories that are similar across cultures. These face expression categories include happy, fear, anger, surprise, grief and sadness.

A related approach is to identify a few clusters of variables or factors that result from multidimensional analysis of questionnaires, and to identify these factors as basic emotions. (Multidimensional analyses such as factor analysis seek to identify a few underlying sources of variance to which a large number of data values such as answers to questions are related.)

One potential problem with some of these approaches is that they risk finding seven plus or minus two categories, which is the maximal number of categories with which humans normally operate, as described in a famous paper by George Miller (1956). A second problem is that there is no special reason why the first few factors (which account for most of the variance) in a factor analysis should provide a complete or principled classification of different emotions, or of their functions. In contrast, the theory described here does produce a principled classification of different emotions based on reinforcement contingencies, the nature of the primary and secondary reinforcers, etc., as set out by Rolls (2014b). Moreover, the present theory links the functions of emotions to the classification produced, by showing how the functions of emotion can be understood in terms of the gene-specified reinforcers that produce different emotions (Rolls 2014b, 2018).

An opposite approach to the dimensional or categorical approach is to attempt to describe the richness of every emotion (Ben-Ze'ev 2001). Although it is important to understand the richness of every emotion, I believe that this is better performed with a set of underlying principles of the type set out above and by Rolls (2014b), rather than without any obvious principles to approach the subtlety of emotions.

LeDoux has described a theory of the neural basis of emotion (LeDoux 1992, 1995, 1996, 2012) that is probably conceptually similar to that of Rolls (Rolls 1999, 2000a, 2005, 2014a, 2018) except that he focuses mostly on the role of the amygdala in emotion (and not on other brain regions such as the orbitofrontal cortex, which are poorly developed in the rat); except that he focuses mainly on fear (based on his studies of the role of the amygdala and related structures in fear conditioning in the rat); and except that he suggests from his neurophysiological findings that an important route for conditioned emotional stimuli to influence behaviour is via the subcortical inputs (especially auditory from the medial part of the medial geniculate nucleus of the thalamus) to the amygdala. In contrast, I suggest that cortical processing to the object representation level before the representation is then sent to areas such as the amygdala and orbitofrontal cortex is normally involved in emotion, as emotions normally occur to objects, faces, etc. and not to spots of light or pure tones, which is what are represented precortically. Further, LeDoux (2012) has emphasized especially reflexes and classically conditioned reflexes such as autonomic responses and freezing, which I argue have adaptive value or in LeDoux's words 'survival value', whereas Rolls' theory is that emotional and motivational states are important intervening states in relation to instrumental actions. The way in which the rodent and amygdala literature has focussed on conditioned responses and not on emotional feelings is described by LeDoux and colleagues (LeDoux, Brown, Pine and Hofmann 2018; LeDoux and Daw 2018; Mobbs et al. 2019; LeDoux 2020; Taschereau-Dumouchel et al. 2022). Indeed, in this more recent research, LeDoux has reached the view that in humans the amygdala is not closely involved in emotional feelings, but does not suggest an alternative. I propose that the key parts of the human brain involved in emotional feelings are the orbitofrontal and anterior cingulate cortex, with the evidence provided above (Rolls 2014b, 2018; Rolls, Deco, Huang and Feng 2023a).

Barrett proposed a theory of constructed emotion (Barrett 2017) in which "the dynamics of the default mode, salience and frontoparietal control networks form the computational core of a brain's dynamic internal working model of the body in the world, entraining sensory and motor systems to create multi-sensory representations of the world at various time scales from the perspective of someone who has a body, all in the service of allostasis." "In other words, allostasis (predictively regulating the internal milieu) and interoception (representing the internal milieu) are at the anatomical and functional core of the nervous system (see also (Kleckner et al. 2017))." Claude Bernard introduced the concept of regulation of the internal milieu (Bernard 1865), and Cannon developed this into the concept of homeostasis (Cannon 1929; Goldstein 2019). Rolls' theory is that many

physiological processes involve regulation, but that emotion is linked to instrumental behavioural actions required to obtain rewards and avoid punishers. Rewards and punishers may be useful for homeostasis, but are also useful for many other functions including for example reproduction, and for example in humans using the reasoning system to perform actions in the interests of phenotypes, which can be other individuals.

Other approaches to emotion are summarized by Keltner et al. (2018).

Acknowledgements This research was supported by the UK Medical Research Council. The participation of many colleagues in the studies cited is sincerely acknowledged. They include Ivan de Araujo, Gordon Baylis, Leslie Baylis, Heather Berlin, Wei Cheng, Hugo Critchley, Gustavo Deco, Jianfeng Feng, Paul Gabbott, Fabian Grabenhorst, Michael Hasselmo, Julia Hornak, Mikiko Kadohisa, Morten Kringelbach, Christiana Leonard, Christian Margot, Ciara McCabe, Francis McGlone, John O’Doherty, Barbara Rolls, Juliet Rolls, Thomas Scott, Zenon Sienkiewicz, Simon Thorpe, Maria Ines Velazco, Justus Verhagen, and Simon Yaxley.

Author contributions Edmund Rolls is the sole author of this paper.

Funding This paper did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability No new data were analysed for this review paper. The effective connectivity, functional connectivity, and diffusion tractography analyses referred to here (Huang et al. 2021; Ma et al. 2022; Rolls et al. 2022a, b, 2023a, b, d, e, f, h, i) were performed with Human Connectome Project data (Glasser et al. 2016b), which are available at the HCP website <http://www.humanconnectome.org/>. The Human Connectome Project Multimodal Parcellation atlas and its availability are described by Glasser et al. (2016a), and the extended version HCPex is described by Huang et al. (2022), and is available at <https://www.oxcns.org>.

Declarations

Conflict of interest None.

References

- Adolphs R, Damasio H, Tranel D, Cooper G, Damasio AR (2000) A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J Neurosci* 20(7):2683–2690
- Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR (2005) A mechanism for impaired fear recognition after amygdala damage. *Nature* 433(7021):68–72
- Adolphs R, Anderson DJ (2018) The neuroscience of emotion. The neuroscience of emotion. Princeton University Press, Princeton
- Aggelopoulos NC, Franco L, Rolls ET (2005) Object perception in natural scenes: encoding by inferior temporal cortex simultaneously recorded neurons. *J Neurophysiol* 93(3):1342–1357. <https://doi.org/10.1152/jn.00553.2004>
- Aggelopoulos NC, Rolls ET (2005) Natural scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. *Eur J Neurosci* 22:2903–2916
- Amaral DG, Price JL, Pitkanen A, Carmichael ST (1992) Anatomical organization of the primate amygdaloid complex. In: Aggleton JP (ed) *The amygdala*. Wiley-Liss, New York, pp 1–66
- Antoniadis EA, Winslow JT, Davis M, Amaral DG (2009) The non-human primate amygdala is necessary for the acquisition but not the retention of fear-potentiated startle. *Biol Psychiatry* 65(3):241–248
- Aron AR, Robbins TW, Poldrack RA (2014) Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn Sci* 18(4):177–185. <https://doi.org/10.1016/j.tics.2013.12.003>
- Balleine BW (2019) The meaning of behavior: discriminating reflex and volition in the brain. *Neuron* 104(1):47–62. <https://doi.org/10.1016/j.neuron.2019.09.024>
- Barat E, Wirth S, Duhamel JR (2018) Face cells in orbitofrontal cortex represent social categories. *Proc Natl Acad Sci USA* 115(47):E11158–E11167. <https://doi.org/10.1073/pnas.1806165115>
- Barbas H (1995) Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. *Neurosci Biobehav Rev* 19:499–510
- Barbas H (2007) Specialized elements of orbitofrontal cortex in primates. *Ann N Y Acad Sci* 1121:10–32
- Barreiros IV, Ishii H, Walton ME, Panayi MC (2021) Defining an orbitofrontal compass: functional and anatomical heterogeneity across anterior-posterior and medial-lateral axes. *Behav Neurosci* 135(2):165–173. <https://doi.org/10.1037/bne0000442>
- Barrett LF (2017) The theory of constructed emotion: an active inference account of interoception and categorization. *Soc Cogn Affect Neurosci* 12(11):1833. <https://doi.org/10.1093/scan/nsx060>
- Baylis LL, Rolls ET, Baylis GC (1995) Afferent connections of the caudolateral orbitofrontal cortex taste area of the primate. *Neuroscience* 64(3):801–812. [https://doi.org/10.1016/0306-4522\(94\)00449-f](https://doi.org/10.1016/0306-4522(94)00449-f)
- Beaver JD, Lawrence AD, Jv D, Davis MH, Woods A, Calder AJ (2006) Individual differences in reward drive predict neural responses to images of food. *J Neurosci* 26:5160–5166
- Bechara A, Damasio H, Tranel D, Damasio AR (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–1295
- Bechara A, Damasio H, Tranel D, Damasio AR (2005) The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers. *Trends Cogn Sci* 9(4):159–162. <https://doi.org/10.1016/j.tics.2005.02.002>. (discussion 162–154)
- Ben-Ze’ev A (2001) The subtlety of emotions. MIT Press, Cambridge
- Berlin H, Rolls ET (2004) Time perception, impulsivity, emotionality, and personality in self-harming Borderline Personality Disorder Patients. *J Pers Disord* 18:358–378
- Berlin H, Rolls ET, Kischka U (2004) Impulsivity, time perception, emotion, and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain* 127:1108–1126
- Berlin H, Rolls ET, Iversen SD (2005) Borderline Personality Disorder, impulsivity and the orbitofrontal cortex. *Am J Psychiatry* 162:2360–2373
- Bermond B, Fasotti L, Nieuwenhuys B, Shuerman J (1991) Spinal cord lesions, peripheral feedback and intensities of emotional feelings. *Cogn Emot* 5:201–220
- Bernard C (1865) *Introduction à l’étude de la médecine expérimentale*, vol 2. JB Baillière, Paris
- Berridge KC (1996) Food reward: brain substrates of wanting and liking. *Neurosci Biobehav Rev* 20(1):1–25
- Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Brain Res Rev* 28:309–369

- Berridge KC, Robinson TE (2003) Parsing reward. *Trends Neurosci* 26:507–513
- Berridge KC, Robinson TE, Aldridge JW (2009) Dissecting components of reward: “liking”, “wanting”, and learning. *Curr Opin Pharmacol* 9(1):65–73
- Berridge KC, Dayan P (2021) Liking. *Curr Biol* 31(24):R1555–R1557. <https://doi.org/10.1016/j.cub.2021.09.069>
- Berthoud HR, Morrison CD, Ackroff K, Sclafani A (2021) Learning of food preferences: mechanisms and implications for obesity and metabolic diseases. *Int J Obes (Lond)* 45(10):2156–2168. <https://doi.org/10.1038/s41366-021-00894-3>
- Bisley JW, Goldberg ME (2010) Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci* 33:1–21. <https://doi.org/10.1146/annurev-neuro-060909-152823>
- Block N (2005) Two neural correlates of consciousness. *Trends Cogn Sci* 9(2):46–52
- Bonnici HM, Maguire EA (2018) Two years later—revisiting autobiographical memory representations in vmPFC and hippocampus. *Neuropsychologia* 110:159–169. <https://doi.org/10.1016/j.neuro-psychologia.2017.05.014>
- Bromberg-Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* 68(5):815–834. <https://doi.org/10.1016/j.neuron.2010.11.022>
- Brown R, Lau H, LeDoux JE (2019) Understanding the higher-order approach to consciousness. *Trends Cogn Sci* 23(9):754–768. <https://doi.org/10.1016/j.tics.2019.06.009>
- Buss DM (1989) Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behav Brain Sci* 12:1–14
- Buss DM (2015) *Evolutionary psychology: the new science of the mind*, 5th edn. Pearson, New York
- Butter CM (1969) Perseveration in extinction and in discrimination reversal tasks following selective prefrontal ablations in *Macaca mulatta*. *Physiol Behav* 4:163–171
- Cabanac M (1992) Pleasure: the common currency. *J Theor Biol* 155(2):173–200
- Cai X, Padoa-Schioppa C (2012) Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J Neurosci* 32(11):3791–3808. <https://doi.org/10.1523/JNEUROSCI.3864-11.2012>
- Cai X, Padoa-Schioppa C (2019) Neuronal evidence for good-based economic decisions under variable action costs. *Nat Commun* 10(1):393. <https://doi.org/10.1038/s41467-018-08209-3>
- Camille N, Tsuchida A, Fellows LK (2011) Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. *J Neurosci* 31(42):15048–15052. <https://doi.org/10.1523/JNEUROSCI.3164-11.2011>
- Canli T, Sivers H, Whitfield SL, Gotlib IH, Gabrieli JD (2002) Amygdala response to happy faces as a function of extraversion. *Science* 296(5576):2191
- Cannon WB (1927) The James-Lange theory of emotion: a critical examination and an alternative theory. *Am J Psychol* 39:106–124
- Cannon WB (1929) *Bodily changes in pain, hunger, fear and rage*, 2nd edn. Appleton, New York
- Cannon WB (1931) Again the James-Lange theory of emotion: a critical examination and an alternative theory. *Psychol Rev* 38:281–295
- Cardinal N, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev* 26:321–352
- Carmichael ST, Price JL (1994) Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *J Comp Neurol* 346(3):366–402. <https://doi.org/10.1002/cne.903460305>
- Carmichael ST, Price JL (1995) Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *J Comp Neurol* 363:642–664
- Carruthers P (2019) *Human and animal minds*. Oxford University Press, Oxford
- Chau BK, Sallet J, Papageorgiou GK, Noonan MP, Bell AH, Walton ME, Rushworth MF (2015) Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. *Neuron* 87:1106–1118. <https://doi.org/10.1016/j.neuron.2015.08.018>
- Cheng W, Rolls ET, Qiu J, Liu W, Tang Y, Huang CC, Wang X, Zhang J, Lin W, Zheng L, Pu J, Tsai SJ, Yang AC, Lin CP, Wang F, Xie P, Feng J (2016) Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression. *Brain* 139(Pt 12):3296–3309. <https://doi.org/10.1093/brain/aww255>
- Cheng W, Rolls ET, Qiu J, Xie X, Wei D, Huang CC, Yang AC, Tsai SJ, Li Q, Meng J, Lin CP, Xie P, Feng J (2018a) Increased functional connectivity of the posterior cingulate cortex with the lateral orbitofrontal cortex in depression. *Transl Psychiatry* 8(1):90. <https://doi.org/10.1038/s41398-018-0139-1>
- Cheng W, Rolls ET, Qiu J, Yang D, Ruan H, Wei D, Zhao L, Meng J, Xie P, Feng J (2018b) Functional connectivity of the precuneus in unmedicated patients with depression. *Biol Psychiatry Cogn Neurosci Neuroimaging* 3(12):1040–1049. <https://doi.org/10.1016/j.bpsc.2018.07.008>
- Ciaramelli E, De Luca F, Monk AM, McCormick C, Maguire EA (2019) What “wins” in VMPFC: scenes, situations, or schema? *Neurosci Biobehav Rev* 100:208–210. <https://doi.org/10.1016/j.neubiorev.2019.03.001>
- Corr PJ, McNaughton N (2012) Neuroscience and approach/avoidance personality traits: a two stage (valuation-motivation) approach. *Neurosci Biobehav Rev* 36(10):2339–2354. <https://doi.org/10.1016/j.neubiorev.2012.09.013>
- Cox J, Witten IB (2019) Striatal circuits for reward learning and decision-making. *Nat Rev Neurosci* 20(8):482–494. <https://doi.org/10.1038/s41583-019-0189-2>
- Critchley HD, Rolls ET (1996) Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J Neurophysiol* 75(4):1673–1686. <https://doi.org/10.1152/jn.1996.75.4.1673>
- Critchley HD, Harrison NA (2013) Visceral influences on brain and behavior. *Neuron* 77:624–638. <https://doi.org/10.1016/j.neuron.2013.02.008>
- Damasio A, Damasio H, Tranel D (2013) Persistence of feelings and sentience after bilateral damage of the insula. *Cereb Cortex* 23(4):833–846. <https://doi.org/10.1093/cercor/bhs077>
- Damasio AR (1994) *Descartes’ error*. Putnam, New York
- Damasio AR (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos Trans R Soc Lond B Biol Sci* 351(1346):1413–1420. <https://doi.org/10.1098/rstb.1996.0125>
- Darwin C (1872) *The expression of the emotions in man and animals*, 3rd edn. University of Chicago Press, Chicago
- Davis M (1992) The role of the amygdala in conditioned fear. In: Aggleton JP (ed) *The amygdala*. Wiley-Liss, New York, pp 255–305
- Davis M (1994) The role of the amygdala in emotional learning. *Int Rev Neurobiol* 36:225–266
- Davis M, Campeau S, Kim M, Falls WA (1995) Neural systems and emotion: the amygdala’s role in fear and anxiety. In: McGaugh JL, Weinberger NM, Lynch G (eds) *Brain and memory: modulation and mediation of neuroplasticity*. Oxford University Press, New York

- Davis M (2011) NMDA receptors and fear extinction: implications for cognitive behavioral therapy. *Dialog Clin Neurosci* 13(4):463–474
- Dawkins MS (2021) *The science of animal welfare, understanding what animals want*. Oxford University Press, Oxford
- Dawkins MS (2023) Farm animal welfare: beyond “natural” behavior. *Science* 379(6630):326–328. <https://doi.org/10.1126/science.ade5437>
- Dawkins R (1976) *The selfish gene*. Oxford University Press, Oxford
- Dawkins R (1982) *The extended phenotype: the gene as the unit of selection*. W.H. Freeman, San Francisco
- Dawkins R (1986) *The blind watchmaker*. Longman, Harlow
- Dawkins R (1989) *The selfish gene*, 2nd edn. Oxford University Press, Oxford
- De Araujo IE, Rolls ET (2004) Representation in the human brain of food texture and oral fat. *J Neurosci* 24(12):3086–3093. <https://doi.org/10.1523/JNEUROSCI.0130-04.2004>
- De Araujo IE, Rolls ET, Kringelbach ML, McGlone F, Phillips N (2003a) Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *Eur J Neurosci* 18(7):2059–2068. <https://doi.org/10.1046/j.1460-9568.2003.02915.x>
- de Araujo IET, Kringelbach ML, Rolls ET, Hobden P (2003b) The representation of umami taste in the human brain. *J Neurophysiol* 90:313–319
- de Araujo IET, Kringelbach ML, Rolls ET, McGlone F (2003c) Human cortical responses to water in the mouth, and the effects of thirst. *J Neurophysiol* 90:1865–1876
- De Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I (2005) Cognitive modulation of olfactory processing. *Neuron* 46(4):671–679. <https://doi.org/10.1016/j.neuron.2005.04.021>
- Deco G, Rolls ET (2003) Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *Eur J Neurosci* 18(8):2374–2390. <https://doi.org/10.1046/j.1460-9568.2003.02956.x>
- Deco G, Rolls ET (2004) A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res* 44(6):621–642. <https://doi.org/10.1016/j.visres.2003.09.037>
- Deco G, Rolls ET (2005a) Neurodynamics of biased competition and co-operation for attention: a model with spiking neurons. *J Neurophysiol* 94:295–313
- Deco G, Rolls ET (2005b) Attention, short-term memory, and action selection: a unifying theory. *Prog Neurobiol* 76:236–256
- Deco G, Rolls ET (2005c) Synaptic and spiking dynamics underlying reward reversal in orbitofrontal cortex. *Cereb Cortex* 15:15–30
- Deco G, Rolls ET, Zihl J (2005) A neurodynamical model of visual attention. In: Itti L, Rees G, Tsotsos J (eds) *Neurobiology of attention*. Elsevier, San Diego, pp 593–599 (**Chapter 597**)
- Deco G, Rolls ET, Albantakis L, Romo R (2013) Brain mechanisms for perceptual and reward-related decision-making. *Prog Neurobiol* 103:194–213. <https://doi.org/10.1016/j.pneurobio.2012.01.010>
- Delgado MR, Jou RL, Phelps EA (2011) Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers. *Front Neurosci* 5:71. <https://doi.org/10.3389/fnins.2011.00071>
- Deng WL, Rolls ET, Ji X, Robbins TW, Banaschewski T, Bokde ALW, Bromberg U, Buechel C, Desrivieres S, Conrod P, Flor H, Frouin V, Gallinat J, Garavan H, Gowland P, Heinz A, Ittermann B, Martinot J-L, Lemaitre H, Nees F, Papadoulos Orfanos D, Poustka L, Smolka MN, Walter H, Whelan R, Schumann G, Feng J (2017) Separate neural systems for behavioral change and for emotional responses to failure during behavioral inhibition. *Hum Brain Mapp* 38:3527–3537
- Dennett DC (1991) *Consciousness explained*. Penguin, London
- Descartes R (1644) *The Philosophical Writings of Descartes* (3 volumes, 1984–1991). Cambridge University Press, Cambridge
- Deutsch JA (1960) *The structural basis of behavior*. University of Chicago Press, Chicago
- DeYoung CG, Beaty RE, Genç E, Latzman RE, Passamonti L, Servaes MN, Shackman AJ, Smillie LD, Spreng RN, Viding E, Wacker J (2022) Personality neuroscience: an emerging field with bright prospect. *Personal Sci* 3:1–21
- Du J, Rolls ET, Cheng W, Li Y, Gong W, Qiu J, Feng J (2020) Functional connectivity of the orbitofrontal cortex, anterior cingulate cortex, and inferior frontal gyrus in humans. *Cortex* 123:185–199. <https://doi.org/10.1016/j.cortex.2019.10.012>
- Ekman P, Levenson RW, Friesen WV (1983) Autonomic nervous system activity distinguishes between the emotions. *Science* 221:1208–1210
- Ekman P (1992) An argument for basic emotions. *Cogn Emot* 6(3–4):169–200
- Ekman P (2003) *Emotions revealed: understanding thoughts and feelings*. Weidenfeld and Nicolson, London
- Everitt BJ, Robbins TW (2013) From the ventral to the dorsal striatum: evolving views of their roles in drug addiction. *Neurosci Biobehav Rev* 37(9 Pt A):1946–1954. <https://doi.org/10.1016/j.neubiorev.2013.02.010>
- Eysenck HJ, Eysenck SBG (1968) *Personality structure and measurement*. R.R. Knapp, San Diego
- Farm Animal Welfare Council (2009) *Farm animal welfare in great britain: past, present and future*
- Fattori P, Breveglieri R, Bosco A, Gamberini M, Galletti C (2017) Vision for prehension in the medial parietal cortex. *Cereb Cortex* 27(2):1149–1163. <https://doi.org/10.1093/cercor/bhv302>
- Feinstein JS, Adolphs R, Damasio A, Tranel D (2011) The human amygdala and the induction and experience of fear. *Curr Biol* 21(1):34–38. <https://doi.org/10.1016/j.cub.2010.11.042>
- Fellows LK (2011) Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. *Ann N Y Acad Sci* 1239:51–58. <https://doi.org/10.1111/j.1749-6632.2011.06229.x>
- Freese JL, Amaral DG (2005) The organization of projections from the amygdala to visual cortical areas TE and V1 in the macaque monkey. *J Comp Neurol* 486:295–317
- Freese JL, Amaral DG (2009) Neuroanatomy of the primate amygdala. In: Whalen PJ, Phelps EA (eds) *The Human amygdala*. Guilford, New York, pp 3–42
- Frijda NH (1986) *The emotions*. Cambridge University Press, Cambridge
- Fuster JM (2014) The prefrontal cortex makes the brain a preadaptive system. *Proc IEEE* 102(4):417–426
- Fuster JM (2015) *The prefrontal cortex*, 5th edn. Academic Press, London
- Galletti C, Fattori P (2018) The dorsal visual stream revisited: stable circuits or dynamic pathways? *Cortex* 98:203–217. <https://doi.org/10.1016/j.cortex.2017.01.009>
- Gamberini M, Passarelli L, Filippini M, Fattori P, Galletti C (2021) Vision for action: thalamic and cortical inputs to the macaque superior parietal lobule. *Brain Struct Funct* 226(9):2951–2966. <https://doi.org/10.1007/s00429-021-02377-7>
- Gazzaniga MS, LeDoux J (1978) *The integrated mind*. Plenum, New York
- Ge T, Feng J, Grabenhorst F, Rolls ET (2012) Componential Granger causality, and its application to identifying the source and mechanisms of the top-down biased activation that controls attention to affective vs sensory processing. *Neuroimage* 59:1846–1858
- Ghashghaei HT, Barbas H (2002) Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* 115(4):1261–1279
- Glascher J, Adolphs R, Damasio H, Bechara A, Rudrauf D, Calamia M, Paul LK, Tranel D (2012) Lesion mapping of cognitive

- control and value-based decision making in the prefrontal cortex. *Proc Natl Acad Sci USA* 109(36):14681–14686. <https://doi.org/10.1073/pnas.1206608109>
- Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC (2016a) A multi-modal parcellation of human cerebral cortex. *Nature* 536(7615):171–178. <https://doi.org/10.1038/nature18933>
- Glasser MF, Smith SM, Marcus DS, Andersson JL, Auerbach EJ, Behrens TE, Coalson TS, Harms MP, Jenkinson M, Moeller S, Robinson EC, Sotiropoulos SN, Xu J, Yacoub E, Ugurbil K, Van Essen DC (2016b) The Human Connectome Project's neuroimaging approach. *Nat Neurosci* 19(9):1175–1187. <https://doi.org/10.1038/nn.4361>
- Glimcher PW, Fehr E (eds) (2013) *Neuroeconomics: decision-making and the brain*, 2nd edn. Academic Press, New York
- Goldman-Rakic PS (1996) The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philos Trans R Soc B* 351:1445–1453
- Goldstein DS (2019) How does homeostasis happen? Integrative physiological, systems biological, and evolutionary perspectives. *Am J Physiol Regul Integr Comp Physiol* 316(4):R301–R317. <https://doi.org/10.1152/ajpregu.00396.2018>
- Gosnell BA, Levine AS (2009) Reward systems and food intake: role of opioids. *Int J Obes (lond)* 33(Suppl 2):S54–58. <https://doi.org/10.1038/ijo.2009.73>
- Grabenhorst F, Rolls ET (2008) Selective attention to affective value alters how the brain processes taste stimuli. *Eur J Neurosci* 27:723–729
- Grabenhorst F, Rolls ET (2010) Attentional modulation of affective vs sensory processing: functional connectivity and a top-down biased activation theory of selective attention. *J Neurophysiol* 104:1649–1660
- Grabenhorst F, Rolls ET (2011) Value, pleasure, and choice in the ventral prefrontal cortex. *Trends Cogn Sci* 15:56–67
- Grabenhorst F, Schultz W (2021) Functions of primate amygdala neurons in economic decisions and social decision simulation. *Behav Brain Res* 409:113318. <https://doi.org/10.1016/j.bbr.2021.113318>
- Grabenhorst F, Rolls ET, Bilderbeck A (2008a) How cognition modulates affective responses to taste and flavor: top down influences on the orbitofrontal and pregenual cingulate cortices. *Cereb Cortex* 18:1549–1559
- Grabenhorst F, Rolls ET, Parris BA (2008b) From affective value to decision-making in the prefrontal cortex. *Eur J Neurosci* 28:1930–1939
- Grabenhorst F, D'Souza A, Parris BA, Rolls ET, Passingham RE (2010a) A common neural scale for the subjective pleasantness of different primary rewards. *Neuroimage* 51:1265–1274
- Grabenhorst F, Rolls ET, Parris BA, D'Souza A (2010b) How the brain represents the reward value of fat in the mouth. *Cereb Cortex* 20:1082–1091
- Grabenhorst F, Baez-Mendoza R, Genest W, Deco G, Schultz W (2019) Primate amygdala neurons simulate decision processes of social partners. *Cell* 177(4):986–998 e915. <https://doi.org/10.1016/j.cell.2019.02.042>
- Grattan LE, Glimcher PW (2014) Absence of spatial tuning in the orbitofrontal cortex. *PLoS ONE* 9(11):e112750. <https://doi.org/10.1371/journal.pone.0112750>
- Gray JA (1970) The psychophysiological basis of introversion-extraversion. *Behav Res Ther* 8:249–266
- Grill HJ, Norgren R (1978) Chronically decerebrate rats demonstrate satiation but not bait shyness. *Science* 201:267–269
- Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35(1):4–26. <https://doi.org/10.1038/npp.2009.129>
- Hadj-Bouziane F, Liu N, Bell AH, Gothard KM, Luh WM, Tootell RB, Murray EA, Ungerleider LG (2012) Amygdala lesions disrupt modulation of functional MRI activity evoked by facial expression in the monkey inferior temporal cortex. *Proc Natl Acad Sci USA* 109(52):E3640–3648. <https://doi.org/10.1073/pnas.1218406109>
- Hamilton WD (1964) The genetical evolution of social behaviour. *J Theor Biol* 7:1–52
- Hamilton WD (1996) *Narrow roads of gene land*. W. H. Freeman, New York
- Hassanpour MS, Simmons WK, Feinstein JS, Luo Q, Lapidus RC, Bodurka J, Paulus MP, Khalsa SS (2018) The insular cortex dynamically maps changes in cardiorespiratory interoception. *Neuropsychopharmacology* 43(2):426–434. <https://doi.org/10.1038/npp.2017.154>
- Hasselmo ME, Sarter M (2011) Modes and models of forebrain cholinergic neuromodulation of cognition. *Neuropsychopharmacology* 36(1):52–73. <https://doi.org/10.1038/npp.2010.104>
- Hasselmo ME, Rolls ET, Baylis GC (1989a) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res* 32(3):203–218
- Hasselmo ME, Rolls ET, Baylis GC, Nalwa V (1989b) Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Exp Brain Res* 75:417–429
- Heimer L, Switzer RD, Van Hoesen GW (1982) Ventral striatum and ventral pallidum. Components of the motor system? *Trends Neurosci* 5:83–87
- Heims HC, Critchley HD, Dolan R, Mathias CJ, Cipolotti L (2004) Social and motivational functioning is not critically dependent on feedback of autonomic responses: neuropsychological evidence from patients with pure autonomic failure. *Neuropsychologia* 42:1979–1988
- Henssen A, Zilles K, Palomero-Gallagher N, Schleicher A, Mohlberg H, Gerboga F, Eickhoff SB, Bludau S, Amunts K (2016) Cytoarchitecture and probability maps of the human medial orbitofrontal cortex. *Cortex* 75:87–112. <https://doi.org/10.1016/j.cortex.2015.11.006>
- Hervig ME, Fiddian L, Piilgaard L, Bozic T, Blanco-Pozo M, Knudsen C, Olesen SF, Alsio J, Robbins TW (2020) Dissociable and paradoxical roles of rat medial and lateral orbitofrontal cortex in visual serial reversal learning. *Cereb Cortex* 30(3):1016–1029. <https://doi.org/10.1093/cercor/bhz144>
- Hohmann GW (1966) Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology* 3:143–156
- Hornak J, Rolls ET, Wade D (1996) Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia* 34:247–261
- Hornak J, Bramham J, Rolls ET, Morris RG, O'Doherty J, Bullock PR, Polkey CE (2003) Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain* 126(Pt 7):1691–1712. <https://doi.org/10.1093/brain/awg168>
- Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, Polkey CE (2004) Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans. *J Cogn Neurosci* 16:463–478
- Huang C-C, Rolls ET, Hsu C-CH, Feng J, Lin C-P (2021) Extensive cortical connectivity of the human hippocampal memory system: beyond the “what” and “where” dual-stream model. *Cereb Cortex* 31:4652–4669. <https://doi.org/10.1093/cercor/bhab113>

- Huang CC, Rolls ET, Feng J, Lin CP (2022) An extended Human Connectome Project multimodal parcellation atlas of the human cortex and subcortical areas. *Brain Struct Funct* 227:763–778. <https://doi.org/10.1007/s00429-021-02421-6>
- Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis Res* 40(10–12):1489–1506
- Iversen SD, Mishkin M (1970) Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Exp Brain Res* 11:376–386
- Izquierdo A (2017) Functional heterogeneity within rat orbitofrontal cortex in reward learning and decision making. *J Neurosci* 37(44):10529–10540. <https://doi.org/10.1523/JNEUROSCI.1678-17.2017>
- James W (1884) What is an emotion? *Mind* 9:188–205
- Kadohisa M, Rolls ET, Verhagen JV (2005a) The primate amygdala: neuronal representations of the viscosity, fat texture, temperature, grittiness and taste of foods. *Neuroscience* 132(1):33–48. <https://doi.org/10.1016/j.neuroscience.2004.12.005>
- Kadohisa M, Rolls ET, Verhagen JV (2005b) Neuronal representations of stimuli in the mouth: the primate insular taste cortex, orbitofrontal cortex, and amygdala. *Chem Senses* 30(5):401–419. <https://doi.org/10.1093/chemse/bji036>
- Kehoe P, Blass EM (1985) Gustatory determinants of suckling in albino rats 5–20 days of age. *Dev Psychobiol* 18(1):67–82. <https://doi.org/10.1002/dev.420180106>
- Keltner D, Oatley K, Jenkins JM (2018) *Understanding emotions*. Wiley, Hoboken
- Kennedy DP, Adolphs R (2011) Reprint of: Impaired fixation to eyes following amygdala damage arises from abnormal bottom-up attention. *Neuropsychologia* 49(4):589–595. <https://doi.org/10.1016/j.neuropsychologia.2011.02.026>
- Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 9(7):940–947
- Kleckner IR, Zhang J, Touroutoglou A, Chanes L, Xia C, Simmons WK, Quigley KS, Dickerson BC, Barrett LF (2017) Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nat Hum Behav* 1:0069. <https://doi.org/10.1038/s41562-017-0069>
- Klein-Flugge MC, Jensen DEA, Takagi Y, Priestley L, Verhagen L, Smith SM, Rushworth MFS (2022) Relationship between nuclei-specific amygdala connectivity and mental health dimensions in humans. *Nat Hum Behav* 6(12):1705–1722. <https://doi.org/10.1038/s41562-022-01434-3>
- Kolb B, Whishaw IQ (2021) *Fundamentals of human neuropsychology*, 8th edn. MacMillan, New York
- Kolling N, Wittmann MK, Behrens TE, Boorman ED, Mars RB, Rushworth MF (2016) Value, search, persistence and model updating in anterior cingulate cortex. *Nat Neurosci* 19(10):1280–1285. <https://doi.org/10.1038/nn.4382>
- Kraut RE, Johnson RE (1979) Social and emotional messages of smiling: an ethological approach. *J Pers Soc Psychol* 37:1539–1553
- Kringelbach ML, Rolls ET (2003) Neural correlates of rapid reversal learning in a simple model of human social interaction. *Neuroimage* 20(2):1371–1383. [https://doi.org/10.1016/S1053-8119\(03\)00393-8](https://doi.org/10.1016/S1053-8119(03)00393-8)
- Kringelbach ML, O'Doherty J, Rolls ET, Andrews C (2003) Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cereb Cortex* 13:1064–1071
- Kuwabara M, Kang N, Holy TE, Padoa-Schioppa C (2020) Neural mechanisms of economic choices in mice. *Elife* 9:e49669. <https://doi.org/10.7554/eLife.49669>
- Lange C (1885) The emotions. In: Dunlap E (ed) *The emotions*. Williams and Wilkins, Baltimore
- LeDoux JE (1992) Emotion and the amygdala. In: Aggleton JP (ed) *The amygdala*. Wiley-Liss, New York, pp 339–351
- LeDoux JE (1995) Emotion: clues from the brain. *Annu Rev Psychol* 46:209–235. <https://doi.org/10.1146/annurev.ps.46.020195.001233>
- LeDoux JE (1996) *The emotional brain*. Simon and Schuster, New York
- LeDoux JE (2000a) Emotion circuits in the brain. *Annu Rev Neurosci* 23:155–184
- LeDoux JE (2000b) The amygdala, fear conditioning and emotion. In: Aggleton JP (ed) *The amygdala: a functional analysis*. Oxford University Press, Oxford
- LeDoux J (2012) Rethinking the emotional brain. *Neuron* 73(4):653–676. <https://doi.org/10.1016/j.neuron.2012.02.004>
- LeDoux JE (2020) Thoughtful feelings. *Curr Biol* 30(11):R619–R623. <https://doi.org/10.1016/j.cub.2020.04.012>
- LeDoux JE, Brown R (2017) A higher-order theory of emotional consciousness. *Proc Natl Acad Sci USA* 114(10):E2016–E2025. <https://doi.org/10.1073/pnas.1619316114>
- LeDoux JE, Daw ND (2018) Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nat Rev Neurosci* 19(5):269–282. <https://doi.org/10.1038/nrn.2018.22>
- LeDoux JE, Pine DS (2016) Using neuroscience to help understand fear and anxiety: a two-system framework. *Am J Psychiatry* 173(11):1083–1093. <https://doi.org/10.1176/appi.ajp.2016.16030353>
- LeDoux J, Brown R, Pine D, Hofmann S (2018) *Know thyself: well-being and subjective experience*. Cerebrum, Dana Foundation, New York
- Leonard CM, Rolls ET, Wilson FAW, Baylis GC (1985) Neurons in the amygdala of the monkey with responses selective for faces. *Behav Brain Res* 15:159–176
- Levenson RW, Ekman P, Friesen WV (1990) Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology* 27:363–384
- Luk CH, Wallis JD (2009) Dynamic encoding of responses and outcomes by neurons in medial prefrontal cortex. *J Neurosci* 29(23):7526–7539
- Luo Q, Ge T, Grabenhorst F, Feng J, Rolls ET (2013) Attention-dependent modulation of cortical taste circuits revealed by Granger causality with signal-dependent noise. *PLoS Comput Biol* 9:e1003265
- Ma Q, Rolls ET, Huang C-C, Cheng W, Feng J (2022) Extensive cortical functional connectivity of the human hippocampal memory system. *Cortex* 147:83–101. <https://doi.org/10.1016/j.cortex.2021.11.014>
- Mackey S, Petrides M (2010) Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *Eur J Neurosci* 32(11):1940–1950. <https://doi.org/10.1111/j.1460-9568.2010.07465.x>
- Maia TV, McClelland JL (2004) A reexamination of the evidence for the somatic marker hypothesis: what participants really know in the Iowa gambling task. *Proc Natl Acad Sci USA* 101(45):16075–16080
- Maia TV, McClelland JL (2005) The somatic marker hypothesis: still many questions but no answers: response to Bechara et al. *Trends Cogn Sci* 9(4):162–164
- Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007m) Medial prefrontal selectivity signalling prediction errors of action values. *Nat Neurosci* 10:647–656
- Matthews G, Gilliland K (1999) The personality theories of H.J.Eysenck and J.A.Gray: a comparative review. *Pers Individ Dif* 26:583–626

- McCabe C, Rolls ET, Bilderbeck A, McGlone F (2008) Cognitive influences on the affective representation of touch and the sight of touch in the human brain. *Soc Cogn Affect Neurosci* 3:97–108
- McCormick C, Ciaramelli E, De Luca F, Maguire EA (2018) Comparing and contrasting the cognitive effects of hippocampal and ventromedial prefrontal cortex damage: a review of human lesion studies. *Neuroscience* 374:295–318. <https://doi.org/10.1016/j.neuroscience.2017.07.066>
- Miller GA (1956) The magic number seven, plus or minus two: some limits on our capacity for the processing of information. *Psychol Rev* 63:81–93
- Miller EK, Lundqvist M, Bastos AM (2018) Working memory 2.0. *Neuron* 100(2):463–475. <https://doi.org/10.1016/j.neuron.2018.09.023>
- Mobbs D, Adolphs R, Fanselow MS, Barrett LF, LeDoux JE, Ressler K, Tye KM (2019) Viewpoints: approaches to defining and investigating fear. *Nat Neurosci* 22(8):1205–1216. <https://doi.org/10.1038/s41593-019-0456-6>
- Moors A, Ellsworth PC, Scherer KR, Frijda NH (2013) Appraisal theories of emotion: state of the art and future development. *Emot Rev* 5:119–124
- Morecraft RJ, Tanji J (2009) Cingulofrontal interaction and the cingulate motor areas. In: Vogt BA (ed) *Cingulate neurobiology and disease*. Oxford University Press, Oxford, pp 113–144
- Morris RW, Dezfouli A, Griffiths KR, Le Pelley ME, Balleine BW (2022) The neural bases of action-outcome learning in humans. *J Neurosci* 42(17):3636–3647. <https://doi.org/10.1523/JNEUROSCI.1079-21.2022>
- Murray EA, Fellows LK (2022) Prefrontal cortex interactions with the amygdala in primates. *Neuropsychopharmacology* 47(1):163–179. <https://doi.org/10.1038/s41386-021-01128-w>
- Murray EA, Rudebeck PH (2018) Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. *Nat Rev Neurosci* 19(7):404–417. <https://doi.org/10.1038/s41583-018-0013-4>
- Nguyen D, Naffziger EE, Berridge KC (2021) Positive affect: nature and brain bases of liking and wanting. *Curr Opin Behav Sci* 39:72–78. <https://doi.org/10.1016/j.cobeha.2021.02.013>
- Noonan MP, Mars RB, Rushworth MF (2011) Distinct roles of three frontal cortical areas in reward-guided behavior. *J Neurosci* 31(40):14399–14412. <https://doi.org/10.1523/JNEUROSCI.6456-10.2011>
- Noonan MP, Chau BKH, Rushworth MFS, Fellows LK (2017) Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. *J Neurosci* 37(29):7023–7035. <https://doi.org/10.1523/JNEUROSCI.0692-17.2017>
- O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* 4:95–102
- O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ (2003) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147–155
- Oatley K, Jenkins JM (1996) *Understanding emotions*. Blackwell, Oxford
- Ongür D, Price JL (2000) The organisation of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex* 10:206–219
- Öngür D, Ferry AT, Price JL (2003) Architectonic division of the human orbital and medial prefrontal cortex. *J Comp Neurol* 460:425–449
- Padoa-Schioppa C (2007) Orbitofrontal cortex and the computation of economic value. *Ann N Y Acad Sci* 1121:232–253. <https://doi.org/10.1196/annals.1401.011>
- Padoa-Schioppa C (2011) Neurobiology of economic choice: a good-based model. *Annu Rev Neurosci* 34:333–359. <https://doi.org/10.1146/annurev-neuro-061010-113648>
- Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441(7090):223–226. <https://doi.org/10.1038/nature04676>
- Padoa-Schioppa C, Assad JA (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat Neurosci* 11(1):95–102
- Padoa-Schioppa C, Cai X (2011) The orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives. *Ann N Y Acad Sci* 1239:130–137. <https://doi.org/10.1111/j.1749-6632.2011.06262.x>
- Padoa-Schioppa C, Conen KE (2017) Orbitofrontal cortex: a neural circuit for economic decisions. *Neuron* 96(4):736–754. <https://doi.org/10.1016/j.neuron.2017.09.031>
- Pager J, Giachetti I, Holley A, Le Magnen J (1972) A selective control of olfactory bulb electrical activity in relation to food deprivation and satiety in rats. *Physiol Behav* 9(4):573–579
- Palouzier-Paulignan B, Lacroix MC, Aime P, Baly C, Caillol M, Congar P, Julliard AK, Tucker K, Fadool DA (2012) Olfaction under metabolic influences. *Chem Sens* 37(9):769–797. <https://doi.org/10.1093/chemse/bjs059>
- Pandya DN, Yeterian EH (1996) Comparison of prefrontal architecture and connections. *Philos Trans R Soc B* 351:1423–1431
- Panksepp J (1998) *Affective neuroscience: the foundations of human and animal emotions*. Oxford University Press, New York
- Panksepp J (2011) The basic emotional circuits of mammalian brains: do animals have affective lives? *Neurosci Biobehav Rev* 35(9):1791–1804. <https://doi.org/10.1016/j.neubiorev.2011.08.003>
- Passingham RE (2021) *Understanding the prefrontal cortex: selective advantage, connectivity and neural operations*. Oxford University Press, Oxford
- Passingham REP, Wise SP (2012) *The neurobiology of the prefrontal cortex*. Oxford University Press, Oxford
- Petrides M, Pandya DN (1995) Comparative architectonic analysis of the human and macaque frontal cortex. In: Boller F, Grafman J (eds) *Handbook of neuropsychology*, vol 9. Elsevier Science, Amsterdam, pp 17–58
- Petrides M, Tomaiuolo F, Yeterian EH, Pandya DN (2012) The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. *Cortex* 48(1):46–57. <https://doi.org/10.1016/j.cortex.2011.07.002>
- Pitcher D, Ungerleider LG (2021) Evidence for a third visual pathway specialized for social perception. *Trends Cogn Sci* 25(2):100–110. <https://doi.org/10.1016/j.tics.2020.11.006>
- Pitcher D, Ianni G, Ungerleider LG (2019) A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. *Sci Rep* 9(1):8242. <https://doi.org/10.1038/s41598-019-44663-9>
- Platt ML, Seyfarth RM, Cheney DL (2016) Adaptations for social cognition in the primate brain. *Philos Trans R Soc Lond B Biol Sci* 371(1687):20150096. <https://doi.org/10.1098/rstb.2015.0096>
- Price JL (2006) Connections of orbital cortex. In: Zald DH, Rauch SL (eds) *The orbitofrontal cortex*. Oxford University Press, Oxford, pp 39–55
- Price JL (2007) Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. *Ann N Y Acad Sci* 1121:54–71
- Quadt L, Critchley HD, Garfinkel SN (2018) The neurobiology of interoception in health and disease. *Ann N Y Acad Sci* 1428(1):112–128. <https://doi.org/10.1111/nyas.13915>

- Quadt L, Critchley H, Nagai Y (2022) Cognition, emotion, and the central autonomic network. *Auton Neurosci* 238:102948. <https://doi.org/10.1016/j.autneu.2022.102948>
- Quirk GJ, Armony JL, Reza JC, Li XF, LeDoux JE (1996) Emotional memory: a search for sites of plasticity. *Cold Spring Harb Symp Quant Biol* 61:247–257
- Reisenzein R (1983) The Schachter theory of emotion: two decades later. *Psychol Bull* 94:239–264
- Ridley M (1993) *The red queen: sex and the evolution of human nature*. Penguin, London
- Robinson TE, Berridge KC (2003) Addiction. *Annu Rev Psychol* 54:25–53
- Rogan MT, Staubli UV, LeDoux JE (1997) Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* 390:604–607
- Rolls ET (1990) A theory of emotion, and its application to understanding the neural basis of emotion. *Cogn Emot* 4:161–190
- Rolls ET (1999) *The brain and emotion*. Oxford University Press, Oxford
- Rolls ET (2000a) *Precis of The brain and emotion*. *Behav Brain Sci* 23(2):177–191. <https://doi.org/10.1017/s0140525x00002429>. (discussion 192–233)
- Rolls ET (2000b) The orbitofrontal cortex and reward. *Cereb Cortex* 10:284–294
- Rolls ET (2000c) Neurophysiology and functions of the primate amygdala, and the neural basis of emotion. In: Aggleton JP (ed) *The amygdala: a functional analysis*, 2nd edn. Oxford University Press, Oxford, pp 447–478
- Rolls ET (2005) *Emotion explained*. Oxford University Press, Oxford
- Rolls ET (2007a) A computational neuroscience approach to consciousness. *Neural Netw* 20(9):962–982. <https://doi.org/10.1016/j.neunet.2007.10.001>
- Rolls ET (2007b) The affective neuroscience of consciousness: higher order linguistic thoughts, dual routes to emotion and action, and consciousness. In: Zelazo P, Moscovitch M, Thompson E (eds) *Cambridge handbook of consciousness*. Cambridge University Press, Cambridge, pp 831–859
- Rolls ET (2008) Emotion, higher order syntactic thoughts, and consciousness. In: Weiskrantz L, Davies MK (eds) *Frontiers of consciousness*. Oxford University Press, Oxford, pp 131–167
- Rolls ET (2010a) Noise in the brain, decision-making, determinism, free will, and consciousness. In: Perry E, Collerton D, Lebeau F, Ashton H (eds) *New horizons in the neuroscience of consciousness*. John Benjamins, Amsterdam, pp 113–120
- Rolls ET (2010b) The affective and cognitive processing of touch, oral texture, and temperature in the brain. *Neurosci Biobehav Rev* 34:237–245
- Rolls ET (2011) Consciousness, decision-making, and neural computation. In: Cutsuridis V, Hussain A, Taylor JG (eds) *Perception-action cycle: models, algorithms and systems*. Springer, Berlin, pp 287–333
- Rolls ET (2012a) Taste, olfactory, and food texture reward processing in the brain and the control of appetite. *Proc Nutr Soc* 71:488–501
- Rolls ET (2012b) *Neuroculture. On the implications of brain science*. Oxford University Press, Oxford
- Rolls ET (2012c) Invariant visual object and face recognition: neural and computational bases, and a model. *VisNet Front Comput Neurosci* 6(35):1–70. <https://doi.org/10.3389/fncom.2012.00035>
- Rolls ET (2013a) A biased activation theory of the cognitive and attentional modulation of emotion. *Front Hum Neurosci* 7:74. <https://doi.org/10.3389/fnhum.2013.00074>
- Rolls ET (2013b) What are emotional states, and why do we have them? *Emot Rev* 5:241–247
- Rolls ET (2014a) Emotion and decision-making explained: précis. *Cortex* 59:185–193
- Rolls ET (2014b) *Emotion and decision-making explained*. Oxford University Press, Oxford
- Rolls ET (2015) Taste, olfactory, and food reward value processing in the brain. *Prog Neurobiol* 127–128:64–90. <https://doi.org/10.1016/j.pneurobio.2015.03.002>
- Rolls ET (2016a) Reward systems in the brain and nutrition. *Annu Rev Nutr* 36:435–470
- Rolls ET (2016b) A non-reward attractor theory of depression. *Neurosci Biobehav Rev* 68:47–58. <https://doi.org/10.1016/j.neubiorev.2016.05.007>
- Rolls ET (2016c) *Cerebral cortex: principles of operation*. Oxford University Press, Oxford
- Rolls ET (2016d) Functions of the anterior insula in taste, autonomic, and related functions. *Brain Cogn* 110:4–19. <https://doi.org/10.1016/j.bandc.2015.07.002>
- Rolls ET (2016e) Brain processing of reward for touch, temperature, and oral texture. In: Olausson H, Wessberg J, Morrison I, McGlone F (eds) *Affective touch and the neurophysiology of CT afferents*. Springer, Berlin, pp 209–225
- Rolls ET (2016f) Motivation explained: ultimate and proximate accounts of hunger and appetite. *Adv Motiv Sci* 3:187–249
- Rolls ET (2017) The roles of the orbitofrontal cortex via the habenula in non-reward and depression, and in the responses of serotonin and dopamine neurons. *Neurosci Biobehav Rev* 75:331–334. <https://doi.org/10.1016/j.neubiorev.2017.02.013>
- Rolls ET (2018) *The brain, emotion, and depression*. Oxford University Press, Oxford
- Rolls ET (2019a) The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Struct Funct* 224(9):3001–3018. <https://doi.org/10.1007/s00429-019-01945-2>
- Rolls ET (2019b) *The orbitofrontal cortex*. Oxford University Press, Oxford
- Rolls ET (2019c) Emotion and reasoning in human decision-making. *Econ Open-Access Open-Assess E-J*. <https://doi.org/10.5018/economics-ejournal.ja.2019-5039>
- Rolls ET (2019d) The orbitofrontal cortex and emotion in health and disease, including depression. *Neuropsychologia* 128:14–43. <https://doi.org/10.1016/j.neuropsychologia.2017.09.021>
- Rolls ET (2020) Neural computations underlying phenomenal consciousness: a Higher Order Syntactic Thought theory. *Front Psychol (consciousness Research)* 11:655. <https://doi.org/10.3389/fpsyg.2020.00655>
- Rolls ET (2021a) Learning invariant object and spatial view representations in the brain using slow unsupervised learning. *Front Comput Neurosci* 15:686239. <https://doi.org/10.3389/fncom.2021.686239>
- Rolls ET (2021b) *Brain computations: what and how*. Oxford University Press, Oxford. <https://doi.org/10.1093/oso/9780198871101.001.0001>
- Rolls ET (2021c) The neuroscience of emotional disorders. In: Heilman KM, Nadeau SE (eds) *Handbook of clinical neurology: disorders of emotion in neurologic disease*, vol 183. Elsevier, Oxford, pp 1–26
- Rolls ET (2021d) Neurons including hippocampal spatial view cells, and navigation in primates including humans. *Hippocampus* 31:593–611. <https://doi.org/10.1002/hipo.23324>
- Rolls ET (2021e) A neuroscience levels of explanation approach to the mind and the brain. *Front Comput Neurosci* 15:649679. <https://doi.org/10.3389/fncom.2021.649679>
- Rolls ET (2021f) Mind causality: a computational neuroscience approach. *Front Comput Neurosci* 15:70505. <https://doi.org/10.3389/fncom.2021.706505>
- Rolls ET (2022a) *The brain, emotion, and depression*, Chinese. East China Normal University Press, Shanghai

- Rolls ET (2022b) The hippocampus, ventromedial prefrontal cortex, and episodic and semantic memory. *Prog Neurobiol* 217:102334. <https://doi.org/10.1016/j.pneurobio.2022.102334>
- Rolls ET (2023a) Hippocampal spatial view cells, place cells, and concept cells: view representations. *Hippocampus* 33(5):667–687. <https://doi.org/10.1002/hipo.23536>
- Rolls ET (2023b) From brain mechanisms of emotion and decision-making to neuroeconomics. In: Teschl M, Kirman A (eds) *The state of mind in economics*. Cambridge University Press, Cambridge
- Rolls ET (2023c) Hippocampal spatial view cells for memory and navigation, and their underlying connectivity in humans. *Hippocampus* 33(5):533–572. <https://doi.org/10.1002/hipo.23467>
- Rolls ET (2023d) *Brain computations and connectivity*. Oxford University Press, Oxford
- Rolls ET (2023e) The orbitofrontal cortex, food reward, body weight and obesity. *Soc Cogn Affect Neurosci* 18(1):nsab044. <https://doi.org/10.1093/scan/nsab044>
- Rolls ET, Deco G (2006) Attention in natural scenes: neurophysiological and computational bases. *Neural Netw* 19:1383–1394
- Rolls ET, Deco G (2010) *The noisy brain: stochastic dynamics as a principle of brain function*. Oxford University Press, Oxford
- Rolls ET, Deco G (2016) Non-reward neural mechanisms in the orbitofrontal cortex. *Cortex* 83:27–38. <https://doi.org/10.1080/23273798.2016.1203443>
- Rolls ET, Grabenhorst F (2008) The orbitofrontal cortex and beyond: from affect to decision-making. *Prog Neurobiol* 86(3):216–244. <https://doi.org/10.1016/j.pneurobio.2008.09.001>
- Rolls ET, McCabe C (2007) Enhanced affective brain representations of chocolate in cravers vs non-cravers. *Eur J Neurosci* 26:1067–1076
- Rolls BJ, Rolls ET (1981) The control of drinking. *Br Med Bull* 37:127–130
- Rolls BJ, Rolls ET (1982) *Thirst*. Cambridge University Press, Cambridge
- Rolls ET, Rolls JH (1997) Olfactory sensory-specific satiety in humans. *Physiol Behav* 61:461–473
- Rolls ET, Webb TJ (2014) Finding and recognising objects in natural scenes: complementary computations in the dorsal and ventral visual systems. *Front Comput Neurosci* 8:85
- Rolls ET, Williams GV (1987) Neuronal activity in the ventral striatum of the primate. In: Carpenter MB, Jayaraman A (eds) *The Basal Ganglia II—structure and function—current concepts*. Plenum, New York, pp 349–356
- Rolls ET, Xiang J-Z (2005) Reward-spatial view representations and learning in the hippocampus. *J Neurosci* 25:6167–6174
- Rolls ET, Judge SJ, Sanghera M (1977) Activity of neurones in the inferotemporal cortex of the alert monkey. *Brain Res* 130:229–238
- Rolls BJ, Wood RJ, Rolls ET (1980a) Thirst: the initiation, maintenance, and termination of drinking. *Prog Psychobiol Physiol Psychol* 9:263–321
- Rolls BJ, Wood RJ, Rolls ET, Lind H, Lind R, Ledingham J (1980b) Thirst following water deprivation in humans. *Am J Physiol* 239:R476–R482
- Rolls BJ, Rolls ET, Rowe EA, Sweeney K (1981a) Sensory specific satiety in man. *Physiol Behav* 27:137–142
- Rolls BJ, Rolls ET, Rowe EA, Sweeney K (1981b) How sensory properties of foods affect human feeding behaviour. *Physiol Behav* 29:409–417
- Rolls BJ, Rowe EA, Rolls ET, Kingston B, Megson A, Gunary R (1981c) Variety in a meal enhances food intake in man. *Physiol Behav* 26:215–221
- Rolls ET, Rolls BJ, Rowe EA (1983a) Sensory-specific and motivation-specific satiety for the sight and taste of food and water in man. *Physiol Behav* 30:185–192
- Rolls ET, Thorpe SJ, Maddison SP (1983b) Responses of striatal neurons in the behaving monkey. 1: Head of the caudate nucleus. *Behav Brain Res* 7:179–210
- Rolls ET, Murzi E, Yaxley S, Thorpe SJ, Simpson SJ (1986) Sensory-specific satiety: food-specific reduction in responsiveness of ventral forebrain neurons after feeding in the monkey. *Brain Res* 368(1):79–86. [https://doi.org/10.1016/0006-8993\(86\)91044-9](https://doi.org/10.1016/0006-8993(86)91044-9)
- Rolls ET, Scott TR, Sienkiewicz ZJ, Yaxley S (1988) The responsiveness of neurones in the frontal opercular gustatory cortex of the macaque monkey is independent of hunger. *J Physiol* 397:1–12
- Rolls ET, Sienkiewicz ZJ, Yaxley S (1989) Hunger modulates the responses to gustatory stimuli of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *Eur J Neurosci* 1(1):53–60. <https://doi.org/10.1111/j.1460-9568.1989.tb00774.x>
- Rolls ET, Hornak J, Wade D, McGrath J (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J Neurol Neurosurg Psychiatry* 57:1518–1524
- Rolls ET, Critchley HD, Mason R, Wakeman EA (1996) Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J Neurophysiol* 75:1970–1981
- Rolls ET, Aggelopoulos NC, Zheng F (2003a) The receptive fields of inferior temporal cortex neurons in natural scenes. *J Neurosci* 23:339–348
- Rolls ET, Kringelbach ML, de Araujo IET (2003b) Different representations of pleasant and unpleasant odors in the human brain. *Eur J Neurosci* 18:695–703
- Rolls ET, O'Doherty J, Kringelbach ML, Francis S, Bowtell R, McGlone F (2003c) Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices. *Cereb Cortex* 13(3):308–317. <https://doi.org/10.1093/cercor/13.3.308>
- Rolls ET, Browning AS, Inoue K, Hernadi S (2005) Novel visual stimuli activate a population of neurons in the primate orbitofrontal cortex. *Neurobiol Learn Mem* 84:111–123
- Rolls ET, Critchley HD, Browning AS, Inoue K (2006) Face-selective and auditory neurons in the primate orbitofrontal cortex. *Exp Brain Res* 170:74–87
- Rolls ET, Grabenhorst F, Margot C, da Silva MAAP, Velazco MI (2008a) Selective attention to affective value alters how the brain processes olfactory stimuli. *J Cogn Neurosci* 20:1815–1826
- Rolls ET, Grabenhorst F, Parris BA (2008b) Warm pleasant feelings in the brain. *Neuroimage* 41:1504–1513
- Rolls ET, McCabe C, Redoute J (2008c) Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cereb Cortex* 18:652–663
- Rolls ET, Grabenhorst F, Deco G (2010a) Decision-making, errors, and confidence in the brain. *J Neurophysiol* 104:2359–2374
- Rolls ET, Grabenhorst F, Deco G (2010b) Choice, difficulty, and confidence in the brain. *Neuroimage* 53(2):694–706. <https://doi.org/10.1016/j.neuroimage.2010.06.073>
- Rolls ET, Joliot M, Tzourio-Mazoyer N (2015) Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage* 122:1–5. <https://doi.org/10.1016/j.neuroimage.2015.07.075>
- Rolls ET, Mills T, Norton A, Lazidis A, Norton IT (2018) Neuronal encoding of fat using the coefficient of sliding friction in the cerebral cortex and amygdala. *Cereb Cortex* 28:4080–4089
- Rolls ET, Cheng W, Du J, Wei D, Qiu J, Dai D, Zhou Q, Xie P, Feng J (2020a) Functional connectivity of the right inferior frontal gyrus and orbitofrontal cortex in depression. *Soc Cogn Affect Neurosci* 15:75–86. <https://doi.org/10.1093/scan/nsaa014>
- Rolls ET, Cheng W, Feng J (2020b) The orbitofrontal cortex: reward, emotion, and depression. *Brain Commun* 2:fcaa196. <https://doi.org/10.1093/braincomms/fcaa196>

- Rolls ET, Vatansever D, Li Y, Cheng W, Feng J (2020c) Rapid rule-based reward reversal and the lateral orbitofrontal cortex. *Cerebral Cortex Commun.* <https://doi.org/10.1093/texcom/tgaa1087>
- Rolls ET, Deco G, Huang C-C, Feng J (2022a) The human language effective connectome. *Neuroimage* 258:119352. <https://doi.org/10.1016/j.neuroimage.2022.119352>
- Rolls ET, Deco G, Huang CC, Feng J (2022b) The effective connectivity of the human hippocampal memory system. *Cereb Cortex* 32:3706–3725. <https://doi.org/10.1093/cercor/bhab442>
- Rolls ET, Wan Z, Cheng W, Feng J (2022c) Risk-taking in humans and the medial orbitofrontal cortex reward system. *Neuroimage* 249:118893. <https://doi.org/10.1016/j.neuroimage.2022.118893>
- Rolls ET, Deco G, Huang C-C, Feng J (2023a) Human amygdala compared to orbitofrontal cortex connectivity, and emotion. *Prog Neurobiol* 220:102385. <https://doi.org/10.1016/j.pneurobio.2022.102385>
- Rolls ET, Deco G, Huang C-C, Feng J (2023b) Multiple cortical visual streams in humans. *Cereb Cortex* 33:3319–3349. <https://doi.org/10.1093/cercor/bhac276>
- Rolls ET, Deco G, Huang C-C, Feng J (2023c) The connectivity of the human frontal pole cortex, and a theory of its involvement in exploit vs explore
- Rolls ET, Deco G, Huang CC, Feng J (2023d) The human orbitofrontal cortex, vmPFC, and anterior cingulate cortex effective connectome: emotion, memory, and action. *Cereb Cortex* 33:330–359. <https://doi.org/10.1093/cercor/bhac070>
- Rolls ET, Deco G, Huang CC, Feng J (2023e) Prefrontal and somatosensory-motor cortex effective connectivity in humans. *Cereb Cortex* 33(8):4939–4963. <https://doi.org/10.1093/cercor/bhac391>
- Rolls ET, Deco G, Huang CC, Feng J (2023f) The human posterior parietal cortex: effective connectome, and its relation to function. *Cereb Cortex* 33(6):3142–3170. <https://doi.org/10.1093/cercor/bhac266>
- Rolls ET, Feng R, Cheng W, Feng J (2023g) Orbitofrontal cortex connectivity is associated with food reward and body weight in humans. *Soc Cogn Affect Neurosci* 18(1):nsab083. <https://doi.org/10.1093/scan/nsab083>
- Rolls ET, Rauschecker JP, Deco G, Huang C-C, Feng J (2023h) Auditory cortical connectivity in humans. *Cereb Cortex.* <https://doi.org/10.1093/cercor/bhac1496>
- Rolls ET, Wirth S, Deco G, Huang C-C, Feng J (2023i) The human posterior cingulate, retrosplenial and medial parietal cortex effective connectome, and implications for memory and navigation. *Hum Brain Mapp* 44:629–655. <https://doi.org/10.1002/HBM.26089>
- Rosenkilde CE, Bauer RH, Fuster JM (1981) Single unit activity in ventral prefrontal cortex in behaving monkeys. *Brain Res* 209:375–394
- Rosenthal DM (2004) Varieties of higher-order theory. In: Gennaro RJ (ed) Higher order theories of consciousness. John Benjamins, Amsterdam, pp 17–44
- Rozin P, Kalat JW (1971) Specific hungers and poison avoidance as adaptive specializations of learning. *Psychol Rev* 78:459–486
- Rudebeck PH, Behrens TE, Kennerley SW, Baxter MG, Buckley MJ, Walton ME, Rushworth MF (2008) Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J Neurosci* 28(51):13775–13785
- Rudebeck PH, Saunders RC, Lundgren DA, Murray EA (2017) Specialized representations of value in the orbital and ventrolateral prefrontal cortex: desirability versus availability of outcomes. *Neuron* 95(5):1208–1220 e1205. <https://doi.org/10.1016/j.neuron.2017.07.042>
- Rushworth MF, Buckley MJ, Behrens TE, Walton ME, Bannerman DM (2007) Functional organization of the medial frontal cortex. *Curr Opin Neurobiol* 17(2):220–227
- Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex and reward-guided learning and decision-making. *Neuron* 70(6):1054–1069. <https://doi.org/10.1016/j.neuron.2011.05.014>
- Rushworth MF, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Curr Opin Neurobiol* 22(6):946–955. <https://doi.org/10.1016/j.conb.2012.04.011>
- Saez RA, Saez A, Paton JJ, Lau B, Salzman CD (2017) Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward. *Neuron* 95(1):70–77 e73. <https://doi.org/10.1016/j.neuron.2017.06.012>
- Saleem KS, Kondo H, Price JL (2008) Complementary circuits connecting the orbital and medial prefrontal networks with the temporal, insular, and opercular cortex in the macaque monkey. *J Comp Neurol* 506(4):659–693. <https://doi.org/10.1002/cne.21577>
- Saleem KS, Miller B, Price JL (2014) Subdivisions and connective networks of the lateral prefrontal cortex in the macaque monkey. *J Comp Neurol* 522(7):1641–1690. <https://doi.org/10.1002/cne.23498>
- Sallet J, Noonan MP, Thomas A, O'Reilly JX, Anderson J, Papageorgiou GK, Neubert FX, Ahmed B, Smith J, Bell AH, Buckley MJ, Roumazeilles L, Cuell S, Walton ME, Krug K, Mars RB, Rushworth MFS (2020) Behavioral flexibility is associated with changes in structure and function distributed across a frontal cortical network in macaques. *PLoS Biol* 18(5):e3000605. <https://doi.org/10.1371/journal.pbio.3000605>
- Sanghera MK, Rolls ET, Roper-Hall A (1979) Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Exp Neurol* 63:610–626
- Schachter S, Singer J (1962) Cognitive, social and physiological determinants of emotional state. *Psychol Rev* 69:387–399
- Scherer KS (2009) The dynamic architecture of emotion: evidence for the component process model. *Cogn Emot* 23:1307–1351
- Schultz W (2013) Updating dopamine reward signals. *Curr Opin Neurobiol* 23(2):229–238. <https://doi.org/10.1016/j.conb.2012.11.012>
- Schultz W (2016a) Dopamine reward prediction error coding. *Dialog Clin Neurosci* 18(1):23–32
- Schultz W (2016b) Dopamine reward prediction-error signalling: a two-component response. *Nat Rev Neurosci* 17(3):183–195. <https://doi.org/10.1038/nrn.2015.26>
- Schultz W (2016c) Reward functions of the basal ganglia. *J Neural Transm (vienna)* 123(7):679–693. <https://doi.org/10.1007/s00702-016-1510-0>
- Schultz W (2017) Reward prediction error. *Curr Biol* 27(10):R369–R371. <https://doi.org/10.1016/j.cub.2017.02.064>
- Segerdahl AR, Mezue M, Okell TW, Farrar JT, Tracey I (2015) The dorsal posterior insula subserves a fundamental role in human pain. *Nat Neurosci* 18(4):499–500. <https://doi.org/10.1038/nn.3969>
- Seligman ME (1970) On the generality of the laws of learning. *Psychol Rev* 77(5):406–418
- Shadlen MN, Kiani R (2013) Decision making as a window on cognition. *Neuron* 80(3):791–806. <https://doi.org/10.1016/j.neuron.2013.10.047>
- Shallice T, Cipolotti L (2018) The prefrontal cortex and neurological impairments of active thought. *Annu Rev Psychol* 69:157–180. <https://doi.org/10.1146/annurev-psych-010416-044123>
- Sharpe MJ, Wikenheiser AM, Niv Y, Schoenbaum G (2015) The state of the orbitofrontal cortex. *Neuron* 88(6):1075–1077. <https://doi.org/10.1016/j.neuron.2015.12.004>
- Small DM, Scott TR (2009) Symposium overview: what happens to the pontine processing? Repercussions of interspecies differences in pontine taste representation for tasting and feeding. *Ann N Y*

- Acad Sci 1170:343–346. <https://doi.org/10.1111/j.1749-6632.2009.03918.x>
- Stefanacci L, Suzuki WA, Amaral DG (1996) Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys. *J Comp Neurol* 375:552–582
- Steiner JE, Glaser D, Hawilo ME, Berridge KC (2001) Comparative expression of hedonic impact: affective reactions to taste by human infants and other primates. *Neurosci Biobehav Rev* 25(1):53–74
- Sternson SM (2013) Hypothalamic survival circuits: blueprints for purposive behaviors. *Neuron* 77(5):810–824. <https://doi.org/10.1016/j.neuron.2013.02.018>
- Stice E, Figlewicz DP, Gosnell BA, Levine AS, Pratt WE (2013) The contribution of brain reward circuits to the obesity epidemic. *Neurosci Biobehav Rev* 37(9 Pt A):2047–2058. <https://doi.org/10.1016/j.neubiorev.2012.12.001>
- Strait CE, Blanchard TC, Hayden BY (2014) Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron* 82(6):1357–1366. <https://doi.org/10.1016/j.neuron.2014.04.032>
- Strait CE, Sleazer BJ, Hayden BY (2015) Signatures of value comparison in ventral striatum neurons. *PLoS Biol* 13:e1002173. <https://doi.org/10.1371/journal.pbio.1002173>
- Strongman KT (2003) *The psychology of emotion*, 5th edn. Wiley, London
- Taschereau-Dumouchel V, Michel M, Lau H, Hofmann SG, LeDoux JE (2022) Putting the “mental” back in “mental disorders”: a perspective from research on fear and anxiety. *Mol Psychiatry* 27(3):1322–1330. <https://doi.org/10.1038/s41380-021-01395-5>
- Teitelbaum P (1974) The biology of drive. In: Ven Der Kloot W, Walcott C, Dabe B (eds) *Readings in behavior*. Holt, Rinehart, Winston, New York, p 199
- Thorpe SJ, Rolls ET, Maddison S (1983) The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res* 49(1):93–115. <https://doi.org/10.1007/BF00235545>
- Tracey I (2017) Neuroimaging mechanisms in pain: from discovery to translation. *Pain* 158(Suppl 1):S115–S122. <https://doi.org/10.1097/j.pain.0000000000000863>
- Tranel D, Bechara A, Denburg NL (2002) Asymmetric functional roles of right and left ventromedial prefrontal cortices in social conduct, decision-making, and emotional processing. *Cortex* 38(4):589–612. [https://doi.org/10.1016/s0010-9452\(08\)70024-8](https://doi.org/10.1016/s0010-9452(08)70024-8)
- Trull TJ, Widiger TA (2013) Dimensional models of personality: the five-factor model and the DSM-5. *Dialog Clin Neurosci* 15(2):135–146
- Tsuchida A, Fellows LK (2012) Are you upset? Distinct roles for orbitofrontal and lateral prefrontal cortex in detecting and distinguishing facial expressions of emotion. *Cereb Cortex* 22(12):2904–2912. <https://doi.org/10.1093/cercor/bhr370>
- Van Hoesen GW (1981) The differential distribution, diversity and sprouting of cortical projections to the amygdala in the rhesus monkey. In: Ben-Ari Y (ed) *The amygdaloid complex*. Elsevier, Amsterdam, pp 77–90
- Vogt BA (ed) (2009) *Cingulate neurobiology and disease*. Oxford University Press, Oxford
- Vogt BA (2019) The cingulate cortex in neurologic diseases: History, Structure, Overview. *Handb Clin Neurol* 166:3–21. <https://doi.org/10.1016/B978-0-444-64196-0.00001-7>
- Vogt BA, Sikes RW (2000) The medial pain system, cingulate cortex, and parallel processing of nociceptive information. *Prog Brain Res* 122:223–235
- Vogt BA, Derbyshire S, Jones AKP (1996) Pain processing in four regions of human cingulate cortex localized with co-registered PET and MR imaging. *Eur J Neurosci* 8:1461–1473
- Völlm BA, de Araujo IET, Cowen PJ, Rolls ET, Kringelbach ML, Smith KA, Jezzard P, Heal RJ, Matthews PM (2004) Methamphetamine activates reward circuitry in drug naïve human subjects. *Neuropsychopharmacology* 29:1715–1722
- Walton ME, Bannerman DM, Alterescu K, Rushworth MF (2003) Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *J Neurosci* 23:6475–6479
- Walton ME, Devlin JT, Rushworth MF (2004) Interactions between decision making and performance monitoring within prefrontal cortex. *Nat Neurosci* 7(11):1259–1265
- Wan Z, Rolls ET, Cheng W, Feng J (2020) Sensation-seeking is related to functional connectivities of the medial orbitofrontal cortex with the anterior cingulate cortex. *Neuroimage* 215:116845. <https://doi.org/10.1016/j.neuroimage.2020.116845>
- Wang XJ (2002) Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36:955–968
- Warlow SM, Berridge KC (2021) Incentive motivation: “wanting” roles of central amygdala circuitry. *Behav Brain Res* 411:113376. <https://doi.org/10.1016/j.bbr.2021.113376>
- Whalen PJ, Phelps EA (eds) (2009) *The human amygdala*. Guilford, New York
- Williams GV, Rolls ET, Leonard CM, Stern C (1993) Neuronal responses in the ventral striatum of the behaving macaque. *Behav Brain Res* 55(2):243–252. [https://doi.org/10.1016/0166-4328\(93\)90120-f](https://doi.org/10.1016/0166-4328(93)90120-f)
- Wilson FAW, Rolls ET (2005) The primate amygdala and reinforcement: a dissociation between rule-based and associatively-mediated memory revealed in amygdala neuronal activity. *Neuroscience* 133:1061–1072
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81(2):267–279. <https://doi.org/10.1016/j.neuron.2013.11.005>
- Wise SP (2008) Forward frontal fields: phylogeny and fundamental function. *Trends Neurosci* 31(12):599–608. <https://doi.org/10.1016/j.tins.2008.08.008>
- Xie C, Jia T, Rolls ET, Robbins TW, Sahakian BJ, Zhang J, Liu Z, Cheng W, Luo Q, Zac Lo C-Y, Wang H, Banaschewski T, Barker G, Bodke ALW, Buchel C, Quinlan EB, Desrivieres S, Flor H, Grigis A, Garavan H, Gowland P, Heinz A, Hohmann S, Ittermann B, Martinot J-L, Martinot M-LP, Nees F, Papadopoulos Orfanos D, Paus T, Poustka L, Frohner JH, Smolka MN, Walter H, Whelan R, Schumann G, Feng J, IMAGEN C (2021) Reward vs non-reward sensitivity of the medial vs lateral orbitofrontal cortex relates to the severity of depressive symptoms. *Biol Psychiatry Cogn Neurosci Neuroimaging* 6:259–269. <https://doi.org/10.1016/j.bpsc.2020.08.017>
- Yamada H, Louie K, Tymula A, Glimcher PW (2018) Free choice shapes normalized value signals in medial orbitofrontal cortex. *Nat Commun* 9(1):162. <https://doi.org/10.1038/s41467-017-02614-w>
- Yaxley S, Rolls ET, Sienkiewicz ZJ (1988) The responsiveness of neurons in the insular gustatory cortex of the macaque monkey is independent of hunger. *Physiol Behav* 42:223–229
- Zaborszky L, Gombkoto P, Varsanyi P, Gielow MR, Poe G, Role LW, Ananth M, Rajebhosale P, Talmage DA, Hasselmo ME,

- Dannenberg H, Minces VH, Chiba AA (2018) Specific basal forebrain-cortical cholinergic circuits coordinate cognitive operations. *J Neurosci* 38(44):9446–9458. <https://doi.org/10.1523/JNEUROSCI.1676-18.2018>
- Zatka-Haas P, Steinmetz NA, Carandini M, Harris KD (2021) Sensory coding and the causal impact of mouse cortex in a visual decision. *Elife* 10:e63163. <https://doi.org/10.7554/eLife.63163>
- Zhang B, Rolls ET, Wang X, Xie C, Cheng W, Feng J (2023) Roles of the medial and lateral orbitofrontal cortex in major depression and its treatment

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.