Discussion

The roles of the orbitofrontal cortex via the habenula in non-reward and depression, and in the responses of serotonin and dopamine neurons

Edmund T. Rolls [a,b,*,1]

[a] Oxford Centre for Computational Neuroscience, Oxford, UK
[b] Department of Computer Science, University of Warwick, Coventry, UK

ARTICLE INFO

ABSTRACT

Cortical regions such as the orbitofrontal cortex involved in reward and in non-reward and which are implicated in depression, and the amygdala, are connected to the habenula via the striatum and pallidum, and via subcortical limbic structures. The habenula in turn projects to the raphe nuclei, the source of the serotonin-containing neurons that project to the forebrain. It is proposed that this provides a route for cortical signals related to reward, and to not obtaining expected rewards, to influence the serotonin-containing neuronal system that is influenced by many antidepressant treatments. This helps to provide a more circuit-based understanding of the brain mechanisms related to depression, and how some treatments influence this system. The habenula also projects via the rostromedial tegmental nucleus to the dopamine-containing neurons, and this, it is proposed, provides a route for reward prediction error signals and other reward- and punishment-related signals of cortical and striatal origin to influence the dopamine system.

© 2017 Elsevier Ltd. All rights reserved.

This paper is a commentary on the paper "A non-reward attractor theory of depression" that focused on orbitofrontal cortex non-reward attractor networks and their potential relevance to depression (Rolls, 2016b). This commentary draws out the implications of that theory of depression for understanding the functions of the habenula in depression, and the subcortical structures to which it projects including the serotonin (5-HT, 5-hydroxytryptamine) and dopamine systems. This commentary and extension of the theory (Rolls, 2016b) is timely in view of the current interest in the functions of the habenula in depression (Fakhoury, 2017; Loonen and Ivanova, 2015, 2016a, 2017).

The habenula does have neurons that respond to non-reward in macaques (Bromberg-Martin and Hikosaka, 2011; Proulx et al., 2014). Loonen et al. draw attention to this, and consider whether the habenula may be involved in depression (Loonen and Ivanova, 2015, 2016a). That is certainly a possibility. However, it is of interest to consider how the orbitofrontal cortex and its related cortical areas may because of their cortical architecture play an important

functional role in non-reward and depression, in a way that may not be possible for the habenula with its different architecture.

Non-reward neurons, that is neurons that respond when the reward obtained (the outcome) is less than that which is expected (the expected reward value), termed negative reward prediction error neurons, were discovered in the orbitofrontal cortex by Thorpe et al. (1983) (see Rolls, 2014). In the lateral habenula, neurons that respond to signaled low reward value or to punishment have been described (Matsumoto and Hikosaka, 2009), and so have neurons that reflect negative reward prediction error (Bromberg-Martin and Hikosaka, 2011). Similar neurons are found in the globus pallidus glutamatergic excitatory habenula-projecting neurons, providing evidence that the necessary computations are not performed in the lateral habenula (Stephenson-Jones et al., 2016).

One key point is that the neocortex with its highly developed local recurrent collateral excitatory connections between the pyramidal cells appears to allow ongoing firing to continue in attractor states, each one of which might represent a short-term memory, an expected value, a long-term memory, the result of a decision just taken, or the recent receipt of non-reward (Rolls, 2016a).

This ability to maintain firing may allow for a stimulus previously associated with a reward to produce expected reward firing, which may continue until the reward is or is not received (the outcome). Another 'non-reward' attractor network may be triggered

* Corresponding author at: Oxford Centre for Computational Neuroscience, Oxford, UK.
  E-mail address: Edmund.Rolls@oxcns.org
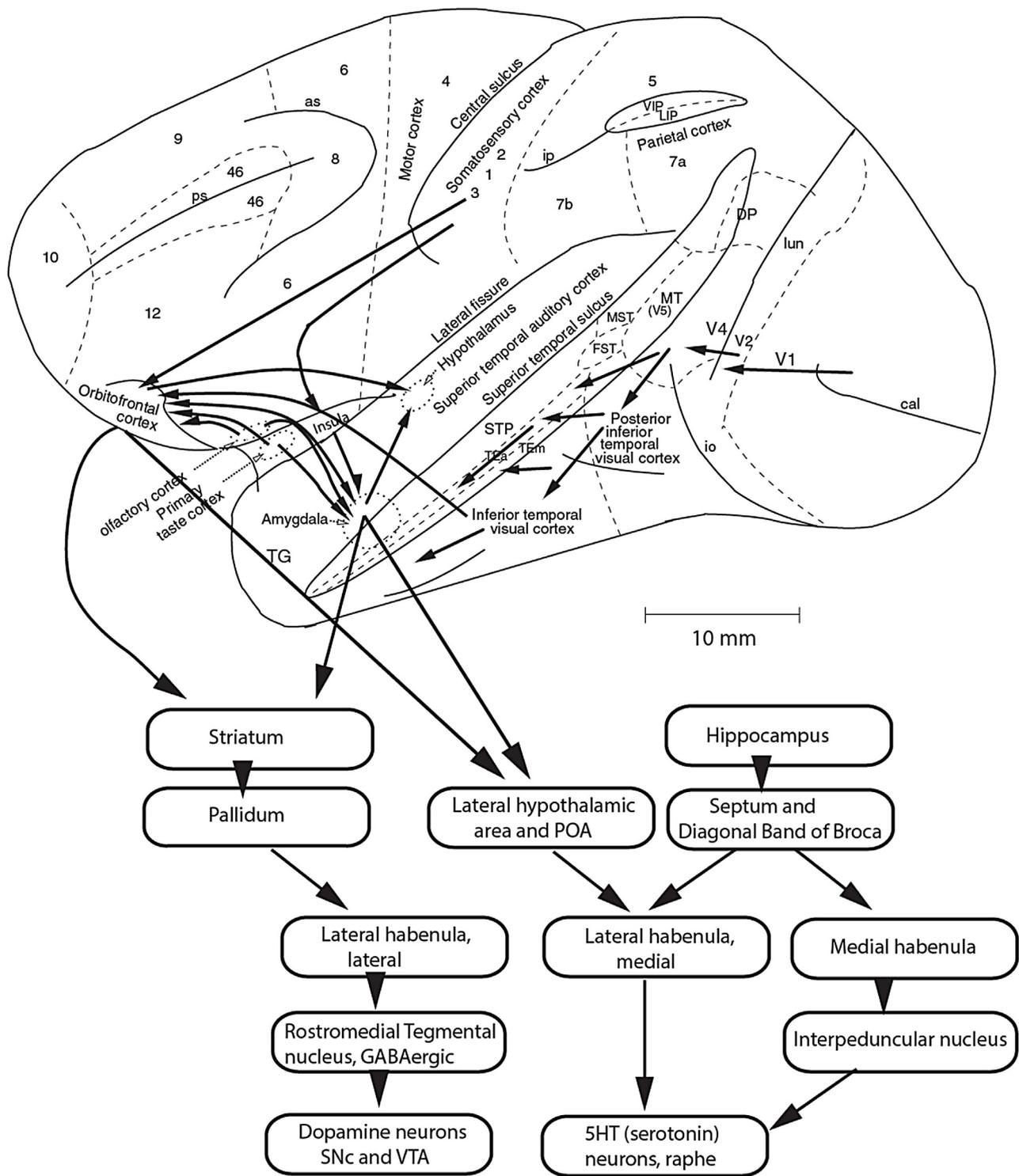 [1] http://www.oxcns.org.

Fig. 1. The orbitofrontal cortex and amygdala systems involved in reward and non-reward can operate via a lateral hypothalamic area/lateral preoptic area (POA) to influence the Lateral Habenula, medial part, which in turn can influence the 5-HT (serotonin) neurons in the raphe nuclei. Many antidepressant drugs may influence this cortical to brainstem pathway by influencing the effects of the 5-HT neurons, which terminate in many brain areas. The hippocampus influence via the septal nuclei and diagonal band of Broca may enable reward context to access the same Lateral Habenula, medial part, to 5-HT-neuron system (Luo et al., 2011; Rolls, 2015). The medial habenula also receives septal inputs, and projects to the interpeduncular nucleus, and thereby to 5-HT neurons (and probably dopamine neurons) (see Fig. 1) (Loonen and Ivanova, 2016b; Proulx et al., 2014). The orbitofrontal cortex, amygdala (and probably anterior cingulate cortex and subgenual cingulate cortex) systems involved in reward and non-reward can operate via a basal ganglia route (striatum, ventral pallidum, and globus pallidus/bed nucleus of the stria terminalis) to influence the Lateral Habenula, lateral part, which in turn via the GABAergic Rostromedial Tegmental nucleus can influence dopamine neurons in the Substantia Nigra pars compacta and ventral Tegmental Area (SNc and VTA). This provides a route for reward, non-reward, and reward prediction error signals of largely cortical origin to influence the dopamine neurons. Details of some of these anatomical connections are provided elsewhere (Loonen and Ivanova, 2016b; Proulx et al., 2014). These connections are shown in the context of some of the pathways involved in reward-related processes and emotion shown on the lateral view of the brain of the macaque monkey in the upper part of the Figure (Rolls, 2014). Connections from the primary taste and olfactory cortices to the orbitofrontal cortex and amygdala are shown. Connections are also shown in the 'ventral visual system' from the visual cortical areas V1 to V2, V4, the inferior temporal visual cortex, etc., with some connections reaching the amygdala and orbitofrontal cortex. In addition,

into activity lasting for several seconds if the expected reward is not received (Rolls, 2014). Possible neuronal network mechanisms to compute this in the orbitofrontal cortex have been described (Rolls and Deco, 2016).

When reward associations with stimuli keep changing, reward-related behavior can switch in one trial, because a rule attractor for which stimulus is currently rewarded holds the current rule in short-term memory (Rolls, 2016a,b; Rolls and Deco, 2016; Thorpe et al., 1983). Again, a cortical attractor network appears to be involved in this type of computation.

Further, in addition to the local attractor in cortical areas such as the lateral orbitofrontal cortex for non-reward, Rolls hypothesizes a long-loop attractor involving language cortical areas, which he proposes contributes to ruminating thoughts in depression, and which feeds back to the non-reward system in the orbitofrontal cortex to maintain its activity (Rolls, 2016b).

None of these computational functions would seem to be supported by the habenula, as it may not have the highly developed recurrent collateral system that is characteristic of the neocortex. Indeed, the lateral habenula is a subcortical structure that receives it inputs from the basal ganglia (the habenula-projecting excitatory glutamatergic globus pallidus neurons (GPh) (Stephenson-Jones et al., 2016)), diagonal band of Broca, bed nucleus of the stria terminalis, and hypothalamus, and projects on to neuronal systems that include the dopamine and serotonin neurons in the brainstem (Loonen and Ivanova, 2015, 2016b; Proulx et al., 2014) (see Fig. 1). The medial habenula receives septal inputs, and projects to the interpeduncular nucleus and 5-HT neurons (see Fig. 1) (Loonen and Ivanova, 2016b; Proulx et al., 2014).

Now in these considerations of the habenula, there is little that is described about from where the habenula receives its non-reward and punishment-related signals. In fact, the likely source is the orbitofrontal cortex and related cortical areas, which have major projections to the striatum, which in turn projects to the globus pallidus/ventral pallidum, lateral hypothalamus etc (Fig. 1) (Haber et al., 2006; Haber and Knutson, 2010; Hong and Hikosaka, 2013). Indeed, in an investigation of the macaque ventral striatum, reward and punishment-related neurons were discovered with properties that were generally similar to those in the orbitofrontal cortex that sends projections to the ventral striatum (Williams et al., 1993). Further, it has always been the hypothesis that the reward and punishment-related signals that reach the dopamine and related neurons have their origin in primates largely in the orbitofrontal cortex and its closely connected areas such as the anterior cingulate cortex and amygdala (Rolls, 2014). In this situation, the habenula can be considered as a diencephalic region that relays information of cortical origin received via the basal ganglia, septum, and hypothalamus, and that projects on to brainstem structures such as the dopamine and serotonin-containing neurons (see Fig. 1). As such, the habenula may be an important route for reward, non-reward, and punisher-related information to influence, via the rostromedial tegmental nucleus, brainstem systems such as the 5-Hydroxy-Tryptamine (serotonin) neurons that in turn can influence mood and depression, and may carry useful signals. Indeed, the habenula provides a route by which the many antidepressants that facilitate 5-HT function (Fakhoury, 2016; Harmer

and Cowen, 2013; Ma, 2015) can produce their effect, by influencing pathways carrying the results of cortical computation about reward and non-reward (Fig. 1). But the habenula is unlikely itself to perform the crucial computational functions, to compute for example non-reward/negative reward prediction error from expected value and outcome value signals, all of which are represented in the orbitofrontal cortex and computed by it from the sensory inputs that it receives (Rolls, 2014, 2016a; Rolls and Grabenhorst, 2008). The orbitofrontal cortex is crucial in these computations, in that the preceding cortical stages do not represent reward value (e.g. insula for taste, inferior temporal visual cortex for objects) (Grabenhorst and Rolls, 2011; Rolls, 2014; Rolls and Grabenhorst, 2008). The orbitofrontal cortex, which represents reward outcome value, expected reward value, and negative reward prediction error (Grabenhorst and Rolls, 2011; Rolls, 2014; Rolls and Grabenhorst, 2008), thus appears to perform these computations of stimulus reward value. The orbitofrontal cortex then projects onto the cingulate cortex (which learns which actions to be performed based on the rewards received), and to the basal ganglia, amygdala, etc (Rolls, 2014).

Evidence on the orbitofrontal cortex in humans, including evidence from humans with depression, does appear to provide support for the hypothesis that the orbitofrontal cortex is involved in depression (Rolls, 2016b). In relation to whether the lateral orbitofrontal cortex is involved in depression, the largest, voxel-level, brain-wide analysis of functional connectivity in depression involving 421 patients with major depressive disorder and 488 controls showed that the lateral orbitofrontal cortex (involved in non-reward (Grabenhorst and Rolls, 2011; Rolls, 2014; Rolls and Grabenhorst, 2008)) has increased functional connectivity in depression with the precuneus, angular gyrus, and middle temporal gyrus (which could relate respectively to the low self-esteem, ruminating thoughts, and negative evaluation of face expressions in depression) (Cheng et al., 2016). In contrast, the medial orbitofrontal cortex, implicated in reward (Grabenhorst and Rolls, 2011; Rolls, 2014; Rolls and Grabenhorst, 2008), has reduced functional connectivity with memory systems in the medial temporal lobe (via perirhinal and entorhinal cortex), which may be related to the bias away from happy to sad memories in depression (Cheng et al., 2016).

Given that there is evidence linking the habenula to depression (Fakhoury, 2017; Loonen and Ivanova, 2015, 2016a, 2017; Proulx et al., 2014), it will be of great interest in future to investigate further how cortical non-reward mechanisms linked to depression are functionally connected via the basal ganglia and subcortical limbic structures to the habenula, and thus to the serotonin and dopamine-containing brainstem neurons, providing one route for cortical systems to influence mood and depression, by the pathways illustrated in Fig. 1.

An important part of the extension to the theory of a non-reward orbitofrontal cortex attractor system for depression (Rolls, 2016b) that is introduced here is that the orbitofrontal cortex via the basal ganglia and habenula may provide the inputs to the dopamine neurons that enable them to respond to positive reward prediction error and to other reward- and punishment-related signals (Bromberg-Martin et al., 2010; Schultz, 2013, 2016) (Fig. 1), for

---

connections from the somatosensory cortical areas BA 1, 2, and 3 that reach the orbitofrontal cortex directly and via the insular cortex, and that reach the amygdala via the insular cortex, are shown. as, arcuate sulcus; cal, calcarine sulcus; cs, central sulcus; lf, lateral (or Sylvian) fissure; lun, lunate sulcus; ps, principal sulcus; io, inferior occipital sulcus; ip, intraparietal sulcus (which has been opened to reveal some of the areas it contains); sts, superior temporal sulcus (which has been opened to reveal some of the areas it contains). AIT, anterior inferior temporal cortex; FST, visual motion processing area; LIP, lateral intraparietal area; MST, visual motion processing area; MT, visual motion processing area (also called V5); PIT, posterior inferior temporal cortex; STP, superior temporal plane; TA, architectonic area including auditory association cortex; TE, architectonic area including high order visual association cortex, and some of its subareas TEa and TEm; TG, architectonic area in the temporal pole; V1–V4, visual areas V1–V4; VIP, ventral intraparietal area; TEO, architectonic area including posterior visual association cortex. The numerals refer to architectonic areas, and have the following approximate functional equivalence: 1,2,3, somatosensory cortex (posterior to the central sulcus); 4, motor cortex; 5, superior parietal lobule; 7a, inferior parietal lobule, visual part; 7b, inferior parietal lobule, somatosensory part; 6, lateral premotor cortex; 8, frontal eye field; 12, part of orbitofrontal cortex; 46, dorsolateral prefrontal cortex. (Habenula2.eps).

the dopamine neurons could not themselves compute these representations for which cortical attractor operations are important as described above.

To summarize, the orbitofrontal cortex, and perhaps closely connected cortical areas such as the anterior cingulate cortex, are key areas involved in the computation of reward value, expected reward value, and reward prediction errors, and these cortical areas are, it is proposed in this paper, the source of these signals that reach the basal ganglia, habenula, and thereby the serotonin neurons implicated in depression, and also the dopamine neurons. Evidence for this includes the following, described above and in more detail elsewhere (Rolls, 2014, 2016a,b). First, the pathways that reach the orbitofrontal cortex provide the information required for these computations (Fig. 1). Second, the orbitofrontal cortex contains neurons in macaques that encode reward value, expected reward value, punishment, expected punishment, and negative reward prediction error ('non-reward neurons'), with consistent fMRI evidence in humans. Third, behaviors that require the rapid re-evaluation of expected reward value and a rapid change of behavior are impaired by damage to the orbitofrontal cortex in primates including humans (Hornak et al., 2004; Rolls, 2016b). Fourth, cortical areas contain the recurrent collateral connections that provide the basis for holding neuronal activity on-line in attractor networks that is required to compute and maintain expected value, expected punishment, and reward prediction errors. Subcortical structures including the basal ganglia, habenula and the serotonin and dopamine neurons do not have this architecture and thus computational ability (Rolls, 2016a). Fifth, there are pathways via the basal ganglia and other subcortical structures for signals computed in the orbitofrontal cortex and related cortical areas to reach the habenula and the serotonin and dopamine-containing neurons (Fig. 1). There may be some transformation of signals on the route from the orbitofrontal cortex to the habenula and brainstem, but for the reasons given, the main computations are likely to be performed in the cortical areas. For these reasons, it will be of interest in future to develop further our understanding of how cortical areas such as the orbitofrontal cortex may be involved in psychiatric disorders including depression (Cheng et al., 2016; Rolls, 2016a,b).

References

Bromberg-Martin, E.S., Hikosaka, O., 2011. Lateral habenula neurons signal errors in the prediction of reward information. Nat. Neurosci. 14, 1209–1216.

Bromberg-Martin, E.S., Matsumoto, M., Hikosaka, O., 2010. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68, 815–834.

Cheng, W., Rolls, E.T., Qiu, J., Liu, W., Tang, Y., Huang, C.C., Wang, X., Zhang, J., Lin, W., Zheng, L., Pu, J., Tsai, S.J., Yang, A.C., Lin, C.P., Wang, F., Xie, P., Feng, J., 2016. Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression. Brain 139, 3296–3309.

Fakhoury, M., 2016. Revisiting the Serotonin Hypothesis: Implications for Major Depressive Disorders. Mol. Neurobiol. 53, 2778–2786.

Fakhoury, M., 2017. The habenula in psychiatric disorders: more than three decades of translational investigation. Neurosci. Biobehav. Rev.

Grabenhorst, F., Rolls, E.T., 2011. Value, pleasure, and choice in the ventral prefrontal cortex. Trends Cogn. Sci. 15, 56–67.

Haber, S.N., Kim, K.S., Mailly, P., Calzavara, R., 2006. Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. J. Neurosci. 26, 8368–8376.

Haber, S.N., Knutson, B., 2010. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35, 4–26.

Harmer, C.J., Cowen, P.J., 2013. 'It's the way that you look at it'–a cognitive neuropsychological account of SSRI action in depression. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20120407.

Hong, S., Hikosaka, O., 2013. Diverse sources of reward value signals in the basal ganglia nuclei transmitted to the lateral habenula in the monkey. Front. Hum. Neurosci. 7, 778.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E.T., Morris, R.G., Bullock, P.R., Polkey, C.E., 2004. Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans. J. Cogn. Neurosci. 16, 463–478.

Loonen, A.J., Ivanova, S.A., 2015. Circuits regulating pleasure and happiness: the evolution of reward-seeking and misery-fleeing behavioral mechanisms in vertebrates. Front. Neurosci. 9, 394.

Loonen, A.J., Ivanova, S.A., 2016a. Circuits regulating pleasure and happiness in major depression. Med. Hypotheses 87, 14–21.

Loonen, A.J., Ivanova, S.A., 2016b. Circuits regulating pleasure and happiness: the evolution of the amygdalar-hippocampal-habenular connectivity in vertebrates. Front. Neurosci. 10, 539.

Loonen, A.J., Ivanova, S.A., 2017. Commentary on a non-reward attractor theory of depression: a proposal to include the habenula connnection. Neurosci. Biobehav. Rev.

Luo, A.H., Tahsili-Fahadan, P., Wise, R.A., Lupica, C.R., Aston-Jones, G., 2011. Linking context with reward: a functional circuit from hippocampal CA3 to ventral tegmental area. Science 333, 353–357.

Ma, Y., 2015. Neuropsychological mechanism underlying antidepressant effect: a systematic meta-analysis. Mol. Psychiatry 20, 311–319.

Matsumoto, M., Hikosaka, O., 2009. Representation of negative motivational value in the primate lateral habenula. Nat. Neurosci. 12, 77–84.

Proulx, C.D., Hikosaka, O., Malinow, R., 2014. Reward processing by the lateral habenula in normal and depressive behaviors. Nat. Neurosci. 17, 1146–1152.

Rolls, E.T., 2014. Emotion and Decision-Making Explained. Oxford University Press, Oxford.

Rolls, E.T., 2015. Limbic systems for emotion and for memory, but no single limbic system. Cortex 62, 119–157.

Rolls, E.T., 2016a. Cerebral Cortex: Principles of Operation. Oxford University Press, Oxford.

Rolls, E.T., 2016b. A non-reward attractor theory of depression. Neurosci. Biobehav. Rev. 68, 47–58.

Rolls, E.T., Deco, G., 2016. Non-reward neural mechanisms in the orbitofrontal cortex. Cortex 83, 27–38.

Rolls, E.T., Grabenhorst, F., 2008. The orbitofrontal cortex and beyond: from affect to decision-making. Prog. Neurobiol. 86, 216–244.

Schultz, W., 2013. Updating dopamine reward signals. Curr. Opin. Neurobiol. 23, 229–238.

Schultz, W., 2016. Dopamine reward prediction-error signalling: a two-component response. Nat. Rev. Neurosci. 17, 183–195.

Stephenson-Jones, M., Yu, K., Ahrens, S., Tucciarone, J.M., van Huijstee, A.N., Mejia, L.A., Penzo, M.A., Tai, L.H., Wilbrecht, L., Li, B., 2016. A basal ganglia circuit for evaluating action outcomes. Nature 539, 289–293.

Thorpe, S.J., Rolls, E.T., Maddison, S., 1983. Neuronal activity in the orbitofrontal cortex of the behaving monkey. Exp. Brain Res. 49, 93–115.

Williams, G.V., Rolls, E.T., Leonard, C.M., Stern, C., 1993. Neuronal responses in the ventral striatum of the behaving macaque. Behav. Brain Res. 55, 243–252.