

Non-Reward Neural Mechanisms in the Orbitofrontal Cortex

Edmund T. Rolls, Oxford Centre for Computational Neuroscience, Oxford, UK*
and University of Warwick, Department of Computer Science, Coventry CV4 7AL, UK

and

Gustavo Deco, Universitat Pompeu Fabra, Theoretical and Computational Neuroscience
Roc Boronat 138, 08018 Barcelona, Spain
and Institutio Catalana de Recerca i Estudis Avancats (ICREA)

May 3, 2016

Key words: emotion, reward, non-reward, reward prediction error, orbitofrontal cortex, attractor network, depression, impulsiveness

*Corresponding author. Oxford Centre for Computational Neuroscience, Oxford, UK. Email: Edmund.Rolls@oxcns.org, Url: <http://www.oxcns.org>

Abstract

Single neurons in the primate orbitofrontal cortex respond when an expected reward is not obtained, and behavior must change. The human lateral orbitofrontal cortex is activated when non-reward, or loss occurs. The neuronal computation of this negative reward prediction error is fundamental for the emotional changes associated with non-reward, and with changing behavior. Little is known about the neuronal mechanism. Here we propose a mechanism, which we formalize into a neuronal network model, which is simulated to enable the operation of the mechanism to be investigated. A single attractor network has a Reward population (or pool) of neurons that is activated by Expected Reward, and maintain their firing until, after a time, synaptic depression reduces the firing rate in this neuronal population. If a Reward Outcome is not received, the decreasing firing in the Reward neurons releases the inhibition implemented by inhibitory neurons, and this results in a second population of Non-Reward neurons to start and continue firing encouraged by the spiking-related noise in the network. If a Reward Outcome is received, this keeps the Reward attractor active, and this through the inhibitory neurons prevents the Non-Reward attractor neurons from being activated. If an Expected Reward has been signalled, and the Reward Attractor neurons are active, their firing can be directly inhibited by a Non-Reward Outcome, and the Non-Reward neurons become activated because the inhibition on them is released. The neuronal mechanism in the orbitofrontal cortex for computing negative reward prediction error are important, for this system may be over-reactive in depression, under-reactive in impulsive behavior, and may influence the dopaminergic ‘prediction error’ neurons.

1 Introduction: non-reward neurons in the orbitofrontal cortex, and the mechanism by which they may be activated

The aim of the research described here is to investigate the neuronal mechanisms that may underlie the computation of non-reward in the primate including human orbitofrontal cortex (Thorpe, Rolls and Maddison 1983, Rolls and Grabenhorst 2008, Rolls 2014). This is a fundamentally important process involved in changing behavior when an expected reward is not received. The process is important in many emotions, which can be produced if an expected reward is not received or lost, with examples including sadness and anger (Rolls 2014). Consistent with this the psychiatric disorder of depression may arise when the non-reward system leading to sadness is too sensitive, or maintains its activity for too long (Rolls 2016b), or has increased functional connectivity (Cheng, Rolls, Qiu, Liu, Tang, Huang, Wang, Zhang, Lin, Zheng, Pu, Tsai, Yang, Lin, Wang, Xie and Feng 2016). Conversely, if the non-reward system is underactive or is damaged by lesions of the orbitofrontal cortex, the decreased sensitivity to non-reward may contribute to increased impulsivity (Berlin, Rolls and Kischka 2004, Berlin, Rolls and Iversen 2005), and even antisocial and psychopathic behavior (Rolls 2014). For these reasons, understanding the mechanisms that underlie non-reward is important not only for understanding normal human behavior and how it changes when rewards are not received, but also for understanding and potentially treating better some emotional and psychiatric disorders.

In the investigation described here, we develop hypotheses about the mechanisms involved in the computation of non-reward in the orbitofrontal cortex (Section 3) based on neurophysiological, functional neuroimaging, and neuropsychological evidence on the orbitofrontal cortex (Section 2). (These hypotheses are based on the functions of the orbitofrontal cortex in primates including humans, given the evidence that most of the orbitofrontal part that is present in rodents is agranular, and does not correspond to most of the primate including human orbitofrontal cortex (Wise 2008, Passingham and Wise 2012, Rolls 2014).) We then develop an integrate-and-fire neuronal network model of how the non-reward signal is computed, and simulate the model to elucidate its properties, and to examine the parameters for the correct operation of the model (Sections 3 and 4). The particular neuronal responses that are modelled are how the non-reward neurons in the orbitofrontal cortex (Thorpe, Rolls and Maddison 1983, Rosenkilde, Bauer and Fuster 1981) are triggered into a high firing rate with continuing activity, if no Reward Outcome is received after an Expected Reward has been indicated. The non-reward neurons should also be activated if after an expected reward signal, a punishment is received. The non-reward neurons should not be triggered if after an expected reward signal, a reward outcome is received Thorpe, Rolls and Maddison (1983).

2 Background to the hypotheses: Neurophysiological, neuroimaging, and neuropsychological evidence on what needs to be modelled to account for non-reward neurons

Some of the background evidence on the role of the primate including human orbitofrontal cortex in non-reward, including crucially what needs to be modelled at the neuronal, mechanism, level, is as follows.

Single neurons in the primate orbitofrontal cortex respond when an expected reward is

not obtained, and behavior must change. This was discovered by Thorpe, Rolls and Maddison (1983), who found that 3.5% of neurons in the macaque orbitofrontal cortex detect different types of non-reward. These neurons thus signal negative reward prediction error, the reward outcome value minus the expected value. For example, some neurons responded in extinction, immediately after a lick had been made to a visual stimulus that had previously been associated with fruit juice reward. Other non-reward neurons responded in a reversal task, immediately after the monkey had responded to the previously rewarded visual stimulus, but had obtained the punisher of salt taste rather than reward, indicating that the choice of stimulus should change in this visual discrimination reversal task. Importantly, at least some of these non-reward neurons continue firing for several seconds when an expected reward is not obtained, as illustrated in Fig. 1. These neurons do not respond when an expected punishment is received, for example the taste of salt from a correctly labelled dispenser. Different non-reward neurons may respond in different tasks, providing the potential for behaviour to reverse in one task, but not in all tasks (Thorpe, Rolls and Maddison 1983, Rolls and Grabenhorst 2008, Rolls 2014). The existence of neurons in the middle part of the macaque orbitofrontal cortex that respond to non-reward (Thorpe, Rolls and Maddison 1983) (originally described by Thorpe, Maddison and Rolls (1979) and Rolls (1981)) is confirmed by recordings that revealed 10 such non-reward neurons (of 140 recorded, or approximately 7%) found in delayed match to sample and delayed response tasks by Joaquin Fuster and colleagues (Rosenkilde, Bauer and Fuster 1981).

The orbitofrontal cortex is likely to be where non-reward is computed, for not only does it contain non-reward neurons, but it also contains neurons that encode the signals from which non-reward is computed. These signals include a representation of Expected Reward, and Reward Outcome (Rolls 2014). Expected reward is represented in the primate orbitofrontal cortex in that many neurons respond to the sight of a food reward (Thorpe, Rolls and Maddison 1983), learn to respond to any visual stimulus associated with a food reward (Rolls, Critchley, Mason and Wakeman 1996), very rapidly reverse the stimulus to which they respond when the reward outcome associated with each stimulus changes (Rolls, Critchley, Mason and Wakeman 1996), and represent reward value in that they stop responding to a visual stimulus when the reward is devalued by satiation (Critchley and Rolls 1996a). Other orbitofrontal cortex neurons represent the expected reward value of olfactory stimuli, based on similar evidence (Rolls and Baylis 1994, Critchley and Rolls 1996b, Rolls, Critchley, Mason and Wakeman 1996, Critchley and Rolls 1996a). Reward Outcome is represented in the orbitofrontal cortex by its neurons that respond to taste and fat texture (Rolls, Yaxley and Sienkiewicz 1990, Rolls and Baylis 1994, Rolls, Critchley, Browning, Hernadi and Lenard 1999, Verhagen, Rolls and Kadohisa 2003, Rolls, Verhagen and Kadohisa 2003b), and do this based on their reward value as shown by the fact that their responses decrease to zero when the reward is devalued by feeding to satiety (Rolls, Sienkiewicz and Yaxley 1989, Rolls, Critchley, Browning, Hernadi and Lenard 1999, Rolls 2015b, Rolls 2016c). Moreover, the primate orbitofrontal cortex is key in these reward value representations, for it receives the necessary inputs from the inferior temporal visual cortex, insular taste cortex, and pyriform olfactory cortex, yet in these preceding areas reward value is not represented (Rolls 2014, Rolls 2015b, Rolls 2015a). There is consistent evidence that Expected Value and Reward Outcome Value are represented in the human orbitofrontal cortex (O'Doherty, Kringelbach, Rolls, Hornak and Andrews 2001, Rolls, O'Doherty, Kringelbach, Francis, Bowtell and McGlone 2003a, Kringelbach, O'Doherty, Rolls and Andrews 2003, Rolls, McCabe and Redoute 2008, Grabenhorst, Rolls and Bilderbeck 2008, Rolls 2014, Rolls 2015b).

In that most neurons in the macaque orbitofrontal cortex respond to reinforcers and punishers, or to stimuli associated with rewards and punishers, and do not respond in relation to responses, the orbitofrontal cortex is closely related to stimulus processing, including the stimuli that give rise to affective states (Rolls 2014). When the orbitofrontal cortex computes errors, it computes mismatches between stimuli that are expected, and stimuli that are obtained, and in this sense the errors are closely related to those required to produce affective states, and to change the reward value representation of a stimulus. This type of error representation provided by the orbitofrontal cortex may thus be different from that represented in the cingulate cortex, in which behavioural responses are represented, where the errors may be more closely related to errors that arise when action–outcome expectations are not met, and where action–outcome rather than stimulus–reinforcer representations need to be corrected (Rushworth, Kolling, Sallet and Mars 2012, Grabenhorst and Rolls 2011, Rolls 2014).

We have also been able to obtain evidence that non-reward used as a signal to reverse behavioral choice is represented in the human orbitofrontal cortex. Kringelbach and Rolls (2003) used the faces of two different people, and if one face was selected then that face smiled, and if the other was selected, the face showed an angry expression. After good performance was acquired, there were repeated reversals of the visual discrimination task. Kringelbach and Rolls (2003) found that activation of a lateral part of the orbitofrontal cortex in the fMRI study was produced on the error trials, that is when the human chose a face, and did not obtain the expected reward. Control tasks showed that the response was related to the error, and the mismatch between what was expected and what was obtained as the reward outcome, in that just showing an angry face expression did not selectively activate this part of the lateral orbitofrontal cortex. An interesting aspect of this study that makes it relevant to human social behavior is that the conditioned stimuli were faces of particular individuals, and the unconditioned stimuli were face expressions. Moreover, the study reveals that the human orbitofrontal cortex is very sensitive to social feedback when it must be used to change behaviour (Kringelbach and Rolls 2003, Kringelbach and Rolls 2004). Correspondingly, it has now been shown in macaques using fMRI that the lateral orbitofrontal cortex is activated by non-reward during reversal (Chau, Sallet, Papageorgiou, Noonan, Bell, Walton and Rushworth 2015).

The non-reward neurons in the orbitofrontal cortex are implicated in changing behavior when non-reward occurs. Monkeys with orbitofrontal cortex damage are impaired on Go/NoGo task performance, in that they Go on the NoGo trials (Iversen and Mishkin 1970), and in an object-reversal task in that they respond to the object that was formerly rewarded with food, and in extinction in that they continue to respond to an object that is no longer rewarded (Butter 1969, Jones and Mishkin 1972, Meunier, Bachevalier and Mishkin 1997). The visual discrimination reversal learning deficit shown by monkeys with orbitofrontal cortex damage (Jones and Mishkin 1972, Baylis and Gaffan 1991, Murray and Izquierdo 2007) may be due at least in part to the tendency of these monkeys not to withhold responses to non-rewarded stimuli (Jones and Mishkin 1972) including objects that were previously rewarded during reversal (Rudebeck and Murray 2011), and including foods that are not normally accepted (Butter, McDonald and Snyder 1969, Baylis and Gaffan 1991). Consistently, orbitofrontal cortex (but not amygdala lesions) impaired instrumental extinction (Murray and Izquierdo 2007). Similarly, in humans, patients with ventral frontal lesions made more errors in the reversal (or in a similar extinction) task, and completed fewer reversals, than control patients with damage elsewhere in the frontal lobes or in other brain regions (Rolls, Hornak, Wade and McGrath 1994), and continued to respond to the previously rewarded

stimulus when it was no longer rewarded. A reversal deficit in a similar task in patients with ventromedial frontal cortex damage was also reported by Fellows and Farah (2003). Further, in patients with well localised surgical lesions of the orbitofrontal cortex (made to treat epilepsy, tumours etc), it was found that they were severely impaired at the reversal task, in that they accumulated less money (Hornak, O’Doherty, Bramham, Rolls, Morris, Bullock and Polkey 2004). These patients often failed to switch their choice of stimulus after a large loss; and often did switch their choice even though they had just received a reward, and this has been quantified in a more recent study (Berlin, Rolls and Kischka 2004). The importance of the failure to rapidly learn about the value of stimuli from negative feedback has also been described as a critical difficulty for patients with orbitofrontal cortex lesions (Fellows 2007, Wheeler and Fellows 2008, Fellows 2011), and has been contrasted with the effects of lesions to the anterior cingulate cortex which impair the use of feedback to learn about actions (Fellows 2011, Camille, Tsuchida and Fellows 2011, Rolls 2014).

Dopamine neurons, some of which appear to respond to positive reward prediction error (when a reward outcome is larger than expected) (Schultz 2013), do not provide the answer to how non-reward is computed for a number of reasons (Rolls 2014). First, they may reflect only weakly any non-reward, by having somewhat reduced firing rates below their already low firing rates. Second, there are no known computations that could be performed by dopamine neurons based on expected reward and reward outcome signals, which are not known to be represented in the midbrain region where the dopamine neurons are represented. Instead, it is suggested that the dopamine neurons reflect at least in part processing of expected reward value, outcome reward value, and negative reward prediction error performed in the orbitofrontal cortex (Rolls 2014). Consistent with this, the orbitofrontal cortex does project to the ventral striatum in which neurons of the type present in the orbitofrontal cortex are found (Rolls and Williams 1987, Williams, Rolls, Leonard and Stern 1993, Rolls, Thorpe and Maddison 1983), and which then projects to the midbrain area where the dopamine neuron cell bodies are located (Rolls 2014).

Given these findings, the aim of the current investigation was to produce hypotheses and a model that could account for the non-reward firing of orbitofrontal cortex neurons, as analyzed by Thorpe, Rolls and Maddison (1983). The particular neuronal responses to model were how the non-reward neurons are triggered into a high firing rate with continuing activity, if no Reward Outcome is received after an Expected Reward has been indicated. The non-reward neurons should also be activated if after an expected reward signal, a punishment is received. The non-reward neurons should not be triggered if after an expected reward signal, a reward outcome is received Thorpe, Rolls and Maddison (1983). The hypothesis and model utilize the properties of Expected Reward and Reward Outcome neurons in the orbitofrontal cortex, as well as the responses of non-reward neurons in the orbitofrontal cortex, as just described.

3 Methods

3.1 Outline of the hypothesis and model

A network with two competing attractor subpopulations of neurons, one for Reward, and one for Non-Reward, is proposed. The single attractor network has one population (or pool) of neurons that is activated by Expected Reward, and maintains its firing until, after a time, synaptic depression reduces the firing rate in this Reward neuronal population. A second population of neurons encodes Non-Reward, and is triggered into high firing rate

continuing activity because of reduced inhibition from the common inhibitory neurons if no Reward Outcome is received to keep the Reward attractor population active. The emergence of activity when it is not being inhibited in the Non-Reward neurons is facilitated by the stochastic spiking-time-related noise in the system (Rolls and Deco 2010). This accounts for Non-Reward signalling in extinction, if an Expected Reward is not followed by a Reward Outcome. If a Reward Outcome is received, this keeps the Reward attractor active, and this through the inhibitory neurons prevents the Non-Reward attractor neurons from being activated. If an Expected Reward has been signalled, and the Reward attractor is active, the Reward Attractor firing can be directly inhibited by a Non-Reward outcome, and the Non-Reward neurons, which signal negative reward prediction error, are activated.

3.2 Attractor Framework

Our aim is to investigate these mechanisms in a biophysically realistic attractor framework, so that the properties of receptors, synaptic currents and the statistical effects related to the probabilistic spiking of the neurons can be part of the model. We use a minimal architecture, a single attractor or autoassociation network (Hopfield 1982, Amit 1989, Hertz, Krogh and Palmer 1991, Rolls and Treves 1998, Rolls and Deco 2002, Rolls 2008). We chose a recurrent (attractor) integrate-and-fire network model which includes synaptic channels for AMPA, NMDA and GABA_A receptors (Brunel and Wang 2001).

Both excitatory and inhibitory neurons are represented by a leaky integrate-and-fire model (Tuckwell 1988). The basic state variable of a single model neuron is the membrane potential. It decays in time when the neurons receive no synaptic input down to a resting potential. When synaptic input causes the membrane potential to reach a threshold, a spike is emitted and the neuron is set to the reset potential at which it is kept for the refractory period. The emitted action potential is propagated to the other neurons in the network. The excitatory neurons transmit their action potentials via the glutamatergic receptors AMPA and NMDA which are both modeled by their effect in producing exponentially decaying currents in the postsynaptic neuron. The rise time of the AMPA current is neglected, because it is typically very short. The NMDA channel is modeled with an alpha function including both a rise and a decay term. In addition, the synaptic function of the NMDA current includes a voltage dependence controlled by the extracellular magnesium concentration (Jahr and Stevens 1990). The inhibitory postsynaptic potential is mediated by a GABA_A receptor model and is described by a decay term.

The single attractor network contains 800 excitatory and 200 inhibitory neurons, which is consistent with the observed proportions of pyramidal cells and interneurons in the cerebral cortex (Abeles 1991, Braitenberg and Schütz 1991). The connection strengths are adjusted using mean-field analysis (Brunel and Wang 2001, Rolls and Deco 2010), so that the excitatory and inhibitory neurons exhibit a spontaneous activity of 3 Hz and 9 Hz, respectively (Wilson, O’Scalaidhe and Goldman-Rakic 1994, Koch and Fuster 1989). The recurrent excitation mediated by the AMPA and NMDA receptors is dominated by the NMDA current to avoid instabilities during the delay periods (Wang 2002).

Our cortical network model features a minimal architecture to investigate stability of recalled memories in a short-term memory period, and consists of two selective pools, a Reward population (S1) and a Non-Reward population (S2) (Fig. 2). We use just two selective pools to eliminate possible disturbing factors. The non-selective pool NS models the spiking of cortical neurons and serves to generate an approximately Poisson spiking dynamics in the

model (Brunel and Wang 2001), which is what is observed in the cortex. The inhibitory pool IH contains the 200 inhibitory neurons. The connection weights between the neurons within each selective pool or population are called the intra-pool connection strengths w_+ . The increased strength of the intra-pool connections is counterbalanced by the other excitatory connections (w_-) to keep the average input constant.

The network receives Poisson input spikes via AMPA receptors which are envisioned to originate from 800 external neurons at an average spontaneous firing rate of 3 Hz from each external neuron, consistent with the spontaneous activity observed in the cerebral cortex (Wilson et al. 1994, Rolls and Treves 1998, Rolls 2008). Given that there are 800 synapses on each neuron in the network for these external inputs, the number of spikes being received by every neurons in the network is 2,400 spikes/s. This external input can be altered for specific neuronal populations to intriduce the effects of external stimuli on the network. In addition to tese external excitatory inputs, each excitatory neuron receives 80 excitatory inputs from other neurons in the same population in which the firing rate is modulated by w_+ , and 720 excitatory inputs from other excitatory neurons (given that there are 800 excitatory neurons) in which the firing rate is modulated by w_- (see Fig. 2). A detailed mathematical description is provided in the Supplementary Material.

3.3 Short term synaptic depression

A synaptic depression mechanism was used following Dayan and Abbott (2001), page 185. In particular, the probability of transmitter release P_{rel} was decreased after each presynaptic spike by a factor $P_{\text{rel}} = P_{\text{rel}} \cdot f_D$ with $f_D = 0.988$. Between presynaptic action potentials the release probability P_{rel} is updated by

$$\tau_P \frac{dP_{\text{rel}}}{dt} = P_0 - P_{\text{rel}} \quad (1)$$

with $P_0 = 1$ and $\tau_P = 1000$ ms.

3.4 Analysis

Simulations were performed with the following protocols. Reference to Figs. 3–5, especially the event part at the bottom of each Figure, may be helpful.

3.4.1 Non-Reward: Extinction

First there was a period of spontaneous firing from 0–500 ms in which the input to the Reward and Non-Reward pools was 2.90 spikes/s per synapse. (There were 800 external synapses onto each neuron, making the mean Poisson firing rate input to each neuron in these pools correspond to 2,320 spikes/s.) This 500 ms period corresponded to the inter-trial period.

At 500 ms the external input to the Reward pool was increased to 3.1 spikes/s per synapse, to reflect an Expected Reward input. (The Expected Reward input might correspond to the sight of food or of a visual stimulus currently associated with the sight of food.) This was maintained at this level until the end of the trial at 5000 ms. Because no Reward Outcome (such as a taste of food in the mouth) was received for the rest of the trial, this was an Extinction trial, in which no reward outcome was delivered. Systematic parameter exploration including mean field analyses showed that a suitable value for w_+ was 2.1 for the

Reward population, for this enabled a high firing rate attractor state to be maintained by that population in the absence of synaptic depression even when the activating stimulus was removed and the external input to the Reward attractor was reset to the default value of 3.0 spikes/s per synapse (Deco and Rolls 2006, Rolls and Deco 2010). That choice defined $w-$ as 0.88, as described above. Systematic parameter exploration showed that a value for $w+$ of 2.22 for the Non-Reward population was in the region where this population of neurons would emerge from its spontaneous firing rate under the influence of the spiking-related noise in the system into a high firing rate attractor state if the Reward neurons were not firing. The two values of $w+$ became fixed parameters that were not altered in any of the simulations. For these extinction simulations, at 500 ms the input to the Non-reward population of neurons was increased to 3.05 spikes/s per synapse, and was maintained constant for the rest of the trial, that is, until 5500 ms.

3.4.2 Expected reward followed by a reward

In this scenario (Fig. 4), after a period of spontaneous activity from 0 until 0.5 s, the Expected Reward input was applied to the Reward attractor population of neurons at 3.1 Hz per synapse. At 2500 ms a Reward Outcome input (corresponding for example to the delivery of a taste reward) was received (3.7 Hz per external synapse, of which there were 800), and this was maintained until the end of the trial. As in the other simulations, the input to the Non-Reward population of neurons was maintained constant at 3.05 Hz per synapse from time = 0.5 s until the end of the trial.

3.4.3 Expected reward followed by punishment

In this scenario (Fig. 5), after a period of spontaneous activity from 0 until 0.5 s, the Expected Reward input was applied to the Reward attractor population of neurons (at 3.1 Hz per synapse). At 2500 ms a Punishment was received (corresponding for example to the delivery of an aversive saline taste instead of the reward of a good taste), and this decreased the Expected Reward input to the Reward population to a low value (2.8 Hz per synapse), and this was maintained until the end of the trial. The input to the Non-Reward population of neurons was maintained constant at 3.05 Hz per synapse from time = 0.5 s until the end of the trial.

4 Results

4.1 Expected Reward followed by Non-Reward: Extinction

Fig. 3 shows the firing rates of the populations of neurons when an Expected Reward input is not followed by a Reward Outcome input, that is, in extinction. $w+$ was 2.1 for the Reward population, and 2.22 for the Non-Reward population to make it more excitable, and $w-$ was 0.88. The Expected Reward input started at 0.5 s and was maintained at a value of 3.1 Hz per synapse until the end of the trial. The external input to the Non-Reward population also increased at the end of the spontaneous period to reflect the fact that a trial was in progress, but its value was 3.05 Hz per synapse as this was an Expected Reward trial, so that the Expected Reward had to be greater than any expected Non-Reward. Fig. 3 shows that the Reward population increased its firing rate peaking at approximately 1300 ms (800

ms after the Expected Reward input started), and then decreased to a low firing rate of approximately 15 spikes/s by 2500 ms because of the synaptic depression that is implemented in the recurrent collateral inputs (and not in the external inputs to the network). At about this time, 2500 ms, the Non-Reward population started to increase its firing rate, because it no longer was receiving inhibition through the inhibitory neurons from the Reward population. The Non-Reward population started firing when the inhibition on it was released because it had a slightly higher than the typical value for $w+$ (i.e. it is 2.22), and because of the spiking-timing-related noise in the system. The Non-reward population maintained its activity for approximately 2 s before synaptic depression in its recurrent collateral connections reduced its firing rate back to baseline.

This simulation thus shows how a Non-Reward population of neurons can be triggered into firing when an Expected Reward is not followed by a Reward Outcome.

4.2 Expected Reward followed by a Reward Outcome

The operation of the network when an Expected Reward is followed by a Reward Outcome is illustrated in Fig. 4. After a period of spontaneous activity from 0 until 0.5 s, the Expected Reward input was applied to the Reward attractor population of neurons (3.1 Hz per synapse). At 2500 ms a Reward Outcome input was received (3.7 Hz per external synapse), and this was maintained until the end of the trial. The input to the Non-Reward population of neurons was maintained constant at 3.05 Hz per synapse from time = 0.5 s until the end of the trial. Fig. 4 shows that the effect of the Reward Outcome was to increase the firing of the Reward population (even though synaptic depression was present in the recurrent collateral synapses). The high firing rate of the Reward population of neurons produced sufficient inhibition via the inhibitory neurons on the Non-Reward population of neurons that they fired only at a low rate of approximately 3 spikes/s.

This simulation thus shows how the Non-Reward population of neurons is not triggered into a high firing when an Expected Reward is followed by a Reward Outcome.

4.3 Expected Reward followed by Punishment Outcome activates the Non-Reward population

The operation of the network when an Expected Reward is followed by a Punishment Outcome is illustrated in Fig. 5. After a period of spontaneous activity from 0 until 0.5 s, the Expected Reward input was applied to the Reward attractor population of neurons (at 3.1 Hz per synapse). At 2500 ms a Punishment was received, and this decreased the Expected Reward input to the Reward population to a low value (2.8 Hz per synapse), and this was maintained until the end of the trial. The input to the Non-Reward population of neurons was maintained constant at 3.05 Hz per synapse from time = 0.5 s until the end of the trial. Fig. 5 shows that when the Expected Reward input to the Reward population is decreased at 2500 ms by the receipt of Punishment, the release of inhibition through the inhibitory neurons caused the Non-Reward population to emerge into a high firing rate state reflecting the non-reward contingency.

This simulation thus shows how the Non-Reward population of neurons is triggered into a high firing when an Expected Reward is followed by a Punishment Outcome.

5 Discussion

The simulations show how Non-Reward neurons could be activated when an Expected Reward is followed by non-reward in the form of no Reward Outcome, or a Punishment Outcome. The mechanism implemented in the simulations is the release of inhibition on a Non-Reward population of neurons when synaptic depression in the recurrent collateral connections decreases the firing of the Reward population of neurons, and the firing of the Non-Reward population is not maintained by the receipt of a Reward Outcome input to maintain the Reward neuronal firing. The simulations are valuable in showing the parametric conditions under which this process can be implemented. In the simulations, the speed of the synaptic depression, and the time before the non-reward population become active, can be influenced by for example the time constant of the synaptic depression. It is useful to note that the same synaptic depression operated in all the excitatory connections of the neurons in the model, and is not a special property of the reward neurons.

Other mechanisms than the synaptic depression used to illustrate the concept of how non-reward neuronal firing is produced can be envisaged. For example, if the trials typically involved a delay of several seconds before the reward outcome was received, then a separate short-term memory set to estimate the typical delay before reward outcome might be used to decrease the expected reward input to the network after the appropriate delay, allowing the non-reward neurons to then fire. What is a key concept in the mechanism proposed is however the balance between the Reward and the Non-Reward attractor populations, which are effectively in competition with each other via the common population of inhibitory neurons. This helps to account for the neurophysiological evidence that if a food reward is moved towards a macaque, then the expected reward neurons in the primate orbitofrontal cortex fire fast, and as soon as the reward stops moving (within sight) towards the macaque, or is slowly removed (by being moved backwards), the non-reward neurons start to fire, and the reward neurons stop firing (Rolls, Thorpe and Maddison 1983). The competing attractor hypothesis for reward and non-reward representations in the orbitofrontal cortex described here provides an account for this important property of the populations of neurons in the primate orbitofrontal cortex. Consistent with this competing attractor hypothesis, there is evidence for reciprocal activity in the human medial orbitofrontal cortex reward area and the lateral orbitofrontal cortex non-reward / loss area (O’Doherty, Kringelbach, Rolls, Hornak and Andrews 2001). In particular, activations in the human medial orbitofrontal increase in proportion to the amount of (monetary) reward received, and decrease in proportion to the amount of (monetary) loss; and in the lateral orbitofrontal cortex, the activations are reciprocally related to these changes (O’Doherty et al. 2001).

The hypothesis that there are timing mechanisms in the prefrontal cortex that do time how long to wait for an expected reward is supported by our finding that patients with orbitofrontal cortex damage and patients with borderline personality disorder not only are impulsive with reduced sensitivity to non-reward, but have a faster perception of time (they underproduce a time interval) (Berlin, Rolls and Iversen 2005, Rolls 2014). We have hypothesized that this faster perception of time may indeed be part of the mechanism by which some patients behave with what is described as impulsivity, because what they expected to happen has not happened yet.

Non-Reward neurons of the type described here may be very important in changing behaviour (Rolls 2014). For example, in a visual discrimination task in which one stimulus is rewarded (for example a triangle, which if selected leads to the delivery of a sweet taste), and

the other is punished (for example a square, which if selected leads to the delivery of aversive salt taste), then the selection switches if the rule reverses, and selection of the triangle leads to a salt taste. This reversal can be very fast, in one trial, and can be non-associative in that after the selection of the triangle led to the delivery of saline, the next choice of the square was to select it, even though its previous presentation had been associated with salt taste (Thorpe, Rolls and Maddison 1983, Rolls 2014). A model for this rule-based reversal has been described (Deco and Rolls 2005c), and the Non-Reward signal needed for that model could be produced by the neuronal network mechanism described here.

The present model, although based on our previous research on attractor based biologically plausible dynamical modelling of computations performed by the cortex (Treves 1993, Battaglia and Treves 1998, Panzeri, Rolls, Battaglia and Lavis 2001, Deco, Ledberg, Almeida and Fuster 2005, Deco and Rolls 2005c, Deco and Rolls 2005b, Deco and Rolls 2005a, Deco and Rolls 2006, Rolls, Grabenhorst and Deco 2010, Deco, Rolls and Romo 2010, Rolls and Deco 2010, Rolls and Deco 2011, Rolls, Webb and Deco 2012, Rolls, Dempere-Marco and Deco 2013, Rolls and Deco 2015a, Rolls and Deco 2015b, Rolls 2016a), is different from earlier models by what is being modelled (the generation of a non-reward signal), by the brain region modelled (the orbitofrontal cortex, and in particular the lateral orbitofrontal cortex), and by the architecture in which Reward Attractor network is activated by an Expected Reward and inhibits a Non-Reward attractor network, which latter only becomes active if the habituating Reward attractor is not maintained in its activity by the Reward Outcome.

The implementation of the model described here used non-overlapping populations of neurons for the Reward and Non-Reward attractors in the single network, for clarity and simplicity. However separate attractors with partially overlapping neurons in the different populations using a sparse distributed representation more like that found in the cerebral cortex (Rolls and Treves 2011, Rolls 2016a) do operate with similar properties (Battaglia and Treves 1998, Rolls, Treves, Foster and Perez-Vicente 1997, Rolls 2012).

Although the system described here was modelled with networks that have realistic dynamics as they move into and out of stable attractor states (Treves 1993, Battaglia and Treves 1998, Panzeri et al. 2001, Deco et al. 2005, Deco and Rolls 2006, Rolls and Deco 2010, Rolls 2016a) another approach to such modelling is a network that when the system does not have sufficient time to fall into a stable attractor state, nevertheless may visit a succession of states in a stable trajectory (Rabinovich, Huerta and Laurent 2008, Rabinovich, Varona, Tristan and Afraimovich 2014). A possible way to understand the operation of such a system is that of a stable heteroclinic channel. A stable heteroclinic channel is defined by a sequence of successive metastable ('saddle') states. Under the proper conditions, all the trajectories in the neighborhood of these saddle points remain in the channel, ensuring robustness and reproducibility of the dynamical trajectory through the state space (Rabinovich et al. 2008, Rabinovich et al. 2014). This approach might be considered in future as a somewhat different implementation of the system described here.

A model of a different process, action-outcome learning, for a different part of the brain, the anterior cingulate cortex, has in common with the present model that an error is signalled if a prediction is not inhibited by the expected outcome (Alexander and Brown 2011). However, the implementation as well as the purpose of that model is quite different, for no biologically plausible networks with attractor dynamics are modelled as here, but instead that is a mathematical model based on reinforcement learning algorithms.

In conclusion, an approach has been described to how the orbitofrontal cortex may compute non-reward from Expected Value and Reward Outcome. A key attribute of the new

approach described here is that there are competing attractors in the orbitofrontal cortex for Reward and Non-Reward, and that these have reciprocal activity because of the competition between them. This can account for the neurophysiological findings that the firing rates of reward and non-reward neurons in the primate orbitofrontal cortex are dynamically reciprocally related to each other (Thorpe, Rolls and Maddison 1983, Rolls 2014). This is seen not only in the activity of different single neurons, some of which respond as a reward approaches, and others of which respond when the reward stops approaching (Thorpe, Rolls and Maddison 1983), but also at the fMRI level, where activations in the medial orbitofrontal cortex increase linearly with the amount of money gained, and activations in the lateral orbitofrontal cortex increase linearly with the amount of money lost (O’Doherty, Kringelbach, Rolls, Hornak and Andrews 2001). Separate populations of reward and non-reward neurons may be needed because in the cerebral cortex the high firing rates of neurons are usually the drivers of outputs (Rolls 2016a, Rolls and Treves 2011), and different output systems must be activated to lead to the different behavioral effects of receiving an expected reward, and of not receiving an expected reward. We know of no other hypothesis and model that accounts for the neurophysiological findings described here. The process that we analyze here is fundamentally important in changing behavior when an expected reward is not received. The process is important in many emotions, which can be produced if an expected reward is not received or lost, with examples including sadness and anger (Rolls 2014). Consistent with this the psychiatric disorder of depression may arise when the non-reward system leading to sadness is too sensitive, or maintains its activity for too long (Rolls 2016b), or has increased functional connectivity (Cheng, Rolls, Qiu, Liu, Tang, Huang, Wang, Zhang, Lin, Zheng, Pu, Tsai, Yang, Lin, Wang, Xie and Feng 2016). Conversely, if the non-reward system is underactive or is damaged by lesions of the orbitofrontal cortex, the decreased sensitivity to non-reward may contribute to increased impulsivity (Berlin, Rolls and Kischka 2004, Berlin, Rolls and Iversen 2005), and even antisocial and psychopathic behavior (Rolls 2014). For these reasons, understanding the mechanisms that underlie non-reward is important not only for understanding normal human behavior and how it changes when rewards are not received, but also for understanding and potentially treating better some emotional and psychiatric disorders. Indeed, a prediction of the current model is that if there are non-reward attractors in the orbitofrontal cortex, and these are over-sensitive in depression, then treatments such as ketamine which by blocking NMDA receptors knock the network out of its attractor state, may in this way reduce depression, and there is already evidence consistent with this (Rolls 2016b, Carlson, Diazgranados, Nugent, Ibrahim, Luckenbaugh, Brutsche, Herscovitch, Manji, Zarate and Drevets 2013, Lally, Nugent, Luckenbaugh, Niciu, Roiser and Zarate 2015).

Acknowledgements. This research was supported by the Oxford Centre for Computational Neuroscience, Oxford, UK. GD was supported by the ERC Advanced Grant: DYSTRUCTURE (n. 295129), and by the Spanish Research Project SAF2010-16085.

References

- Abeles, A. (1991). *Corticonics*, Cambridge University Press, New York.
- Alexander, W. H. and Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor, *Nature Neuroscience* **14**: 1338–44.
- Amit, D. J. (1989). *Modeling Brain Function. The World of Attractor Neural Networks.*, Cambridge University Press, Cambridge.
- Battaglia, F. and Treves, A. (1998). Stable and rapid recurrent processing in realistic autoassociative memories, *Neural Computation* **10**: 431–450.
- Baylis, L. L. and Gaffan, D. (1991). Amygdalectomy and ventromedial prefrontal ablation produce similar deficits in food choice and in simple object discrimination learning for an unseen reward, *Experimental Brain Research* **86**: 617–622.
- Berlin, H., Rolls, E. T. and Kischka, U. (2004). Impulsivity, time perception, emotion, and reinforcement sensitivity in patients with orbitofrontal cortex lesions, *Brain* **127**: 1108–1126.
- Berlin, H., Rolls, E. T. and Iversen, S. D. (2005). Borderline Personality Disorder, impulsivity, and the orbitofrontal cortex, *American Journal of Psychiatry* **58**: 234–245.
- Braitenberg, V. and Schütz, A. (1991). *Anatomy of the Cortex*, Springer-Verlag, Berlin.
- Brunel, N. and Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition, *Journal of Computational Neuroscience* **11**: 63–85.
- Butter, C. M. (1969). Perseveration in extinction and in discrimination reversal tasks following selective prefrontal ablations in *Macaca mulatta*, *Physiology and Behavior* **4**: 163–171.
- Butter, C. M., McDonald, J. A. and Snyder, D. R. (1969). Orality, preference behavior, and reinforcement value of non-food objects in monkeys with orbital frontal lesions, *Science* **164**: 1306–1307.
- Camille, N., Tsuchida, A. and Fellows, L. K. (2011). Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage, *Journal of Neuroscience* **31**: 15048–15052.
- Carlson, P. J., Diazgranados, N., Nugent, A. C., Ibrahim, L., Luckenbaugh, D. A., Brutsche, N., Herscovitch, P., Manji, H. K., Zarate, C. A., J. and Drevets, W. C. (2013). Neural correlates of rapid antidepressant response to ketamine in treatment-resistant unipolar depression: a preliminary positron emission tomography study, *Biological Psychiatry* **73**: 1213–21.
- Chau, B. K., Sallet, J., Papageorgiou, G. K., Noonan, M. P., Bell, A. H., Walton, M. E. and Rushworth, M. F. (2015). Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques, *Neuron* **87**: 1106–1118.

- 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. and Feng, J. (2016). Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression.
- Critchley, H. D. and Rolls, E. T. (1996a). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex, *Journal of Neurophysiology* **75**: 1673–1686.
- Critchley, H. D. and Rolls, E. T. (1996b). Olfactory neuronal responses in the primate orbitofrontal cortex: analysis in an olfactory discrimination task, *Journal of Neurophysiology* **75**: 1659–1672.
- Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience*, MIT Press, Cambridge, MA.
- Deco, G. and Rolls, E. T. (2005a). Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons, *Journal of Neurophysiology* **94**: 295–313.
- Deco, G. and Rolls, E. T. (2005b). Sequential memory: a putative neural and synaptic dynamical mechanism, *Journal of Cognitive Neuroscience* **17**: 294–307.
- Deco, G. and Rolls, E. T. (2005c). Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex, *Cerebral Cortex* **15**: 15–30.
- Deco, G. and Rolls, E. T. (2006). A neurophysiological model of decision-making and Weber’s law, *European Journal of Neuroscience* **24**: 901–916.
- Deco, G., Ledberg, A., Almeida, R. and Fuster, J. (2005). Neural dynamics of cross-modal and cross-temporal associations, *Experimental Brain Research* **166**: 325–336.
- Deco, G., Rolls, E. T. and Romo, R. (2010). Synaptic dynamics and decision-making, *Proceedings of the National Academy of Sciences* **107**: 7545–7549.
- Fellows, L. K. (2007). The role of orbitofrontal cortex in decision making: a component process account, *Annals of the New York Academy of Sciences* **1121**: 421–430.
- Fellows, L. K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage, *Annals of the New York Academy of Sciences* **1239**: 51–58.
- Fellows, L. K. and Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm, *Brain* **126**: 1830–1837.
- Grabenhorst, F. and Rolls, E. T. (2011). Value, pleasure, and choice systems in the ventral prefrontal cortex, *Trends in Cognitive Sciences* **15**: 56–67.
- Grabenhorst, F., Rolls, E. T. and Bilderbeck, A. (2008). How cognition modulates affective responses to taste and flavor: top-down influences on the orbitofrontal and pregenual cingulate cortices, *Cerebral Cortex* **18**: 1549–1559.
- Hertz, J., Krogh, A. and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*, Addison Wesley, Wokingham, U.K.

- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities, *Proc. Nat. Acad. Sci. USA* **79**: 2554–2558.
- Hornak, J., O’Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R. and Polkey, C. E. (2004). Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans, *Journal of Cognitive Neuroscience* **16**: 463–478.
- Iversen, S. D. and Mishkin, M. (1970). Perseverative interference in monkey following selective lesions of the inferior prefrontal convexity, *Experimental Brain Research* **11**: 376–386.
- Jahr, C. and Stevens, C. (1990). Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics, *Journal of Neuroscience* **10**: 3178–3182.
- Jones, B. and Mishkin, M. (1972). Limbic lesions and the problem of stimulus–reinforcement associations, *Experimental Neurology* **36**: 362–377.
- Koch, K. W. and Fuster, J. M. (1989). Unit activity in monkey parietal cortex related to haptic perception and temporary memory, *Experimental Brain Research* **76**: 292–306.
- Kringelbach, M. L. and Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction, *Neuroimage* **20**: 1371–1383.
- Kringelbach, M. L. and Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology, *Progress in Neurobiology* **72**: 341–372.
- Kringelbach, M. L., O’Doherty, J., Rolls, E. T. and Andrews, C. (2003). Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness, *Cerebral Cortex* **13**: 1064–1071.
- Lally, N., Nugent, A. C., Luckenbaugh, D. A., Niciu, M. J., Roiser, J. P. and Zarate, C. A., J. (2015). Neural correlates of change in major depressive disorder anhedonia following open-label ketamine, *Journal of Psychopharmacology* **29**: 596–607.
- Meunier, M., Bachevalier, J. and Mishkin, M. (1997). Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys, *Neuropsychologia* **35**: 999–1015.
- Murray, E. A. and Izquierdo, A. (2007). Orbitofrontal cortex and amygdala contributions to affect and action in primates, *Annals of the New York Academy of Sciences* **1121**: 273–296.
- O’Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J. and Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex, *Nature Neuroscience* **4**: 95–102.
- Panzeri, S., Rolls, E. T., Battaglia, F. and Lavis, R. (2001). Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons, *Network: Computation in Neural Systems* **12**: 423–440.
- Passingham, R. E. P. and Wise, S. P. (2012). *The Neurobiology of the Prefrontal Cortex*, Oxford University Press, Oxford.

- Rabinovich, M., Huerta, R. and Laurent, G. (2008). Transient dynamics for neural processing, *Science* **321**: 48–50.
- Rabinovich, M. I., Varona, P., Tristan, I. and Afraimovich, V. S. (2014). Chunking dynamics: heteroclinics in mind, *Frontiers in Computational Neuroscience* **8**: 22.
- Rolls, E. T. (1981). Processing beyond the inferior temporal visual cortex related to feeding, learning, and striatal function, in Y. Katsuki, R. Norgren and M. Sato (eds), *Brain Mechanisms of Sensation*, Wiley, New York, chapter 16, pp. 241–269.
- Rolls, E. T. (2008). *Memory, Attention, and Decision-Making. A Unifying Computational Neuroscience Approach*, Oxford University Press, Oxford.
- Rolls, E. T. (2012). Advantages of dilution in the connectivity of attractor networks in the brain, *Biologically Inspired Cognitive Architectures* **1**: 44–54.
- Rolls, E. T. (2014). *Emotion and Decision-Making Explained*, Oxford University Press, Oxford.
- Rolls, E. T. (2015a). Functions of the anterior insula in taste, autonomic, and related functions, *Brain and Cognition* p. doi: 10.1016/j.bandc.2015.07.002.
- Rolls, E. T. (2015b). Taste, olfactory, and food reward value processing in the brain, *Progress in Neurobiology* **127–128**: 64–90.
- Rolls, E. T. (2016a). *Cerebral Cortex: Principles of Operation*, Oxford University Press, Oxford.
- Rolls, E. T. (2016b). A non-reward attractor theory of depression, *Neuroscience and Biobehavioral Reviews*.
- Rolls, E. T. (2016c). Reward systems in the brain and nutrition, *Annual Review of Nutrition* **36**: 14.
- Rolls, E. T. and Baylis, L. L. (1994). Gustatory, olfactory and visual convergence within the primate orbitofrontal cortex, *Journal of Neuroscience* **14**: 5437–5452.
- Rolls, E. T. and Deco, G. (2002). *Computational Neuroscience of Vision*, Oxford University Press, Oxford.
- Rolls, E. T. and Deco, G. (2010). *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*, Oxford University Press, Oxford.
- Rolls, E. T. and Deco, G. (2011). Prediction of decisions from noise in the brain before the evidence is provided, *Frontiers in Neuroscience* **5**: 33.
- Rolls, E. T. and Deco, G. (2015a). Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain, *Brain Research* **1621**: 316–334.
- Rolls, E. T. and Deco, G. (2015b). A stochastic neurodynamics approach to the changes in cognition and memory in aging, *Neurobiology of Learning and Memory* **118**: 150–161.

- Rolls, E. T. and Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making, *Progress in Neurobiology* **86**: 216–244.
- Rolls, E. T. and Treves, A. (1998). *Neural Networks and Brain Function*, Oxford University Press, Oxford.
- Rolls, E. T. and Treves, A. (2011). The neuronal encoding of information in the brain, *Progress in Neurobiology* **95**: 448–490.
- Rolls, E. T. and Williams, G. V. (1987). Neuronal activity in the ventral striatum of the primate, in M. B. Carpenter and A. Jayaraman (eds), *The Basal Ganglia II – Structure and Function – Current Concepts*, Plenum, New York, pp. 349–356.
- Rolls, E. T., Thorpe, S. J. and Maddison, S. P. (1983). Responses of striatal neurons in the behaving monkey. 1. Head of the caudate nucleus, *Behavioural Brain Research* **7**: 179–210.
- Rolls, E. T., Sienkiewicz, Z. J. and Yaxley, S. (1989). Hunger modulates the responses to gustatory stimuli of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey, *European Journal of Neuroscience* **1**: 53–60.
- Rolls, E. T., Yaxley, S. and Sienkiewicz, Z. J. (1990). Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey, *Journal of Neurophysiology* **64**: 1055–1066.
- Rolls, E. T., Hornak, J., Wade, D. and McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage, *Journal of Neurology, Neurosurgery and Psychiatry* **57**: 1518–1524.
- Rolls, E. T., Critchley, H. D., Mason, R. and Wakeman, E. A. (1996). Orbitofrontal cortex neurons: role in olfactory and visual association learning, *Journal of Neurophysiology* **75**: 1970–1981.
- Rolls, E. T., Treves, A., Foster, D. and Perez-Vicente, C. (1997). Simulation studies of the CA3 hippocampal subfield modelled as an attractor neural network, *Neural Networks* **10**: 1559–1569.
- Rolls, E. T., Critchley, H. D., Browning, A. S., Hernadi, A. and Lenard, L. (1999). Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex, *Journal of Neuroscience* **19**: 1532–1540.
- Rolls, E. T., O’Doherty, J., Kringelbach, M. L., Francis, S., Bowtell, R. and McGlone, F. (2003a). Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices, *Cerebral Cortex* **13**: 308–317.
- Rolls, E. T., Verhagen, J. V. and Kadohisa, M. (2003b). Representations of the texture of food in the primate orbitofrontal cortex: neurons responding to viscosity, grittiness, and capsaicin, *Journal of Neurophysiology* **90**: 3711–3724.
- Rolls, E. T., McCabe, C. and Redoute, J. (2008). Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task, *Cerebral Cortex* **18**: 652–663.

- Rolls, E. T., Grabenhorst, F. and Deco, G. (2010). Decision-making, errors, and confidence in the brain, *Journal of Neurophysiology* **104**: 2359–2374.
- Rolls, E. T., Webb, T. J. and Deco, G. (2012). Communication before coherence, *European Journal of Neuroscience* **36**: 2689–2709.
- Rolls, E. T., Dempere-Marco, L. and Deco, G. (2013). Holding multiple items in short term memory: a neural mechanism, *PLoS One* **8**: e61078.
- Rosenkilde, C. E., Bauer, R. H. and Fuster, J. M. (1981). Single unit activity in ventral prefrontal cortex in behaving monkeys, *Brain Research* **209**: 375–394.
- Rudebeck, P. H. and Murray, E. A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior, *Journal of Neuroscience* **31**: 10569–10578.
- Rushworth, M. F., Kolling, N., Sallet, J. and Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems?, *Current Opinion in Neurobiology* **22**: 946–955.
- Schultz, W. (2013). Updating dopamine reward signals, *Current Opinion in Neurobiology* **23**: 229–238.
- Thorpe, S. J., Maddison, S. and Rolls, E. T. (1979). Single unit activity in the orbitofrontal cortex of the behaving monkey, *Neuroscience Letters* **S3**: S77.
- Thorpe, S. J., Rolls, E. T. and Maddison, S. (1983). Neuronal activity in the orbitofrontal cortex of the behaving monkey, *Experimental Brain Research* **49**: 93–115.
- Treves, A. (1993). Mean-field analysis of neuronal spike dynamics, *Network* **4**: 259–284.
- Tuckwell, H. (1988). *Introduction to Theoretical Neurobiology*, Cambridge University Press, Cambridge.
- Verhagen, J. V., Rolls, E. T. and Kadohisa, M. (2003). Neurons in the primate orbitofrontal cortex respond to fat texture independently of viscosity, *Journal of Neurophysiology* **90**: 1514–1525.
- Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits, *Neuron* **36**: 955–968.
- Wheeler, E. Z. and Fellows, L. K. (2008). The human ventromedial frontal lobe is critical for learning from negative feedback, *Brain* **131**: 1323–1331.
- Williams, G. V., Rolls, E. T., Leonard, C. M. and Stern, C. (1993). Neuronal responses in the ventral striatum of the behaving macaque, *Behavioural Brain Research* **55**: 243–252.
- Wilson, F. A. W., O’Scalaidhe, S. P. and Goldman-Rakic, P. (1994). Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex, *Proceedings of the National Academy of Science* **91**: 4009–4013.
- Wise, S. P. (2008). Forward frontal fields: phylogeny and fundamental function, *Trends in Neuroscience* **31**: 599–608.

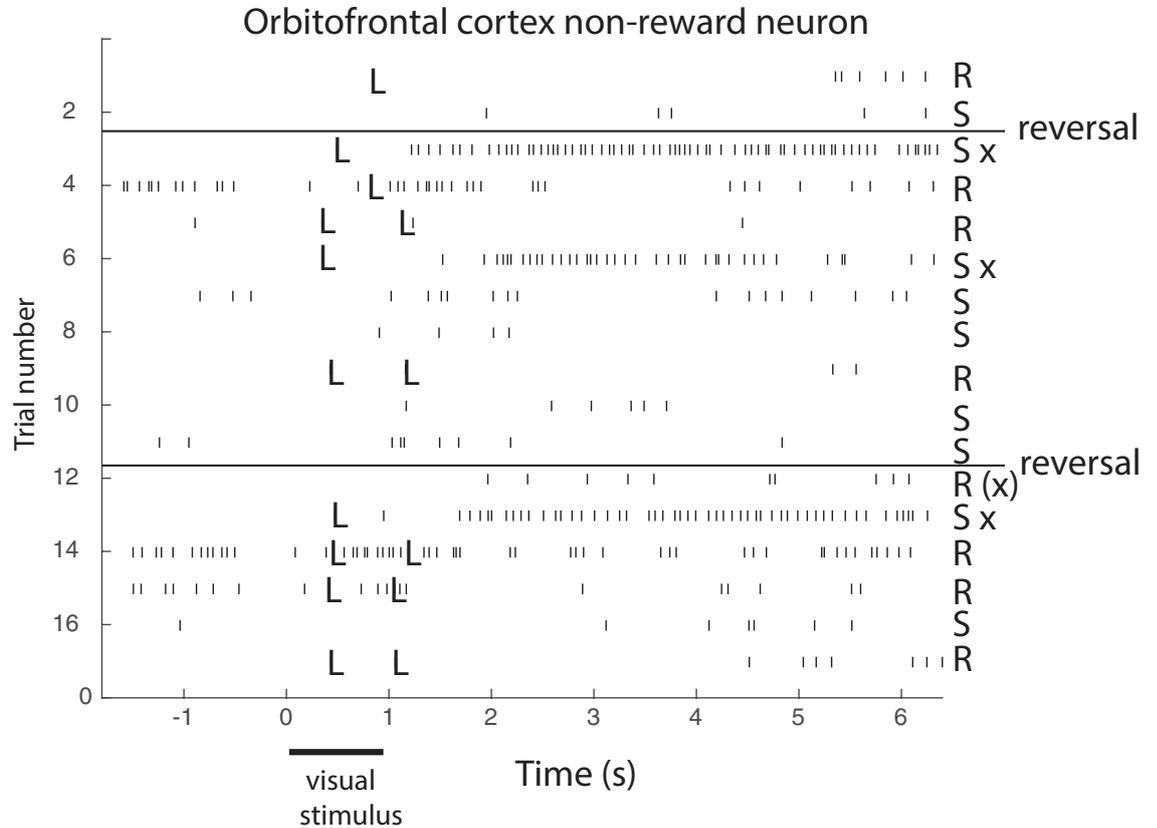


Figure 1: Error neuron: Responses of an orbitofrontal cortex neuron that responded only when the monkey licked to a visual stimulus during reversal, expecting to obtain fruit juice reward, but actually obtained the taste of aversive saline because it was the first trial of reversal (trials 3, 6, and 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 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, Rolls and Maddison 1983)

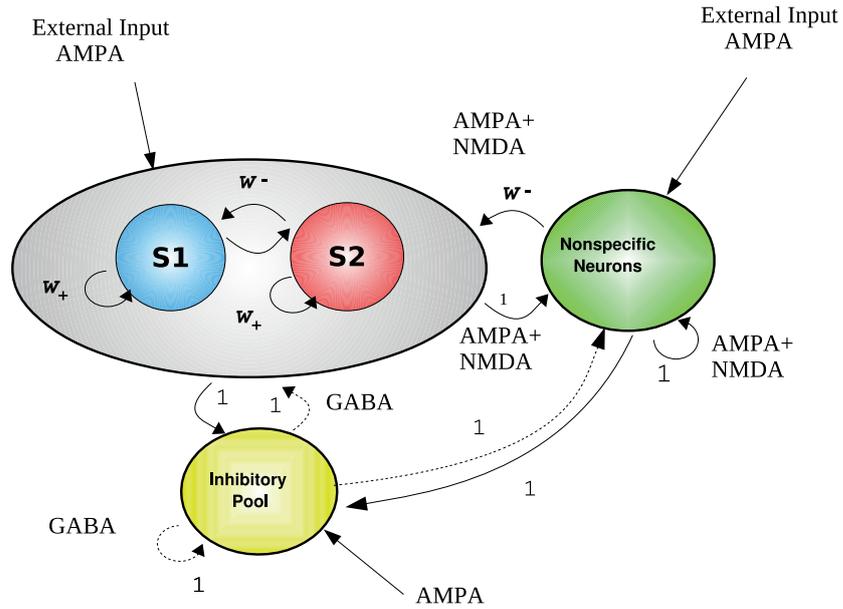


Figure 2: The attractor network model. The excitatory neurons are divided into two selective pools S1 (termed the Reward Attractor population of neurons) and S2 (termed the Non-Reward Attractor population of neurons) (with 80 neurons each) with strong intra-pool connection strengths w_+ and one non-selective pool (NS) (with 640 neurons). The value of w_+ for S2 is a little higher than that for S1, so that S2 tends to emerge into a high firing rate if activity is not maintained in S1 after an Expected Reward input has been received (see text and Supplementary Material). The other connection strengths are 1 or weak w_- . The network contains 1000 neurons, of which 800 are in the excitatory pools and 200 are in the inhibitory pool IH. Every neuron in the network also receives inputs from 800 external neurons, and these neurons increase their firing rates to apply a stimulus to one of the pools S1 or S2. The Supplementary Material contains the synaptic connection matrices.

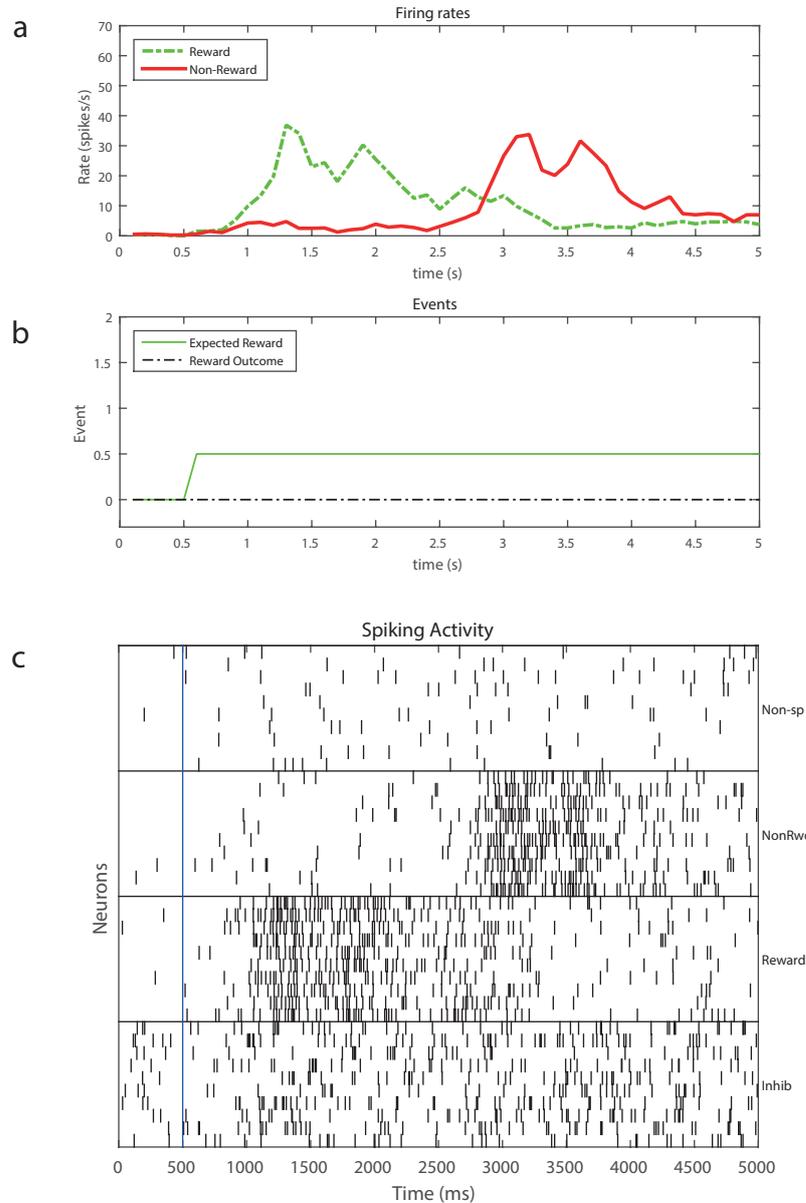


Figure 3: The operation of the network when an Expected Reward is not obtained (extinction). a. The firing rates of the Reward population of neurons and the Non-Reward population of neurons during a 5 s trial. b. After a period of spontaneous activity from 0 until 0.5 s, the Expected Reward input was applied to the reward attractor population of neurons (3.1 Hz per synapse), and maintained at that level for the remainder of the trial. No Reward Outcome input was received. The input to the Non-Reward population of neurons was maintained constant at 3.05 Hz per synapse from time = 0.5 s until the end of the trial. Note: there were 800 external synapses onto each neuron through which the external inputs were applied with the firing rates specified for each synapse.) c. Rastergrams showing for each of the four populations of neurons, Non-Specific, Non-Reward, Reward, and Inhibitory, the spiking of ten neurons chosen at random from each population. Each small vertical line represents a spike from a neuron. Each horizontal row shows the spikes of one neuron. The different neurons are from the same trial.

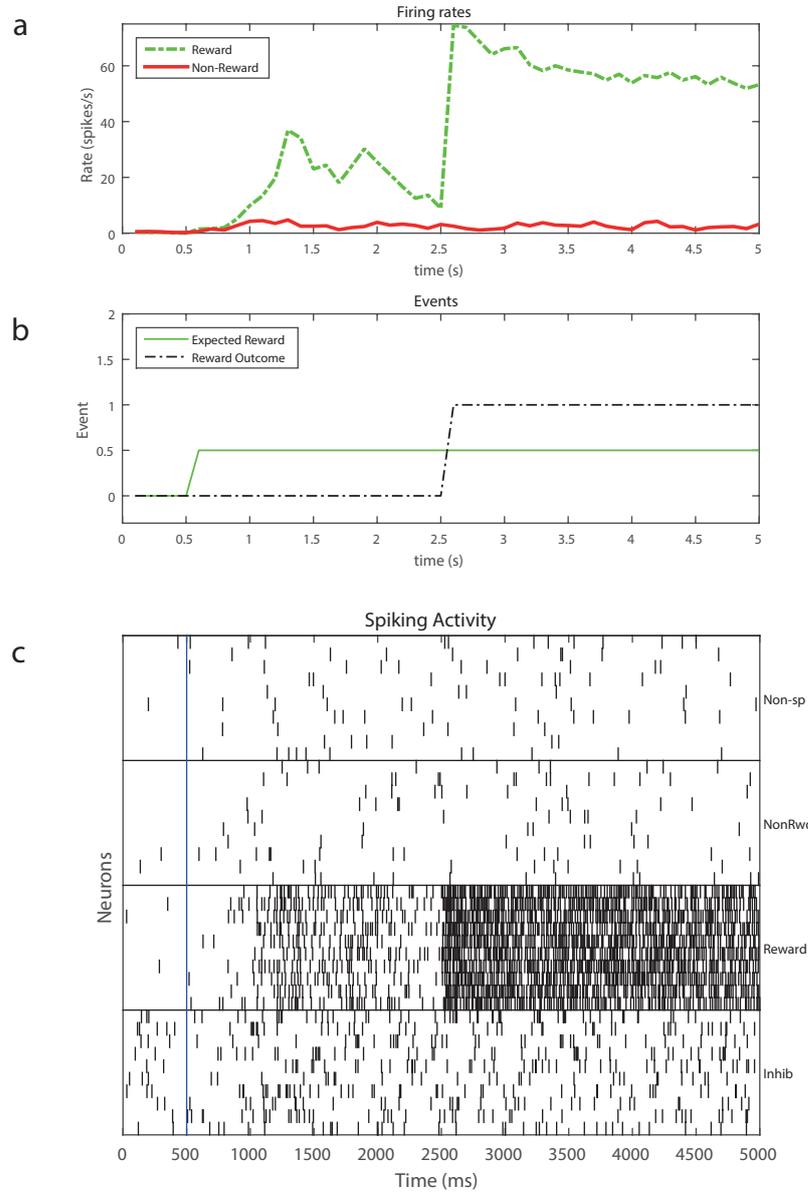


Figure 4: The operation of the network when an Expected Reward is followed by a Reward Outcome. a. The firing rates of the Reward population of neurons and the Non-Reward population of neurons during a 5 s trial. b. After a period of spontaneous activity from 0 until 0.5 s, the expected reward input was applied to the reward attractor population of neurons (3.1 Hz per external synapse). At 2500 ms a Reward Outcome input was received (3.7 Hz per synapse), and this was maintained until the end of the trial. The input to the Non-Reward population of neurons was maintained constant at 3.05 Hz per synapse from time = 0.5 s until the end of the trial. c. Rastergrams showing for each of the four populations of neurons, Non-Specific, Non-Reward, Reward, and Inhibitory, the spiking of ten neurons chosen at random from each population. Each small vertical line represents a spike from a neuron. Each horizontal row shows the spikes of one neuron.

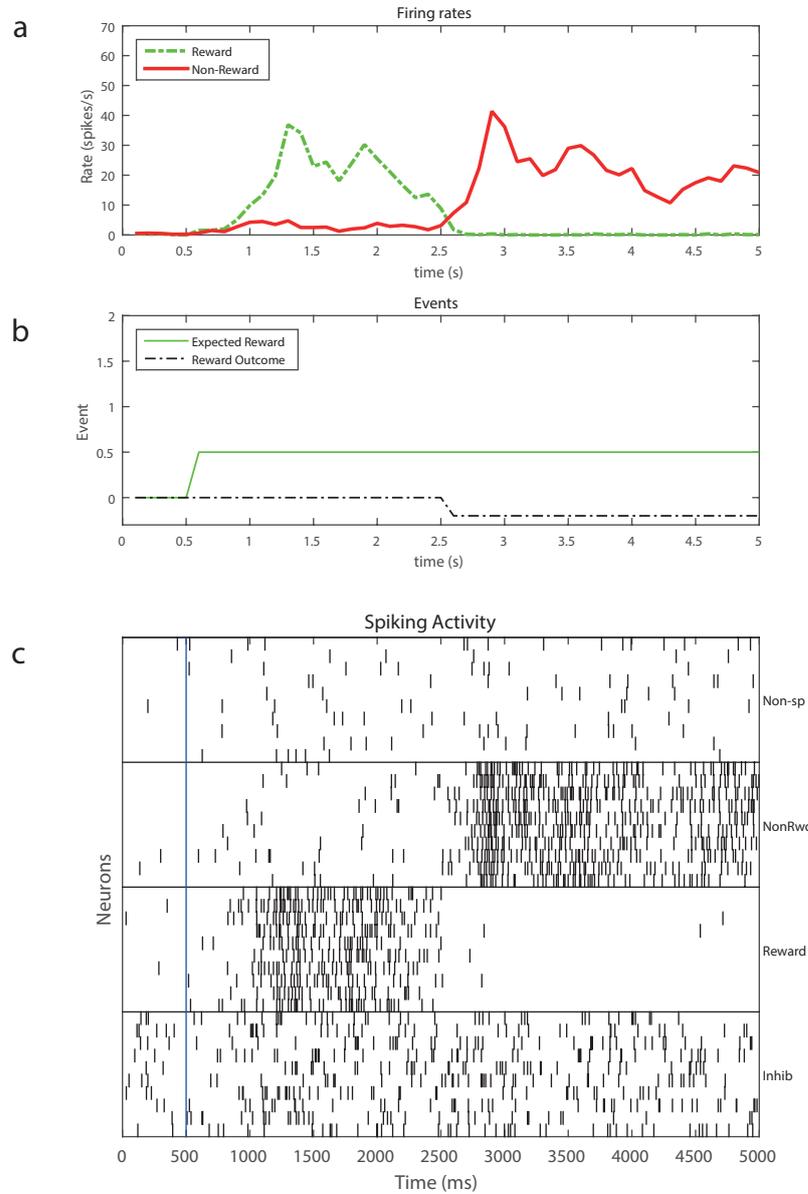


Figure 5: The operation of the network when an Expected Reward is followed by a Punishment. a. The firing rates of the Reward population of neurons and the Non-Reward population of neurons during a 5 s trial. b. After a period of spontaneous activity from 0 until 0.5 s, the Expected Reward input was applied to the Reward attractor population of neurons (3.1 Hz per synapse). At 2500 ms a Punishment was received, and this decreased the Expected Reward input to the Reward population to a low value (2.8 Hz per synapse), and this was maintained until the end of the trial. The input to the Non-Reward population of neurons was maintained constant at 3.05 Hz per synapse from time = 0.5 s until the end of the trial. c. Rastergrams showing for each of the four populations of neurons, Non-Specific, Non-Reward, Reward, and Inhibitory, the spiking of ten neurons chosen at random from each population. Each small vertical line represents a spike from a neuron. Each horizontal row shows the spikes of one neuron.