

lateral firing, rather than variability in the external inputs to the network, is what makes the major contribution to the noise in the network (Deco and Rolls, 2006). Thus, once the firing in the recurrent collaterals is spike-implemented by integrate-and-fire neurons, the probabilistic behavior seems inevitable, even up to quite large attractor network sizes.

The graded nature of the sparse distributed representations in the cortex tends to increase the noise (Webb, Rolls, Deco and Feng, 2011). Representations in the cortex are often distributed with graded firing rates in the neuronal populations. The firing rate probability distribution of each neuron to a set of stimuli is often exponential or gamma (see Section C.3.1.1) (Webb et al., 2011). In integrate-and-fire simulations of an attractor decision-making network, we showed that the noise is indeed greater for a given sparseness of the representation for graded, exponential, than for binary firing rate distributions. The greater noise was measured by faster escaping times from the spontaneous firing rate state when the decision cues are applied, and this corresponds to faster decision or reaction times. The greater noise was also evident as less stability of the spontaneous firing state before the decision cues are applied. The implication is that spiking-related noise will continue to be a factor that influences processes such as decision-making, signal detection, short-term memory, and memory recall even with the quite large networks found in the cerebral cortex. In these networks there are several thousand recurrent collateral synapses onto each neuron. The greater noise with graded firing rate distributions has the advantage that it can increase the speed of operation of cortical circuitry (Webb, Rolls, Deco and Feng, 2011).

Dilution of the connectivity within an attractor network can have the effect of decreasing the noise compared to a fully connected network with the same number of connections onto each neuron and therefore the same memory capacity (Rolls and Webb, 2012). The connectivity of the cerebral cortex is diluted, with the probability of excitatory connections between even nearby pyramidal cells rarely more than 0.1, and in the hippocampus 0.04 (Chapter 9 and Rolls (2016b)). To investigate the extent to which this diluted connectivity affects the dynamics of attractor networks in the cerebral cortex, we simulated an integrate-and-fire attractor network taking decisions between competing inputs with diluted connectivity of 0.25 or 0.1, and with the same number of synaptic connections per neuron for the recurrent collateral synapses within an attractor population as for full connectivity. The results indicated that there was less spiking-related noise with the diluted connectivity, in that the stability of the network when in the spontaneous state of firing increased, and the accuracy of the correct decisions increased. The decision times were a little slower with diluted than with complete connectivity. Given that the capacity of the network is set by the number of recurrent collateral synaptic connections per neuron, on which there is a biological limit, the findings indicate that the stability of cortical networks, and the accuracy of their correct decisions or memory recall operations, can be increased by utilizing diluted connectivity and correspondingly increasing the number of neurons in the network, with little impact on the speed of processing of the cortex (Rolls and Webb, 2012).

18.2.2 Attractor networks, energy landscapes, and stochastic neurodynamics

Attractor networks can be used for short-term memory and long-term memory, and for decision making (Section B.3). This section considers approaches to the stability of these networks. This section can be thought of as explaining the stability of memory networks, though the principles apply just as much to the decision-making networks considered in Section 11.5.1.

Autoassociation attractor systems (described in Section B.3) can have two types of stable fixed points: a spontaneous state with a low firing rate, and one or more attractor states with

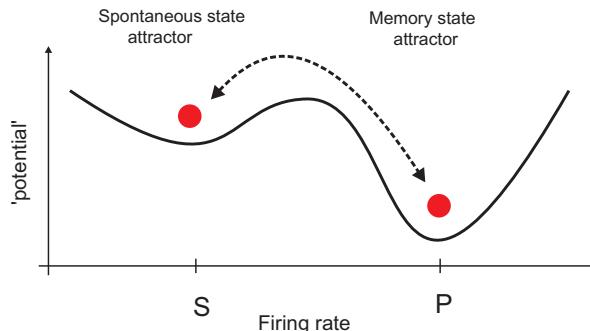


Fig. 18.1 Stability and noise in an attractor network: an energy landscape. The noise influences when the system will jump out of the spontaneous firing stable (low energy) state S, and whether it jumps into the high firing rate state labelled P (with persistent or continuing firing in a state which is even more stable with even lower energy), which might correspond to a short-term memory, or to a decision.

high firing rates in which the positive feedback implemented by the recurrent collateral connections maintains a high firing rate. We sometimes refer to this latter state as the persistent state, because the high firing normally persists to maintain a set of neurons active, which might implement a short-term memory, or the recall of a long-term memory that persists for a short time.

The stable points of the system can be visualized in an energy landscape (see Fig. 18.1). The area in the energy landscape within which the system will move to a stable attractor state is called its basin of attraction. The attractor dynamics can be pictured by energy landscapes, which indicate the basins of attraction by valleys, and the attractor states or fixed points by the bottom of the valleys (see Fig. 18.1).

The stability of an attractor is characterized by the average time in which the system stays in the basin of attraction under the influence of noise. The noise provokes transitions to other attractor states. One source of noise results from the interplay between the Poissonian character of the spikes and the finite-size effect due to the limited number of neurons in the network.

Two factors determine the stability. First, if the depths of the attractors are shallow (as in the left compared to the right valley in Figure 18.1), then less force is needed to move a ball from one valley to the next. Second, high noise will make it more likely that the system will jump over an energy boundary from one state to another. We envision that the brain as a dynamical system has characteristics of such an attractor system including statistical fluctuations (see further Rolls and Deco (2010), where the effects of noise are defined quantitatively). The noise could arise not only from the probabilistic spiking of the neurons which has significant effects in finite size integrate-and-fire networks (Deco and Rolls, 2006), but also from any other source of noise in the brain or the environment (Faisal et al., 2008), including the effects of distracting stimuli.

In an attractor network in which a retrieval cue is provided to initiate recall but then removed, a landscape can be defined in terms of the synaptic weights. An example is shown in Fig. 18.2a. The basins in the landscape can be defined by the strengths of the synaptic weights which describe the stable operating points of the system, where the depth of the basins can be defined in terms of the synaptic weight space, in terms defined by an associative rule operating on the firing rates of pairs of neurons during the learning as follows

$$w_{ij} = y_i y_j \quad (18.1)$$

where y_i is the firing rate of the postsynaptic neuron, y_j is the firing rate of the presynaptic neuron, and w_{ij} is the strength of the synapses connecting these neurons.

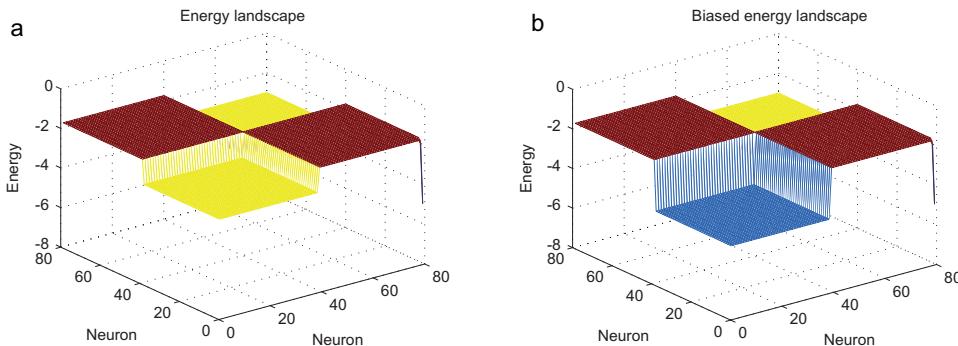


Fig. 18.2 Stability and noise in an attractor network. (a) Energy landscape without any differential bias applied. The landscape is for a network with neurons 1–40 connected by strengthened synapses so that they form attractor 1, and neurons 41–80 connected by synapses strengthened (by the same amount) so that they form attractor 2. The energy basins in the two-dimensional landscape are calculated by Equation 18.2. In this two-dimensional landscape, there are two stable attractors, and each will be reached equally probably under the influence of noise. This scenario might correspond to decision-making where the input λ_1 to attractor 1 has the same value as the input λ_2 to attractor 2, and the network is equally likely under the influence of the noise to fall into attractor 1 representing decision 1 as into attractor 2 representing decision 2. (b) Energy landscape with bias applied to neurons 1–40. This makes the basin of attraction deeper for attractor 1, as calculated with Equation 18.2. Thus, under the influence of noise caused by the randomness in the firing of the neurons, the network will reach attractor 1 more probably than it will reach attractor 2. This scenario might correspond to decision-making, where the evidence for decision 1 is stronger than for decision 2, so that a higher firing rate is applied as λ_1 to neurons 1–40. The scenario might also correspond to memory recall, in which memory 1 might be probabilistically more likely to be recalled than memory 2 if the evidence for memory 1 is stronger. Nevertheless, memory 2 will be recalled sometimes in what operates as a non-deterministic system. (From Rolls, E. T. and Deco, G. (2010) *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford University Press: Oxford.)

Hopfield (1982) showed how many stable states a simple attractor system might contain, and this is the capacity of the network described in Section B.3.3.7. He showed that the recall process in his attractor network can be conceptualized as movement towards basins of attraction, and his equation defines the energy at a given point in time as being a function of the synaptic weights and the current firing rates as follows

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} (y_i - \langle y \rangle)(y_j - \langle y \rangle). \quad (18.2)$$

where y_i is the firing rate of the postsynaptic neuron, y_j is the firing rate of the presynaptic neuron, w_{ij} is the strength of the synapse connecting them, and $\langle y \rangle$ is the mean firing rate of the neurons. I note that the system defined by Hopfield had an energy function, in that the neurons were connected by symmetric synaptic weights (produced for example by associative synaptic modification of the recurrent collateral synapses) and there was no self-coupling (Hertz, Krogh and Palmer, 1991; Moreno-Bote, Rinzel and Rubin, 2007; Hopfield and Herz, 1995). (Equation 18.2 can be understood as showing that the system will be stable if high firing rate neurons are connected by strong synaptic weights, and a stable fixed point can be thought of as having low energy so that the system does not jump somewhere else in the energy landscape.)

The situation is more complicated in an attractor network if it does not have a formal energy function. One such condition is when the connectivity is randomly diluted, for then the synaptic weights between pairs of neurons will not be symmetric. Indeed, in general, neuronal systems do not admit such an energy function. (This is the case in that it is not in general possible to define the flow in terms of the gradient of an energy function. Hopfield defined first an energy function, and from there derived dynamics.) However, such diluted

connectivity systems can still operate as attractor systems (Treves, 1993, 1991a,b; Treves and Rolls, 1991; Treves, Rolls and Simmen, 1997; Rolls and Treves, 1998; Battaglia and Treves, 1998a), and the concept of an energy function and landscape is useful for discussion purposes. In practice, a Lyapunov function can be used to prove analytically that there is a stable fixed point such as an attractor basin (Khalil, 1996), and even in systems where this can not be proved analytically, it may still be possible to show numerically that there are stable fixed points, to measure the flow towards those fixed points which describes the depth of the attractor basin as we have done for this type of network (Loh, Rolls and Deco, 2007a), and to use the concept of energy or potential landscapes to help visualize the properties of the system (Rolls and Deco, 2010).

If an external input remains on during the retrieval process, this will influence the energy function of such a network, and its stable points, as implied by Equation 18.2, and as illustrated in Fig. 18.2b. In this situation, the external inputs bias the stable points of the system. Indeed, in this situation, a landscape, though not necessarily formally an energy landscape, can be specified by a combination of the synaptic weights and external inputs that bias the firing rates. The noise introduced into the network by for example the random neuronal spiking can be conceptualized as influencing the way that the system flows across this fixed landscape shaped by the synaptic weights, and by the external inputs if they remain on during operation of the network, in what is referred to as a ‘clamped’ condition, the normal condition that applies during decision-making (see Section 11.5.1).

In more detail, the flow, which is the time derivative of the neuronal activity, specifies the landscape in an attractor system. The flow is defined in the mean field analysis in terms of the effects of the synaptic weights between the neurons and the external inputs (Loh, Rolls and Deco, 2007a; Rolls and Deco, 2010). The flow is the force that drives the system towards the attractor given a parameter value in phase space, i.e. the firing rates of the pools (populations) of neurons. This is measured by fixing the value of the firing rate of the selective pool and letting the other values converge to their fixed point. The flow can then be computed with this configuration (Mascaro and Amit, 1999). This landscape is thus fixed by the synaptic and the external inputs. The noise, produced for example by the almost Poissonian spiking of the neurons, can be conceptualized as influencing the way that the system flows across this fixed landscape. Moreover, the noise can enable the system to jump over a barrier in this fixed landscape, as illustrated in Figs. 18.1 and 11.33.

In Fig. 18.2a (and in Fig. 11.33) the decision basins of attraction are equally deep, because the inputs λ_1 and λ_2 to the decision-making network are equal, that is, ΔI the difference between them is zero. If λ_1 is greater than λ_2 , the basin will be deeper for λ_1 . The shape of the landscape is thus a function of the synaptic weights and the biassing inputs to the system. This is illustrated in Fig. 18.2b. Noise can be thought of as provoking movements across the ‘effective energy landscape’ conceptualized in this way.

The way in which we conceptualise the operation of an attractor network used for noise-driven stochastic decision-making, stimulus detection, etc, is as follows. The noise in the system (caused for example by statistical fluctuations produced by the Poisson-like neuronal firing in a finite-sized system as described in Section 18.2.1) produces changes in neuronal firing. These changes may accumulate stochastically, and eventually may become sufficiently large that the firing is sufficient to produce energy to cause the system to jump over an energy barrier (see Fig. 18.3). Opposing this noise-driven fluctuation will be the flow being caused by the shape and depth of the fixed energy landscape defined by the synaptic weights and the applied external input bias or biases. The noisy statistical fluctuation is a diffusion-like process. If the spontaneous firing rate state is stable with the decision cues applied (see Fig. 18.4 middle), eventually the noise may provoke a transition over the energy barrier in an escaping time, and the system will drop, again noisily, down the valley on the other side of the hill.

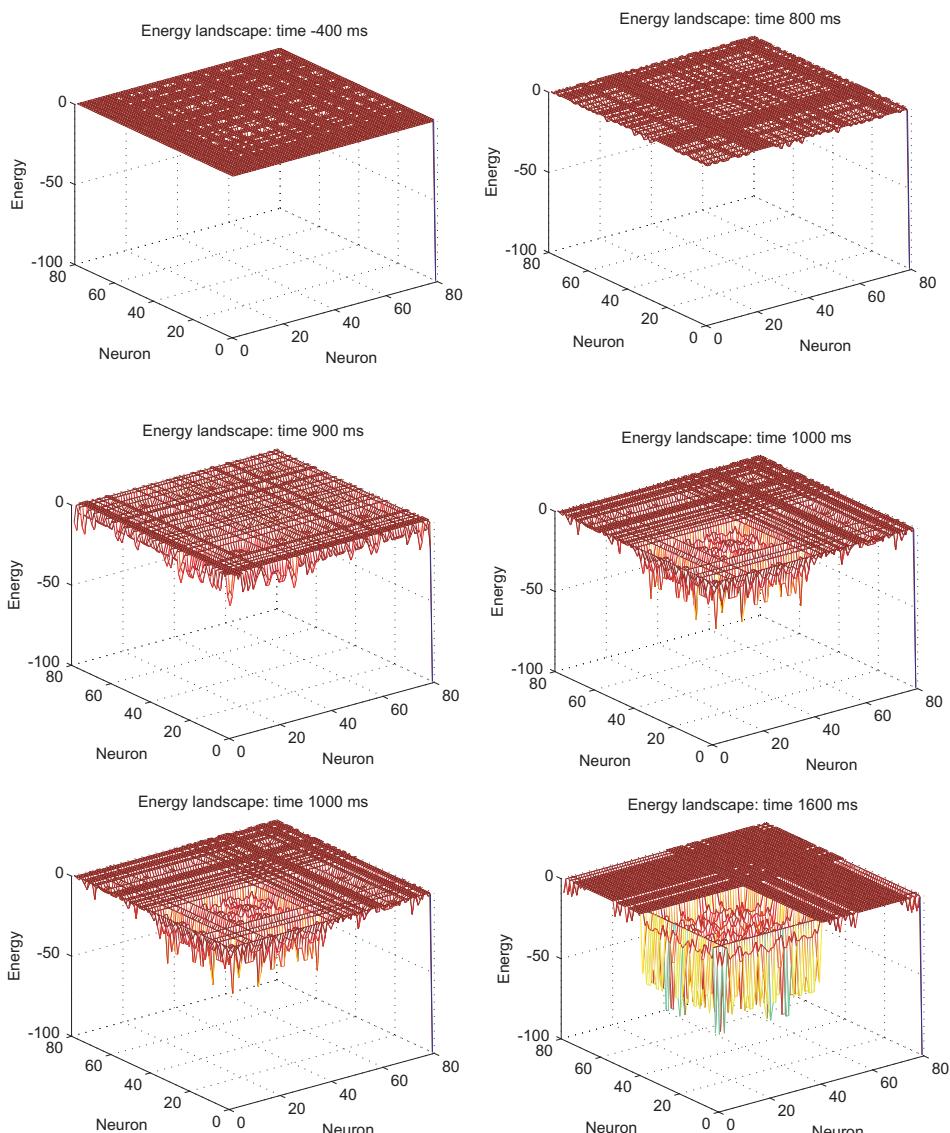


Fig. 18.3 Stability and noise in an attractor network. Energy states shown at different times after the decision cues were applied. Neurons 1–40 are in attractor 1 and are connected by strong weights with each other; and neurons 41–80 are in attractor 2 and are connected by strong weights with each other. The energy state is defined in Equation 18.2, and the energy between any pair of neurons is a product of the firing rates of each neuron and the synaptic weight that connects them. These are the energy states for the trial shown in Fig. 11.35b and c on page 550. Time 0 is the time when the decision stimuli were applied (and this corresponds to time 2 s in Fig. 11.35). (From Rolls, E. T. and Deco, G. (2010) *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford University Press: Oxford.)

The rate of change of the firing rate is again measured by the flow, and is influenced by the synaptic weights and applied biases, and by the statistical fluctuations. In this scenario, the reaction times will depend on the amount of noise, influenced by the size of the network, and by the fixed ‘effective energy landscape’ as determined by the synaptic weights and applied biasing inputs λ , which will produce an escaping time as defined further in the Appendix of Rolls and Deco (2010).

If the spontaneous state is not stable (see Fig. 18.4 right), the reaction times will be influenced primarily by the flow as influenced by the gradient of the energy landscape, and by the noise caused by the random neuronal firings. A noise-produced escaping time from a stable spontaneous state attractor will not in this situation contribute to the reaction times.

The noise-driven escaping time from the stable spontaneous state is important in understanding long and variable reaction times, and such reaction times are present primarily in the scenario when the parameters make the spontaneous state stable, as described further by Marti et al. (2008). While in a spontaneous stable state the system may be thought of as being driven by the noise, and it is primarily when the system has reached a ridge at the edge of the spontaneous valley and the system is close to a bifurcation point into a high firing rate close to the ridge that the attractor system can be thought of as accumulating evidence from the input stimuli (Deco et al., 2007). While in the spontaneous state valley (see Fig. 18.4 middle), the inputs can be thought of as biasing the ‘effective energy landscape’ across which the noise is driving the system stochastically.

An interesting aspect of the model is that the recurrent connectivity, and the relatively long time constant of the NMDA receptors (Wang, 2002), may together enable the attractor network to accumulate evidence over a long time period of several hundred milliseconds. Important aspects of the functionality of attractor networks are that they can accumulate and maintain information.

18.2.3 A multistable system with noise

In the situation illustrated in Figs. 18.1 and 11.33, there is multistability, in that the spontaneous state and a large number of high firing rate persistent states are stable. More generally, and depending on the network parameters including the strengths of the inputs, a number of different scenarios can occur. These are illustrated in Fig. 18.4. Let us consider the activity of a given neuronal population while inputs are being applied.

In Fig. 18.4 (left) we see a situation in which only the spontaneous state S is stable. This might occur if the external inputs λ_1 and λ_2 are weak.

On the right we have a situation in which our neuronal population is either in a high firing rate stable state C2, or in a low firing rate state C1 because another population is firing fast and inhibiting our neuronal population. There is no stable spontaneous state.

In the middle of Fig. 18.4 we see a situation in which our population may be either in C1, or in C2, or in a spontaneous state of firing S when no population has won the competition. We emphasize that this can be a scenario even when the decision cues λ_1 and λ_2 are being applied during the decision-making period. We refer to this system as a multistable system.

The differences between these scenarios are of interest in relation to how noise influences the decision-making. In the scenario shown in the middle of Fig. 18.4 we see that there are three stable states when the inputs λ_1 and λ_2 are being applied, and that it is the stochastic noise that influences whether the system jumps from the initial spontaneous state to a high firing rate state in which one of the decision-state populations fires fast, producing either C2 if our population wins, or C1 if our population loses. The statistical properties of the noise (including its amplitude and frequency spectrum), and the shape of the different basins in the energy landscape, influence whether a decision will be taken, the time when it will be

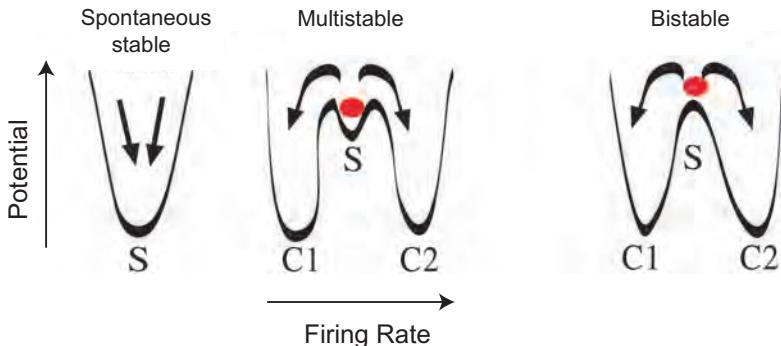


Fig. 18.4 Computational principles underlying the different dynamical regimes of the decision-making attractor network (see text). The x -axis represents the neuronal activity of one of the populations (ν_i) and the landscape represents an energy landscape ('potential') regulating the evolution of the system. S is a stable state of spontaneous activity, C2 is a high firing rate state of this neuronal population corresponding to the decision implemented by this population, and C1 is a low firing rate state present when the other population wins the competition.

taken, and which high firing rate decision attractor wins. In contrast, in the scenario shown in Fig. 18.4 (right) the energy landscape when the stimuli are being applied is such that there is no stable spontaneous state, so the system moves to one of the high firing rate decision attractors without requiring noise. In this case, the noise, and the shape of the energy landscape, influence which high firing rate decision state attractor will win.

A more detailed analysis suggests that there are two scenarios that are needed to understand the time course of processes such as decision-making and memory retrieval (which are very similar processes involving an attractor network driven by one input, or by several inputs) (Marti, Deco, Mattia, Gigante and Del Giudice, 2008).

First, in the scenario investigated by Wang (2002), the spontaneous state is unstable when the decision cues are applied. The network, initially in the spontaneous state, is driven to a competition regime by an increase of the external input (that is, upon stimulus presentation) that destabilizes the initial state. The decision process can then be seen as the relaxation from an unstable stationary state towards either of the two stable decision states (Fig. 18.4 (right)). When the system is completely symmetric (i.e. when there is no bias in the external inputs that favours one choice over the other), this destabilization occurs because the system undergoes a pitchfork bifurcation for sufficiently high inputs. The time spent by the system to evolve from the initial state to either of the two decision states is determined by the actual stochastic trajectory of the system in the phase space. In particular, the transition time increases significantly when the system wanders in the vicinity of the saddle that appears when the spontaneous state becomes unstable. Reaction times in the order of hundreds of ms may be produced in this way, and are strongly influenced by the long time constants of the NMDA receptors (Wang, 2002). The transition can be further slowed down by setting the external input slightly above the bifurcation value. This tuning can be exploited to obtain realistic decision times.

Second, there is a scenario in which the stimuli do not destabilize the spontaneous state, but rather increase the probability for a noise-driven transition from a stable spontaneous state to one of the decision states (see Fig. 18.4 (centre) and Section 18.2.3). Due to the presence of finite-size noise in the system there is a nonzero probability that this transition occurs and hence a finite mean transition rate between the spontaneous and the decision states. It has been shown that in this scenario mean decision times tend to the Van't Hoff-Arrhenius exponential dependence on the amplitude of noise in the limit of infinitely large networks. As a consequence, in this limit, mean decision times increase exponentially with the size of the network (Marti, Deco, Mattia, Gigante and Del Giudice, 2008). Further, the decision events

money (O'Doherty et al., 2001a) or not winning (Xie et al., 2021) (Fig. 11.9), and by many other aversive stimuli (Grabenhorst and Rolls, 2011) (Fig. 11.18). Further evidence that the orbitofrontal cortex is involved in changing rewarded behavior when non-reward is detected is that damage to the human orbitofrontal cortex impairs reward reversal learning, in that the previously rewarded stimulus is still chosen during reversal even when no reward is being obtained (Rolls, Hornak, Wade and McGrath, 1994a; Hornak, O'Doherty, Bramham, Rolls, Morris, Bullock and Polkey, 2004; Fellows and Farah, 2003; Fellows, 2011). Further, the right lateral orbitofrontal cortex is strongly activated by non-reward in a one-trial rule-based reward reversal task (Rolls, Vatansever, Li, Cheng and Feng, 2020e), which is the same brain region with increased functional connectivity in depression as described below.

Now it is well established that not receiving expected reward, or receiving unpleasant stimuli or events, can produce feelings of depression (Beck, 2008; Drevets, 2007; Harmer and Cowen, 2013; Price and Drevets, 2012; Pryce et al., 2011; Eshel and Roiser, 2010; Rolls, 2018a). A clear example is that if a member of the family dies, then this is the removal of reward (in that we would work to try to avoid this), and the result of the removal of the reward can be depression. More formally, in terms of learning theory, the omission or termination of a reward can give rise to sadness or depression, depending on the magnitude of the reward that is lost, if there is no action that can be taken to restore the reward (Rolls, 2005a, 2013d, 2014a, 2018a) (Fig. 11.20). If an action can be taken, then frustration and anger may arise to the same reinforcement contingency (Rolls, 2014a, 2018a, 2019e,f, 2023b) (Section 11.3.8). This relates the current approach to the learned helplessness approach to depression, in which depression arises because no actions are being taken to restore rewards (Forgeard et al., 2011; Pryce et al., 2011; Maier and Seligman, 2016).

18.5.2 A non-reward attractor theory of depression

The theory has been proposed that in depression, the lateral orbitofrontal cortex non-reward / punishment attractor network system is more easily triggered, and maintains its attractor-related firing for longer (Rolls, 2016e, 2017c, 2019f, 2018a). The greater attractor-related firing of the non-reward / punishment system triggers negative cognitive states held on-line in other cortical systems such as the language system and in the dorsolateral prefrontal cortex which is implicated in attentional control. These other cortical systems then in turn have top-down effects on the orbitofrontal non-reward system that bias it in a negative direction (Rolls, 2013a) (see Section 4.3.5 and Fig. 4.16), and thus increase the sensitivity of the lateral orbitofrontal cortex to non-reward and maintain its overactivity (Rolls, 2016e) (Fig. 18.12). It is proposed that the interaction of non-reward and language / attentional brain systems of these types accounts for the ruminating and continuing depressive thoughts, which occur as a result of a positive feedback cycle between these types of brain system (Rolls, 2016e).

Indeed, we have shown that cognitive states can have 'top-down' effects on affective representations in the orbitofrontal cortex (De Araujo, Rolls, Velazco, Margot and Cayeux, 2005; Grabenhorst, Rolls and Bilderbeck, 2008a; McCabe, Rolls, Bilderbeck and McGlone, 2008; Rolls, 2013a). Further, top-down selective attention can also influence affective representations in the orbitofrontal cortex (Rolls et al., 2008b; Grabenhorst and Rolls, 2008; Ge et al., 2012; Luo et al., 2013; Rolls, 2013a), and paying attention to depressive symptoms when depressed may in this way exacerbate the problems in a positive feedback way.

More generally, the presence of the cognitive ability that is afforded by language to think ahead and see the implications of recent events may be a computational development in the brain that exacerbates the vulnerability of the human brain to depression (Rolls, 2014a, 2018a). For example, with language we can think ahead and see that perhaps the loss of an

Interaction of non-reward and language networks in depression

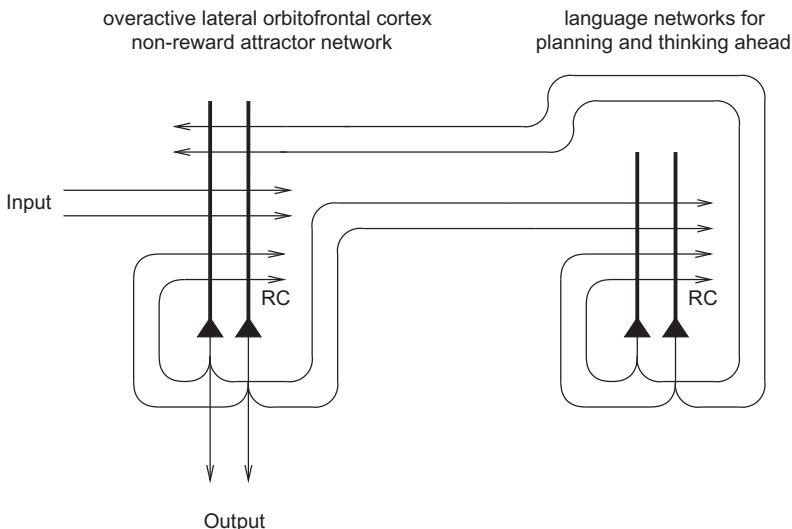


Fig. 18.12 Interaction of orbitofrontal cortex non-reward networks with language networks in depression. Illustration of how an overactive non-reward attractor network in the lateral orbitofrontal cortex could send excitatory information forward to networks for language and planning ahead; which could in turn send excitatory 'top-down' feedback back down to the orbitofrontal non-reward network to maintain its over-activity. It is suggested that such a system with mutual 'long loop' re-excitation contributes to the persistent ruminating thoughts in depression. (After Rolls, E. T. (2016) A non-reward attractor theory of depression. *Neuroscience and Biobehavioral Reviews* 68: 47–58. © Elsevier.)

individual in one's life may be long-term, and this thought and its consequences for our future can become fully evident.

The theory is that one way in which depression could result from over-activity in this lateral orbitofrontal cortex system is if there is a major negatively reinforcing life event that produces reactive depression and activates this system, which then becomes self-re-exciting based on the cycle between the lateral orbitofrontal cortex non-reward / punishment attractor system and the cognitive / language system, which together operate as a systems-level attractor (Fig. 18.12). (The generic cortical architecture for such reciprocal feedforward and feedback excitatory effects is illustrated by Rolls (2016b).)

The theory is that a second way in which depression might arise is if this lateral orbitofrontal cortex non-reward / punishment system is especially sensitive in some individuals. This might be related for example to genetic predisposition, or to the effects of stress (Gold, 2015). In this case, the orbitofrontal system would over-react to normal levels of non-reward or punishment, and start the local attractor circuit in the lateral orbitofrontal cortex (Section 11.3.4) (Rolls, 2016e; Rolls and Deco, 2016), which in turn would activate the cognitive system, which would feed back to the over-reactive lateral orbitofrontal cortex system to maintain now a systems-level attractor with ruminating thoughts. This is described as a 'systems-level' attractor because it includes mutual excitations between different brain areas.

Given that the activations of the lateral and medial orbitofrontal cortex often appear to be reciprocally oppositely related (O'Doherty, Kringelbach, Rolls, Hornak and Andrews, 2001a; Rolls, Kringelbach and De Araujo, 2003c; Xie, Jia, Rolls, Robbins, Sahakian, Zhang, Liu, Cheng, Luo, Zac Lo, Wang, Banaschewski, Barker, Bodke, Buchel, Quinlan, Desrivieres, Flor, Grigis, Garavan, Gowland, Heinz, Hohmann, Ittermann, Martinot, Martinot, Nees, Papadopoulos Orfanos, Paus, Poustka, Frohner, Smolka, Walter, Whelan, Schumann, Feng and IMAGEN, 2021; Zhang, Rolls, Wang, Xie, Cheng and Feng, 2023a), the other part of the theory of depression is that in depression there may be underactivity, under-sensitivity, or

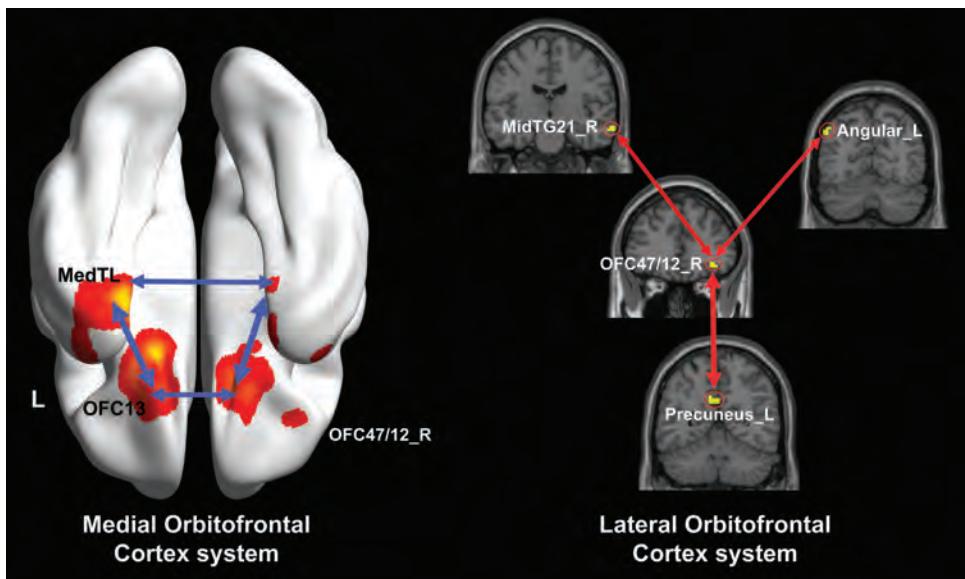


Fig. 18.13 Resting state functional connectivity in depression. The medial orbitofrontal cortex has reduced functional connectivity (blue) in depression with medial temporal lobe memory systems. The lateral orbitofrontal cortex has increased functional connectivity (red) in depression with the angular gyrus, precuneus, and middle temporal gyrus. MedTL – medial temporal lobe from the parahippocampal gyrus to the temporal pole; MidTG21R – middle temporal gyrus area 21 right; OFC13 – medial orbitofrontal cortex area 13; OFC12/47R – lateral orbitofrontal cortex area 12/47 right. The lateral orbitofrontal cortex cluster in OFC12/47 is visible on the ventral view of the brain anterior and lateral to the OFC13 clusters. (After Cheng,W., Rolls,E.T., Qiu,J., Liu,W., Tang,Y., Huang,C-C., Wang,XF., Zhang,J., Lin,W., Zheng,L., Pu,JC., Tsai,S-J., Yang,AC., 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. *Brain* 139: 3296–3309. © Oxford University Press.)

under-connectivity of the (reward-related) medial orbitofrontal cortex in depression (Rolls, 2016e, 2018a; Uran et al., 2022). The theory is further that under-responsiveness of the medial orbitofrontal cortex could contribute to other aspects of depression, such as anhedonia.

18.5.3 The orbitofrontal cortex, and the theory of depression

There is some evidence for altered structure and function of the lateral orbitofrontal cortex in depression (Drevets, 2007; Ma, 2015; Price and Drevets, 2012; Zhang, Rolls, Wang, Xie, Cheng and Feng, 2023a). For example, reductions of grey-matter volume and cortex thickness have been demonstrated in the posterolateral OFC (BA 12/47, caudal BA 11 and the adjoining BA 45), and also in the subgenual cingulate cortex (BA 24, 25) (Drevets, 2007; Nugent et al., 2006; Schmaal et al., 2017), and in the medial orbitofrontal cortex (Schmaal et al., 2017). In depression, there is increased cerebral blood flow in areas that include the ventrolateral orbitofrontal cortex (which is a prediction of the theory), and also in regions such as the subgenual cingulate cortex and amygdala, and these increases appear to be related to the mood change, in that they become more normal when the mood state remits (Drevets, 2007; Lally, Nugent, Luckenbaugh, Nicu, Roiser and Zarate, 2015; Zhang, Rolls, Wang, Xie, Cheng and Feng, 2023a).

In the first brain-wide voxel-level resting state functional connectivity neuroimaging analysis of depression (with 421 patients with major depressive disorder and 488 controls), we have found that one major circuit with altered functional connectivity involved the medial orbitofrontal cortex BA 13, which had reduced functional connectivity in depression with memory systems in the parahippocampal gyrus and medial temporal lobe (Cheng, Rolls, Qiu, Liu, Tang, Huang, Wang, Zhang, Lin, Zheng, Pu, Tsai, Yang, Lin, Wang, Xie and Feng,

2016a) (Fig. 18.13). The lateral orbitofrontal cortex BA 12/47, involved in non-reward and punishing events, did not have this reduced functional connectivity with memory systems, so that there is an imbalance in depression towards decreased reward-related memory system functionality.

A second major circuit change was that the lateral orbitofrontal cortex area BA 12/47 had increased functional connectivity with the precuneus, the angular gyrus, and the temporal visual cortex BA 21 (Cheng et al., 2016a) (Fig. 18.13). This enhanced functional connectivity of the non-reward/punishment system (BA 12/47) with the precuneus (involved in the sense of self and agency), and the angular gyrus (involved in language) is thus related to the explicit affectively negative sense of the self, and of self-esteem, in depression.

The differences in orbitofrontal connectivity with these brain regions were related to the depression by evidence that the symptoms of depression were correlated with these differences of functional connectivity; and that the lateral orbitofrontal cortex functional connectivity links described were less high if the patients were receiving antidepressant medication (Cheng et al., 2016a).

The reduced functional connectivity of the medial orbitofrontal cortex, implicated in reward, with memory systems provides a new way of understanding how memory systems may be biased away from pleasant events in depression. The increased functional connectivity of the lateral orbitofrontal cortex, implicated in non-reward and punishment, with areas of the brain implicated in representing the self, language, and inputs from face and related perceptual systems provides a new way of understanding how unpleasant events and thoughts, and lowered self-esteem, may be exacerbated in depression (Cheng et al., 2016a; Rolls et al., 2018a).

These differences of functional connectivity are related to the orbitofrontal cortex attractor theory of depression, because increased functional connectivity of the non-reward lateral orbitofrontal cortex would increase the stability and persistence of its negative attractor mood-related states; and decreased functional connectivity of the reward-related medial orbitofrontal cortex would decrease the stability and persistence of its positive mood states (Rolls, 2018a).

These advances have stimulated many other large-scale voxel-level investigations of functional connectivity in depression, described in Section 18.5.4. A recent large-scale activation investigation of the role of the orbitofrontal cortex in depressive symptoms is described in Section 18.5.5. Implications for treatments are considered in Section 18.5.6.

18.5.4 Altered connectivity of the orbitofrontal cortex in depression

Further investigations have provided more evidence for different functional connectivity of the orbitofrontal cortex in depression, as described next, with a summary of some of the differences shown in Fig. 18.14 (Rolls, Cheng and Feng, 2020a). **These differences of functional connectivity all relate to altered stability of the circuits involved, which help to account for some of the symptoms of depression.**

18.5.4.1 Precuneus: higher connectivity with the lateral orbitofrontal cortex

The precuneus is a medial parietal cortex region implicated in the sense of self and agency, autobiographical memory, spatial function, and navigation (Cavanna and Trimble, 2006; Fretton et al., 2014). The retrosplenial cortex provides connections to and receives connections from the hippocampal system, connecting especially with the parahippocampal gyrus areas TF and TH, and with the subiculum (Bubb et al., 2017; Kobayashi and Amaral, 2007; Rolls et al., 2023i) (Fig. 9.1). The precuneus can be conceptualized as providing access to the hippo-

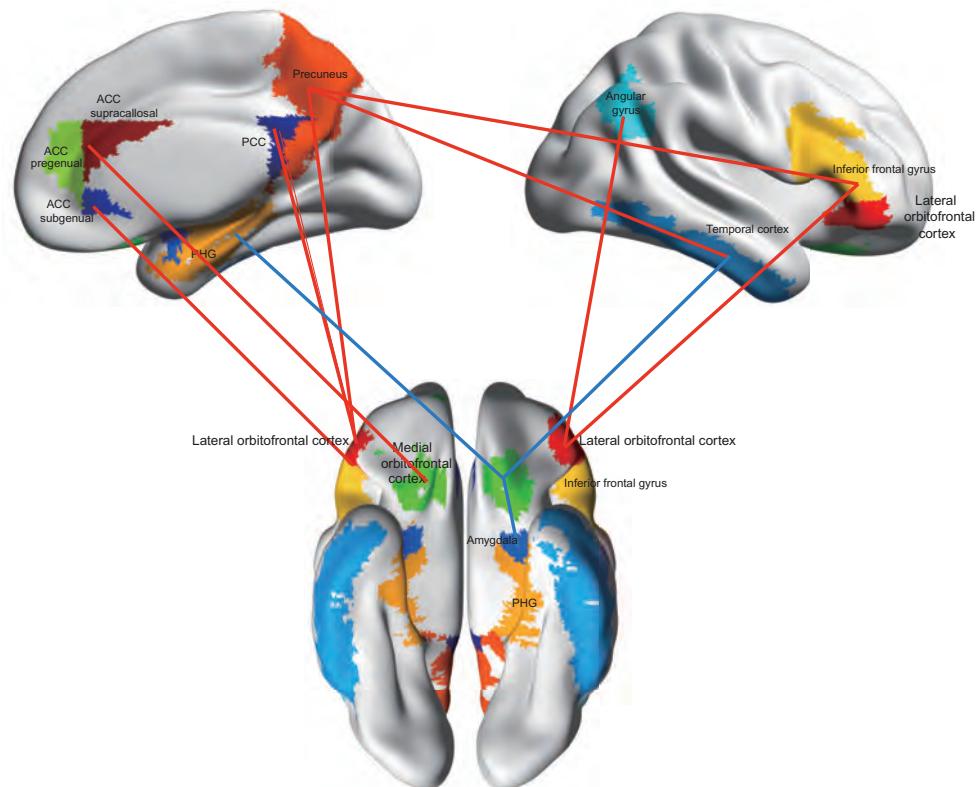


Fig. 18.14 Functional connectivity (FC) differences of the medial and lateral orbitofrontal cortex in major depressive disorder. Higher functional connectivity in depression is shown in red, and includes higher functional connectivity of the non-reward / punishment-related lateral orbitofrontal cortex with the precuneus, posterior cingulate cortex (PCC), subgenual anterior cingulate cortex (ACC subgenual), angular gyrus, and inferior frontal gyrus. Lower functional connectivity in depression is shown in blue, and includes lower functional connectivity of the medial orbitofrontal cortex with the parahippocampal gyrus memory system (PHG), amygdala, and temporal cortex. 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 2 (Rolls, Joliot and Tzourio-Mazoyer, 2015), although the investigations that form the basis for the summary were at the voxel level. (After Rolls, E. T., Cheng, W. and Feng, J. (2020) The orbitofrontal cortex: reward, emotion, and depression. *Brain Communications* 2: fcaa196.)

campus for spatial and related information from the parietal cortex (given the rich connections between the precuneus and parietal cortex) (Rolls, 2018a; Rolls et al., 2023i).

To further analyze the functioning of the precuneus in depression, resting state functional connectivity was measured in 282 patients with major depressive disorder and 254 controls (Cheng, Rolls, Qiu, Yang, Ruan, Wei, Zhao, Meng, Xie and Feng, 2018c). In 125 patients not receiving medication, voxels in the precuneus had significantly higher functional connectivity with the lateral orbitofrontal cortex (Fig. 18.14). In patients receiving medication, the functional connectivity between the lateral orbitofrontal cortex and precuneus was decreased back towards that in the controls (Cheng et al., 2018c).

These findings support the theory that the non-reward system in the lateral orbitofrontal cortex has increased effects on areas in which the self is represented including the precuneus, which could relate to the low self-esteem in depressed patients (Rolls, 2016e, 2018a; Rolls et al., 2020a).

18.5.4.2 Parahippocampal gyrus / medial temporal lobe memory system, and temporal lobe visual cortex: lower connectivity with the medial orbitofrontal cortex

We found that voxels in the medial orbitofrontal cortex had lower functional connectivity with the parahippocampal gyrus / medial temporal lobe memory system (Cheng et al., 2016a) (Fig. 18.13), and interpreted this as resulting in fewer happy memories being recalled, as the medial orbitofrontal cortex has activations that correlate with subjective pleasantness, as described above, and the parahippocampal gyrus is a pathway in the hippocampal episodic memory system (Kesner and Rolls, 2015; Rolls, 2016b, 2018b; Rolls and Wirth, 2018; Rolls et al., 2023e, 2022a; Rolls, 2023c). The reduced connectivity with temporal cortex areas in which objects and faces are represented was interpreted as contributing to the reduced positive valuation of signals involved in emotion such as the sight of face expressions, and of people (Hasselmo et al., 1989a; Critchley et al., 2000).

In a further analysis, which also investigated the effects of antidepressant medication (Rolls, Cheng, Du, Wei, Qiu, Dai, Zhou, Xie and Feng, 2020b), medial orbitofrontal cortex voxels had lower functional connectivity with temporal cortex areas, the parahippocampal gyrus, fusiform gyrus, and supplementary motor area, and medication did not result in these functional connectivities being closer to controls. This is consistent with the anhedonia of depression and reduced happy memories being related to these low functional connectivities of the medial orbitofrontal cortex with temporal lobe and memory systems. What is especially interesting is that these low functional connectivities are not normalized by treatment with antidepressant drugs (Rolls et al., 2020b), suggesting that one goal of future treatment for depression might be to increase the functionality of the medial orbitofrontal cortex (Rolls et al., 2020a).

18.5.4.3 Posterior cingulate cortex: higher functional connectivity with the lateral orbitofrontal cortex in depression

The posterior cingulate cortex is a region with strong connectivity in primates with the entorhinal cortex and parahippocampal gyrus (areas TF and TH), and thus with the hippocampal memory system (Vogt, 2009; Bubb et al., 2017; Rolls and Wirth, 2018; Rolls, 2019c) (Fig. 9.1). The posterior cingulate cortex also has connections with the orbitofrontal cortex (Vogt and Pandya, 1987; Vogt, 2009; Hsu et al., 2020), and the posterior cingulate cortex has high functional connectivity with the parahippocampal regions that are involved in memory (Cheng et al., 2018b; Rolls, 2019c). The posterior cingulate region (including the retrosplenial cortex) is consistently engaged by a range of tasks that examine episodic memory including autobiographical memory, and imagining the future; and also spatial navigation and scene processing (Leech and Sharp, 2014; Auger and Maguire, 2013). Self-reflection and self-imagery activate the ventral part of the posterior cingulate cortex (vPCC, the part with which we will be mainly concerned here) (Kircher et al., 2002, 2000; Johnson et al., 2002; Sugiura et al., 2005; Rolls, 2019c; Rolls et al., 2023i).

To analyze the functioning of the posterior cingulate cortex in depression, we performed a fully voxel-level resting-state functional connectivity neuroimaging analysis of depression of the posterior cingulate cortex, with 336 patients with major depressive disorder and 350 controls (Cheng, Rolls, Qiu, Xie, Wei, Huang, Yang, Tsai, Li, Meng, Lin, Xie and Feng, 2018b). In depression, the posterior cingulate cortex had significantly higher functional connectivity with the lateral orbitofrontal cortex (Fig. 18.14). In patients receiving medication, the functional connectivity between the lateral orbitofrontal cortex and the posterior cingulate cortex was decreased back towards that in the controls.

These findings are consistent with the hypothesis that the non-reward system in the lateral orbitofrontal cortex has increased effects on memory systems, which contribute to the rumination about sad memories and events in depression (Cheng et al., 2018b).

18.5.4.4 Anterior cingulate cortex: higher connectivity with the orbitofrontal cortex in depression

The orbitofrontal cortex projects to the anterior cingulate cortex (Vogt, 2009; Bubb et al., 2017; Rolls, 2019c; Vogt, 2019b; Rolls et al., 2023c; Zhang et al., 2023a). As shown in Chapter 12, the supracallosal anterior cingulate cortex is activated by many aversive stimuli, and has strong connectivity with the lateral orbitofrontal cortex (Rolls, Cheng, Gong, Qiu, Zhou, Zhang, Lv, Ruan, Wei, Cheng, Meng, Xie and Feng, 2019; Hsu, Rolls, Huang, Chong, Lo, Feng and Lin, 2020; Du, Rolls, Cheng, Li, Gong, Qiu and Feng, 2020). The pregenual anterior cingulate cortex is activated by many pleasant, rewarding, stimuli, and has strong functional connectivity with the medial orbitofrontal cortex (Rolls et al., 2019; Hsu et al., 2020; Du et al., 2020; Rolls et al., 2023c). However, the dorsal or supracallosal anterior cingulate cortex appears to be involved in learning actions to obtain rewards (action-outcome learning), where the outcome refers to the reward or punisher for which an action is being learned (Rolls, 2019c). In contrast, the medial orbitofrontal cortex is implicated in reward-related processing and learning, and the lateral orbitofrontal cortex in non-reward and punishment-related processing and learning. These involve stimulus-stimulus associations, where the second stimulus is a reward (or its omission), or a punisher (Rolls, 2019f,e). Now given that emotions can be considered as states elicited by rewarding and punishing stimuli, and that moods such as depression can arise from prolonged non-reward or punishment (see above and Chapter 11), the part of the brain that processes these stimulus-reward associations, the orbitofrontal cortex, is more likely to be involved in depression than the action-related parts of the cingulate cortex, although the anterior cingulate cortex and other regions related to action such as the right inferior frontal gyrus could contribute to depression as it is an output region for the orbitofrontal cortex (Rolls et al., 2023c,b).

However, the subgenual (or subcallosal) cingulate cortex has been implicated in depression, and electrical stimulation in that region may relieve depression (Lujan et al., 2013; Laxton et al., 2013; Lozano et al., 2012; Hamani et al., 2011, 2009; Mayberg, 2003) (although it has not been possible to confirm this in a double-blind study (Holtzheimer et al., 2017)). The subgenual cingulate cortex is also implicated in autonomic function (Gabbott et al., 2003), and this could be related to some of the effects found in this area that are related to depression. Whether the subgenual cingulate cortex is activated because of inputs from the orbitofrontal cortex, or performs separate computations is not yet clear. Further, the possibility is considered that electrical stimulation of the subcallosal region, which includes parts of the ventromedial prefrontal cortex (Laxton et al., 2013), that may relieve depression, may do so at least in part by activating connections involving the orbitofrontal cortex, other parts of the anterior cingulate cortex, and the striatum (Lujan et al., 2013; Hamani et al., 2009; Johansen-Berg et al., 2008; Elias et al., 2022a; Zhang et al., 2023a).

In a study of depression, overall the anterior cingulate cortex (ACC) had significantly higher functional connectivity with the orbitofrontal cortex, inferior frontal gyrus, superior parietal lobule, and with early cortical visual areas (Fig. 18.14) (Rolls, Cheng, Gong, Qiu, Zhou, Zhang, Lv, Ruan, Wei, Cheng, Meng, Xie and Feng, 2019). A pregenual ACC subdivision had high functional connectivity with medial orbitofrontal cortex areas, and a supracallosal ACC subdivision had high functional connectivity with the lateral orbitofrontal cortex and inferior frontal gyrus. The high FC in depression between the lateral orbitofrontal cortex and the subcallosal (subgenual) parts of the ACC may provide a mechanism for more non-reward information transmission to the ACC, contributing to depression. The high functional

connectivity between the medial orbitofrontal cortex and supracallosal ACC in depression may also contribute to depressive symptoms. These higher functional connectivities in un-medicated patients were ameliorated by treatment with antidepressants (Rolls et al., 2019).

18.5.4.5 Inferior frontal gyrus: increased connectivity with the lateral orbitofrontal cortex in depression

The lateral orbitofrontal cortex projects to the inferior frontal gyrus (Hsu et al., 2020; Rolls et al., 2023c,b), and very interestingly higher functional connectivity was found in depression of voxels in the right inferior frontal gyrus with voxels in the lateral and medial orbitofrontal cortex, cingulate cortex, inferior and middle temporal gyrus and temporal pole, the angular gyrus, precuneus, hippocampus and mid- and superior frontal gyrus connectivity with the orbitofrontal cortex, temporal lobe areas, the parahippocampal gyrus and hippocampus, and motor areas (Fig. 18.14) (Rolls, Cheng, Du, Wei, Qiu, Dai, Zhou, Xie and Feng, 2020b). In medicated patients these functional connectivities of the inferior frontal gyrus were lower and towards those in controls.

The hypothesis was proposed (Rolls et al., 2020b,a) that one way in which the orbitofrontal cortex influences behavior in depression is via the right inferior frontal gyrus, which projects in turn to premotor cortical areas (Du, Rolls, Cheng, Li, Gong, Qiu and Feng, 2020; Rolls, Deco, Huang and Feng, 2023f). Consistent with the consequent hypothesis that the inferior frontal gyrus route may allow non-reward signals to have too great an effect to inhibit behavior in depression, lesions of the right inferior frontal gyrus impair stopping in the stop-signal task, and produce impulsiveness (Aron, Robbins and Poldrack, 2014; Dalley and Robbins, 2017). Also consistent with the hypothesis, successful stopping in the stop-signal task is associated with high activation of the inferior frontal gyrus and lateral orbitofrontal cortex (Deng, Rolls et al (2017)).

18.5.4.6 Amygdala: reduced connectivity with the orbitofrontal cortex in depression

The amygdala is involved in emotion, though as shown in Chapter 11, it may be overshadowed in humans by the orbitofrontal cortex (LeDoux and Pine, 2016; Rolls, 2019e; Rolls et al., 2023b). In a large scale study of depression, amygdala voxels had decreased functional connectivity with the medial orbitofrontal cortex (involved in reward); the lateral orbitofrontal cortex (involved in non-reward and punishment); temporal lobe areas (involved in visual and auditory perception including face expression analysis (Perrett, Rolls and Caan, 1982; Rolls, 2011d, 2012d; Leonard, Rolls, Wilson and Baylis, 1985); and the parahippocampal gyrus (involved in memory) (Fig. 18.14) (Cheng, Rolls, Qiu, Xie, Lyu, Li, Huang, Yang, Tsai, Lyu, Zhuang, Lin, Xie and Feng, 2018a). This disconnectivity of the amygdala may contribute to particularly some of the changed behavioral and autonomic responses in depression, and perhaps not to the depressed subjective feelings (Rolls et al., 2023b).

18.5.4.7 Sleep, depression, and increased lateral orbitofrontal cortex connectivity

Sleep is frequently impaired in depression (Becker et al., 2017). To advance understanding of the brain regions involved in sleep and depression, the relation between functional connectivity, depressive symptoms (the Adult Self-Report Depressive Problems scores), and poor sleep quality was measured in 1017 participants from the general population in the Human Connectome Project (Cheng, Rolls, Ruan and Feng, 2018d). The brain areas with increased functional connectivity of these common links related to both sleep and depressive scores included the lateral orbitofrontal cortex; the dorsolateral prefrontal cortex; the anterior and posterior cingulate cortex; the insula; the parahippocampal gyrus and hippocampus; the

amygdala; the temporal cortex; and the precuneus. A mediation analysis showed that these functional connectivities in the brain contribute to the relation between depression and poor sleep quality.

Evidence was also found in this general population that the Depressive Problems scores were correlated with functional connectivities between areas that included the lateral orbitofrontal cortex, cingulate cortex, precuneus, angular gyrus, and temporal cortex (Cheng, Rolls, Ruan and Feng, 2018d). Part of the importance of this is that it provides strong support for a role of the lateral orbitofrontal cortex in depression in a general population in the U.S.A. in which a tendency to have depressive problems could be assessed. This cross-validation in a completely different population and in people not selected to have depression (Cheng et al., 2018d) provides support for the theory that the lateral orbitofrontal cortex is a key brain area that might be targeted in the search for treatments for depression (Rolls, 2016e). (Further validation is provided by a European cohort of 1,877 adolescents (Xie et al., 2021).) Low sleep duration and high depressive scores are also related to lower cortical area of the orbitofrontal cortex in a sample of 11,067 participants (Cheng, Rolls, Gong, Du, Zhang, Zhang, Li and Feng, 2021).

18.5.4.8 Effective connectivity in depression

Effective connectivity measures the effect of one brain region on another in a particular direction, and can in principle therefore provide information related to the causal processes that operate in brain function, that is, how one brain region influences another.

In a resting state fMRI investigation, effective connectivity directed to the medial orbitofrontal cortex from areas including the parahippocampal gyrus, temporal pole, inferior temporal gyrus, and amygdala was decreased in depression (Rolls, Cheng, Gilson, Qiu, Hu, Ruan, Li, Huang, Yang, Tsai, Zhang, Zhuang, Lin, Deco, Xie and Feng, 2018a). This implies less strong positive driving influences of these input regions on the medial and middle orbitofrontal cortex, regions implicated in reward, and thus helps to elucidate part of the decreased feelings of happy states in depression (Rolls, 2016e). The links from temporal cortical areas to the precuneus were increased in depression, and this may relate to representations of the sense of self (Cavanna and Trimble, 2006), which become more negative in depression (Rolls, 2016e; Cheng et al., 2016a). The lateral orbitofrontal cortex, implicated in non-reward and punishment, had an increased level of activity as reflected in the analysis in the depressed group. In addition, activity in the analysis was also higher in the right and left hippocampus of patients with depression, implying heightened memory-related processing (Rolls et al., 2018a).

18.5.5 Increased activations to non-reward of the lateral orbitofrontal cortex, and decreased sensitivity to reward of the medial orbitofrontal cortex, are related to depression scores

In 1140 adolescents at age 19 and 1877 at age 14 in the monetary incentive delay task, we found that the medial orbitofrontal cortex had graded increases in activation as the reward (Win) value increased (Xie, Jia, Rolls et al (2021)). The lateral orbitofrontal cortex had graded increases of activation as the reward value dropped to zero (the No-Win condition) (Fig. 11.9).

In a subgroup with a high score on the Adolescent Depression Rating Scale at age 19 and 14, the medial orbitofrontal cortex activations had reduced sensitivity to the different reward conditions; and the lateral orbitofrontal cortex activation showed high activation to the No-Win (i.e. Non-reward) condition (Xie et al., 2021) (Fig. 18.15). These new findings provide support for the hypothesis that those with symptoms of depression have increased sensitivity to non-reward in the lateral orbitofrontal cortex, and decreased sensitivity for differences

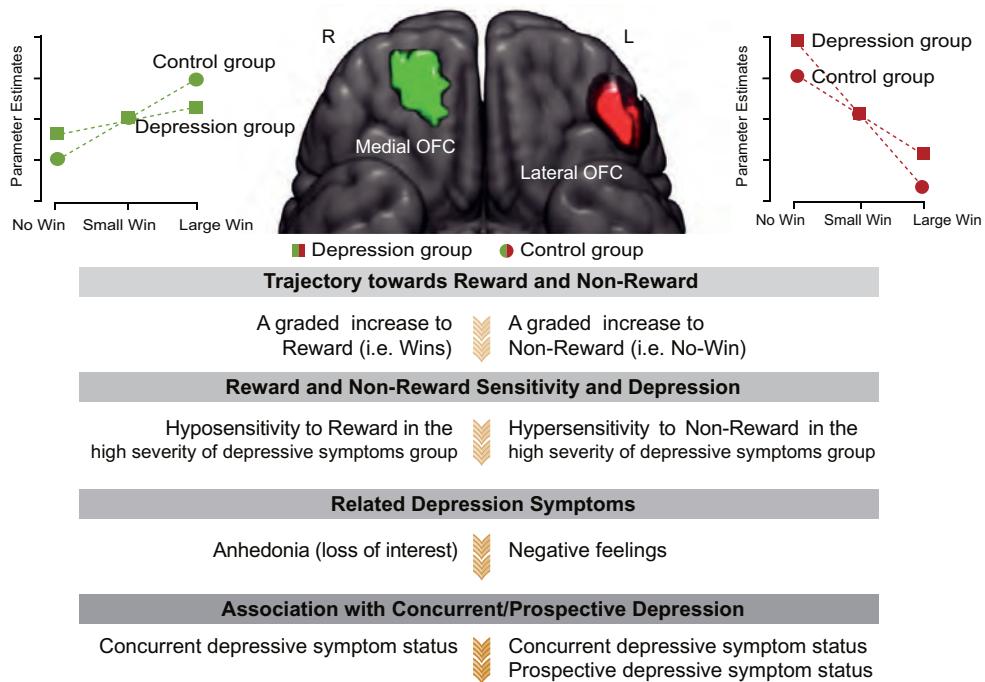


Fig. 18.15 Reduced sensitivity to reward of the medial orbitofrontal cortex, and increased sensitivity of the lateral orbitofrontal cortex to non-reward (No Win) in individuals with high scores on the Adolescent Depression Rating Scale (squares) compared to a control group (circles) in the Monetary Incentive Delay task. The summary is based on findings in 1140 adolescents at age 19 and 1877 at age 14. L, left; R, right. (After Xie,C., Jia,T., Rolls, E. T. et al (2021) Reward versus Nonreward sensitivity of the Medial versus Lateral Orbitofrontal Cortex relates to the severity of depressive symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 6: 259-269.)

in reward of the medial orbitofrontal cortex. Moreover, these differences are evident at an age as early as 14 years old (Xie et al., 2021). This increase in Non-reward sensitivity of the lateral orbitofrontal cortex in depression, and decreased Reward sensitivity of the medial orbitofrontal cortex, may act together with the altered functional connectivity of these regions just described, to make some individuals susceptible to depression (Rolls et al., 2020a; Zhang et al., 2023a).

It is hypothesized that as part of the process of evolution, variation of the sensitivity of individuals to specific types of Reward and Non-Reward may be present (Rolls, 2014a). Individuals with high sensitivity to Non-Reward may be susceptible to depression, and individuals with low sensitivity to Non-Reward may be impulsive because they are little affected by non-reward (Rolls, 2014a, 2018a). Individuals with high sensitivity to Reward may be sensation-seekers (Wan, Rolls, Cheng and Feng, 2020) and risk-takers (Rolls, Wan, Cheng and Feng, 2022c) (with increased functional connectivity of the medial orbitofrontal cortex with for example the anterior cingulate cortex, and for that reason also impulsive (Wan, Rolls, Cheng and Feng, 2020; Rolls, Wan, Cheng and Feng, 2022c)); and individuals with low sensitivity to Reward may have reduced goal-seeking behavior and reduced motivation (Rolls, 2014a, 2018a; Zhang, Rolls, Wang, Xie, Cheng and Feng, 2023a). These types of natural variation may be important foundations for different types of personality (Rolls, 2014a, 2012e, 2023b).

18.5.6 Implications, and possible treatments, and subtypes of depression

The differences of functional connectivity described in this section (18.5) all relate to altered stability of the circuits involved, which help to account for some of the symptoms of depression. I now consider some of the implications of these advances (Rolls, Cheng and Feng, 2020a; Zhang, Rolls, Wang, Xie, Cheng and Feng, 2023a).

Because the lateral orbitofrontal cortex responds to many punishing and non-rewarding stimuli (Grabenhorst and Rolls, 2011; Rolls, 2014a,b, 2019e; Zhang et al., 2023a; Rolls, 2023b) that are likely to elicit autonomic/visceral responses, as does the supracallosal anterior cingulate cortex, and in view of connections from these areas to the anterior insula which is implicated in autonomic/visceral function (Critchley and Harrison, 2013; Rolls, 2016c), the anterior insula would also be expected to be overactive in depression, which it is (Drevets, 2007; Hamilton et al., 2013; Ma, 2015).

A notable discovery is that ketamine, a N-methyl-D-aspartate (NMDA) receptor antagonist, in subanaesthetic doses, produces rapid (within hours) antidepressant responses in patients who are resistant to typical antidepressants, and that the effects may last for two weeks or longer (Zanos and Gould, 2018; Iadarola et al., 2015; Maltbie et al., 2017; Zanos et al., 2016; Wilkinson et al., 2018; Rhee et al., 2022). Clinically, ketamine may be useful with a single dose, or doses may be repeated (Rhee et al., 2022). The short-term effects of ketamine include blocking excitatory NMDA receptors on cortical pyramidal cells which reduces the excitatory effect produced by the excitatory transmitter glutamate; and blocking excitatory receptors on GABA inhibitory neurons, which will tend to decrease GABAergic neuron firing, resulting in a potential increase in pyramidal cell firing. However, ketamine produces further effects, such as inducing synaptogenesis on excitatory neurons, increased glutamate transmission, reversing the synaptic deficits caused by chronic stress, and effects mediated by a ketamine metabolite hydroxynorketamine (Duman and Aghajanian, 2012; Ghasemi et al., 2017; Zorumski et al., 2016; Abdallah et al., 2016; Aleksandrova et al., 2017; Zanos and Gould, 2018; Zhang et al., 2023a). Another way in which ketamine may be effective in depression is by reducing inflammatory processes, which are sometimes related to depression (Ghasemi et al., 2017).

Ketamine may act in part by quashing the attractor state in the lateral orbitofrontal cortex at least for a number of days. Evidence consistent with this is that the activity (metabolism) of the inferior frontal gyrus and orbitofrontal cortex was decreased by a single dose of ketamine, which reduced the anhedonia of depression (see further below) (Lally et al., 2015). This NMDA receptor blocker may act at least in part by decreasing the high firing rate state of attractor networks by reducing transmission in the recurrent collateral excitatory connections between the neurons (Rolls, 2016b; Rolls, Loh, Deco and Winterer, 2008d; Rolls and Deco, 2010; Rolls, 2012c; Deco, Rolls, Albantakis and Romo, 2013; Rolls and Deco, 2015b). Treatment with conventional antidepressant drugs (e.g. SSRIs) decreases the activity and connectivity of this lateral orbitofrontal cortex system (Ma, 2015; Rolls et al., 2020a; Rolls, 2018a; Rolls et al., 2020a). Electroconvulsive therapy may have antidepressant effects, may and might knock the non-reward system out of its attractor state.

Given that a ketamine metabolite, hydroxynorketamine, may be related to the antidepressant effects of ketamine and may act via facilitating effects mediated by AMPA receptors (Zanos et al., 2016), the effects of ketamine might be mediated by increasing the activity of the medial orbitofrontal cortex reward-related system (which tends to be reciprocally inversely related to the lateral orbitofrontal cortex non-reward system), or the functional connectivity of the medial orbitofrontal cortex reward system with the hippocampal system which is reduced in depression (Fig. 18.13). There is now some evidence for an action of ketamine to

increase activity or connectivity in the medial orbitofrontal cortex, as follows (Zhang, Rolls, Wang, Xie, Cheng and Feng, 2023a).

A pharmacological investigation tested 10 treatment resistant depression (TRD) patients with a monetary incentive delay task before and after ketamine infusions. The study showed that ketamine administration induced a sustained increase in medial orbitofrontal cortex and nucleus accumbens activation during both the anticipation of positive reward cues and during the receipt of positive feedback (the outcome phase). The enhanced reactivity of the medial orbitofrontal cortex and nucleus accumbens was accompanied by decreased depression symptoms and better behavioral performance to positive items (Sterpenich et al., 2019). Consistent with this, in one of the first clinical studies to investigate the anti-anhedonic mechanisms of ketamine in depression, 52 TRD patients received a single ketamine infusion. The primary outcome indicated that ketamine rapidly reduced levels of anhedonia, with a substantial effect within 40 minutes which lasted up to 3 days post-infusion (Lally et al., 2015). Importantly, in a subgroup of patients who received PET scans, the single ketamine infusion decreased glucose metabolism within medial orbitofrontal cortex regions, and the reduced magnitude of the orbitofrontal cortex metabolism was positively correlated with the changes in the anhedonia symptoms (i.e., changes in SHAPS scores) (Lally et al., 2015). In addition, a double-blind, placebo-controlled, crossover trial in TRD and healthy controls also showed that ketamine infusion increased the functional connectivity between the right medial orbitofrontal cortex and ventral rostral putamen for the TRD patients toward the levels observed in healthy controls, and the increased magnitude of orbitofrontal cortex-putamen connectivity was correlated with improved levels of anhedonia symptoms (i.e., reductions in SHAPS scores) (Mkrchian, Evans, Kraus, Yuan, Kadriu, Nugent, Roiser and Zarate, 2021). Taken together, these findings provide evidence that ketamine has acute and chronic anti-anhedonic effects in major depressive disorder by acting in part to restore activity and sensitivity to reward, and increasing functional connectivity, of the medial orbitofrontal cortex.

A very interesting possibility arises. Ketamine may act at least in part to increase sensitivity to reward and connectivity of the medial orbitofrontal cortex, and thereby ameliorate the anhedonia and motivational symptoms of depression (Zhang, Rolls, Wang, Xie, Cheng and Feng, 2023a). Conventional antidepressants such as the selective serotonin reuptake inhibitors may act in part on the non-reward-related lateral orbitofrontal cortex to decrease its elevated sensitivity to non-reward and connectivity in depression, and thereby treat the symptoms of sadness related to non-reward over-efficacy (Rolls et al., 2020a; Zhang et al., 2023a). It is suggested that low doses of both types of treatment would be useful to explore.

Deep brain stimulation of the orbitofrontal cortex may be useful in the treatment of mood disorders and depression. The macaque orbitofrontal cortex is a key brain site at which deep brain electrical stimulation is rewarding (Rolls et al., 1980b; Rolls, 2005a, 2019e). Electrical stimulation of the human orbitofrontal cortex can also produce reward and raise mood (Rao et al., 2018), and many of the sites were in the middle part of the orbitofrontal cortex, areas 13 and 11, which are categorised as medial orbitofrontal cortex, the area activated by rewards (Rolls, 2019e). It is likely that these medial orbitofrontal cortex sites will produce better reward in humans than stimulation in the lateral orbitofrontal cortex BA12/47, for these lateral sites are activated by unpleasant stimuli and by not obtaining expected rewards. The medial (/ middle) orbitofrontal cortex may for the reasons described here and elsewhere (Rolls, 2019e) be a key area of interest for deep brain stimulation to help relieve depression (Rolls et al., 2020a).

The anterior cingulate cortex, including the subcallosal cingulate cortex, is a key brain region to which the orbitofrontal cortex projects (Hsu et al., 2020; Du et al., 2020; Rolls et al., 2019; Rolls, 2019e; Rolls et al., 2023c). It is possible that brain stimulation of the subcallosal cingulate cortex might be useful in the treatment of at least some patients with

depression (Riva-Posse et al., 2018; Holtzheimer et al., 2017; Dunlop et al., 2017; Lujan et al., 2013; Johansen-Berg et al., 2008), and it is possible that the subcallosal cingulate stimulation affects pathways that connect with the orbitofrontal cortex (Riva-Posse et al., 2018; Dunlop et al., 2017; Lujan et al., 2013; Johansen-Berg et al., 2008; Elias et al., 2022b,a). Given that the anterior cingulate cortex is an output region of the orbitofrontal cortex, it may be that treatments of the orbitofrontal cortex, where emotion-related processing is implemented, may be a better target for potential treatments for depression (Rolls et al., 2020a).

This non-reward / punishment attractor network sensitivity theory of depression has implications for treatments. These implications can be understood and further explored in the context of investigations of the factors that influence the stability of attractor neuronal networks with integrate-and-fire neurons with noise introduced by the close to Poisson spiking times of the neurons (Section 18.2) (Wang, 2002; Rolls, 2016b; Deco et al., 2009; Rolls and Deco, 2010; Deco et al., 2013; Loh et al., 2007a; Rolls and Deco, 2015b; Rolls, 2016b).

One implication is that antianxiety drugs, by increasing inhibition, might reduce the stability of the high firing rate state of the non-reward attractor, thus acting to quash the depression-related attractor state.

A second implication is that it might be possible to produce agents that decrease the efficacy of NMDA receptors in the lateral orbitofrontal cortex, thereby reducing the stability of the depression-related attractor state. The evidence that there are genes that are selective for NMDA receptors for the neurons in different populations is that there are separate knock-outs for NMDA receptors in the CA3 and CA1 regions of the hippocampus (Nakazawa et al., 2002; Tonegawa et al., 2003; Nakazawa et al., 2003, 2004). The present theory suggests that searching for ways to influence the attractor networks in the lateral orbitofrontal cortex by decreasing excitatory or increasing inhibitory transmission in this region may be of considerable interest. It should be noted that the present theory is a theory specifically of non-reward and punishment-related attractor networks in the lateral orbitofrontal cortex and related areas in relation to depression, and of reduced efficacy of the medial orbitofrontal cortex, and that alterations of attractor networks in other cortical areas may be related to other psychiatric disorders (Rolls, 2012c).

In terms of the implications of the attractor-based aspect of the present theory, an important point is that the attractor dynamics must be kept stable in the face of the randomness or noise introduced into the system by the almost Poisson firing times of neurons for a given mean firing rate. Moreover, the spontaneous firing rate state of the non-reward attractor must be maintained stable when no non-reward inputs are present (or otherwise the non-reward attractor would jump into a high firing rate non-reward state for no external reason, contributing to depression). The inhibitory transmitter GABA may be important in maintaining this type of stability (Rolls and Deco, 2010). Moreover, the high firing rate state produced by non-reward must not reach too high a firing rate, as this would cause overstability of the non-reward / depression state. In a complementary way, if the high firing rate attractor state is insufficiently high, then that attractor state might be unstable, and the individual might be relatively insensitive to non-reward, not depressed, and impulsive because of not responding sufficiently to non-reward or punishment. The excitatory transmitter glutamate acting at NMDA or AMPA receptors may be important in setting the stability of the high firing rate attractor state. In this respect and in this sense, the tendency to become depressed or to be impulsive may be reciprocally related to each other. Predictions for treatments follow from understanding these noisy attractor-based dynamics (Rolls and Deco, 2010; Rolls, 2012c).

There is growing interest in possible subtypes of depression, for it may be possible to treat different subtypes differently, for example by targeting different brain systems, or by different types of cognitive therapy (Drysdale et al., 2017). In one investigation, the subtypes appeared to reflect different combinations of depression and anxiety (Drysdale et al., 2017).