Modelling the Effect of Dorsal Raphe Serotonin Neurons on Patience for Future Rewards

Marc Sutherland and Bernd Porr(✉)

University of Glasgow, University Avenue, Glasgow, Scotland
Bernd.Porr@glasgow.ac.uk

Abstract. Serotonin is a neurotransmitter that is implicated in many basic human functions and behaviours and is closely associated with happiness, depression and reward processing. In particular it appears to be involved in suppressing responses to distracting stimuli while waiting for a delayed reward. Here we present a system level model of the limbic system which is able to generate a serotonin (5-hydroxytryptamine [5HT]) signal so that a simulated animal waits for a delayed reward. We propose that the 5HT signal is computed by a network involving the medial Orbital Frontal Cortex (mOFC), medial Pre Frontal Cortex (mPFC), Dorsal Raphe Nucleus (DRN)and the Nucleus Accumbens Core (NAcc). The serotonin signal encodes pre-reward liking, motivation throughout the trial and delayed reward waiting. We have successfully replicated the behaviour and dynamics of laboratory studies. With the help of this model we can predict that low levels of serotonin indirectly cause less encountered rewards because the animal gives up too early.

Keywords: Serotonin · Dopamine · Reward · Inhibition · Waiting

1 Introduction

The neurotransmitter serotonin is considered to be involved in the regulation of a number of behaviours principally involving aggression, aversive learning, impulsivity, attention, decision making, and reward [2]. It is implicated in many psychiatric disorders including depression, panic attacks, anxiety and obsessive compulsions [3]. In spite of serotonin's implication in a wide array of fundamental behaviours, the explicit circuitry that regulates serotonin producing neurons continues to be insufficiently understood [4]. And despite a considerable amount of research, the challenge of creating a unified theory of serotonin function persists [5].

We present a biologically inspired, systems level model which combines dopamine and serotonin networks to actualize learning and reversal learning in a simulated reward seeking task. Higher levels of serotonin allow the agent to remain at the reward site long enough to receive the reward without being distracted by competing attractions. Lower levels of serotonin mean that the agent does not wait long enough for the reward if it is delayed, and thus receives less rewards [1].

2 Task and Simulated Agent

The model is tested on a food and water seeking task based on the experiment conducted by Miyazaki [1]. A straightforward scenario is used in which an agent has a choice of two potential reward sites, see Fig. 1. The reward is only present at one site, see Fig. 1 "cake". The animal is simulated to move based on Braitenberg behaviour, calculated from distance from left and right eye to the Conditioned Stimulus (CS). At the beginning of the task the agent wanders around the designated area and approaches the different sites by chance. Once the agent has learned where the reward is located it will consistently go to that reward site. If the reward is toggled to the other site, reversal learning will lead the agent to eventually return to the haphazard wandering stage, until it happens on the reward again by chance. The reward is not presented at the site immediately in all cases, the agent may have to wait until it is delivered.

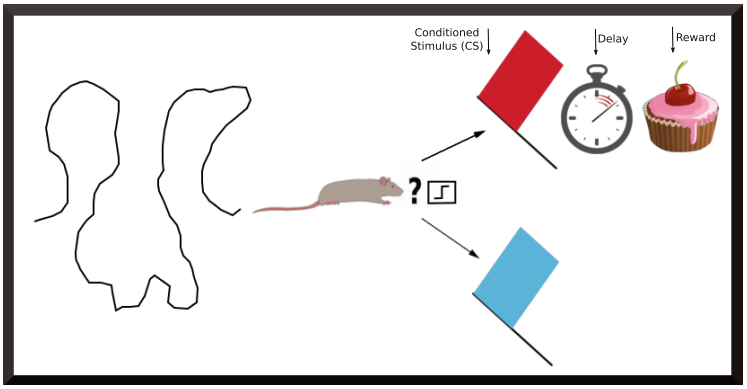

Fig. 1. Simulation environment

Mouse Simulation Environment: At the beginning of the task, approach behaviour is governed by a proximal signal which is determined by the angle of sight and distance between the rat and a red or blue flag. As the rat learns which flag harbours the reward, the flags become conditioned stimuli and approach behaviour is governed by a distal signal when the rat has the flag in its line of vision. When the rat is in position to receive the reward it must wait until the reward is delivered. A learning weight related to reward delivery increases as the reward is delivered consistently and diminishes as the reward is omitted. If the reward weight falls back to zero, the agent returns to its wandering activity and approach behaviour is again governed by a proximal signal.

3 The Role of Serotonin

The DRN signal in our model is composed of three main aspects. First, serotonin neurons signal pre-reward motivation to access the reward associated with

a CS. Nakamura argues that the DRN signals reward value associated with current behaviour and that the reaction to CS signals motivation to access the received reward [21]. Bromberg-Martin et al.'s 2010 study found that DRN neurons systematically encoded behaviour tasks in terms of their capacity to provide future rewards [22]. A view supported by Homberg who argues that there is some evidence that 5HT could signal the possibility of future reward [18] (see Fig. 2A). Secondly, serotonin also signals reward receipt. Nakamura's posits that the reaction to the received reward demonstrates appreciation with some neurons exhibiting a preference for large rewards while other neurons exhibit a preference for smaller rewards [21,25] (see Fig. 2C). Lastly specific DRN neurons fire when the agent is in position, waiting to receive the reward (see Fig. 2B). These neurons fire until the completion of the task, when the reward is finally presented to the subject. If the level of neuronal firing diminishes before the agent has received the reward, the agent will leave the site to start a new search [1]. Therefore, the agent's ability to wait is dependent on its serotonin level. Lower firing rates mean that the agent will move away before the reward is presented.

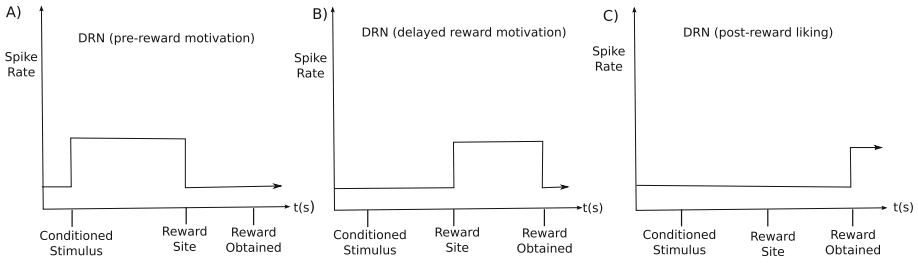

Fig. 2. DRN serotonin signals

4 Model Description

Limbic System Model: The model consists of a serotonin pathway which encodes pre-reward liking, motivation throughout the trial and delayed reward waiting. Dopamine (DA) pathways have also been created according to the standard model of DA action. These consist of reward, reward prediction and reward omission pathways which are capable of generating a dopamine liking signal to promote reward seeking action and also a dopamine reward prediction error.

Abbreviations: l-OFC - lateral Orbital Frontal Cortex, m-OFC - medial Orbital Frontal Cortex, m-PFC - medial Pre-Frontal Cortex, DRN - Dorsal Raphe Nucleus, l-shell - lateral shell of the Nucleus Accumbens, m-shell - medial shell of the Nucleus Accumbens, core - core of the Nucleus Accumbens, dl-VP - dorso-lateral Ventral Pallidum, m-VP - medial Ventral Pallidum, EP - Entopenduncular Nucleus, LHb - Lateral Habenula, RMTg - Rostral Medial Tegmental Nucleus, LH - Lateral Hypothalamus, VTA - Ventral Tegmental Area.

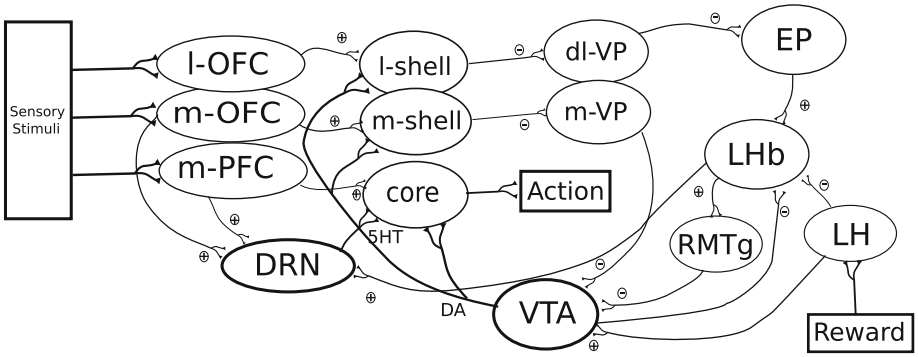

Fig. 3. Limbic system model

4.1 Reward Circuit

The dopamine circuits presented below are based on the standard prediction error paradigm by Schultz, Montague and Dayan [28]. The reward pathway is activated when the agent receives a primary reward (see Fig. 4A). This is modelled as beginning from the Lateral Hypothalamus (LH) which sends a strong inhibitory projection to the Lateral Habenula and an excitatory projection to the Ventral Tegmental Area (VTA) [6,7] (see Fig. 3).

4.2 Reward Prediction Circuit

The reward prediction circuit is activated when a conditioned stimulus (CS) associated with reward is observed (see Fig. 4B). The circuit creates a dopamine burst in response to the CS and suppresses the dopamine burst that would be

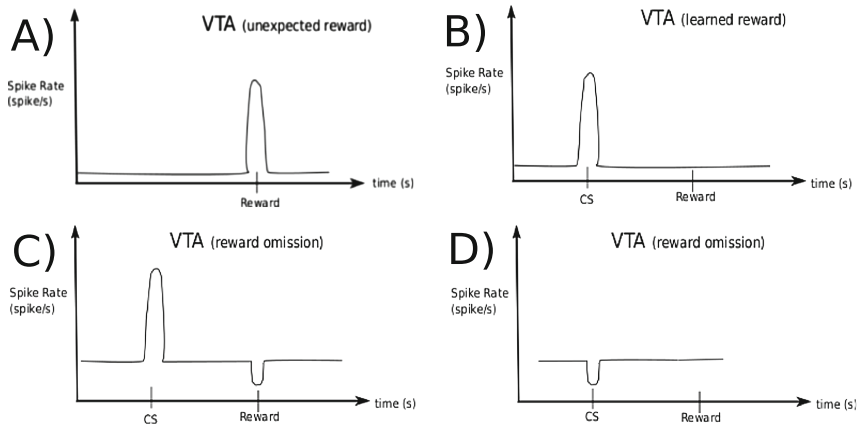

Fig. 4. Reward omission circuits

created by the LH upon reward receipt. The circuit starts at the mOFC, which generates persistent activity in response to each CS (see Fig. 1). An excitatory efferent from the m-OFC to the m-shell is modelled to undergo long term potentiation (LTP) if a dopamine burst coincides with the falling edge of activity on this connection [8]. If a dopamine dip coincides with the falling edge then long term depression (LTD) is modelled to take place. LTP allows the CS signal to pass through the inhibitory m-shell to the m-VP connection [9] and from there to the disinhibitory m-VP to VTA projection [7] (see Fig. 3). After sufficient LTP this will cause a DA burst when the CS is observed. Sustained CS activity also creates a sustained increase in GABAergic projections [10] which suppress DA bursts in the VTA due to reward receipt, as the reward is now predicted.

4.3 Reward Omission Circuit

This circuit is activated when a forecasted reward is omitted (see Fig. 4C) and also when a CS which is associated with reward omission is observed (see Fig. 4D). The signal starts from the lOFC, top left of the diagram, and comprises a range of signals depending on learning weight wlOFC, ranging from a short burst at learning weight zero to persistent activity at learning weight one. The appearance of a CS creates lOFC activity which projects to the l-shell [11] and is modelled to experience LTP due to a combination of the falling edge of the lOFC signal and a DA burst. This excludes activation of the circuit by a novel stimulus or a stimulus that has not previously led to a reward. When a CS that is associated with reward is presented, the signal is passed to the EP via the vl-VP by inhibition/disinhibition [9,12]. If no reward is delivered as expected, the VTA DA activity falls. The EP projects to the LHb [13,14] at the falling edge of the signal. The EP innervates the LHb which sends a glutamergic signal to the RMTg [15] which then inhibits the VTA [16], causing a dip in DA projections (see Fig. 3). The signal is also propagated when a CS that predicts omission is observed.

4.4 Serotonin Circuit

Serotonin is widely implicated in reward seeking behaviour. Nakamura et al.'s 2008 study of the primate dorsal raphe nucleus found that DRN responded tonically to both stimulus and reward and reliably encoded the value of the received reward, whether it was expected or not [21]. Based on Nakamura we proffer that the 5HT signal is computed by a network involving the medial Orbital Frontal Cortex (mOFC), medial Pre Frontal Cortex (mPFC), Dorsal Raphe Nucleus (DRN) and the Nucleus Accumbens Core (Core) (see Fig. 3). The lateral Orbital Frontal Cortex (lOFC) links specific stimuli to certain reward and failure results whereas the mOFC and mPFC are involved in appraising reward value, decision making, inhibition and choice across subsequent decisions [8]. Observance of the conditioned stimulus causes the signal to start at the lOFC and then transfer through the mOFC to the mPFC [17]. The mPFC in turn

innervates the DRN [18,19]. Finally the signal terminates at the NAcc Core, controlling actions [20].

5 Results

We first present the DA signal results and then the 5HT results.

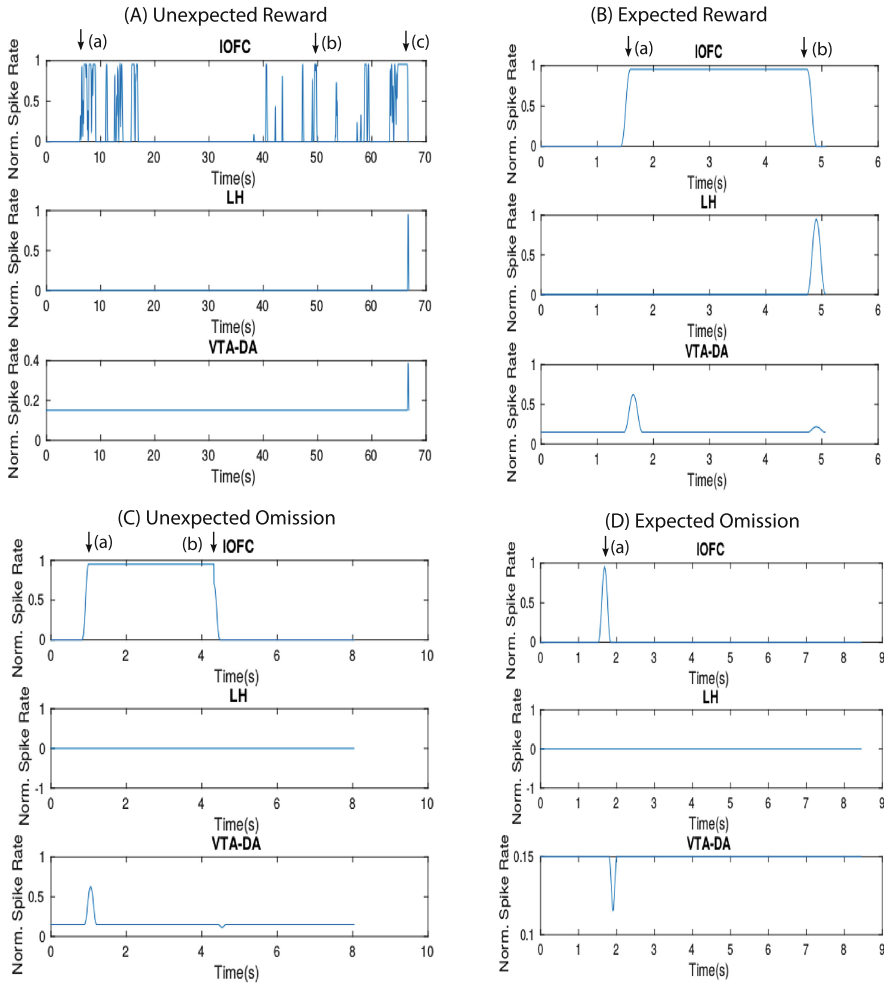

Fig. 5. Dopamine results

5.1 Dopamine Results

Unexpected Reward: At the start of the trial the agent wanders around the task area in a haphazard manner. In this wandering phase the stimulus comes

in and out of view. Viewing the stimulus causes the lOFC spike rate to increase, examples of which are points (a) and (b) in Fig. 5A. As the stimulus is not yet associated with reward the agent does not move towards it when it is viewed. Eventually the agent's wandering leads it by chance to the stimulus where it finds there is a reward. At point (c) in Fig. 5A LH spikes at reward presentation causing VTA DA to also increase it's spike rate as presented in Fig. 4A.

Expected Reward: At the start of the task the agent begins its wandering phase, to examine the site. At point (a) in Fig. 5B there is visual onset of the Conditioned Stimulus. As the CS is at this point associated with reward VTA DA increases its spike rate and the agent moves towards the CS to obtain the reward. At point (b) the agent receives the reward causing LH to increase its spiking rate. VTA GABA inhibits the release of VTA DA at reward presentation, hence the rate of spiking has diminished at this point as presented in Fig. 4B.

Unexpected Omission: The agent has now learned to associate the Conditioned Stimulus with reward. After the initial wandering stage, at point (a) in Fig. 5C, it sees the CS and moves towards it. At point (b) it arrives at the CS and when that reward is unexpectedly omitted, it causes a dip in the base firing rate of the VTA DA as displayed in Fig. 4C.

Expected Omission: The agent has now learned to associate the Conditioned Stimulus with an omission of reward. At the start of the task there is the initial wandering stage, then at point (a) in Fig. 5D, the agent sees the CS. This causes a dip in the VTA DA spiking rate, as presented in Fig. 4D.

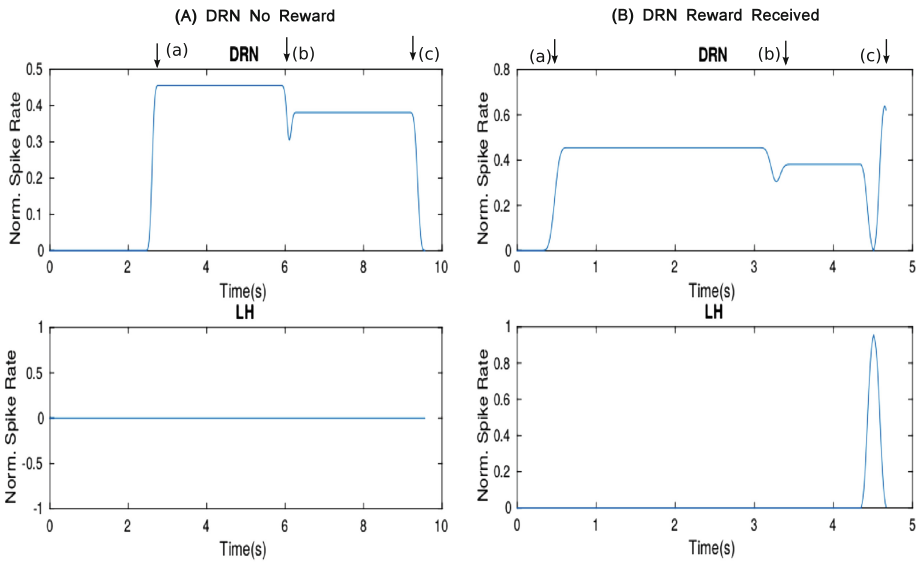

Fig. 6. DRN results

5.2 DRN Results

DRN No Reward: The task starts with the agent wandering around the task area. At point(a) in Fig. 6A the agent sees the conditioned stimulus and moves towards it. The DRN neurons increase spiking related to pre-reward motivation to achieve a reward (see Fig. 2A). At point (b) the agent is at the reward site. DRN spiking is maintained to inhibit leaving the site before the reward is delivered (see Fig. 2B). Eventually spiking reduces and as no reward has been provided, the agent leaves the site to begin a new reward search, point (c).

DRN Reward Received: The task starts with the agent wandering around the task area. At point (a) in Fig. 6B the agent sees the conditioned stimulus and moves towards it. DRN neurons increase spiking related to pre-reward liking (see Fig. 2A). At point (b) the agent is at the reward site. DRN spiking is maintained to inhibit leaving the site before the reward is delivered (see Fig. 2B). When the reward is delivered at point (c), the neurons which maintained delayed reward waiting cease firing and an alternative set of serotonin neurons fire, signalling post-reward liking (see Fig. 2C) [1].

6 Discussion

The work we present here builds on the limbic model inspired by the seminal work of Papez, Yakovlev and MacLean. Current dominant models of serotonin function focus on reward seeking behaviour, inhibition, perseveration and the processing of aversive cues. Nakamura argues that the DRN signals reward value associated with current behaviour and that the reaction to CS signals motivation to access the reward and the reaction to the received reward demonstrates appreciation [21,25]. Dayan's model proposes that a reduction in 5HT leads to behavioural disinhibition which is interconnected to an increased sensitivity to aversive cues and large negative prediction errors [2]. Cools asserts that serotonin has the opposite function to dopamine in that it deals with aversive cues and inhibiting behaviour [27]. Seymour argues that the depletion of serotonin produces perseverative responding, inducing the agent to persistently respond to a previously rewarding stimulus that offer diminishing returns, no returns or even negative outcomes [29].

The novel model of serotonin function we propose builds on the models above. We have shown that serotonin plays a critical role in reward seeking activity; enhanced spiking rates signal motivation to achieve a reward, motivation to wait for a reward and also appreciation of the achieved reward. Lower levels of serotonin in the agent would mean less motivation, less patience and lower reward appreciation, therefore less rewards, which would have a significant impact on the agent's well-being and mood. Bromberg-Martin et al.'s 2010 study asserts that serotonin controls motivation and reward seeking. High levels of serotonin led the case studies to wait for larger delayed rewards. Lower levels made the monkeys impulsively choose the smaller more readily available reward [22]. Higher 5HT helps the agent to stay focussed and become less distracted, allowing them to

exploit a resource rather than set off to explore before it has been fully exploited [18]. Robinson et al. state that different subtypes of serotonin receptor control varying forms of impulsive behaviour. They claim there are at least 15 subtypes. Their study found that a reduction of forebrain 5HT led to impulsive responding in rats and argue that their findings add to a growing body of evidence for multiple neurotransmitter systems that regulate impulsive behaviour which includes serotonin, dopamine, noradrenergic and histaminergic systems [26]. In a 2011 paper Cools et al. discuss the fact that the depletion of 5HT is characterised by both impulsive behaviour and depression. They consider this fact incongruous as depression is associated with reduced behavioural vitality. They posit that 5HT's link to depression may be indirect and caused by associative learning or the disinhibition of unpleasant thoughts [27]. It would appear that serotonin does have an indirect effect on depression as anti-depressants that target depression only start to affect mood after a considerable length of time. One could also argue that increased impulsivity could lead to a negative mood if the impulsive behaviour was having negative outcomes.

7 Model Equations

The following equations show how the signals are generated in different sections of the limbic system model.

G(t) is a Gaussian filter which smooth out transitions in the raw signals.

$$LH(t) = G(t) * rewardSet(t) \tag{1}$$

$$VTA_{DA}(t) = \begin{cases} \frac{LH(t)}{VTA_{GABA}(t)} + EP(t) + G(t) * reward_{LTP} * w_{mshell}(t), & mOFC_{diff}(t) > 0 \\ \frac{LH(t)}{VTA_{GABA}(t)} + EP(t), & mOFC_{diff}(t) < 0 \end{cases} \tag{2}$$

reward_{LTP} denotes when the long term potentiation in the reward prediction circuit has reached a level significant enough to allow the signal to pass through the circuit.

w_{mshell} is the learning weight associated with reward prediction.

$\text{mOFC}_{\text{diff}} > 0$ denotes a rising edge of the mOFC signal.

$\text{mOFC}_{\text{diff}} < 0$ denotes a falling edge of the mOFC signal.

$$EP(t) = \begin{cases} G(t) * -0.2 * omission_{LTP}(t) * LH_{inhibition}(t), & lOFC_{diff}(t) < 0 \& lOFC(t) < 0.1 \\ 0, & lOFC_{diff}(t) > 0 \parallel lOFC(t) > 0.1 \end{cases} \tag{3}$$

omission_{LTP} denotes when the long term potentiation in the reward omission circuit has reached a level significant enough to allow the signal to pass through the circuit.

$\text{lOFC}_{diff} > 0$ denotes a rising edge of the lOFC signal.

$\text{lOFC}_{diff} < 0$ denotes a falling edge of the lOFC signal.

$$lOFC(t) = lOFC_{pa}(t) * w_{lOFC}(t) * G(t)(lOFC_burst(t) * (1 - w_{lOFC}(t)) \tag{4}$$

The lOFC signal is a combination of a permanent lOFC signal and a lOFC burst signal, controlled by an omission learning weight.

$$mOFC(t) = delay(lOFC_{pa}(t)) \tag{5}$$

The mOFC signal originates in the lOFC and is therefore a delayed version of that signal.

$$mPFC(t) = \begin{cases} G(t) * 0.5 + LH, & mOFC_{diff}(t) > 0 \\ LH(t), & mOFC_{diff}(t) < 0 \end{cases} \tag{6}$$

$$DRN(t) = G(t) * (mOFC(t) + mPFC(t)) \tag{7}$$

References

1. Miyazaki, K., Miyazaki, K.W., Doya, K.: Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. J. Neurosci. **31**(2), 469–479 (2011)
2. Dayan, P., Huys, Q.J.M.: Serotonin, inhibition, and negative mood. PLoS Comput. Biol. **4**(2), e4 (2008)
3. Dorocic, I.P., Furth, D., Xuan, Y., Johansson, Y., Pozzi, L., Silberberg, G., Carlen, M., Meletis, K.: A whole-brain atlas of inputs to serotonergic neurons of the dorsal and median raphe nuclei. Neuron **83**, 663–678 (2014)
4. Sparta, D.R., Stuber, G.D.: Cartography of serotonergic circuits. Neuron **83**(3), 513–515 (2014)
5. Ranade, S.P., Mainen, Z.F.: Transient firing of dorsal raphe neurons encodes diverse and specific sensory, motor, and reward events. J. Neurophysiol. **102**(5), 3026–3037 (2009)
6. Nakamura, K., Ono, T.: Lateral hypothalamus neuron involvement in integration of natural and artificial rewards and cue signals. J. Neurophysiol. **55**(1), 163–181 (1986)
7. Sesack, S.R., Grace, A.A.: Cortico-basal ganglia reward network: microcircuitry. Neuropsychopharmacology **35**(1), 27–47 (2010). Official publication of the American College of Neuropsychopharmacology
8. Noonan, M.P., Kolling, N., Walton, M.E., Rushworth, M.F.S.: Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. Eur. J. Neurosci. **35**(7), 997–1010 (2012)
9. Humphries, M.D., Prescott, T.J.: The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. Progress Neurobiol. **90**(4), 385–417 (2010)
10. Van Bockstaele, E.J., Pickel, V.M.: GABA-containing neurons in the ventral tegmental area project to the nucleus accumbens in rat brain. Brain Res. **682**(1–2), 215–221 (1995)
11. Brog, J.S., Salyapongse, A., Deutch, A.Y., Zahm, D.S.: The patterns of afferent innervation of the core and shell in the 'accumbens' part of the rat ventral striatum. J. Comp. Neurol. **338**, 255–278 (1993)
12. Basar, K., Sesia, T., Groenewegen, H., Steinbusch, H.W.M., Visser-Vandewalle, V., Temel, Y.: Nucleus accumbens and impulsivity. Progress Neurobiol. **92**(4), 533–557 (2010)

13. Rajakumar, N., Elisevich, K., Flumerfelt, B.A.: Compartmental origin of the striato-entopeduncular projection in the rat. J. Comp. Neurol. **331**(2), 286–296 (1993)
14. Hong, S., Hikosaka, O.: The globus pallidus sends reward-related signals to the lateral habenula. Neuron **60**(4), 720–729 (2008)
15. Barrot, M., Sesack, S.R., Georges, F., Pistis, M., Hong, S., Jhou, T.C.: Braking dopamine systems: a new GABA master structure for mesolimbic and nigrostriatal functions. J. Neurosci. **32**(41), 14094–14101 (2012). The Official Journal of the Society for Neuroscience
16. Bourdy, R., Barrot, M.: A new control center for dopaminergic systems: Pulling the VTA by the tail. Trends Neurosci. **35**(11), 681–688 (2012)
17. Hoover, W.B., Vertes, R.P.: Projections of the medial orbital and ventral orbital cortex in the rat. J. Comp. Neurol. **519**(18), 3766–3801 (2011)
18. Homberg, J.R.: Serotonin and decision making processes. Neurosci. Biobehav. Rev. **36**(1), 218–236 (2012)
19. Juckel, G., Mendlin, A., Jacobs, B.L.: Electrical stimulation of rat medial prefrontal cortex enhances forebrain serotonin output: implications for electroconvulsive therapy and transcranial magnetic stimulation in depression. Neuropsychopharmacology **21**(3), 391–398 (1999)
20. Vertes, R.P., Linley, S.B.: Efferent, afferent connections of the dorsal, median raphe nuclei in the rat. In: Monti, J.M., Pandi-Perumal, S.R., Jacobs, B.L., Nutt, D.J. (eds.) Serotonin, Sleep: Molecular, Functional and Clinical Aspects. Birkhäuser, Switzerland (2008)
21. Nakamura, K., Matsumoto, M., Hikosaka, O.: Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. J. Neurosci. **28**(20), 5331–5343 (2008). The Official Journal of the Society for Neuroscience
22. Bromberg-Martin, E.S., Hikosaka, O., Nakamura, K.: Coding of task reward value in the dorsal raphe nucleus. J. Neurosci. **30**(18), 6262–6272 (2010). The Official Journal of the Society for Neuroscience
23. Cardinal, R.N.: Neural systems implicated in delayed and probabilistic reinforcement. Neural Netw. **19**, 1277–1301 (2006)
24. Evenden, J.: Impulsivity: a discussion of clinical and experimental findings. J. Psycopharmacology **13**(2), 180–192 (1999)
25. Nakamura, K.: The role of the doral raphe nucleus in reward-seeking behaviour. Front. Integr. Neurosci. 60 (2013)
26. Robinson, E.S.J., Dalley, J.W., Theobald, D.E.H., Glennon, J.C., Pezze, M.A., Murphy, E.R., Robbins, T.W.: Opposing roles for 5-HT2A and 5-HT2C receptors in the nucleus accumbens on inhibitory response control in the 5-choice serial reaction time task. Neuropsychopharmacology **33**(10), 27–47 (2010). Official Publication of the American College of Neuropsychopharmacology
27. Cools, R., Nakamura, K., Daw, N.D.: Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology **36**(1), 98–113 (2011). Official publication of the American College of Neuropsychopharmacology
28. Schulz, W., Dayan, P., Montague, P.R.: A neural substrate of prediction and reward. Science **275**, 1593–1599 (1997)
29. Seymour, B., Daw, N., Roiser, J., Dayan, P., Dolan, R.: Serotonin selectively modulates reward value in human decision-making. J. Neurosci. **32**(17), 5833–5842 (2012)