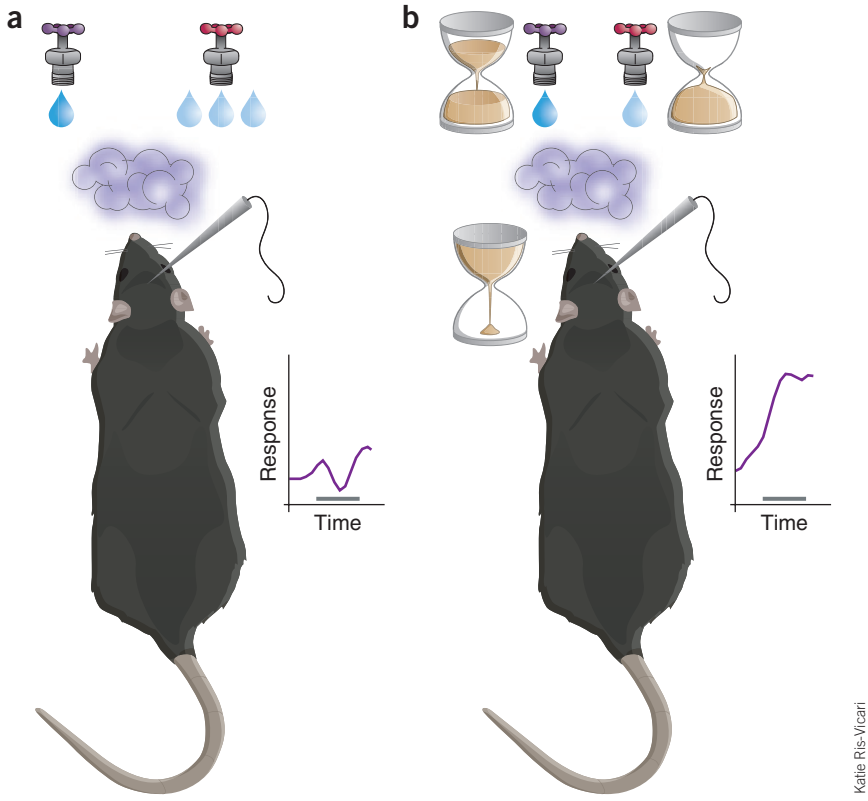
Dopamine: at the intersection of reward and action

Nathaniel D Daw

A new paper reports that dopaminergic neurons initially responded optimistically in rats given free choice between two rewards, as though the animal had chosen the better reward, even on trials when it failed to do so. These findings suggest that current computational theories of dopaminergic function may need to be revised.

How can we make sense of a brain system that is implicated not just in the reinforcing qualities of a fine wine, but also in the movements required to lift the glass? Both reward and motion are among the striking assortment of crucial functions that seem to depend on the neuromodulator dopamine. One influential attempt at uniting these observations[1,2] is the proposal that the system's role is actually in the link between the two: learning about how actions lead to reward (called 'reinforcement learning')[3]. By recording from dopaminergic neurons in rats learning to choose between responses that were rewarded with different amounts of sucrose or after different delays, Roesch *et al.*[4] have tested, and confirmed, a raft of important predictions of this theory. The results demonstrate the unifying nature of the hypothesis and also raise new puzzles for the modelers to chew on.

In each trial of the experiment, rats sampled one of three odor cues and then chose between sucrose wells to the left and right of the odor port. Two of the odors signaled 'forced choices' of one side or the other: if the rat chose the appropriate well, it was rewarded. Trials of this sort taught the rat about the rewards available at each port. This experience was used on other trials, when the third odor signaled a 'free choice' of either well to obtain its reward. In any particular block of trials, one well was better than the other; in successive blocks, the wells rotated between being associated with large versus small rewards or with small rewards delivered earlier versus later. (Lab animals, like humans, prefer their gratification sooner, a phenomenon known as temporal discounting.)



Katie Ris-Vicari

Figure 1 The response of dopamine neurons to a cue predicting the same reward depends on the alternative that might have been received. The presented odor cues a forced choice for one drop of sucrose after a short delay. (**a**) In blocks in which these trials were intermixed with an odor signaling a larger reward of three drops of sucrose, dopamine neurons showed little response. (**b**) In blocks in which the alternative was one drop of sucrose, but with a longer delay, dopamine neurons were strongly excited by the cue for the same reward. This modulation suggests that the response to the cue carries a prediction error that signals whether reward expectations after receiving the cue are better or worse than the average expectations preceding it.

In short, to choose the best rewards, the animals were required to relearn which well was richer. Computational modelers hypothesize that dopaminergic neurons drive just this kind of reinforcement learning by reporting a 'reward prediction error'[1,2,5] measuring the mismatch between predicted and obtained rewards. A classic example of such a prediction error is the finding that primate dopaminergic neurons can be excited or inhibited (or unaffected) by the receipt of an identical reward, depending on whether the reward is more or less than (or equal to) what had been expected[6].

The author is at the Center for Neural Science and Department of Psychology, New York University, 4 Washington Place, Room 809, New York, New York 10003, USA.
e-mail: daw@cns.nyu.edu

Most of the interesting issues about reward prediction concern anticipatory neural responses, here to odor cues signaling future reward. Dopaminergic neurons also respond to such cues; a simple, but deep, and unifying idea of the models is to view these responses as just another example of the same prediction error signal. Cues produce error by signaling that prospects for future reward are better or worse than had been expected. So, by providing information about future prospects, cues can serve, effectively, as substitutes for the reinforcers that they predict—driving prediction error and learning about the prospects of still earlier cues. The unexpected receipt of money, for instance, is not primary reinforcement, but it signals the future ability to purchase rewards and is a potent activator of the dopaminergic system[7]. Conversely, a server prematurely clearing your not-yet-empty glass of beer causes an unexpected decline in your future reward prospects.

Roesch *et al.*[4] discovered a surprisingly intuitive confirmation of this interpretation by comparing the response to a cue signaling a forced choice of a single drop of sucrose between different blocks (**Fig. 1**). The response was large in blocks when this was the better option: when the alternative was the sucrose after a longer delay. As both choices are forced unpredictably in such blocks, the animal's average expectation should be intermediate before sampling the cue. Because it signals that reward is nearer than previously expected, the cue should then be a pleasant surprise. In contrast, the response to the same cue for the same reward was smaller in blocks when it signaled the poorer alternative—that is, when the alternative was three drops. These findings mirror the modulation of the response to a primary reward by expectations about it, the hallmark of a prediction error.

The experiment also confirmed a related prediction about dopaminergic responses to cues; because animals treat delayed rewards as less valuable, cue responses should be modulated by the proximity of the reward that they predict. In blocks pitting short against long delays, the longer delay varied between 1 and 7 s. The dopaminergic response to the cue forcing the choice of this wait (an increasingly unwelcome surprise) fell off systematically as the delay increased.

Most important for understanding dopamine's suspected role in the interface between reward and action is how the neurons respond to free choices. A second unifying idea of prediction error models is that your possible actions are themselves analogous to cues, in that they also have consequences for future reward. Learning these consequences can help you to decide which action to choose. Furthermore, if chosen or even potential actions, like cues, signal changes in future expectations of reinforcement, then the resulting prediction errors can be used to learn about preceding actions and events. This would allow the system to tackle complex, sequential decision tasks, such as chess, in which decisions typically lead to yet more decisions.

But what, then, is the value of a free choice between two actions, when the decision may not yet be made? The headline result of Roesch *et al.*[4] is that neurons reacted optimistically to the free-choice cue, with a large excitation that was indistinguishable from that for the forced choice of the better outcome. Evidently, the system reports prediction error as though the chance to choose is worth the same as the better choice. Notably, it does this even though animals actually ended up choosing the worse option about 30% of the time. As the animals left the odor port, however, the responses changed to reflect the choice to come (that is, they declined sharply preceding choices of the worse option). The latter finding again points to the relationship between actions and cues; here, perhaps, the intention to choose a well furnishes a sort of internal cue providing new information to the dopaminergic neurons about future rewards. In other words, it seems as if the dynamics of the response may reflect the formation of the decision.

The initially optimistic valuation of the free choice cue, as equivalent to the better choice, is itself anticipated by computational models; indeed, a prominent reinforcement learning algorithm called Q-learning[8] works just this way (such a prediction error allows learning about prior events to proceed as if the better choice had been taken, even if the subject actually tries something else). This is the first time that this phenomenon has been demonstrated experimentally. Indeed, a similar experiment with monkeys[9] previously supported a different conclusion; there, dopaminergic neurons responded to a free choice by immediately signaling the value of whichever option would actually be chosen.

What could account for these different findings? Ignoring some potentially important differences in task design, this may relate to another puzzle, about the differential function of the separate subdivisions of the dopaminergic system that were recorded in the two studies. A standard view is that one division (the ventral tegmental area) targets brain areas that specialize in learning the values of cues, *per se*, whereas the other (substantia nigra pars compacta) targets neighboring areas more concerned with the analogous values of actions[10]. However, the usual reinforcement learning model that motivates this division of labor, called the actor/critic[11], does not quite correspond with either of the error signals reported, instead predicting yet a third variation[12]. It therefore falls on modelers to design a new functional breakdown to rationalize the difference in observed error signals or on experimentalists to trace the difference to some other aspect of the tasks.

Finally, with dopaminergic neurons now being recorded during complex, learned decision-making tasks of this sort, the next step in understanding their role in reward and action should be connecting their responses more closely to choice behavior. For instance, does dopaminergic error signaling drive measurable trial-by-trial changes in behavioral choice preferences? Does the decline in responses for cues predicting delayed rewards accurately predict the decline in the behavioral propensity to choose a delayed option? Using human functional neuroimaging, Kable *et al.*[13] (also in this issue) report a result suggesting the affirmative. They show that the blood oxygenation response in several brain areas (including key dopaminergic targets) tracks the decline in value of monetary rewards as they are offered at greater delays; across subjects, the steepness of this neural decline predicts individual differences in how patient or impatient subjects were in choosing between immediate and delayed rewards.

In these respects, the Roesch *et al.*[4] experiment is all the more promising for having been conducted with rats. Although basic mechanisms of learning are largely conserved between species, the humble rodent has historically been the subject for many of the most detailed behavioral studies of learning. This exquisite behavioral heritage makes the rat a very attractive species for dopaminergic recordings. This offsets the great downside of this species, which is that investigators have had a punishing time finding dopaminergic cells to record, perhaps due in part to the very small brain areas involved. Gruelingly, Roesch *et al.*[4] averaged about two prediction-error neurons per animal, a similar yield to that reported by another laboratory[14]. Hopefully, the technical expertise developed in these early attempts will ease future studies. Another, potentially complementary, approach is instead to measure dopaminergic release chemically at the neurons' targets, using fast-scan voltammetry in behaving rats[15]. In either case, the tight interplay between theoretical and experimental work exemplified by the Roesch *et al.* study[4] promises more progress in reconciling the many facets of dopamine.

1. Houk, J.C., Adams, J.L. & Barto, A.G. in *Models of Information Processing in the Basal Ganglia* (eds. Houk, J.C., Davis, J.L. & Beiser, D.G.) Ch. 13, 249–270 (MIT Press, Cambridge, Massachusetts, 1995).
2. Schultz, W., Dayan, P. & Montague, P.R. *Science* **275**, 1593–1599 (1997).
3. Sutton, R.S. & Barto, A.G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
4. Roesch, M.R., Calu, D.J., & Schoenbaum, G. *Nat. Neurosci.* **10**, 1615–1624 (2007).

5. Bayer, H.M. & Glimcher, P.W. *Neuron* **47**, 129–141 (2005).
6. Tobler, P.N., Fiorillo, C.D. & Schultz, W. *Science* **11**, 1642–1645 (2005).
7. Pessiglione, M. *et al. Nature* **442**, 1042–1045 (2006).
8. Watkins, C. *Learning from Delayed Rewards*. Thesis, Cambridge Univ. (1989).
9. Morris, G., Nevet, A., Arkadir, E. & Bergman, H. *Nat. Neurosci.* **9**, 1057–1063 (2006).
10. O'Doherty, J. *et al. Science* **304**, 452–454 (2004).
11. Joel, D., Niv, Y. & Rupin, E. *Neural Netw.* **15**, 535–547 (2002).
12. Niv, Y., Daw, N.D. & Dayan, P. *Nat. Neurosci.* **9**, 987–988 (2006).
13. Kable, J.W. & Glimcher, P.W. *Nat. Neurosci.* **10**, 1625–1633 (2007).
14. Pan, W.X., Schmidt, R., Wickens, J.R. & Hyland, B.I.J. *J. Neurosci.* **25**, 6235–6242 (2005).
15. Day, J.J., Roitman, M.F., Wightman, R.M. & Carelli, R.M. *Nat. Neurosci.* **10**, 1020–1028 (2007).

Microgliosis: the questions shape the answers

Richard M Ransohoff

Brain microglia increase their numbers in response to threats. Some of these cells were thought to enter the CNS from the blood, but two new studies suggest that experimental confounds could in part account for such results.

Microglia, the guard cells of the CNS, are the earliest sensors of all forms of pathological incursion. In response to pathogens or damage, they alter their morphology, surface phenotype and gene expression, and their numbers increase markedly, which is termed microgliosis. To clarify the origins of microgliosis, previous studies used bone-marrow chimeras produced by irradiation, which yielded answers with potential confounds, including alteration of blood-brain barrier function by the radiation. Two new reports in this issue, Ajami *et al.*[1] and Mildner *et al.*[2], revisit the mechanisms of microgliosis, using differing, ingenious experimental strategies that overlapped to produce a coherent body of data. These findings not only point toward exciting new directions in studying microglia and using these cells for therapeutic applications, but also prompt a re-evaluation of the published literature.

Irradiation bone-marrow chimerism involves killing bone-marrow cells with radiation and replacing them with labeled bone-marrow cells that are derived from another animal, allowing investigators to determine which cells in the recipient animal are derived from bone marrow. This technique was used to identify the CNS cells that expressed class II major histocompatibility complex (MHC) antigens[3]. Radiation chimerism was nicely adapted for studies of this type, as the polymorphic MHC antigens provide excellent differentiation between the host and donor cells. Early researchers noted that parenchymal microglia did not have the markers of transferred cells, but perivascular microglia (also called perivascular macrophages) did[4].

The technique was soon adapted to address key questions in CNS immunity and inflammation. The most widely used disease model was experimental autoimmune encephalomyelitis (EAE), which involves immunization of rodents with myelin protein fragments to mimic multiple sclerosis. In some studies, animals were immunized to generate myelin-reactive T cells, which were infused into naive recipients—an approach called adoptive transfer. A seminal study used irradiation chimerism and adoptive-transfer EAE to demonstrate that perivascular macrophages of the CNS could stimulate myelin-reactive T lymphocytes enough to cause demyelination. Follow-up studies showed that perivascular macrophages of the CNS are necessary and sufficient for adoptive transfer EAE[5–7].

Microglia have both protective and deleterious activities, which researchers have attempted to understand through extensive studies using irradiation chimerism. Parenchymal microglia retain the characteristics of the host, whereas perivascular macrophages are replaced by donor cells in these chimeras. Further studies, using cells marked with polymorphic genetic determinants or transgenic reporters[8–10], compared the responses of donor and host cells to varied insults. This approach was an attractive method for compensating for the lack of markers with which to differentiate blood-derived microglia-like cells from resident microglia. These studies suggested that microglia might be implicated in neurodegenerative diseases, such as Huntington's disease, Alzheimer's disease and amyotrophic lateral sclerosis. Chimeras between wild-type and knockout animals were used to address the functions of individual gene products, either in the blood and bone marrow or in the CNS, in neuroinflammatory or neurodegenerative disease models (**Fig. 1**). Given the utility of this approach and its extensive application, it is important to focus on the confounds that might be introduced by irradiation chimerism.

Ajami *et al.*[1] and Mildner *et al.*[2] provide a comprehensive dissection of several possible complications of the model (**Table 1**). First, the use of bone-marrow cells to reconstitute

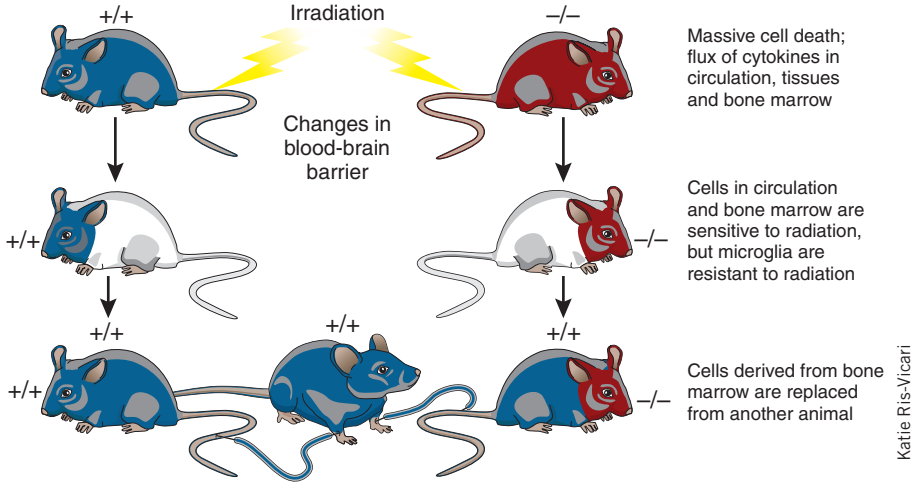


Figure 1 Irradiation chimerism can generate a mouse in which genetic characteristics of circulating microglial precursors differ from those of brain microglia (right mouse with +/+ circulating cells and –/– brain cells). A matched control that has undergone the chimerism procedure is also shown (left mouse with +/+ circulating cells and +/+ brain cells).

The author is in the Neuroinflammation Research Center, Department of Neurosciences, Lerner Research Institute, NC30, 9500 Euclid Avenue, Cleveland Clinic, Cleveland, Ohio 44195, USA.
e-mail: ransohr@ccf.org