HIPPOCAMPAL REPLAY

Spoiled for choice, pressed for time

A new theory derives the sequential nature of hippocampal replay from first principles and, moreover, predicts the specific patterns of replay that are actually observed in multiple different experiments.

John Widloski and David J. Foster

Hippocampal place cells, the subject of the 2014 Nobel Prize in Physiology, live a secret life. Originally thought to fire action potentials dutifully only within their place fields, they revel in periods of promiscuous propagation. When a rat pauses during exploration of a maze, its place cells, initially paused in their labor as faithful reporters of the animal's current location, suddenly come alive in bursts of activity, which zip sequentially from cell to cell up to 20 times faster than normal. One moment, a sequence depicts a series of places the animal is about to visit, as if rehearsing the journey. The next moment, a sequence travels just as swiftly backwards through the animal's past, as if ruminating over the choices it has made. These activity patterns in the rat's brain are commonly referred to as awake replay, and in this issue of *Nature Neuroscience*, Mattar and Daw present the first theoretical account of why replay patterns take the forms that they do[1].

For behavioral neurophysiologists, stumbling onto the phenomenology of awake replay has been like falling into a box of chocolates. After first discovering that awake replay goes backwards[2], and then that it can go forward as well as backwards[3], it was discovered that it can go both ways at a choice-point in a maze[4,5], and it can also join together different experiences to find shortcuts[4]. Awake replay contributes to decisions[6,7] and can depict the precise trajectory that the animal is about to take all the way to a remembered goal location[8]. It is also exquisitely sensitive to the learned shape of the maze the rat is running on[9], and when a rat discovers unexpected changes in reward, there are corresponding changes in the numbers of awake replays that get produced[10,11].

Can you have too much of a good thing? This assortment of results has exposed the absence of a theoretical framework to make sense of all the data. For example, the distinction between forward and backwards replay has been confusing, with some researchers ascribing them different roles and others preferring to ignore backwards replay altogether. Now, in work of extraordinary elegance, Mattar and Daw provide exactly the sort of theoretical framework that the field has been looking for. There are two major accomplishments. First, they derive replay from first principles, giving what is sometimes called a 'normative' account. That is, they start from the Darwinian injunction–eat and don't get eaten–and from there derive replay sequences as the optimal order in which to sample and learn from place memories to maximize future rewards and minimize future costs. Second, they demonstrate that their framework can account for almost all of the results discovered about replay in the last decade. Taken altogether, it is an astonishing achievement.

So how do they do it? They begin by defining the fundamental unit of experience as a movement between two neighboring locations, given an action choice at the first location and with a resulting outcome. Animals use such experiences to learn to improve the action choices they make, to increase the amounts of reward they will obtain in an environment. By using an algorithm called Q learning[12], they model the outcomes not just as immediate rewards or costs, but also including the long-term expectations about what returns will accrue in the future. Every time a unit of experience is used by the Q learning algorithm to make an incremental improvement in action choices, this is called a 'backup'. During behavior, the backups can be made from the actual moves the animal makes through the world. But backups can also be made offline, that is, when the animal is not actively experiencing the movements but rather replaying them while being paused somewhere else. Mattar and Daw are agnostic about whether offline backups are specific previous experiences recalled from memory or simulated experiences using an internal model of the world. The key question they ask is: given the short period of time that an animal pauses in a maze and the very large number of possible backups from all over the maze, which backups should be performed and in what order? Mattar and Daw construct a new quantity, the expected value of a backup (EVB), which is the increase in returns that a backup yields. Their premise is that backups with the highest EVBs should be made first. They then show that EVB factors into two components: gain and need. Gain depends on how much the backup changes actions in the backup location. Need is the probability that the backup location will ever actually be visited. Both factors are important: a backup that doesn't change actions is not worth doing, but neither is a backup relating to a situation that will never occur. Need spreads out in front of the animal (Fig. 1a), the influence of gain spreads out behind (Fig. 1b), and they lead to the performance of backups in sequences moving either ahead of or behind the animal, respectively.

Armed with this deceptively simple framework, Mattar and Daw take a victory march through a decade of replay results, demonstrating one by one that the particular pattern of backups in the optimal order matches the empirical data. Their approach also resolves several apparent inconsistencies in the empirical record. For example, when rats happen upon an unexpected reduction in reward, there is a decrease in the number of replays[11] (Fig. 1c). From the classic perspective on replay as a memory process, this is very surprising: why should an unexpected negative event be less memorable than an unexpected positive event? However, Mattar and Daw reach into the experimental details to find that in this experiment, the actions required of the animal did not change, even when the reward was removed, because the rat was required to keep visiting the unrewarded location in order to obtain chocolate later, at the opposite end of the track. Under these conditions, their model produces the reduction in replays as observed, but also makes a prediction: if the rat is allowed to change its actions in response to a negative event so as to avoid it, then there should be a large increase in gain associated with the new actions and therefore an increase in the numbers of replays. An experiment with this logic was recently performed, with exactly the predicted result[13] (Fig. 1d).

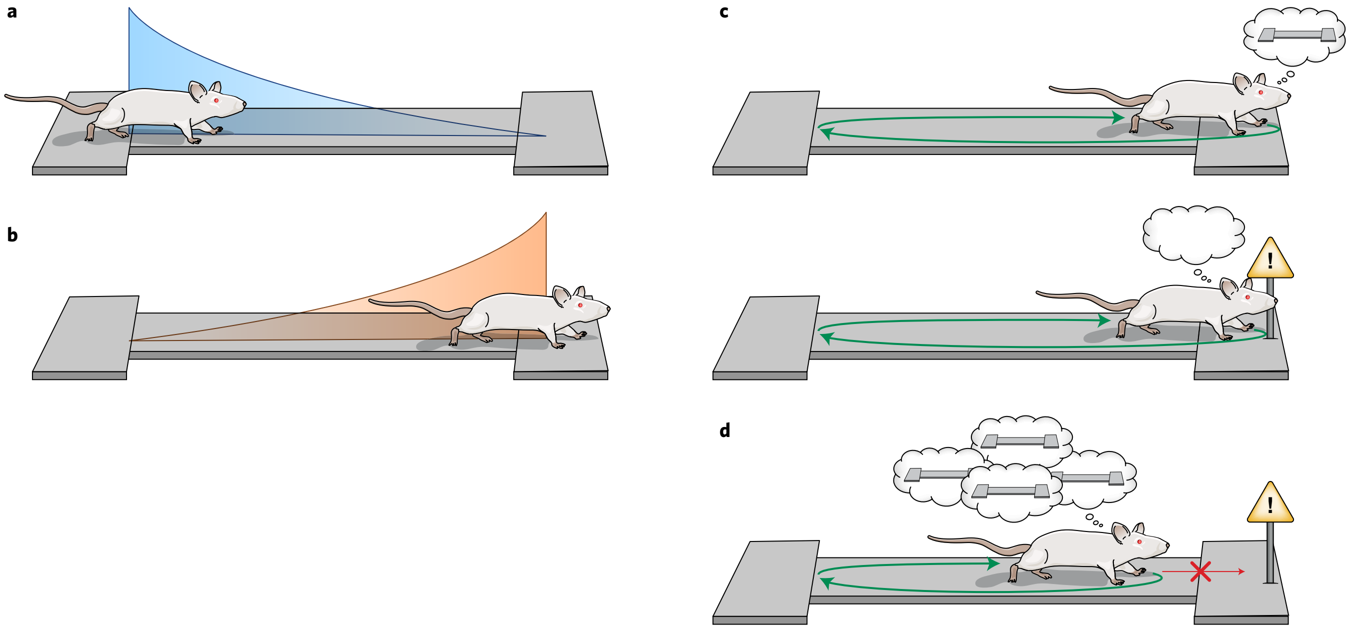

Fig. 1 | Model components; and predicted responses to a negative outcome in two contrasting scenarios. a, Blue wave depicts need, which peaks just in front of the rat and falls off along predicted future locations. **b**, Orange wave depicts the influence of gain, which peaks immediately behind the animal and moves sequentially with successive backups back along the past trajectory. **c**, When a rat encounters an unexpectedly negative outcome (shown by the hazard sign), but needs to keep its behavior the same (green path), the result is fewer replays, as shown by the empty thought bubble. This matches the experimental results of Ambrose et al.[11]. **d**, By contrast, when a rat encounters an unexpectedly negative outcome, but is able to change its actions to stop moving forward (red path) and instead turn around earlier (green path), the result is an increase in replays. This matches the experimental results of Wu et al.[13.]

This pioneering study raises several questions for both theoretical and experimental future work. First, while the need term is easy to compute as learned predictions about future location occupancies, the gain term can only be computed by actually performing all the backups and picking the one that is best. So the utility of the model lies in identifying the optimality of the replay sequences that are observed experimentally; the actual implementation could be based on different mechanisms entirely. Many models have considered how to generate replay sequences[14]. The interesting question is whether these models, or others to be developed, can provide principled approximations to the optimal scheduling of Mattar and Daw that would, for example, generalize to as-yet-untested situations. Likewise, it remains to be seen whether replay itself will continue to behave in the way predicted by Mattar and Daw, since the phenomenon has so far been characterized only in a few simple tasks and in typically small experimental spaces. Second, the rollout of forward replays is a little different in the model than that of backwards replays. For example, forward replays appear to occur only after predictions of future returns have converged on near final values. This matches the experimental data quite well, but it does mean that from the perspective of learning, forward replay remains enigmatic. Third, long sequences, which the authors call “depth-first” backup sequences, as opposed to shorter “breadth-first” sequences, may not be quite as robust in the model as experimentally observed. They depend critically on establishing well-worn pathways of need, which in more open areas may be harder to establish, due to the greater number of possible trajectories.

Replay is a growing area of experimental study and there are still lots of chocolates in the box. Now we have something else: a key card to tell us which chocolates are which. Further experiments will tell whether the key card generalizes to new chocolates yet to be discovered. In the meantime, Mattar and Daw have provided an important conceptual link between replay and offline learning. These chocolates are actually good for you!❐

John Widloski and David J. Foster*
Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, California, USA.
**e-mail: davidfoster@berkeley.edu*

Published online: 22 October 2018
https://doi.org/10.1038/s41593-018-0260-8

References

1. Mattar, M. G. & Daw, N. D. *Nat. Neurosci.* https://doi.org/10.1038/s41593-018-0232-z (2018).
2. Foster, D. J. & Wilson, M. A. *Nature* **440**, 680–683 (2006).
3. Diba, K. & Buzsáki, G. *Nat. Neurosci.* **10**, 1241–1242 (2007).
4. Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. *Neuron* **65**, 695–705 (2010).
5. Johnson, A. & Redish, A. D. *J. Neurosci.* **27**, 12176–12189 (2007).
6. Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. *Science* **336**, 1454–1458 (2012).
7. Singer, A. C., Carr, M. F., Karlsson, M. P. & Frank, L. M. *Neuron* **77**, 1163–1173 (2013).
8. Pfeiffer, B. E. & Foster, D. J. *Nature* **497**, 74–79 (2013).
9. Wu, X. & Foster, D. J. *J. Neurosci.* **34**, 6459–6469 (2014).
10. Singer, A. C. & Frank, L. M. *Neuron* **64**, 910–921 (2009).
11. Ambrose, R. E., Pfeiffer, B. E. & Foster, D. J. *Neuron* **91**, 1124–1136 (2016).
12. Watkins, C. J. C. H. & Dayan, P. *Mach. Learn.* **8**, 279–292 (1992).
13. Wu, C. T., Haggerty, D., Kemere, C. & Ji, D. *Nat. Neurosci.* **20**, 571–580 (2017).
14. Foster, D. J. *Annu. Rev. Neurosci.* **40**, 581–602 (2017).

Competing interests

The authors declare no competing interests.