Implicit and Explicit Learning Mechanisms Meet in Monkey Prefrontal Cortex

Matthew V. Chafee[1,*] and David A. Crowe[2]
[1]Associate Professor, Department of Neuroscience, University of Minnesota, 321 Church Street SE, Minneapolis, MN 55455, USA
[2]Associate Professor, Department of Biology, Augsburg University, Minneapolis, MN, USA
*Correspondence: chafe001@umn.edu
https://doi.org/10.1016/j.neuron.2017.09.049

In this issue, Loonis et al. (2017) provide the first description of unique synchrony patterns differentiating implicit and explicit forms of learning in monkey prefrontal networks. Their results have broad implications for how prefrontal networks integrate the two learning mechanisms to control behavior.

Nearly every neuroscientist has heard of HM, the epilepsy patient who underwent a bilateral removal of his medial temporal lobes and, as a result, could no longer form new long-term memories. Most neuroscientists also know that this was not entirely true—he *could* form some new memories, but only certain kinds. HM couldn't create new memories of events or facts, things that we can explicitly call to mind, but he could learn difficult new motor tasks. He famously improved his performance of tracing a star in a mirror, without afterward ever remembering that he'd done the task. This and similar results led to the striking conclusion that our ability to learn is not unitary—that we have generally separate systems for implicit learning, such as HM could achieve following his injury, and explicit learning, which HM could no longer do.

If the medial temporal lobes are required for explicit learning, what circuit then mediates implicit learning? One traditional view is that learning of automatic motor responses to stimuli (habits) involves reward-driven synaptic plasticity in the striatum. In this view, positive reward prediction error induces the release of dopamine in the nigrostriatal pathway and gates synaptic plasticity in the striatum (Averbeck and Costa, 2017). The adjusted synapses in the striatum then act as a filter operating on descending corticostriatal input to help select and initiate previously rewarded actions, via feedback to motor planning and execution areas, including the prefrontal cortex. As complex as this circuit is, the view is probably too simplistic. Recent lesion studies by Averbeck and colleagues in the nonhuman primate have provided evidence that a broader network involving the amygdala is also critically involved in implicit learning (Costa et al., 2016), and certainly the cerebellum plays a central role in implicit learning in skill acquisition as well.

The information coded into the nervous system by explicit learning is thought to involve rapid synaptic plasticity within a network of areas including the prefrontal cortex and the hippocampus, as in the case of episodic memory (Eichenbaum, 2017). However, in some instances, explicit learning can utilize trial-and-error feedback. Operating in a complex environment, the human brain extracts rules and tests strategies to predict the outcomes of actions, and uses trial-and-error feedback to adjust these strategies to improve behavioral outcomes over time. Work by Lee and colleagues, for example, has characterized neural representations of trial outcomes in monkey prefrontal cortex and their utilization for strategy adjustment in dynamic decision-making tasks (Abe and Lee, 2011).

How are these memory systems integrated to produce the effective control of behavior? Each stimulus we confront may invoke both habitual responses and long-term memories, and these may recruit conflicting actions. A yellow traffic light might invoke acceleration (as a habitual response to the yellow light) or braking (if you remember that this intersection is favored by local law enforcement based on a prior episode). How might these conflicting action plans be reconciled?

Enter Loonis, Miller, and colleagues (Loonis et al., 2017). In this issue of *Neuron*, these authors contrast unique patterns of network synchrony between the prefrontal cortex and striatum during implicit learning, and between the prefrontal cortex and hippocampus during explicit learning.

Loonis et al. make use of an interesting aspect of implicit learning behavior to identify the learning required by their behavioral paradigms as explicit or implicit. Evidence has been accumulating over the last 20 years that performance in some implicit tasks benefits from what is known as errorless learning, a training regimen designed to minimize the errors a subject makes (Maxwell et al., 2001). For example, using this technique to teach a person to shoot a basketball free throw, a subject would start with close shots and move farther away as his or her performance improved (as opposed to "traditional" training in which the subject shoots continually from the far distance, accumulating errors). This type of training often results in better subsequent performance and better retention of skill than traditional training. One explanation for the success of this technique is that errors lead the subject to explicit counter-strategies that interfere with implicit learning, worsening performance. Loonis et al. observed that monkeys performing a saccade task had lower error rates after correct than error trials, consistent with the data on errorless learning in implicit tasks. In contrast, monkeys in two match tasks performed equally well after error and correct trials. Using this behavioral measure, they categorized their tasks as explicit (match tasks) and implicit (saccade task).

Loonis et al. analyzed the synchrony of local field potential (LFP) signals recorded

during these two forms of learning, across different recording sites, by computing the pairwise phase consistency (PPC). PPC measures the degree to which oscillations at the two sites exhibit a consistent phase relation (e.g., are synchronized). They focused their analysis on LFP signals occurring after the response that encoded the outcome of the trial (correct or incorrect). Neural signals encoding trial outcomes are of particular interest to the distinction between implicit and explicit forms of learning because they utilize feedback information differently to adjust behavior.

This led to several interesting observations. First, prefrontal cortex networks exhibited a much more pronounced error-related negativity (ERN) following error feedback on the explicit match tasks in comparison to the implicit saccade task. This is particularly intriguing as a larger ERN magnitude is correlated with explicit awareness of errors on a trial-by-trial basis in humans (Scheffers and Coles, 2000). It is also interesting because the monkeys' decisions on future trials were more strongly influenced by errors on the explicit tasks than the implicit tasks. This provides a neural signal to indicate that errors were more fully processed by prefrontal cortex in the explicit task, and a compelling convergence therefore of behavioral and neurophysiological evidence to support the conclusion that these tasks did in fact recruit explicit learning. Second, the pattern of LFP synchrony associated with outcome encoding in prefrontal networks differed starkly in the explicit and implicit tasks. In the explicit tasks, processing of correct trial feedback was associated with a long-lasting increase in alpha/beta (10–30 Hz) synchrony, whereas processing of error feedback was associated with a shorter increase in delta/theta (3–7 Hz) synchrony. In the implicit task, processing of correct feedback was associated, in contrast, with a long-lasting increase in delta/theta synchrony. These distinct patterns of prefrontal cortex network synchrony associated with outcome encoding in explicit and implicit tasks changed over the course of learning, but with different time courses, further differentiating them.

One exciting possibility raised by these findings is that LFP signals encoding outcomes may constitute teaching signals that could potentially gate synaptic plasticity to improve information processing and performance over trials. Several questions follow. First, where do these signals originate (who is teaching whom)? Considering the explicit tasks, one possibility is that trial outcome signals originate in the hippocampus and are transmitted to prefrontal cortex to train prefrontal cortical circuits. This is broadly consistent with some models of long-term memory consolidation in which output from hippocampus trains cortical circuits as memories are repeatedly recalled so that the information becomes encoded by cortical circuits ultimately and is no longer hippocampally dependent. If hippocampus provides the teaching signal in the form of trial outcome information transmitted from hippocampus to prefrontal cortex, one would predict that hippocampus and prefrontal cortex would become synchronized around the time of trial feedback. Loonis and colleagues show this is the case (Loonis et al., 2017). Further, hippocampus alpha/beta LFP signals drive prefrontal LFP signals in the same band at this time in the trial (Brincat and Miller, 2015). However, the present study also shows that the synchrony established between prefrontal cortex and hippocampus is considerably weaker than the synchrony occurring within prefrontal cortex itself (Loonis et al., 2017), suggesting that LFP oscillatory signals encoding trial feedback in prefrontal cortex are not entirely driven by input from hippocampus. Interestingly, during implicit learning, synchrony is stronger between the prefrontal cortex and the striatum than it is within the prefrontal cortex itself. That suggests that interactions between the prefrontal cortex and striatum are particularly robust and may drive oscillatory activity within the prefrontal cortex, consistent with the operation of a teaching signal originating in the striatum and resonating throughout striatal-thalamo-cortical loops. Ultimately, the presence of oscillatory synchrony between connected structures does not itself reveal the nature of the information transmitted between them, which can be measured by detecting correlated fluctuations in coded information at a rapid timescale (Crowe et al., 2013). This analytical approach could potentially recover the direction in which trial outcome information flows between prefrontal cortex, the striatum, and hippocampus. That in turn could resolve who is teaching whom under explicit and implicit learning conditions.

Although there is converging behavioral and neurophysiological evidence that the two types of tasks employed by Loonis et al. engage implicit and explicit learning, the evidence is still indirect, and the distinction between explicit and implicit learning may not be the only possible source of differences in neural synchrony patterns across the tasks. For example, there are differences in the computation that the motor system must perform. In the implicit saccade task, there is a one-to-one mapping between stimuli and response direction. In the explicit match tasks, the mapping between stimuli and response directions is one-to-several. These differences in motor programming requirements could contribute to the differences in synchrony patterns observed. By the same token, this may be the key functional distinction between implicit and explicit learning. There may be no need to formulate explicit representations of stimulus associations when the required response to a stimulus is fixed; rather, it may be better (most efficient) to automatically (implicitly) associate the required response with the stimulus. When the required response to a stimulus is variable, explicit representations of associations (between stimuli, for example) provide the necessary information to select the correct response when the motor options later become available.

The question of whether a monkey has explicit (e.g., reportable) access to information in its own brain is a question of direct relevance to the distinction between explicit and implicit learning and to the study by Loonis et al. It is also, of course, in the absence of language, a deeply thorny question. However, behavioral paradigms have been developed that could provide a viable approach. Monkeys performing a task developed by Kiani and Shadlen (2009) report their confidence in internally encoded information. This is a form of explicit access. Confronted by a difficult perceptual decision task, the monkeys are provided with an opt-out target they can select for certain but small reward.

This is analogous to a human confronted by the question "how sure are you about your decision?" responding "not very" in the case they select the opt-out choice. Providing an opt-out target in the context of the match tasks used by Loonis et al. could provide a behavioral readout of whether explicit retrieval of a learned stimulus association was *gauged to be successful by the monkey itself*. Correlation of that behavioral choice with fluctuations in the strength of synchrony patterns in prefrontal cortex networks over trials could begin to link neural and behavioral correlates of explicit learning more directly.

Another important question is whether synchrony patterns that encode trial outcomes in the prefrontal-hippocampus and prefrontal-striatum networks can be doubly dissociated in relation to explicit and implicit forms of learning. This will require recording in each prefrontal network during both forms of learning.

Finally, if trial outcome signals are teaching signals in prefrontal cortex networks, it will be important to show that their strength on trial *t* predicts the probability of behavioral correction and enhanced neural signaling on trial *t+1*—that is, that the outcome signals drive and are correlated with learning over trials.

In sum, the study by Loonis et al. delineates several aspects of the relation between prefrontal cortex network synchrony, implicit learning, and explicit learning that will become likely targets for future research. In that respect, this study takes a pioneering and impressive first step.

REFERENCES

Abe, H., and Lee, D. (2011). Neuron *70*, 731–741.

Averbeck, B.B., and Costa, V.D. (2017). Nat. Neurosci. *20*, 505–512.

Brincat, S.L., and Miller, E.K. (2015). Nat. Neurosci. *18*, 576–581.

Costa, V.D., Dal Monte, O., Lucas, D.R., Murray, E.A., and Averbeck, B.B. (2016). Neuron *92*, 505–517.

Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald, A.W., 3rd, and Chafee, M.V. (2013). Nat. Neurosci. *16*, 1484–1491.

Eichenbaum, H. (2017). Nat. Rev. Neurosci. *18*, 547–558.

Kiani, R., and Shadlen, M.N. (2009). Science *324*, 759–764.

Loonis, R.F., Brincat, S.L., Antzoulaos, E.G., and Miller, E.K. (2017). Neuron *96*, this issue, 521–534.

Maxwell, J.P., Masters, R.S., Kerr, E., and Weedon, E. (2001). Q. J. Exp. Psychol. A *54*, 1049–1068.

Scheffers, M.K., and Coles, M.G. (2000). J. Exp. Psychol. Hum. Percept. Perform. *26*, 141–151.