Esposti, F., Johnston, J., Rosa, J.M., Leung, K.-M., and Lagnado, L. (2013). Neuron *79*, this issue, 97–110.

Gastinger, M.J., Bordt, S.A., Bernal, M.P., and Marshak, D.W. (2005). Curr. Eye Res. *30*, 1089–1095.

Huang, L., Maaswinkel, H., and Li, L. (2005). J. Physiol. *569*, 939–948.

Li, L., and Dowling, J.E. (1998). Vis. Neurosci. *15*, 851–857.

Maaswinkel, H., and Li, L. (2003). J. Exp. Biol. *206*, 2201–2209.

Repérant, J., Ward, R., Miceli, D., Rio, J.P., Médina, M., Kenigfest, N.B., and Vesselkin, N.P. (2006). Brain Res. Brain Res. Rev. *52*, 1–57.

Stell, W.K., Walker, S.E., Chohan, K.S., and Ball, A.K. (1984). Proc. Natl. Acad. Sci. USA *81*, 940–944.

Stell, W.K., Walker, S.E., and Ball, A.K. (1987). Ann. N Y Acad. Sci. *519*, 80–96.

Umino, O., and Dowling, J.E. (1991). J. Neurosci. *11*, 3034–3046.

Walker, S.E., and Stell, W.K. (1986). Brain Res. *384*, 262–273.

Zucker, C.L., and Dowling, J.E. (1987). Nature *330*, 166–168.

How Did the Chicken Cross the Road? With Her Striatal Cholinergic Interneurons, Of Course

Geoffrey Schoenbaum,[1,2,3,*] Thomas A. Stalnaker,[1] and Yael Niv[4,*]
[1]Behavioral Neuroscience Branch, NIDA/IRP/NIH/DHHS, Baltimore, MD 21224, USA
[2]Department of Anatomy and Neurobiology, University of Maryland, School of Medicine, 20 Penn Street, HSF-2 S251, Baltimore, MD 21201, USA
[3]Solomon Snyder Department of Neuroscience, Johns Hopkins University, Baltimore, MD 21218, USA
[4]Department of Psychology & Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
*Correspondence: geoffrey.schoenbaum@nih.gov (G.S.), yael@princeton.edu (Y.N.)
http://dx.doi.org/10.1016/j.neuron.2013.06.033

Recognizing when the world changes is fundamental for normal learning. In this issue of *Neuron*, Bradfield et al. (2013) show that cholinergic interneurons in dorsomedial striatum are critical to the process whereby new states of the world are appropriately registered and retrieved during associative learning.

Recognizing when the world has changed—and when it has not—is a fundamental yet much ignored component of associative learning. Imagine relocating to Sydney, Australia. While much there might be familiar, one prominent difference is of life-or-death import: the cars come from the right. If you don't learn to look right-left-right before crossing, your visit might be quite short. On the other hand, since you plan to venture to proper-side-of-the-road-driving countries periodically, it would behoove you to also maintain your previous left-right-left behavior, applying that when appropriate. Optimally, rather than overwriting your original strategy for crossing the street, upon experiencing the strange driving habits in your new hometown, you would form a new "state" of "I am in Sydney" and learn new mappings from actions to goals ("policies" in the jargon of reinforcement learning, "action-outcome associations" in terms of learning theory) relevant to that state. Linking these learned policies to the new state would, conveniently, protect the old policies linked to the old state from being overwritten, so that behavior could be modified quickly if the old state were to reappear.

As this example illustrates, appropriate recognition of when to form new states to which to attach information is vital to adaptive behavior. In this issue of *Neuron*, Bradfield and colleagues (Bradfield et al., 2013) use a series of complex yet highly controlled behavioral manipulations to show that input from a part of the thalamus, the parafascicular nucleus, onto cholinergic interneurons in the posterior compartment of the dorsomedial striatum (pDMS), is critical to the appropriate creation of new states during learning. Note that we use "state" here to refer to a high-order representation of the environment in which actions are being chosen—a notion that encompasses the animal learning theory terms of "context," "discriminative stimulus," and "occasion setter" as well as the statistical learning theory term "latent cause" (Gershman and Niv, 2010), but is different from common usage of the term in reinforcement learning.

In the first phase of training, Bradfield et al. (2013) taught rats to associate two levers with two different, but equally valued, rewards (pellets or sucrose). Satiating the rats on one of the two outcomes (a so-called "devaluation test") selectively reduced responding on the lever leading to that outcome and not on the lever leading to the other outcome. Notably, this was the case both for intact rats and for rats in which cholinergic signaling in the pDMS, an area previously shown to be necessary for goal-directed behavior (Yin et al., 2005), was disrupted via several different manipulations (Figure 1, left). This intact initial learning,

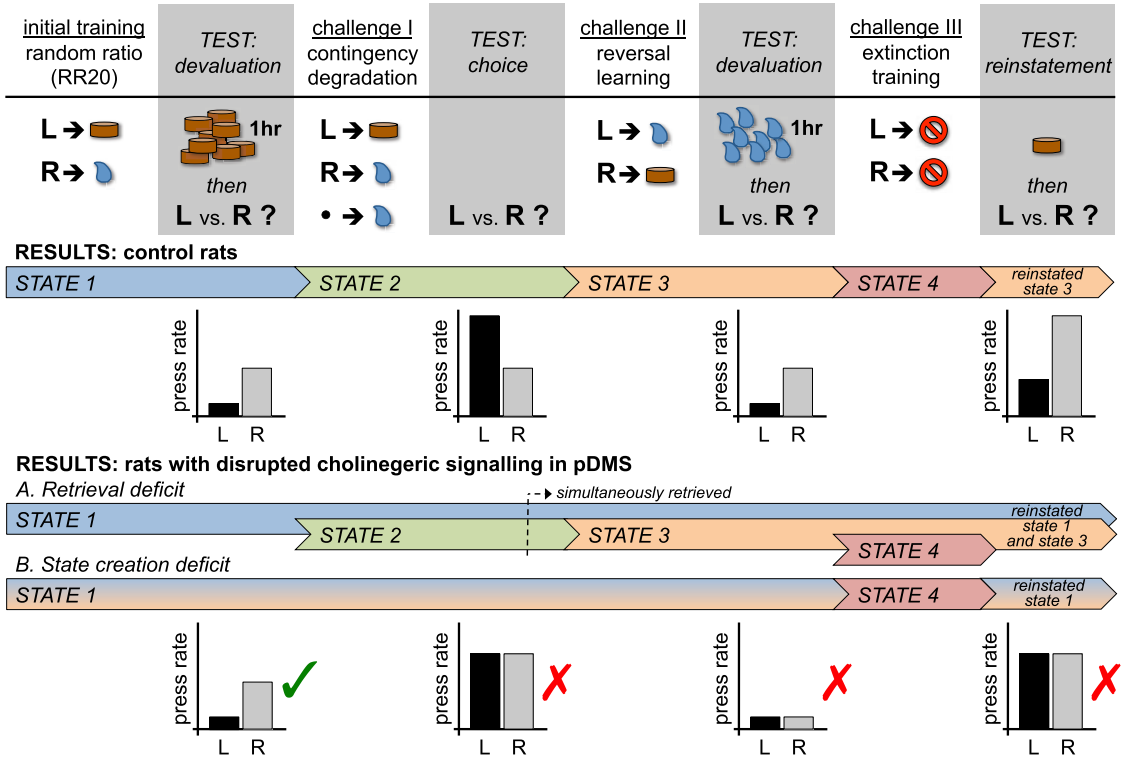

Figure 1. Illustration of Bradfield et al.'s Experimental Design, Results, and Interpretation

Reward for the two actions were a food pellet or sucrose solution (top). All tests were conducted without reward, retraining was administered after tests as necessary, ● denotes no action, and all conditions were counterbalanced (see Bradfield et al., 2013). Selective test responding only on the lever that had previously been mapped to the nondevalued/nondegraded/reinstated outcome in the control group suggests that each change in task contingencies was encoded by the rats as a new state (middle). In contrast, disruption of cholinergic activity in pDMS resulted in nonspecific degradation, devaluation, and reinstatement of both actions, but only after initial learning (bottom). This could be ascribed to (A) a retrieval deficit that caused multiple states to be retrieved throughout or (B) a deficit in creating new states when the identity but not the value of outcomes was changed, such that training in challenges I and II was combined with the initial training.

and sensitivity to devaluation, demonstrated that behavior was indeed goal directed (i.e., guided by an expectation of the specific outcome and its unique attributes [Dickinson and Balleine, 1994]) and that intact cholinergic activity in the pDMS is *not* necessary for this fundamental learning process.

But what happens if the world changes? This was tested in a second phase of training, in which rats faced three challenges (Figure 1, top), each designed to test how changes in the associative structure of the environment would be incorporated into the earlier learning. The first involved contingency degradation—the outcome associated with one of the levers was presented for free, meaning that rats no longer needed to work to receive that reward. The second, reversal learning, involved switching the outcomes associated with each lever, followed by another devaluation test of the effects of satiety on responding. The third, extinction, involved removal of all outcomes for actions, followed by a "reinstatement test" in which one of the outcomes was delivered to test whether it could reinstate pressing on the lever most recently associated with that outcome. In each challenge, the critical question was whether rats would appropriately create new states in which to represent the new environmental contingencies. If so, each challenge should selectively affect responding on the lever most recently associated with the degraded, devalued or reinstated outcome. Any nonspecific effects on both levers would suggest that something had gone awry.

The results were amazingly clear-cut: in each case, intact rats exhibited selective effects on subsequent testing (Figure 1, middle), whereas rats in which cholinergic signaling in the pDMS had been disrupted showed intermediate or nonspecific effects on testing (Figure 1, bottom). Importantly, this same behavioral pattern was induced by bilateral lesions of the parafascicular nucleus, crossed lesions of the parafascicular nucleus and the pDMS, or a pharmacological manipulation that disrupted cholinergic signaling in the pDMS only during the learning phases. This exhaustive characterization of the phenomenon shows both that it is reliable and that it depends on cholinergic signaling *at the time of learning*, with the latter explicitly confirmed using an immunohistochemistry tool specific to cholinergic interneurons that was recently developed by the Balleine lab (Bertran-Gonzalez et al., 2012). In addition, these results were shown to be specific to cholinergic disruption in the pDMS as they was not reproduced by manipulations of cholinergic function in the anterior portion of the dorsomedial striatum.

What Does All This Mean?

While these results are extremely elegant in their consistency and convergence,

they are not entirely straightforward to interpret theoretically. What exactly has gone wrong with the state generation process due to the cholinergic manipulations? Here, the comprehensive set of metaphoric hoops through which the rats were made to jump becomes key to narrowing down the options, highlighting the utility of using the incisive behavioral manipulations that animal learning theorists have spent decades developing.

To understand what went wrong, it is useful to first review what aspects of learning were not disrupted by cholinergic manipulations: in addition to intact goal-directed learning, the comprehensive battery of tests shows that new state formation was not completely abolished. This is evident in the test following the third challenge, extinction training, in which exposure to one of the outcomes led to reinstatement of responding. Reinstatement indicates that extinction training did not simply overwrite and erase previous associations between actions and outcomes (Gershman et al., 2010), but rather reward omission caused rats in both groups to create a new state (Figure 1, state 4). However, reinstatement in cholinergically impaired rats was far from normal: these rats reinstated *both* actions (Figure 1, right).

Retrieval Deficit

One possible explanation for this pattern of results (option A in Figure 1, bottom) is that upon reinstatement the rats erroneously retrieved two states—the most recent, postreversal state (state 3 in Figure 1), in which the right lever was mapped to sucrose and the left to pellets, and the state from initial training (state 1 in Figure 1), in which the lever to reward mapping was the reverse. This may, in fact, sound familiar to world travelers: a foolproof policy for safe street-crossing in some countries is to look left-right-left-right repeatedly, that is, to act upon *both* pre-travel and in-travel states.

Such a retrieval deficit could also explain the lack of specificity of the post-reversal devaluation test, in which cholinergically impaired rats devalued *both* actions rather than only the one associated with the satiated outcome (Figure 1, third column). Finally, it can also explain the intermediate level of responding in the contingency degradation test (Figure 1, second column) by assuming that the new state (state 2, in which not pressing was associated with the outcome) was retrieved together with the old state (state 1).

The deficit in reinstatement was observed even when cholinergic function was disrupted only during learning, yet this does not rule out a retrieval deficit, as retrieval of the appropriate states is also necessary during learning. That is, in order to learn, on every trial, the rat must retrieve and update associations within the current state. If multiple states were retrieved and updated during learning, the rat would show a non-selective response in the reinstatement test even though normal cholinergic function had been restored. Importantly, under this interpretation, new state formation is intact; however, retrieval of appropriate states is disrupted or at least less selective.

State Creation Deficit

A second possible explanation (option B in Figure 1, bottom) is that the rats with disrupted cholinergic function might have been able to form a new state in extinction but not in the other challenges. Why would this happen? To answer this, it is useful to ask how the brain knows that a new state should be formed in the first place. One impetus for state creation is significant differences between the current situation and past experience (Gershman et al., 2010). According to this idea, prediction errors—differences between what is expected (driving is on the right of the road, mass transportation is called "subway," etc.) and what is currently experienced (cars are on the left, the underground train is "the city circle")—drive state formation. Importantly, these prediction errors include both errors in predicted value (the city circle is not cheap), and errors in predicted identity (would you expect "the city circle" to indicate an underground train system?). The former are typically termed reward prediction errors (though we use "value," as changes in rewarding events can also induce identity prediction errors), and Bradfield et al. (2013) refer to the latter as "state prediction errors," though we prefer "identity," as any sort of error could lead to recognition of state change.

Bradfield et al.'s first two manipulations—contingency degradation and reversal learning—involved only identity prediction errors, since the underlying value of the reward associated with lever pressing did not change. However, the last manipulation introduced value prediction errors since the reward was entirely omitted. If cholinergic transmission in the striatum is important for detecting, representing, or learning from identity prediction errors, one would expect to see no new state formation in the first two manipulations due to the cholinergic manipulation, but intact state formation during extinction learning. Thus, like a retrieval deficit, a selective effect on the formation of new states following identity prediction errors would also produce the observed pattern of results (Figure 1, bottom).

What's Ach Got to Do with It?

Though relatively little is known about the function of cholinergic striatal interneurons, what we know so far relates nicely to these two interpretations. For example, one can easily imagine a key role for striatal acetylcholine (Ach) in retrieval: cholinergic interneurons are inhibitory, tonically active, and innervate (and receive input from) a large number of medium spiny neurons (Zhou et al., 2002). This places this local modulatory system in a prime position to provide network-wide inhibition, promoting retrieval of only the relevant state at each point in time (Apicella, 2007). By reducing cholinergic tone, Bradfield et al. (2013) could have thus caused rats to retrieve multiple states during decision making and learning, thereby supporting the first interpretation above. On the other hand, cholinergic interneurons also respond to important events with phasic changes in firing that are notably unrelated to value prediction errors (Morris et al., 2004). Do these responses relate instead to identity prediction errors? This has yet to be tested, and would support the second interpretation.

However, even without complete understanding of the striatal circuitry and its reliance on acetylcholine, the powerful toolkit provided by traditional animal learning theory could be used to test and differentiate the above two hypotheses. One key experiment would be to train rats to associate the two levers with reward of decidedly different magnitude and then put them through Bradfield

et al.'s series of tests. If the deficit depended on the need to learn from identity prediction errors, behavior should now be impervious to cholinergic interventions in the pDMS, since all three manipulations would involve value as well as identity prediction errors. If, on the other hand, the problem was one of retrieval, then the rats' responding should still reflect the erroneous association of both levers with both outcomes, with response rates postreversal evidencing similar predictions for both levers. Of course, single unit recordings would still be useful for understanding the relationship between either of these roles and the precise firing patterns of the neurons, as well as the dynamics of learning in the striatal network that gives rise to these functions (and associated deficits). However, it is always inspiring to see well-controlled behavioral designs reveal underlying neural processes, even absent electrodes.

ACKNOWLEDGMENTS

The authors' work was funded by the NIDA Intramural Research Program (G.S. and T.A.S.) and the Human Frontiers Science Program (Y.N.).

REFERENCES

Apicella, P. (2007). Trends Neurosci. *30*, 299–306.

Bertran-Gonzalez, J., Chieng, B.C., Laurent, V., Valjent, E., and Balleine, B.W. (2012). PLoS ONE *7*, e53195.

Bradfield, L.A., Bertran-Gonzalez, J., Chieng, B., and Balleine, B.W. (2013). Neuron *79*, this issue, 153–166.

Dickinson, A., and Balleine, B.W. (1994). Anim. Learn. Behav. *22*, 1–18.

Gershman, S.J., and Niv, Y. (2010). Curr. Opin. Neurobiol. *20*, 251–256.

Gershman, S.J., Blei, D.M., and Niv, Y. (2010). Psychol. Rev. *117*, 197–209.

Morris, G., Arkadir, D., Nevet, A., Vaadia, E., and Bergman, H. (2004). Neuron *43*, 133–143.

Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005). Eur. J. Neurosci. *22*, 513–523.

Zhou, F.M., Wilson, C.J., and Dani, J.A. (2002). J. Neurobiol. *53*, 590–605.

Pursuing the Link between Neurons and Behavior

Douglas A. Ruff[1] and Marlene R. Cohen[1,*]
[1]Department of Neuroscience and Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA 15260, USA
*Correspondence: cohenm@pitt.edu
http://dx.doi.org/10.1016/j.neuron.2013.06.045

Hohl et al. (2013) found that fluctuations in neuronal responses in the middle temporal area (MT) are correlated with variability in smooth pursuit eye movements. The pattern of neuron-behavior correlations constrains models of how sensory neurons guide behavior and establishes pursuit as an attractive model system for studying how sensory neurons guide behavior.

The way humans and animals respond to any sensory stimulus is unreliable. For example, an animal being pursued by a predator might sometimes run away and might other times lie still and hide. Some of this behavioral variability might come from variability in the way sensory stimuli are encoded in the brain. Neuronal responses are also variable: a given neuron in visual cortex, for example, will respond differently each time an animal views the same visual stimulus.

Over the past two decades, experimenters have capitalized on this variability to establish a link between the activity of neurons in different brain areas and specific behaviors. The earliest such study measured the relationship between motion-direction-selective neurons in the middle temporal area (MT) and monkeys' decisions in a motion-direction discrimination task that required the animals to determine in which of two opposite directions a random dot stimulus was moving (Britten et al., 1996). On repeated presentations of an identical stimulus, fluctuations in the activity of single MT neurons were weakly but consistently correlated with the monkeys' decisions. On trials in which a neuron tuned for upward motion fired more than its average, the monkey was more likely to report seeing upward than downward motion.

Since that initial study, correlations between the fluctuations in the responses of individual neurons and behavior (typically called choice probability for discrimination tasks or detect probability for detection tasks) have been observed in a variety of sensory areas and behavioral tasks (for review, see Nienborg et al., 2012; Parker and Newsome, 1998). The existence of such neuron-behavior correlations, when combined with data from more causal experimental methods like pharmacology, lesions, or electrical stimulation, can provide evidence that those neurons are part of the neural mechanisms underlying specific percepts or behaviors (Parker and Newsome, 1998).

Using neuron-behavior correlations (or other experimental methods) to infer the computation that downstream areas perform to decode sensory information from areas like MT has been much more difficult, however. This difficulty has at least three sources. (1) The relationship between any one neuron's activity and behavior is typically weak and noisy. This is expected because a large number of