et al.[6] and Burgos-Robles *et al.*[7] suggest that aversive information is transferred by both PL- and IL-projecting BLA cells. Indeed, it is likely that the manipulation of aversive information instrumental in extinguishing fear associations occurs primarily in the mPFC, which then projects back to the BLA. Thus, the extinction-related pattern of activity in IL-projecting BLA cells may emerge as a result of reciprocal connectivity from the IL back down to these cells.

Together, these two papers suggest that amygdala projections to the prefrontal cortex convey the identity of aversive stimuli. This has important implications for our understanding of psychiatric disease. Post-traumatic stress disorder, in particular, is associated with hyperactivity in the amygdala, which may indicate unconstrained output of aversive associations. Hyperactive inputs into the prefrontal cortex could underlie resistance of fear memories to extinction in patients, which may reinforce persistent hyperactivity downstream in the amygdala. Finding and fine-tuning the most appropriate circuit-breakers in this loop could constitute an important step forward in developing treatment.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Salzman, C.D. & Fusi, S. *Annu. Rev. Neurosci.* **33**, 173–202 (2010).
2. Burgos-Robles, A., Vidal-Gonzalez, I. & Quirk, G.J. *J. Neurosci.* **29**, 8474–8482 (2009).
3. Milad, M.R. & Quirk, G.J. *Nature* **420**, 70–74 (2002).
4. Likhtik, E., Stujenske, J.M., Topiwala, M.A., Harris, A.Z. & Gordon, J.A. *Nat. Neurosci.* **17**, 106–113 (2014).
5. Bordi, F., LeDoux, J., Clugnet, M.C. & Pavlides, C. *Behav. Neurosci.* **107**, 757–769 (1993).
6. Klavir, O., Prigge, M., Sarel, A., Paz, R. & Yizhar, O. *Nat. Neurosci.* **20**, 836–844 (2017).
7. Burgos-Robles, A. *et al. Nat. Neurosci.* **20**, 824–835 (2017).
8. McGarry, L.M. & Carter, A.G. *J. Neurosci.* **36**, 9391–9406 (2016).
9. Cho, J.H., Deisseroth, K. & Bolshakov, V.Y. *Neuron* **80**, 1491–1507 (2013).
10. Klavir, O., Genud-Gabai, R. & Paz, R. *Neuron* **80**, 1290–1300 (2013).
11. Herry, C. *et al. Nature* **454**, 600–606 (2008).
12. Morgan, M.A. & LeDoux, J.E. *Behav. Neurosci.* **109**, 681–688 (1995).
13. Bukalo, O. *et al. Sci. Adv.* **1**, e1500251 (2015).
14. Senn, V. *et al. Neuron* **81**, 428–437 (2014).
15. Sotres-Bayon, F., Sierra-Mercado, D., Pardilla-Delgado, E. & Quirk, G.J. *Neuron* **76**, 804–812 (2012).

Pinging the brain to reveal hidden memories

Rosanne L Rademaker & John T Serences

Keeping a picture in mind requires many brain cells to actively communicate ... or does it? There might be more to working memory than neuronal chatter, and silent processes could be hiding right beneath the surface.

It is common folklore to liken the mind to water, probably because the mind has traditionally been a viewed as vast and unknowable—an entity of impenetrable depths. As modern neuroscience is slowly lifting the veil on our mind's innermost workings, such longstanding intuitions may prove to have some merit, albeit in surprising new ways. New work reported in this issue of *Nature Neuroscience*[1] suggests that at least one central component of everyday cognition, namely working memory, relies on brain processes hiding right beneath the surface. The study grants a sneak peek into relatively uncharted brain mechanisms that do not rely on active neural firing, and it demonstrates how such 'activity-silent' hidden states relate to human behavior.

Every time you remember a snippet of information over a short bit of time, your brain relegates this information to working memory. There the snippet endures despite being detached from the outer world, existing in the mind alone. Neural spiking has long been assumed to be the common currency of the brain and, by association, the substrate of working memory. Indeed, extracellular recordings in primates have shown that when a monkey remembers a picture over a brief delay, single neurons in its frontal and parietal cortex exhibit sustained patterns of activity[2,3].

A seemingly open-and-shut case, the traditional notion of sustained spiking in frontal areas as neural substrate for working memory has been seriously challenged by neuroimaging studies[4,5] decoding memory contents from brain areas where spikes are not generally observed during such tasks[6,7]. This includes primary sensory regions such as area V1, the first port of entry for visual information in cortex. This suggests that spiking activity may not constitute the whole story when it comes to working memory. Below the directly observable surface could lie an activity-silent state of working memory, possibly in the form of short-term synaptic plasticity[8–11]. The supposition of hidden states begs the question of how one would even go about finding something that is by definition hidden. This is where Wolff *et al.* deploy an ingenious tactic: they 'ping' the brain[1,10].

A ship using active sonar will emit pings of sound to reveal what lies below the surface by sensing how underwater objects reflect the sound waves back, a method known as 'pinging'. The authors employ this analogy to explain how they reveal hidden states during working memory. The idea is simple: if working memories are indeed hiding in an activity-silent network of altered synaptic weights, one can ping that network by pushing a wave of activity through it. Activity will more easily propagate through parts of the network with stronger synaptic weights, and recording patterns of activity after a ping flushes information confined in the network into the open.

What does a ping to the brain look like? For the human participants tested by Wolff *et al.*[1], the ping consisted of three big circles shown side by side on a computer screen. The circles were either plain white or filled with black-and-white dartboards (**Fig. 1**). The precise nature of a brain ping probably doesn't matter, as long as it targets the network doing the remembering and bears no systematic relationship to the thing being remembered. The ping used by Wolff *et al.*[1] targets the visual system of participants trying to keep a picture in working memory for a couple of seconds.

Participants briefly viewed two striped circles, one presented on either side of a computer screen (**Fig. 1**) and committed the orientation in each circle to working memory. Participants concluded the task by indicating if a later orientation was rotated with respect to the remembered orientation. This exercise was repeated many times while participants' brain activity was measured with electroencephalography (EEG). Each visual orientation elicited a unique pattern of activity across posterior electrodes. And because similar orientations elicit similar patterns, a comparison of all patterns can be used to decode a specific orientation.

In a first experiment, the two orientations were remembered for about a second before a small arrow appeared, pointing to one side of the screen. Preceding the arrow, information

Rosanne L. Rademaker and John T. Serences are in the Psychology Department, University of California, San Diego, La Jolla, California, USA.
e-mail: rrademaker@ucsd.edu or jserences@ucsd.edu

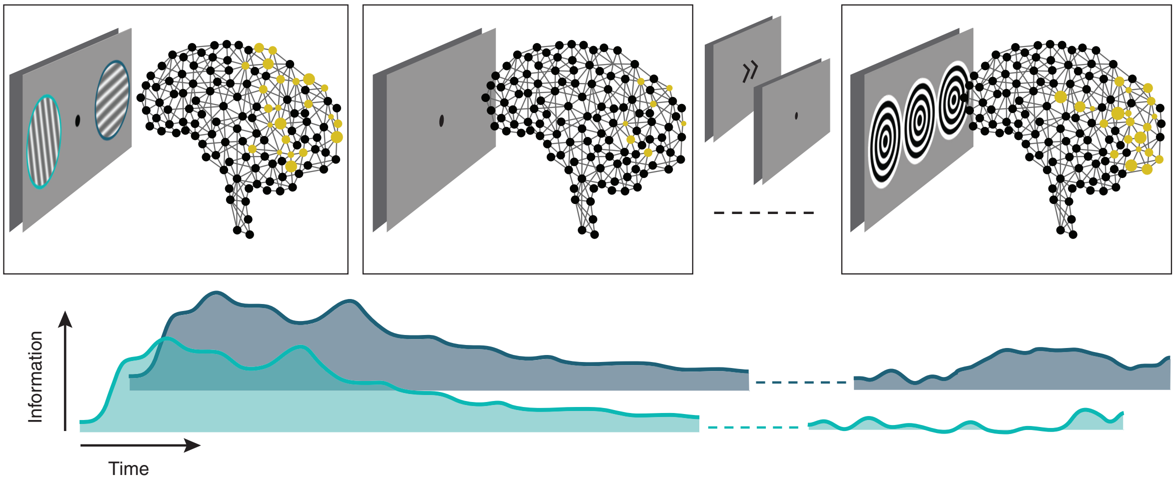

Figure 1 How to ping a brain. Summary of the first experiment by Wolff *et al.*[1]. Top: participants see two striped circles to keep in working memory (left), activating a sensory-related brain network (yellow nodes). Bottom: the information graph is color-coded to match the colored rings outlining the striped circles in the top panel and shows how much information was present for each. (Colored rings were not visible during the actual experiment.) Right after the two orientations are shown, information can be read out from the recorded brain activity equally well for both orientations. As the working memory delay progresses (top middle), the amount of information fizzles out. After about a second, an arrow tells participants which orientation they will need to recall later. Here the arrow points to the right side of the screen, which means that the orientation previously shown on the left is now irrelevant. When the ping appears another second later (right), information reemerges from the neural signal for the orientation with continued relevance but not the irrelevant orientation. The information pattern after the ping (right; yellow nodes) differs fundamentally from the pattern at the time when the orientations were actually perceived.

about both orientations was found in the neural signal, tapering off over time. The arrow designated only one of the two orientations as having continued relevance, i.e., as being the one that would be queried later. When a subsequent ping was shown, a swell of information about the still-relevant orientation ensued. The orientation that had lost its relevance, however, left no trace. This implies that hidden memory states exist and that they harbor a striking amount of flexibility, as information that no longer serves a behavioral goal can be rapidly purged.

So what about behavior? Participants did better on the memory task when the ping exposed more information about the relevant orientation. Conversely, they did worse when the ping exposed more information about the irrelevant orientation. This lends legitimacy to the hidden state idea, with the ping being able to selectively reveal information about the neural underpinnings of working memory.

In a second experiment both memory orientations were probed consecutively in a fixed order. As soon as the orientations appeared, participants knew which of the two they would have to prioritize for recall and which to keep on the back burner for later. Both orientations were decodable almost immediately following presentation, with the prioritized orientation more prominently so. In fact, not long into the delay the deprioritized orientation seemed to have fallen off the proverbial stove altogether, as its information level dropped to chance. A ping presented during this period exposed evidence for the prioritized orientation and, more interestingly, also for the deprioritized orientation. The latter suggests that working memory for temporarily deprioritized, or unattended, information is still stored in a hidden state, even if it cannot be detected from overt brain activity measurements. This dissociation with attention is of particular importance because attention has long been thought necessary for the maintenance of information in working memory[12].

Intuitively, one might suspect the ping of reactivating a sensory-like neural signal. After all, activity was measured from electrodes over the back of the head, implying participation of sensory areas. However, pinging the brain did not evoke a pattern resembling the pattern elicited when the two orientations were originally viewed. What's more, the quality of the sensory-evoked information did not predict behavior in the same way the ping-evoked information did. The ping was shown to neither transform nor interact with the hidden memory state; it merely exposed it[1]. The authors conclude that the hidden state differs fundamentally from a literal reactivation of a sensory representation[13–15].

So what does a hidden state actually represent? The key lies in between the activity a ping sends into the hidden state and the activity coming back out the other end; it lies in between the sound leaving a boat and the echo rising back up from the depths. This is also where the analogy breaks down: in active sonar we know the signal leaving the boat (a wave of sound) and what it does under water (bounce off the ocean floor). With respect to the brain we are less sure of the input and of how exactly sensory signals travel through a network. Even if the input were known, it would be incredibly hard to predict the input's interaction with the hidden state. It is not likely to simply bounce back like sonar; instead it is probably susceptible to thresholds and other nonlinear interactions. While we can infer what the ocean floor looks like from sonar, in the brain we can only say that input and output are systematically related. This is sufficient to provide information about an image held in working memory, and it might be how brain regions downstream of sensory cortex read out what is inside the hidden state. As for the format of the hidden code itself, it might ultimately prove very challenging to draw direct inferences using the pinging approach.

While this set of experiments does present compelling evidence for the existence of activity-silent hidden memory states, the evidence is indirect and inferred from the echo of a ping. To provide direct evidence, one would need a definitive measure of hidden representations, such as a method allowing large-scale access to subthreshold modulations. As it stands, it is equally plausible that a very subtle active trace is present during working memory and that the effect of the ping is to amplify a lingering active representation. Even though the present hypothesis is very intriguing, directly measuring hidden states and distinguishing them from states that are merely 'hidden' from the method being employed to measure neural activity will remain a major challenge for neuroscientists in the future.

All things considered, the work by Wolff *et al.*[1] represents a great plunge into the depths of the nebulous states underlying human cognition. For the very first time, hidden states have been shown to relate to human behavior in a working memory task. It will be interesting to see whether pinging the brain can result in further significant discoveries as scientists venture further into this uncharted territory.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Wolff, M.J., Jochim, J., Akyrek, E.G. & Stokes, M.G. *Nat. Neurosci.* **20**, 864–871 (2017).
2. Goldman-Rakic, P.S. *Neuron* **14**, 477–485 (1995).
3. Riley, M.R. & Constantinidis, C. *Front. Syst. Neurosci.* **9**, 181 (2016).
4. Harrison, S.A. & Tong, F. *Nature* **458**, 632–635 (2009).
5. Serences, J.T., Ester, E.F., Vogel, E.K. & Awh, E. *Psychol. Sci.* **20**, 207–214 (2009).
6. Mendoza-Halliday, D., Torres, S. & Martinez-Trujillo, J.C. *Nat. Neurosci.* **17**, 1255–1262 (2014).
7. van Kerkoerle, T., Self, M.W. & Roelfsema, P.R. *Nat. Commun.* **8**, 13804 (2017).
8. Mongillo, G., Barak, O. & Tsodyks, M. *Science* **319**, 1543–1546 (2008).
9. Stokes, M.G. *Trends Cogn. Sci.* **19**, 394–405 (2015).
10. Wolff, M.J., Ding, J., Myers, N.E. & Stokes, M.G. *Front. Syst. Neurosci.* **9**, 123 (2015).
11. Serences, J.T. *Vision Res.* **128**, 53–67 (2016).
12. Awh, E. & Jonides, J. *Trends Cogn. Sci.* **5**, 119–126 (2001).
13. Sprague, T.C., Ester, E.F. & Serences, J.T. *Neuron* **91**, 694–707 (2016).
14. Rose, N.S. *et al. Science* **354**, 1136–1139 (2016).
15. Ester, E.F., Anderson, D.E., Serences, J.T. & Awh, E. *J. Cogn. Neurosci.* **25**, 754–761 (2013).