

Cognition and Behavior

Registered Report: Transcriptional Analysis of Savings Memory Suggests Forgetting is Due to Retrieval Failure

Tania Rosiles, Melissa Nguyen, Monica Duron, Annette Garcia, George Garcia, Hannah Gordon, Lorena Juarez,  Irina E. Calin-Jageman, and  Robert J. Calin-Jageman

<https://doi.org/10.1523/ENEURO.0313-19.2020>

Neuroscience Program, Dominican University, River Forest, Illinois 60305

Abstract

There is fundamental debate about the nature of forgetting: some have argued that it represents the decay of the memory trace, others that the memory trace persists but becomes inaccessible because of retrieval failure. These different accounts of forgetting lead to different predictions about savings memory, the rapid re-learning of seemingly forgotten information. If forgetting is because of decay, then savings requires re-encoding and should thus involve the same mechanisms as initial learning. If forgetting is because of retrieval failure, then savings should be mechanistically distinct from encoding. In this registered report, we conducted a preregistered and rigorous test between these accounts of forgetting. Specifically, we used microarray to characterize the transcriptional correlates of a new memory (1 d after training), a forgotten memory (8 d after training), and a savings memory (8 d after training but with a reminder on day 7 to evoke a long-term savings memory) for sensitization in *Aplysia californica* ($n=8$ samples/group). We found that the reactivation of sensitization during savings does not involve a substantial transcriptional response. Thus, savings is transcriptionally distinct relative to a newer (1-d-old) memory, with no coregulated transcripts, negligible similarity in regulation-ranked ordering of transcripts, and a negligible correlation in training-induced changes in gene expression ($r=0.04$ 95% confidence interval (CI) $[-0.12, 0.20]$). Overall, our results suggest that forgetting of sensitization memory represents retrieval failure.

Key words: latent memory; long-term memory; sensitization; transcription

Significance Statement

Understanding the nature of forgetting is important because both excessive and insufficient forgetting are related to profound disruptions of mental health. This registered report provides molecular data indicating that forgetting of long-term sensitization in *Aplysia* represents retrieval failure, contributing new evidence toward resolving a long-standing debate over the neural mechanisms of forgetting.

Introduction

Long-term memory is characterized both by its duration and by its dependence on changes in gene expression (Goelet et al., 1986). Although long-term memories can last a lifetime, much of what we initially commit to long-term memory is forgotten, becoming progressively less likely to be recalled (Bahrick, 1984). Forgetting plays an essential role in memory function as both excessive and

insufficient forgetting are related to profound disruptions of mental health (Tröster et al., 1993; Ally et al., 2013; Fitzgerald et al., 2013; Mary et al., 2013).

Currently, there is fundamental disagreement about the nature of forgetting, with one review concluding that “we do not know why or how the brain actually forgets” (p 113, Hardt et al., 2013). Some have argued that forgetting occurs because of decay of the memory trace and, thus,

Received June 10, 2020; accepted July 30, 2020; First published September 14, 2020.

The authors declare no competing financial interests.

Author contributions: I.E.C.-J. and R.J.C.-J. designed research; G.G., H.G., A.G., M.D., A.G., T.R., and M.N. performed research; I.E.C.-J. and R.J.C.-J. analyzed data; I.E.C.-J. and R.J.C.-J. wrote the paper.

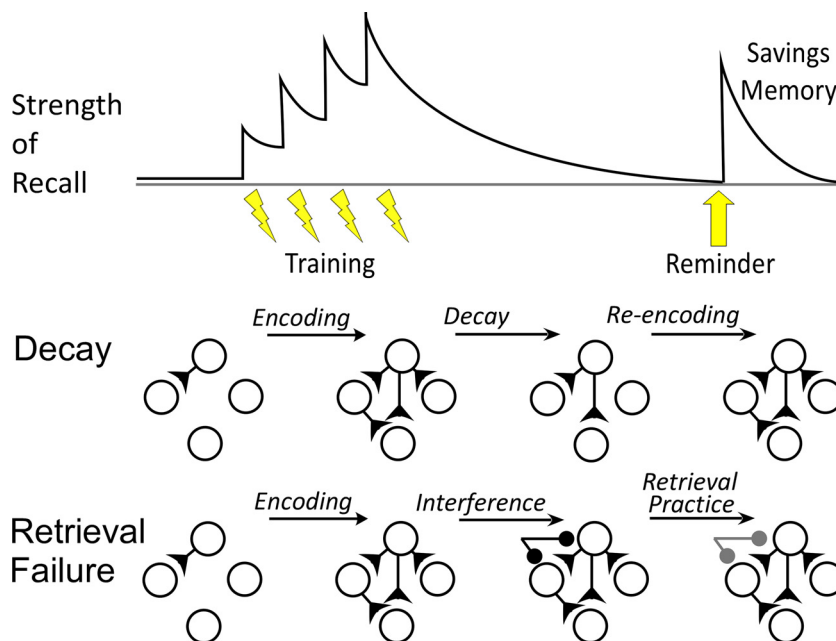


Figure 1. Savings memory according to decay and retrieval-failure accounts of forgetting. Top, Savings memory. Repeated rounds of training (lightning bolts) increase strength of recall, but in the absence of additional practice, forgetting occurs, indicated by a decline in the strength of recall toward zero. Nevertheless, a brief reminder can re-instate recall; this is known as savings memory. Middle, In decay theories of forgetting, initial learning changes synaptic connectivity and strength forming a memory trace. Over time, however, these changes decay away, leading to forgetting (reduced recall). During savings memory, the memory trace must be almost entirely rebuilt. Savings is thus predicted to use the same transcriptional mechanisms that initially created the memory trace. Bottom, In retrieval-failure theories of forgetting, forgetting is due not to decay but to interference from other memories. For example, additional learning could inhibit (dark circles) the otherwise intact memory trace. In this framework, savings involves repairing retrieval mechanisms (e.g., downregulating inhibition). Thus, savings is predicted to be transcriptionally distinct from initial memory storage.

represents a failure of memory maintenance. In stark contrast, others have suggested that forgetting is merely a retrieval failure, and that the original memory trace persists, perhaps indefinitely (for review, see Wixted, 2004; Davis and Zhong, 2017). For example, forgetting could be because of inhibitory processes that repress otherwise intact memory traces (Barron et al., 2017).

Here, we conduct an experiment to shed light on the nature of forgetting by studying savings memory, the rapid re-acquisition of seemingly forgotten information. Ebbinghaus first characterized savings memory (Ebbinghaus, 1885). He learned lists of nonsense words to perfection, waited until he could no longer recall the words, and then re-learned the lists to perfection. He found that it always took less training to re-learn the lists compared with the original acquisition. Since that pioneering demonstration, savings memory has been demonstrated with multiple learning paradigms (Nelson, 1985) and in a variety of species (Antzoulatos et al., 2006;

Philips et al., 2006; Menges et al., 2015), suggesting that it is a core feature of long-term memory.

Decay-failure and retrieval-failure accounts of forgetting make contrasting predictions about the transcriptional correlates of savings memory (Fig. 1). If memory traces decay, then savings is a re-construction of the original information facilitated by the remnants of the original memory trace. Under this account, savings is predicted to be mechanistically similar to initial memory storage and should evoke a transcriptional state similar to what is observed during new learning. If, on the other hand, forgetting involves only retrieval failure, then savings does not require rebuilding the original memory trace. In this case, savings memory would be mechanistically and transcriptionally distinct from initial memory storage.

In this registered report, we tested the decay-failure and retrieval-failure accounts of savings memory (Fig. 2). Specifically, we used microarray to characterize the transcriptional changes that accompany long-term sensitization in *Aplysia californica* for a newly stored memory (1 d after training), a forgotten memory (8 d after training), and a savings memory (8 d after training but with a reminder on day 7).

Long-term sensitization in *Aplysia* requires changes in transcription (Sutton et al., 2001), and new memory storage is associated with regulation of over 1000 transcripts within 1 d of training (Conte et al., 2017). As sensitization memory is forgotten most of this transcriptional response

This work was supported by the National Institute of Mental Health Grant 2R15MH107892-02.

Correspondence should be addressed to Robert J. Calin-Jageman at rcalinjageman@dom.edu.

<https://doi.org/10.1523/ENEURO.0313-19.2020>

Copyright © 2020 Rosiles et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

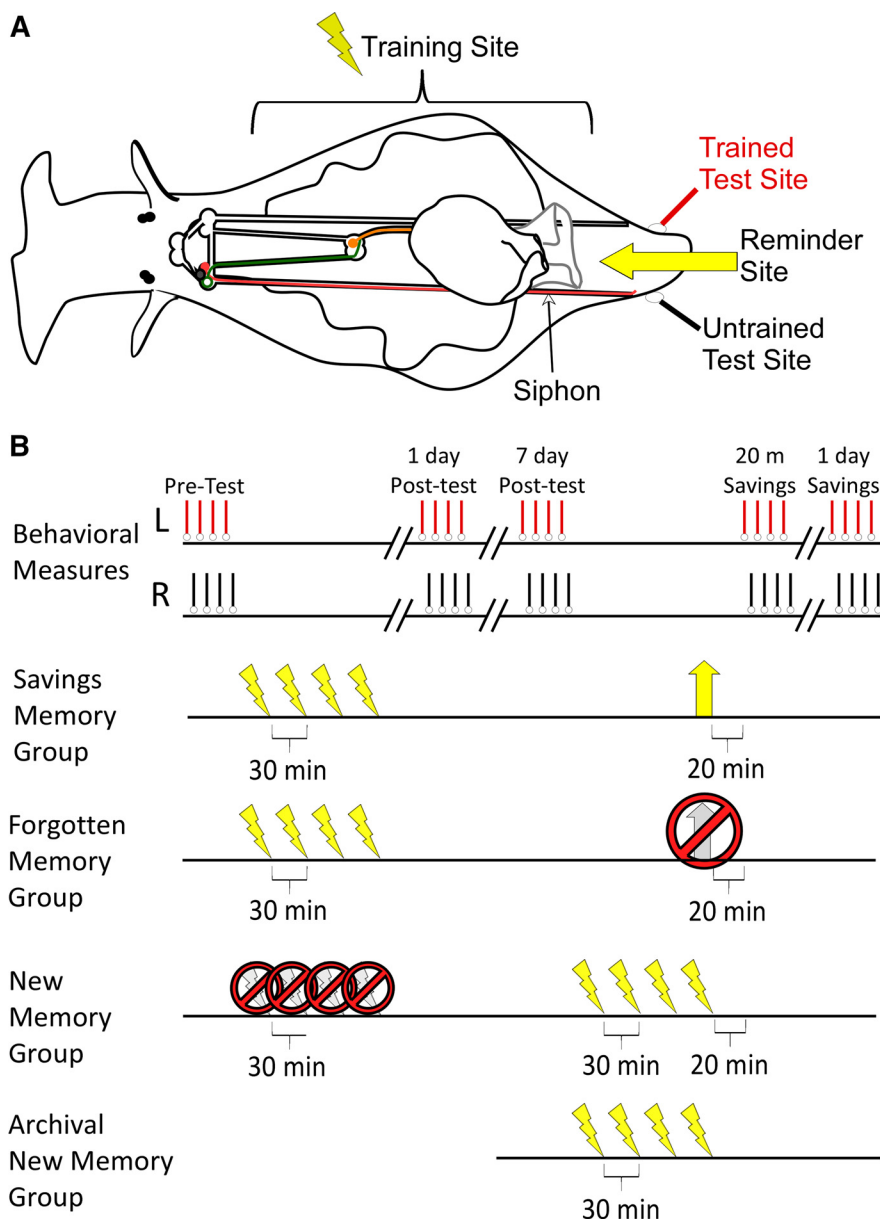


Figure 2. Long-term sensitization in *Aplysia*. **A**, Overview of the behavioral paradigm. Sensitization is an increase in responsiveness because of noxious stimulation. To produce long-term sensitization in *Aplysia*, animals were exposed to four rounds of painful shock to one side of the body (training site: lightning bolt). The effect of training was monitored by measuring the duration of the T-SWR, a defensive withdrawal of the siphon evoked by an innocuous stimulus to the left or right side of the tail (test site). To document savings, a reminder was administered to the midline of the tail (reminder site, wide arrow). In naive animals this reminder produces short-term but not long-term sensitization. **B**, Behavioral protocols. Behavioral measures were the same in all experimental groups: T-SWR responses were evoked on the left and right tail (red and black hash lines). Responses were measured at pre-test, 1 d after training, 7 d after training, 20 min after the reminder, and 1 d after the reminder (day 8 from training). Animals differed in their experimental treatments. For the savings-memory group, animals received standard sensitization training after pre-test measures (four strong shocks, 30 min apart, lightning bolts). Then, 7 d after training animals received the reminder (wide arrow) to evoke savings memory. For the forgotten-memory group, the treatment was the same except animals were given sham reminder (gray arrow crossed out), leaving the sensitization memory dormant. Finally, for new-memory group, animals initially received sham sensitization training (gray lightning bolts crossed out) and then received real sensitization training after the 7-d post-tests. All animals were harvested for microarray analysis after the 1-d savings tests (day 8 from start). Thus, all animals received the same behavioral testing but, when harvested, were expressing different states of sensitization memory: new (1 d since training), forgotten (8 d since training with no reminder), or savings (8 d since training and 1 d since the reminder). In addition, we analyzed an archival dataset from Conte et al. (2017), which roughly replicates the new-memory group (harvested 1 d after sensitization training). We used these archival data to benchmark assessments of similarity.

fades; 7 d after training only approximately seven transcripts remain regulated (Patel et al., 2018; Perez et al., 2018). Here, we asked whether savings memory reactivates the transcriptional changes observed with new memory formation (as predicted if forgetting is because of decay) or if savings has a distinct transcriptional profile (as predicted if forgetting is because of retrieval failure).

Comparing transcriptional states with microarray can be difficult because poor signal-to-noise and sampling error can produce spurious dissimilarity. These problems can be overcome in the learning paradigm we selected. First, sensitization-related transcription in *Aplysia* can be characterized from isolated ganglia that contain neurons known to help encode sensitization memory, providing a strong learning-related transcriptional signal (Herdegen et al., 2014a). Second, learning is expressed on only the trained side of the body (Scholz and Byrne, 1987), enabling a powerful within-subjects comparison. Thus, microarray experiments of moderate size (eight samples/group) can attain high power and convergent validity (Herdegen et al., 2014b). Third, we have previously analyzed the transcriptional correlates of newly stored sensitization memory (1 d after training), providing a set of benchmarks for analyzing the correlates of savings memory (Conte et al., 2017).

Leveraging the advantages of the sensitization paradigm we conducted a rigorous experiment to provide compelling evidence about the nature of forgetting. We found that the transcriptional correlates of savings memory are distinct from new memory formation suggesting that forgetting of sensitization is because of retrieval failure.

Materials and Methods

We conducted this study as a registered report. First, we developed and publicly posted a behavioral protocol, quality controls, and a behavioral analysis plan (<https://osf.io/z2uck>; May 28, 2018). Then, we began initial collection of behavioral data and sample preparation. Once we had enough behavioral data to be confident our design was feasible, we developed a microarray analysis plan and script and submitted a registered report proposal. After review and in-principle acceptance we publicly pre-registered the study (<https://osf.io/fqh8j>; September 11, 2019), completed behavioral data collection, and conducted the planned microarray analysis.

Open data and materials

Our preregistration, analysis scripts, and all data for this project are available on the Open Science Framework (<https://osf.io/z2uck/>). The microarray data are also posted to NCBI's Gene Expression Omnibus (GEO: GSE152045, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152045>).

Animals

Animals (75–125 g) were obtained from the RSMAS National Resource for *Aplysia* and maintained at 16°C in one of two 90-gallon aquariums with continuously

circulating artificial sea water (Instant Ocean, Aquarium Systems Inc.).

Long-term sensitization training

A 1-d long-term sensitization training protocol (Fig. 2) was used (Bonnick et al., 2012). Training consisted of four rounds of noxious shock applied at 30-min intervals to one side of the body with a hand-held electrode. Each round of shock consisted of 10 pulses (60-Hz biphasic) of 500-ms duration at a rate of 1 Hz and an amplitude of 90 mA. Side of training was counterbalanced. This training protocol produces memory that is strongly expressed for several days but which fades in most animals within one week (Perez et al., 2018).

The savings-memory and forgotten-memory groups received long-term sensitization training immediately after pre-tests, on the first day of the protocol. In contrast, the new-memory group initially received sham training. This consisted of the same procedure but the constant-current stimulator was set to deliver 0 mA of current. Animals were otherwise handled in the same way and, in general, were run mixed with batches of animals from the other conditions. For the new-memory group, real sensitization training was finally applied after the 7-d tests, 1 d before harvesting tissue.

Reminders to elicit savings

To elicit savings, animals in the savings group received a reminder shock (Philips et al., 2006; Perez et al., 2018). The reminder was delivered 7 d after training, when most animals show essentially no remaining sensitization memory (<25% increase relative to pre-test). The reminder consisted of two moderate shocks (60-Hz biphasic DC pulse for 2 s at 20 mA of constant current) applied to the midline of the tail with a 15-min rest between the shocks. The reminder produces short-term but not long-term sensitization in naive animals. In previously trained animals the reminder reveals a long-lasting unilateral savings memory, with tail-elicited siphon-withdrawal reflex (T-SWR) durations increasing for at least 1 d after the reminder but only on the previously trained side (Perez et al., 2018).

Animals in the forgotten-memory group received a sham reminder, where the same protocol was applied, but the constant-current stimulator was dialed to deliver 0 mA of current.

Animals in the new-memory group did not receive a reminder or a sham reminder but, instead, received their sensitization training while the other groups received reminders.

Behavioral measurement

As a behavioral outcome, we measured the duration of the T-SWR (see Walters and Erickson, 1986). The reflex was evoked by applying a weak shock to one side of the tail using a hand-held stimulator (60-Hz biphasic DC pulse for 500 ms at 2 mA of constant current). T-SWR behavior was measured as the duration of withdrawal from

the moment of stimulation to the first sign of siphon relaxation.

Measurements were made blind to experimental condition. For each time point (pre-test, 1-d, 7-d, 20-min savings, and 1-d savings) behavioral responsiveness was characterized by a series of eight responses evoked on alternating sides of the body at a 10-min ISI. Scores were split by side of stimulation (trained vs untrained) and averaged (four responses/side for each time point characterized).

Isolation and processing of pleural ganglia RNA

We compared gene expression from pleural ganglia on the trained versus untrained side of the animal. The pleural ganglia contain the ventro-caudal (VC) nociceptors (Walters et al., 2004) which contribute input to the T-SWR circuit as well as several T-SWR interneurons (Mackey et al., 1987; Buonomano et al., 1992; Cleary and Byrne, 1993). The VCs are essential for encoding long-term sensitization memories. Gene expression measured in whole pleural ganglia correlates strongly with expression measured from isolated VC clusters (Conte et al., 2017).

To control for lateralized gene expression, samples from two animals trained on opposite sides were pooled.

To analyze transcription, pleural ganglia RNA were isolated immediately after the long-term savings test, 8 d after protocol start. Animals were anesthetized with an injection of isotonic $MgCl_2$ (50% of body weight), and an incision was then made along the ventral midline to expose the CNS. As dissection can alter gene expression (Alberini et al., 1994), we extracted ganglia rapidly (<5 min per animal) and transferred them immediately to Trizol (Invitrogen) for homogenization.

Tissue was homogenized using the Bullet Blender (NextAdvance) and RNA extracted using Direct-Zol Mini RNA kit (Zymo). Quantity and quality of RNA was assessed using the NanoDrop 1000 (Thermo Scientific).

Reverse-transcription quantitative PCR (qPCR)

Reverse transcription was performed using Maxima cDNA kit with DsDNase (Thermo Scientific). Quantitative PCR was conducted using Maxima SYBR Green/Fluorescein qPCR Master Mix (Thermo Scientific) and the MyIQ real time PCR system (Bio-Rad). Primers were validated for correct PCR efficiency; exact sequences are provided in Patel et al. (2018), their Supplemental Table 1. qPCR samples were analyzed in duplicate or triplicate and the relative amounts of each transcript were determined using the ddCT method and the Bio-rad IQ5 gene expression analysis. All qPCR expression levels were normalized to levels of histone H4, a transcript which is stable during LTS training.

Sample size determination

We set a target of eight biological replicates per group. This sample-size exceeds the consensus recommendation of at least five biological replicates per group for microarray analysis (Pavlidis et al., 2003; Tsai et al., 2003; Allison et al., 2006). Moreover, previous transcriptional analyses (Herdegen et al., 2014b) using this learning paradigm has shown that 8 samples per group can achieve

very low estimated false-positive rates (1–2%) and strong convergent validity with qPCR conducted in independent samples ($r^2 = 0.60$ – 0.79). Finally, we tested our analysis script with real data of known levels of similarity and found that this sample-size was sufficient to distinguish between highly similar, moderately similar, and orthogonal sets of regulated transcripts.

Archival data to benchmark similarity

Our microarray analyses compared gene expression in the new-memory, forgotten-memory, and savings-memory groups. To help provide context for these comparisons, we also re-analyzed gene expression from a previous study examining transcriptional changes 1 d after long-term sensitization (Conte et al., 2017; GEO: GSE95596). We refer to this as the “archival new-memory group.” These archival data essentially replicate the new-memory group, although they did not involve as many rounds of T-SWR measurement nor sham training (see Fig. 2).

Microarray processing

We used the Aplysia Tellabs Array (ATA; GEO: GPL18666) to characterize changes in gene expression because of long-term sensitization training (Herdegen et al., 2014b). This array includes 26,149 distinct probes representing all known sources of *A. californica* ESTs and mRNAs at the time of design (January 2012). Based on estimates from previous microarray designs (Moroz et al., 2006), the ATA should cover >50–60% of all CNS-expressed transcripts.

Microarray processing was completed by Mogene Inc. A two-color approach was used, with each array hybridized to a paired trained and untrained sample. Specifically, each of the 24 arrays compared expression from the trained side of a left-trained and right-trained animal to expression from the untrained sides of the same animals. Experimental condition (savings, new-memory, forgotten-memory) was balanced across slides ($n = 8$ /group). In addition, dye color was counterbalanced across training conditions.

Sample integrity was determined by Bioanalyzer RNA 6000, Pico total RNA protocol; 300 ng of total RNA was amplified and labeled with Cy3 or Cy5 using the Agilent Quick Amp Two-Color Labeling kit. Dye incorporation and yield was determined by Nanodrop. Samples were hybridized to the microarray slide at 65°C and 10 rpm for 17 h. Slides were scanned on an Agilent C scanner at 3- μ m resolution. Data were extracted using Agilent Feature Extraction software, version 11.5. All labeling and post-labeling processing was conducted in an ozone regulated environment, monitored at <5 ppb.

Statistical analyses

In our statistical analyses we focus on effect sizes and 95% confidence intervals (CIs; Calin-Jageman and Cumming, 2019a,b). These can easily be converted to hypothesis tests. If the null hypothesis is not in the 95% CI, the test is significant at $\alpha = 0.05$; otherwise the test is not significant.

Behavioral analysis

Behavioral responses were averaged by time point and side of testing (trained or untrained). Change scores were then calculated by subtracting pre-test scores. At each time point paired comparisons were made between the average change on the trained side and the average change on the untrained side: $M_{\text{diff}} = (M_{\text{trained_change}} - M_{\text{untrained_change}})$. The 95% CI for this contrast was then calculated. This is equivalent to estimating the interaction between training and time with a 2×2 within-subjects ANOVA. Along with raw-score effect sizes we report standardized effect size estimates (Cohen's d). These are corrected for bias (Hedges, 1981) and calculated so that positive values represents a stronger increase in response on the trained side (sensitization).

Microarray

Microarray data were analyzed using limma (Smyth, 2005; Ritchie et al., 2015) from the Bioconductor suite of tools (Gentleman et al., 2004) for R (Ihaka and Gentleman, 1996). Median expression values were analyzed (Zahurak et al., 2007). These were corrected for background using the normexp+offset algorithm recommended for Agilent microarrays by Ritchie et al. (2007). Expression was then normalized using the loess function (Smyth and Speed, 2003). Where multiple probes were used to measure the same EST or mRNA, these were averaged (Holmes et al., 2014).

Identification of regulated transcripts

Within each experimental group, trained and control expression was compared by computing a log-fold change (LFC) score indicating the ratio of expression from the trained to control sides (base 2). Changes in expression were flagged by using the treat function from limma (McCarthy and Smyth, 2009) to test for regulation significantly greater than 10% in either direction (an interval null from -10% to 10%) with an empirical Bayes-moderated t test (Smyth, 2004). Benjamini-Hochberg correction was used to maintain a 5% overall false discovery rate (Benjamini and Hochberg, 1995). All the CIs reported for individual transcripts reflect the same correction for multiple comparisons.

Check for completeness of gene lists

We estimated the proportion of true nulls in each condition using the propTrueNull function (Ritchie et al., 2015) using the convex decreasing densities approach developed by Langaas et al. (2005). We then calculated the false negative rate as $1 - \% \text{regulated} - \% \text{truenull}$. Based on previous analyses we established a criterion of false negative rates $<4\%$ for subsequent comparisons of transcription to be considered valid.

Degree of overlapping regulation

For the genes regulated in the new-memory group (1 d after training) we examined the proportion (P) also regulated in the forgotten-memory, savings-memory, and archival new-memory groups. We then estimated the

difference in overlap as a memory is reactivated during savings: $P_{\text{diff}} = P_{\text{savings_overlap}} - P_{\text{forgotten_overlap}}$.

To follow-up on this analysis of overlap we tested formally for differences in regulation between the forgotten-memory and savings-memory groups among the transcripts regulated 1 d after training. This is equivalent to testing each transcript for a training \times condition interaction. We again used an interval null of $\pm 10\%$.

Similarity of ranked transcript lists

We used the OrderedList package for R (Yang et al., 2019) to calculate similarity scores based on overlap across ordered ranks of transcripts. For each condition, transcripts were first ranked by strength of evidence (p value) and sign of regulation (up or down). We then computed and assessed overlap across the top and bottom ~ 1000 transcripts in each list.

Correlations in training-evoked expression

We also examined correlations in LFC scores. The critical question to examine was similarity to the new-memory group, so we first restricted down to the transcripts flagged as clearly regulated in this group. Then we calculated the correlation in LFC to the savings ($r_{\text{new_savings}}$), forgotten-memory ($r_{\text{new_forgotten}}$), and archival new-memory groups ($r_{\text{new_archivalnew}}$). We calculated each relationship with a simple Pearson's r and with correction for potential measurement error using the genuine association of gene-expression profiles function (genas) in limma (Ritchie et al., 2015).

As our primary outcome we examined whether the savings-memory and forgotten-memory groups differed in similarity to the new-memory group. To do this we calculated the difference in correlations across these groups ($r_{\text{diff}} = r_{\text{new_forgotten}} - r_{\text{new_savings}}$) using the paired.r function from the psych package in R (Revelle, 2018). We expected that if savings reactivates transcriptional mechanisms of memory storage then it should show stronger similarity to the new-memory group, leading to a positive value of r_{diff} . We pre-established the following interpretations based on analysis of previous data: strong increase in similarity if $r_{\text{diff}} \geq 0.5$, moderate increase in similarity if $r_{\text{diff}} \geq 0.25$ but < 0.50 , little-to-no increase in similarity if $r_{\text{diff}} < 0.25$.

Data collection and quality controls

We collected data from 98 *Aplysia*. With pairing left-trained and right-trained animals this provided 49 RNA samples (15 assigned to the savings-memory group, 16 in the new-memory group, and 18 in the forgotten-memory group). Making a fair test between decay and retrieval-failure accounts of forgetting requires that the transcriptional analysis proceeds with samples that exemplify each memory state. Therefore, before data collection, we established a strong set of quality controls and posted them publicly to the Open Science Framework (<https://osf.io/z2uck/wiki/Experimental%20Protocol/>; posted on May 28, 2018).

First, we checked behavioral data to ensure each sample selected for microarray exemplified the desired memory state:

(1) To ensure training effectiveness we required all animals to show robust but unilateral long-term sensitization on measures taken 1 d after their training (>30% increase in T-SWR duration on the side of training and less than a 30% change on the untrained side). All but two animals met this criterion; the samples they were part of were discarded (one from the savings-memory group, one from the forgotten-memory group).

(2) There is some variation in *Aplysia* in forgetting. To ensure that the savings and forgotten-memory groups truly represent a forgotten-memory state we required that animals in these groups show negligible behavioral sensitization by the 7-d tests, indicated by T-SWR durations that have returned to within 25% of pre-test. Two animals from the forgotten-memory group did not meet this criterion (they showed lingering sensitization), so the two samples they were part of were discarded.

(3) We required animals in the savings group to show savings memory after the reminder, defined as having T-SWR durations during the savings test that had increased over baseline more on the previously-trained side than on the previously-untrained side. All animals assigned to the savings group met this criterion.

(4) We checked to ensure that habituation from repeated testing did not contaminate our comparison, requiring that T-SWR measures on the 7-d test were within 30% of baseline on the untrained side. Animals from one sample in the forgetting condition did not meet this criterion. However, in coding this criterion we accidentally applied it to the trained side (which all samples passed) and this error was not detected until after this sample had been included in the microarray analysis. As reported in the exploratory analysis section, excluding this sample did not impact the results of the study.

We also checked the quantity and quality of isolated RNA for each sample, discarding any samples with a very low or uneven yield, poor quality, and/or genomic contamination. Based on this, we discarded an additional 12 samples (six from the new-memory group and six from the forgotten-memory group).

Finally, to ensure samples had been properly processed, we used quantitative real-time PCR to check for upregulation of well-defined transcriptional markers of sensitization training:

(1) For the new-memory group, we checked for upregulation of the transcript encoding ApBiP (GenBank: NM_001204652; Kuhl et al., 1992). This transcript is strongly and consistently upregulated 1 d after sensitization training (Conte et al., 2017). As expected, there was strong upregulation of ApBiP in samples from the new-memory group (LFC = 1.49 95% CI [0.79, 2.18], $n = 10$). However, one sample unexpectedly showed lower expression on the trained side. We discarded this sample.

(2) We have not previously examined transcriptional correlates of savings, but we reasoned that transcripts which remain persistently regulated during forgetting should still be regulated during savings. Thus, we checked

the savings-memory group for regulation of the transcript encoding the peptide neurotransmitter Phe-Met-Arg-Phe NH₂ (FMRFa; GenBank: M11283.1; Schaefer et al., 1985). This transcript is strongly upregulated within 1 d of sensitization training and continues to be upregulated for more than one week (Patel et al., 2018). As expected, we observed strong upregulation of FMRFa in this group (LFC = 1.28 95% CI [0.86, 1.69], $n = 14$). We found two samples, however, with decreased FMRFa expression on the trained side; these were discarded.

(3) We did not specify a positive control for the forgotten-memory group, but as expected, there was upregulation of FMRFa in this group (LFC = 0.61 95% CI [0.01, 1.21], $n = 9$).

For each group we selected eight valid samples, leaving one to four samples per group for potential use with qPCR validation.

Results

First, we report the behavioral data, which were collected primarily before preregistration to ensure the samples would adequately represent the different stages of memory. Then, we report our preregistered microarray analysis; all and only planned analyses are reported. Finally, we report additional exploratory analyses that were not part of our preregistered plan.

Behavioral validation

We confirmed that samples selected for microarray showed the expected trends in sensitization memory. In the savings group ($n = 16$ animals to provide 8 samples; Fig. 3A), training produced long-term sensitization, expressed as a large but unilateral increase in T-SWR duration when tested 1 d later. T-SWR responses increased by 5.4 s on the trained side (95% CI [4.7, 6.1]) but showed no change on the untrained side ($M_{\text{untrained_change}} = 0.0$ s 95% CI [-0.4, 0.4]). Thus, comparing changes on the trained and untrained side indicated a very large training effect ($M_{\text{diff}} = 5.4$ s 95% CI [4.7, 6.1], $d = 4.7$ 95% CI [3.7, 6.5]). Although initial learning was strong, sensitization was then forgotten, as by the 7-d post-tests responses were slightly below pre-test on both sides ($M_{\text{trained_change}} = -0.2$ s 95% CI [-0.7, 0.2]; $M_{\text{untrained_change}} = -0.6$ s 95% CI [-1.0, -0.2]), so there was only a weak residual training effect ($M_{\text{diff}} = 0.3$ s 95% CI [-0.2, 0.9], $d = 0.4$ 95% CI [-0.2, 1.1]). Despite this apparent forgetting, all animals showed robust savings memory, as a reminder evoked a long-term re-expression of sensitization on the previously untrained side. Specifically, 1 d after the reminder (1-d savings test) T-SWR responses were moderately increased on the previously trained side ($M_{\text{trained_change}} = 2.0$ s 95% CI [1.7, 2.4]) but continued to be slightly below pre-test on the previously untrained side; $M_{\text{untrained_change}} = -1.0$ s 95% CI [-1.3, -0.6]), reinstating a relatively large training effect indicative of savings memory ($M_{\text{diff}} = 3.0$ s 95% CI [2.6, 3.5], $d = 4.3$ 95% CI [3.3, 6.0]). Savings memory was also evident when normalized to 7-d post-test scores ($M_{\text{diff}} = 1.9$ s 95% CI [1.2, 2.5], $d = 1.3$ 95% CI [0.9, 1.8]).

In the forgotten-memory group ($n = 16$ animals to provide 8 samples; Fig. 3B), sensitization training also produced robust but unilateral increases in T-SWR duration

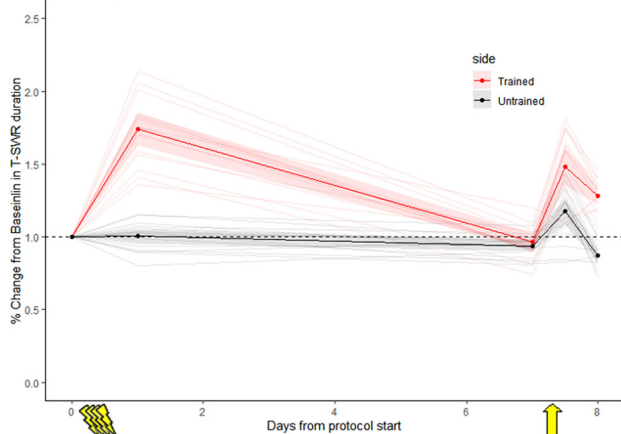
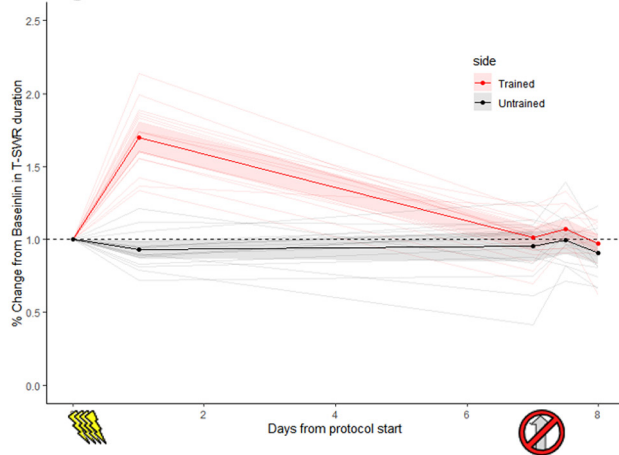
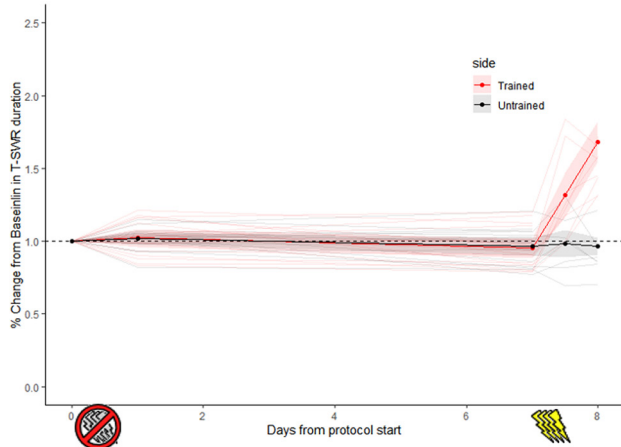
A Savings**B Forgotten****C New**

Figure 3. Behavioral changes in the savings, forgetting, and new-memory conditions. This figure shows T-SWR duration as a % change from pre-test on both the trained (red) and untrained (black) sides. Dark lines with dots represent group means; shading indicates 95% CI of the mean. Individual animals are represented by the light lines. **A**, Savings-memory group, which received real training (lightning bolts) after pre-tests and a reminder (yellow arrow) after the 7-d tests. All animals were expressing a long-term savings memory when harvested on day 8. **B**, Forgotten-memory group, which received real training after pre-tests but a sham reminder (crossed-out

lightning bolts) after the 7-d tests. All animals showed apparent forgetting when harvested. **C**, New-memory group, which received sham training (crossed-out lightning bolts) after pre-tests and then real training after the 7-d tests. All animals were expressing a new (1-d-old) memory when harvested. (1-d post-tests: $M_{\text{diff}} = 7.2$ s 95% CI [6.0, 8.5], $d = 3.1$ 95% CI [2.3, 4.2]). Sensitization was largely, but perhaps not completely, forgotten within one week (7-d post-tests: $M_{\text{diff}} = 0.9$ s 95% CI [-0.1, 1.9], $d = 0.4$ 95% CI [0.0, 0.8]). The sham reminder did not produce a lasting change in behavior; long-term savings tests showed only weak behavioral expression of sensitization memory ($M_{\text{diff}} = 0.8$ s 95% CI [0.0, 1.6], $d = 0.5$ 95% CI [0.0, 1.1]).

Finally, in the new-memory group ($n = 16$ animals to provide 8 samples; Fig. 3C), sham training did not alter behavior, so there was no training effect at the 1-d post-tests ($M_{\text{diff}} = 0.0$ s 95% CI [-0.7, 0.7], $d = 0.0$ 95% CI [-0.7, 0.7]) nor at the 7-d post-tests ($M_{\text{diff}} = 0.0$ s 95% CI [-0.6, 0.6], $d = 0.0$ 95% CI [-0.5, 0.4]). Real sensitization training administered after the 7-d post-tests produced the expected unilateral sensitization, with a large training effect on measures taken the next day (savings post-test: $M_{\text{diff}} = 6.4$ 95% CI [5.1, 7.6], $d = 3.3$ 95% CI [2.5, 4.6]).

Overall, animals selected for microarray exhibited clear and consistent behavioral patterns representative of new, forgotten, and savings stages of sensitization memory.

Planned microarray analysis

The decay account of forgetting predicts that savings recapitulates most of the transcriptional response required to store a memory (savings will be similar to the new-memory group). The retrieval-failure account of forgetting predicts that savings will have distinct transcriptional mechanisms (savings will not be similar to the new-memory group). To test these predictions, we measured the similarity of microarray results between the groups. We assessed similarity in three different ways: (1) as the degree of overlap among transcripts flagged as regulated, (2) as the consistency of rank-order gene lists, and (3) as the linear correlation between LFCs in expression.

How similar is savings memory to new memory?**Overlapping regulation approach**

As an initial way to characterize similarity we flagged clearly regulated transcripts at each epoch of memory (new, forgotten, and savings) and then calculated the overlap in flagged transcripts across conditions. We defined clearly regulated transcripts as those that showed significantly more than a 10% change in expression, with adjustment to maintain a 5% overall false discovery rate. This is a stringent criterion likely to miss some regulated transcripts, but our goal for this initial analysis was to compare results without noise from potentially negligible changes in expression (see next 3 sections for more sensitive comparisons and also for exploratory analyses based on less stringent criteria). Table 1 gives the counts of transcripts flagged in each group. Extended Data

continued

arrow) after the 7-d tests. All animals showed apparent forgetting when harvested. **C**, New-memory group, which received sham training (crossed-out lightning bolts) after pre-tests and then real training after the 7-d tests. All animals were expressing a new (1-d-old) memory when harvested.

Table 1: Counts of clearly regulated transcripts and overlap of regulation

Group	Upregulated	Downregulated	Estimated false negative rate	Proportion of new memory transcripts coregulated
New memory	131	17	<1%	
Forgotten memory	0	0	<1%	0.00 95% CI [0.00, 0.03]
Savings memory	0	0	<1%	0.00 95% CI [0.00, 0.03]
Archival new memory	798	463	<1%	0.95 95% CI [0.91, 0.98]

A complete table of microarray results is provided in Extended Data Table 1-1.

Table 1-1 gives full results for each transcript in each condition and is also posted to the Open Science Framework (<https://osf.io/z2uck/>).

In the new-memory group, there were 148 transcripts that were clearly regulated. The LFC for each of these transcripts is plotted by condition in Figure 4. As expected, nearly all these transcripts had been previously linked with the maintenance phase of sensitization memory (Liu et al., 1997; Conte et al., 2017) and represented predicted proteins with diverse functions, including signaling (ApTBL-1, GenBank: U57369.1, LFC = 0.45 95% CI [0.09, 0.79]), protein production (an eIF2 subunit, GenBank: EB232654.1; LFC = 0.57 95% CI [0.40, 0.74]), the unfolded protein response (GCN-1 like, GenBank: EB230807.1, LFC = 0.53 95% CI [0.34, 0.72]), cytoskeletal function (septin-7-like, GenBank: EB260579.1; LFC = 0.58 95% CI [0.29, 0.87]), and transport (a Dynein β chain component, GenBank: GD206216.1, LFC = -0.37

95% CI [-0.63, -0.12]). Flagged transcripts also correctly included the positive control, ApBiP (LFC = 0.79 95% CI [0.43, 1.16]).

Transcriptional regulation dissipated over time, with no transcripts flagged as clearly regulated in the forgotten-memory condition. This can be seen in Figure 4, where most transcripts regulated in the new-memory condition collapsed toward 0 in the forgotten-memory condition. Thus, defined in terms of overlapping regulation, the forgotten and new phases of memory showed no similarity (Table 1, column showing proportion of overlap). This was expected, as previous studies have also shown that the transcriptional response to sensitization training mostly fades over time and that the very few transcriptional changes that persist are difficult to detect with an array-wide screen (Patel et al., 2018; Perez et al., 2018). Indeed, although we had confirmed upregulation of FMRFa via qPCR in these samples (see Materials and Methods), for

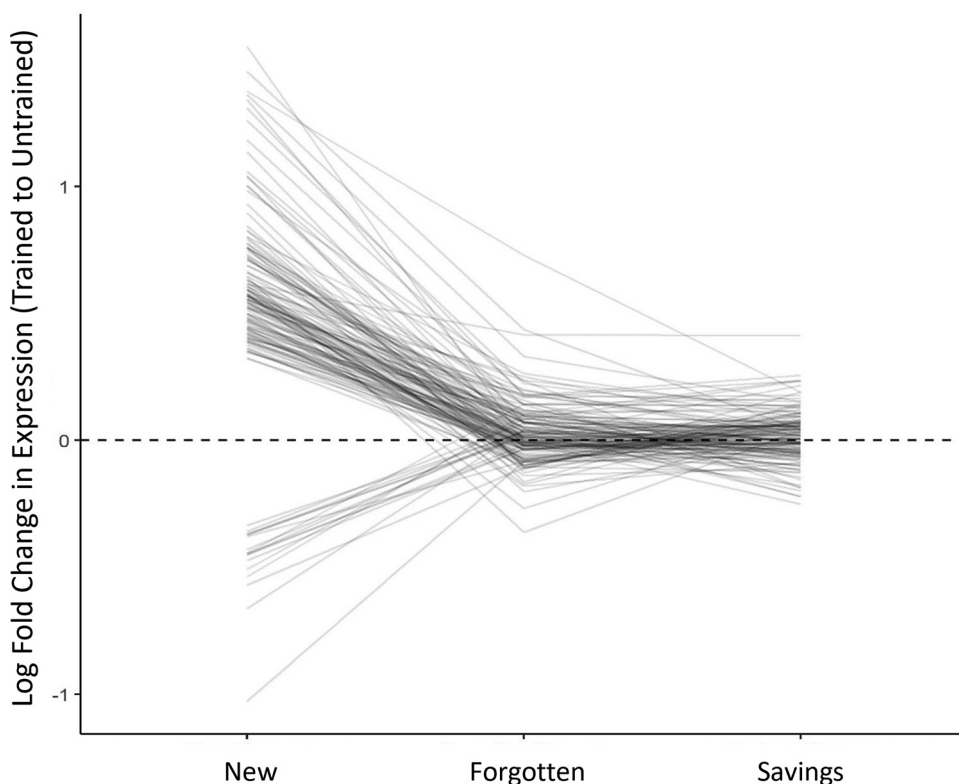


Figure 4. Fate of transcripts regulated during initial learning. This graph shows mean LFC (trained vs untrained) for each of the 148 transcripts flagged as clearly regulated in the new-memory group, tracking their expression during forgetting and savings. The dashed line at 0 represents no change in expression (when trained and untrained expression are the same, their ratio is 1, which gives an LFC of 0). Although these transcripts were clearly regulated 1 d after sensitization (new memory), none were clearly regulated during forgetting or savings.

the microarray analysis this transcript did not meet the threshold for being flagged as clearly regulated (LFC = 0.41 95% CI [0.19, 0.72], $p = 0.02$ before correction, but $p = 1$ after correction).

The critical question was if reactivating the memory during savings would reinstate the transcriptional regulation observed with a new memory. It did not. In the savings-memory group, there were also no transcripts flagged as clearly regulated. This missed at least some transcripts, as even the positive control, FMRFa, did not make the cutoff (LFC = 0.67 95% CI [0.48, 0.86], $p = 0.0001$ before correction, but $p = 1$ after correction). Still, this indicates that among unequivocally regulated transcripts there was no overlap with the new-memory group. In Figure 4, this is shown as the lack of perturbation from the forgetting to the savings conditions. Thus, on the basis of this (admittedly crude) measure of similarity, the savings condition produced no reactivation of transcription, a finding that supports the retrieval-failure account of forgetting.

This was not because of an inability to detect similarity with this approach, as we could detect strong overlap between samples given similar treatments. Specifically, 141 of the 148 transcripts clearly regulated in the new-memory group were also flagged as clearly regulated in the archival new-memory group, which had also been harvested 1 d after sensitization training (Conte et al., 2017).

We also conducted a formal condition by training interaction analysis but did not find any transcripts showing a statistically significant change in training effect from the forgotten to savings stages of memory.

How similar is savings memory to new memory? Gene ranking approach

Overlap of gene lists is not always a sensitive measure of similarity, as it depends on somewhat arbitrary significance classifications. Indeed, our stringent criteria clearly did not capture all regulated transcripts.

As a more sensitive way to measure similarity in transcriptional states, we compared rankings of transcripts across conditions using the using the `OrderedList` package in R (Yang et al., 2019). This allowed us to evaluate similarity in gene rankings across the 1000 most upregulated and downregulated transcripts in each condition regardless of statistical significance status. We compared the transcriptional state in the new-memory group to the forgotten-memory and savings-memory groups and benchmarked against the archival new-memory group.

Analysis of similarity by ranking also indicated that transcriptional regulation fades as a sensitization memory is forgotten. Comparing the new-memory to the forgotten-memory conditions showed only very weak similarity. Figure 5A, left, shows the observed levels of similarity relative to what is expected based on a random shuffle of gene lists. As can be seen, similarity tracks only slightly higher than the average expected by chance. Figure 5A, right, then compares the similarity score observed with the distribution of scores created with the random shuffles. This shows that the level of similarity is marginal (similarity score = 449.0, $p = 0.06$). Thus, as forgetting

progresses, the pattern of regulation shown during encoding is largely lost (although perhaps not completely).

Again, our critical question is what happens as sensitization memory is re-expressed via savings? Figure 5B shows this result, comparing the new-memory and savings groups. The savings group actually showed weaker similarity to the new-memory group (similarity score = 319.9, $p = 0.41$). Note that although the scale of the similarity scores is arbitrary, it can be meaningfully compared across these analyses. Based on gene rankings, savings does not seem to appreciably reactivate the pattern of regulation observed as a memory is stored, a result more consistent with the retrieval-failure account of forgetting. Again, this was not because of a lack of sensitivity of our approach, as we could detect similarity between samples treated with similar protocols. Specifically, Figure 5C compares the new-memory group with the archival new-memory group; this shows very strong similarity in gene list rankings (similarity score = 10,873, $p < 0.0001$).

How similar is savings-memory to new memory?

Correlational approach

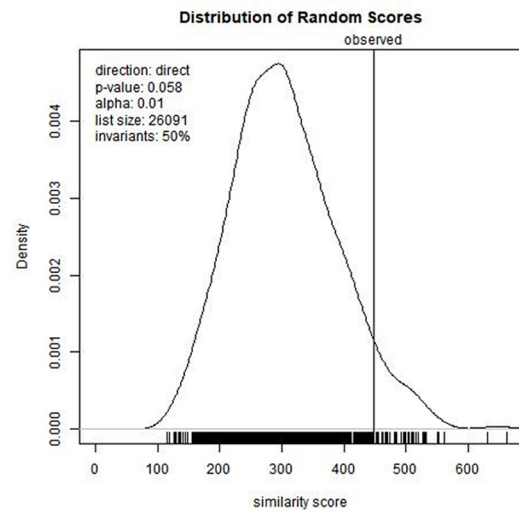
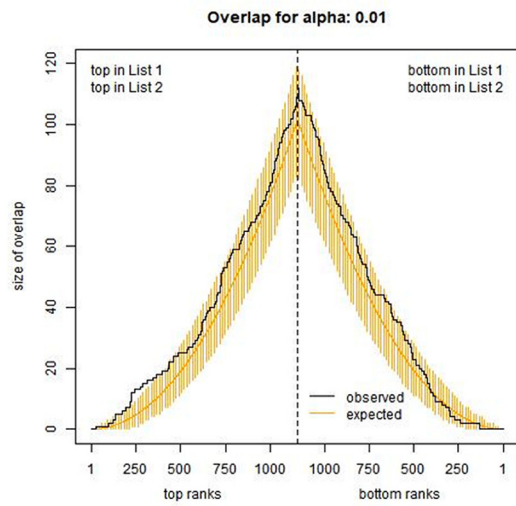
One weakness of quantifying similarity based on rankings is that it omits the magnitude of regulation in the calculation of similarity. Thus, we made a third and final set of similarity measurements by examining correlations between changes in gene expression across conditions. Because we anticipated this to be the most sensitive and complete measure of sensitivity, we preregistered this measure as our primary outcome.

To calculate correlations, we first subsetted to the 148 transcripts clearly regulated in the new-memory group. We then examined the correlation in LFCs in these transcripts in the new-memory and savings-memory groups. We did this using a direct linear correlation (Pearson's r) and with correction for possible measurement error using the `genuine association (genas)` function in the `limma` package for R.

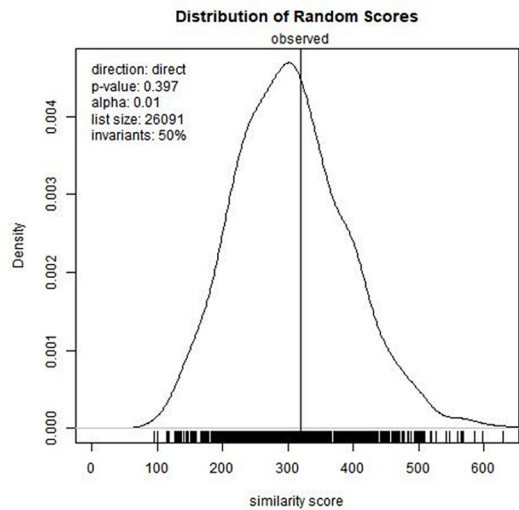
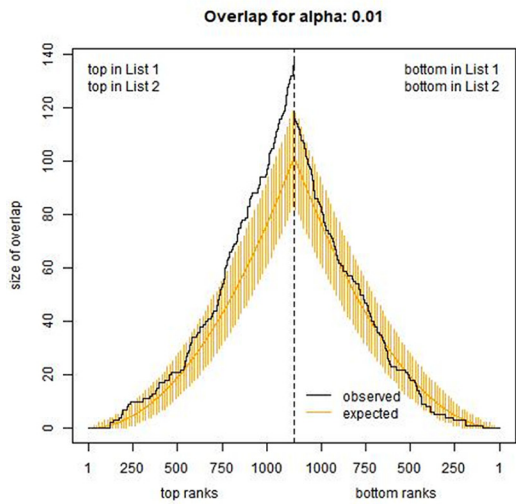
Measuring similarity via correlation also indicated that the transcriptional changes that accompany encoding are largely (but not entirely) dissolved as forgetting proceeds: $r = 0.23$ 95% CI [0.07, 0.38]; $r_{\text{corrected}} = 0.31$, $p = 0.03$ (Fig. 6A). These data are compatible with weak to modest levels of shared variance.

The critical question was if savings would reactivate encoding-related transcriptional changes. By this metric, the answer was again no. The correlation in regulation between the new and savings conditions was very weak when calculated on raw data ($r = 0.04$ 95% CI [-0.13, 0.20]; Fig. 6B) and only marginal when corrected for possible measurement error ($r_{\text{corrected}} = 0.36$, $p = 0.09$). Calculated in raw scores, there was actually a decrease in correlation strength from forgetting to savings ($r_{\text{savings-forgotten}} = -0.19$, $p = 0.05$). Calculated with correction for measurement error, there was a negligible increase in correlation strength ($r_{\text{savings-forgotten}} = 0.05$, $p = 0.30$). This finding is most consistent with forgetting as retrieval failure. This was not because of lack of sensitivity, as we observed a strong correlation among samples treated similarly. Specifically, the correlation between the new-memory and archival new-memory groups was $r = 0.95$ 95% CI [0.93, 0.96]

A Compare New to Forgotten



B Compare New to Savings



C Compare New to Archival New

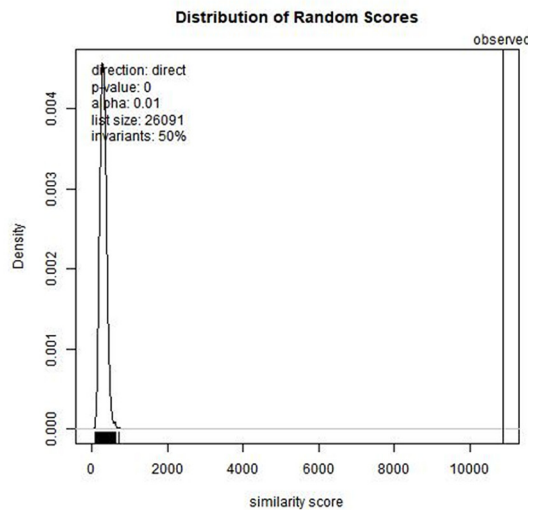
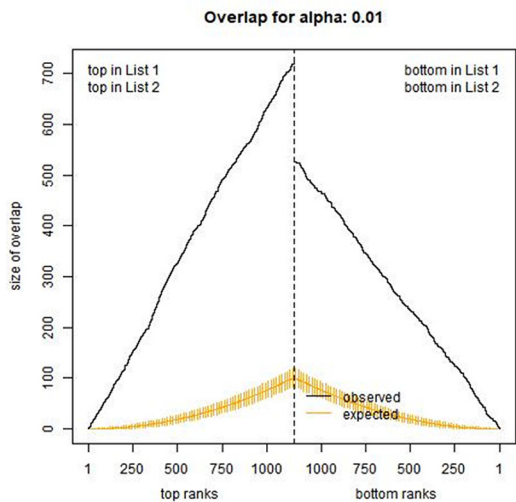


Figure 5. Similarity of conditions measured by rank-ordering of transcripts. These figures show similarity in the transcriptional states

continued

between the new-memory group and the forgotten-memory group (**A**), the savings-memory group (**B**), and the archival new-memory group (**C**). Panels on left compare observed similarity by gene-list length (black line) relative to the range of similarity observed with random shuffles of gene lists (yellow bars). Panels on right show overall similarity score for top ~1000 genes against the distribution of scores from random shuffles.

from raw scores and even higher ($r_{\text{corrected}} = 0.99$, $p < 0.001$) when corrected for possible measurement error (Fig. 6C).

Exploratory analyses

Comparison with previous results

We had previously characterized changes in gene expression that occur 1 d after sensitization training, identifying 1259 clearly regulated transcripts (the archival new-memory group drawn from Conte et al., 2017). We explored the degree to which these transcripts were similarly regulated in this new study. Specifically, we tested for regulation of just these putative memory-related transcripts in the new-memory condition, which was also harvested 1 d after sensitization training (although with additional pre-testing and sham training). We considered transcripts similarly regulated if they showed a statistically significant change in expression (null of no change), with correction to maintain a 5% false discovery rate.

We found strong consistency of results, with 77% of previously-identified transcripts qualifying as regulated in this focused test (972 of the 1259). Moreover, there was a strong correlation in regulation ($r = 0.93$ 95% CI [0.92, 0.94] across LFC scores, $N = 1258$). This is consistent with the planned analyses reported above, which showed high similarity between these datasets. It is notable, however, that for the current study only 148 of these transcripts were flagged when screened over the whole array. Part of this discrepancy is because of the aggressive corrections required for a larger screen (whole array of 26,091 transcripts vs a focused test of 1198 transcripts). In addition, we noticed a general increase in within-group variation (noise) in the current study. Specifically, 70% of the transcripts identified in the archival new-memory group showed higher variance in the current study (larger standard deviation in LFC across samples). Overall, variance increased by 30% 95% CI [27%, 34%], raising thresholds for flagging a transcript as clearly regulated. This increase in noise did not seem to be because of an outlier or bad microarray sample (see next section). It may have been because of the longer duration of the protocol in the current study, which included sham training and three additional rounds of behavioral measurements (compare new-memory to archival new-memory protocols in Fig. 2). Consistent with this possibility, there was modest habituation evident in these animals just before training (see Fig. 3).

Figure 7 shows the fate of the transcripts flagged in the archival new-memory group, plotting LFCs in the new, forgotten, and savings conditions. This graphically shows the strong consistency of expression between the archival and current new-memory conditions. Even with this broader set of transcripts, it is clear that regulation collapses during forgetting (most transcripts decline toward 0 in the forgotten-memory condition) and that savings does

not strongly perturb gene expression (most transcripts remain near 0 in the savings-memory condition).

We have also previously examined changes in gene expression 7 d after sensitization training (Patel et al., 2018; Perez et al., 2018), finding via microarray and qPCR that there is persistent regulation in seven transcripts, all of which are also regulated during initial memory maintenance (1 d after training). To explore consistency with these previous results we tested for regulation just in these transcripts in both the forgotten-memory and savings-memory conditions (both of which were harvested 8 d after training). For this analysis we considered a transcript to be consistently regulated if there was a statistically significant change in expression (null of no change) with correction for multiple comparisons. Table 2 summarizes the results and compares them with previous findings. Overall, there was fair consistency, with four of seven transcripts showing regulation in the forgotten-memory condition, and two of seven in the savings-memory condition. None of these had been flagged in the array-wide screen reported above because of the lower-power of this analysis, but these focused tests again confirm that some transcriptional changes persist beyond the behavioral expression of sensitization memory. This focused test did not, however, reveal a strong change in regulation with savings memory.

How robust are results to analysis parameters?

As the above considerations make clear, microarray results can depend critically on how the analysis is conducted. Although our plan was vetted through peer review and preregistered, it still represents only one reasonable analysis approach out of many. Thus, to examine the analysis generalizability of our results we conducted an exploratory multiverse analysis.

(1) To check for outsized influence of outliers we varied the inclusion of each the 32 microarray samples.

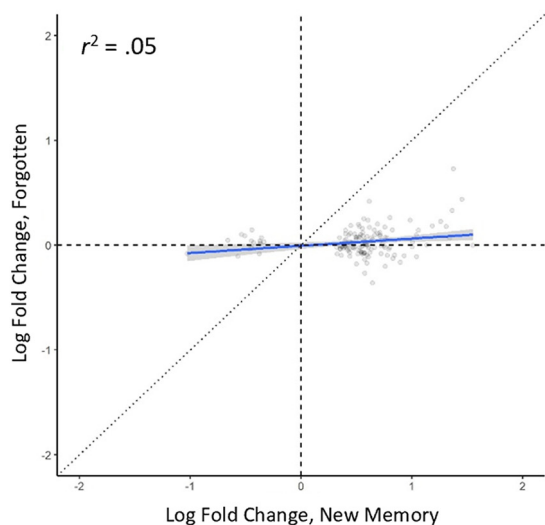
(2) As the new-memory group formed the focal point for comparisons, we ensured results would generalize by varying whether comparisons were made to the new-memory group or to the archival new-memory group.

(3) To ensure our results were not incomplete due to our discovery-based microarray approach we ran the analysis not only broadly on all 26091 unique transcripts on the array but also narrowly on 1198 transcripts we have previously identified as regulated by sensitization.

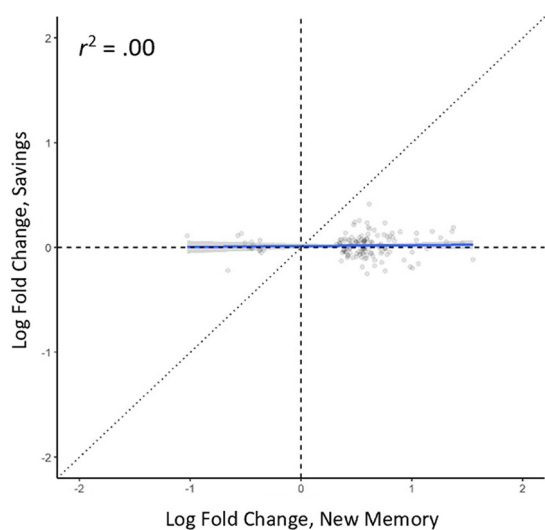
(4) Finally, we varied the stringency with which we flagged regulated transcripts, varying both the use of corrections for multiple comparison and the stringency of the null hypothesis (interval null of $\pm 10\%$ vs a standard null of no change).

These variations provided 528 different analysis specifications. We examined how results varied across these different analysis specifications.

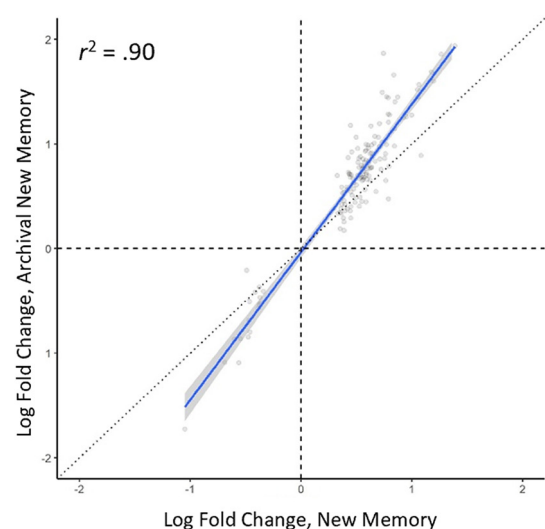
A Compare New to Forgotten



B Compare New to Savings



C Compare New to Archival New



Similarity based on overlap consistently supported forgetting as retrieval failure. The proportion of overlap between savings and the new-memory conditions was always modest, ranging from 0 up to 0.12. Moreover, overlap was always similar to or weaker than in the forgotten-memory conditions: $P_{\text{diff}} \leq 0.006$ in all specifications.

Similarity based on ordered lists was also very stable, with no specifications indicating a statistically significant positive association in regulation between the new and savings phases of memory.

Similarity based on correlation also consistently supported forgetting as retrieval failure. The correlation in regulation between the savings-memory and new-memory condition was never more than moderate (r values ranged from -0.15 to 0.26). Moreover, the relationship was never substantially stronger than what was observed at forgetting: $r_{\text{savings-forgotten}}$ was negative in 319 of 528 analyses, between 0 and 0.1 in 200 analyses, and between 0.1 and 0.21 in nine analyses. Thus, no analysis specification yielded an increase in similarity that met our preregistered criteria of at least 0.25 to indicate modest support for forgetting as decay.

Finally, leaving out each sample did not reveal any outlier that produced a strong or consistent influence on measures of similarity. In particular, we examined the effect of dropping the one sample from the forgotten-memory condition which had shown some habituation one week after training. This did not systematically alter any of the similarity measurements. Overall, our multiverse analysis suggests that our findings are robust to a variety of reasonable analysis specifications.

Discussion

Sensitization training produces complex waves of transcriptional change. This transcriptional response is essential for creating a long-term memory, as blocking transcription during training abolishes long-term memory (Sutton et al., 2001), the long-term facilitation that mediates sensitization (Montarolo et al., 1986; Castellucci et al., 1989), and the structural correlates of long-term memory (Bailey et al., 1992). Moreover, several of the transcripts initially activated by sensitization training are essential for inducing the long-term facilitation thought to contribute to the behavioral expression of sensitization (e.g., C/EBP: Alberini et al., 1994; CREB1: Dash et al., 1990). This does not establish that transcriptional changes maintain long-term sensitization memory, but it is clear that transcriptional changes are required to activate maintenance mechanisms.

continued

Figure 6. Similarity of conditions measured by correlation in LFCs. These figures show the correlation in LFC between the new-memory group and the forgotten-memory group (A), the savings-memory group (B), and the archival new-memory group (C). Only the 148 transcripts clearly regulated in the new-memory group are shown. The black dots are individual transcripts; the blue line is the line of best fit with shading indicating the 95% CI; the diagonal line shows a 1:1 relationship that would occur with perfect similarity.

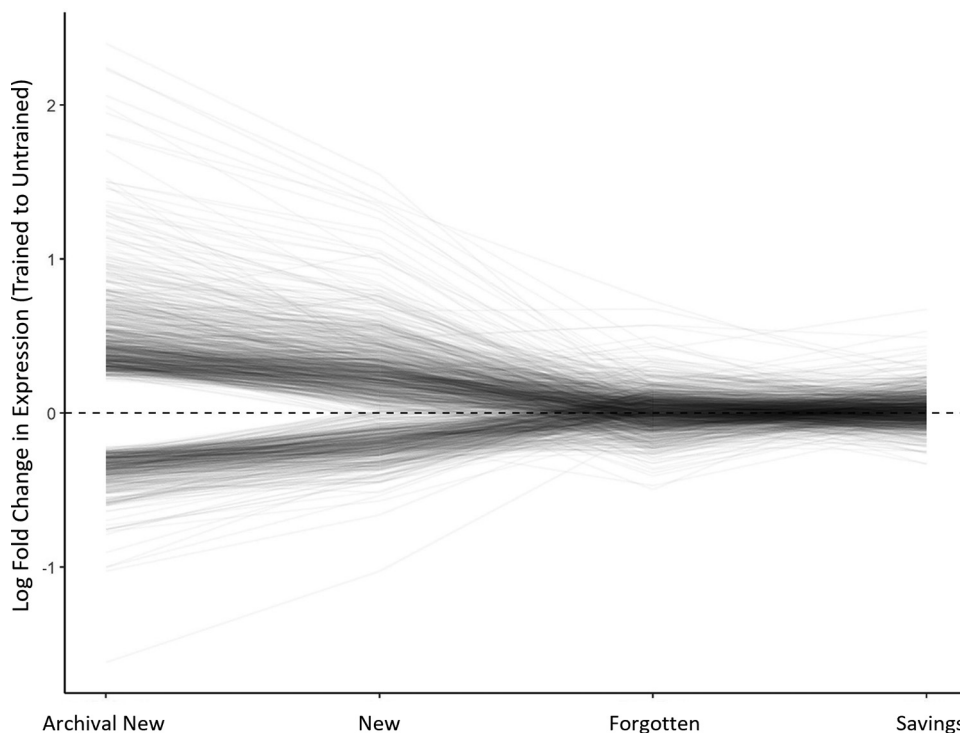


Figure 7. Exploratory analysis of fate of transcripts previously identified as regulated 1 d after sensitization training. This graph shows mean LFC (trained vs untrained) for 1198 transcripts which showed clear regulation in the archival new-memory condition (Conte et al., 2017), tracking their expression in the current new, forgetting, and savings groups. The dashed line at 0 represents no change in expression. Comparing the archival new-memory group to the current new-memory group shows some regression to the mean but that most transcripts show similar regulation. These transcriptional changes fade during forgetting (most transcripts collapse toward 0), and there is no obvious perturbation in expression when savings memory is induced.

As sensitization is forgotten, the transcriptional changes produced by training also fade away. Here, we find that sensitization memory can be persistently reactivated (savings memory) without reactivating storage-related transcriptional changes. Savings memory seems mechanistically distinct from encoding a new memory, few transcripts are coregulated at both phases of memory and overall transcriptional states show only negligible similarity.

One limitation of this study is that power to detect clearly regulated transcripts was lower than anticipated. This means that our finding of no transcriptional changes during savings is tenuous; savings may activate some transcriptional changes that were missed.

This shortcoming did not limit our ability to compare transcriptional states. First, we planned for assessments of similarity based on global patterns of regulation, not just on the statistical-significance status of transcripts. We found that our measures of similarity could detect when samples had been treated alike, always showing very strong similarity between the new memory and archival new-memory conditions. In addition, our findings are robust to a range of reasonable analysis specifications, including ones which more aggressively classify transcripts as regulated.

Overall, our results are not consistent with decay models of forgetting. If the sensitization memory trace substantially decays during forgetting, savings would require

Table 2: Transcriptional changes after forgetting of sensitization

Transcript	Description	Previous result	Forgetting	Savings
Z15041.1	ApBiP	0.63 ± 0.46	-0.04 ± 0.71	0.07 ± 0.31
M11283.1	FMRFa	0.53 ± 0.51	0.41 ± 0.26	0.67 ± 0.18
EB257711.1	LOC101857556	0.52 ± 0.46	0.26 ± 0.24	0.11 ± 0.32
EB254334.1	Transcribed locus	0.51 ± 0.50	0.73 ± 0.31	0.19 ± 0.58
FF066943.1	LOC106013098	0.36 ± 0.42	0.57 ± 0.13	0.49 ± 0.14
EB243511.1	Transcribed locus	-0.23 ± 0.50	0.00 ± 0.22	-0.21 ± 0.24
EB342172.1	Transcribed locus	-0.62 ± 0.35	-0.17 ± 0.23	0.08 ± 0.25

This table shows results for a set of seven transcripts previously identified via microarray and qPCR as being regulated after forgetting of sensitization. The previous result column is taken from Patel et al. (2018); it is qPCR data from pleural ganglia harvested 7 d after training, when sensitization had been forgotten. The forgetting and savings columns microarrays are from the current study, both of which were harvested 8 d after training. Numeric results are LFCs in expression ± the 95% margin of error (expanded for multiple comparisons).

at least partly rebuilding it. We found no evidence for this: long-term savings does not reactivate the transcriptional changes observed with an initial long-term memory. This suggests that the memory trace for sensitization remains reasonably intact and that forgetting of sensitization in *Aplysia* is because of retrieval failure.

This conclusion may seem puzzling: how could the memory trace be intact if there are almost no transcriptional changes that persist during forgetting and savings? The answer may be that the memory trace is molecularly sparse. For example, a recent screen found only two clear transcriptional changes in the hippocampus one week after fear conditioning, a time point when behavioral expression of the memory would still be quite strong (Mizuno et al., 2020). Thus, although learning initially produces widespread transcriptional changes, these may be refined to a relatively small core for maintenance. It is also possible that the maintenance mechanisms are not transcriptional at all, but merely require transcription to be initiated. For example, one (controversial) possibility is that the memory trace for sensitization is epigenetic, and requires transcriptional changes only for behavioral expression (Pearce et al., 2017). We are now working to determine whether the long-term expression of savings requires any transcriptional changes by injecting a transcriptional inhibitor just before the reminder shock used to evoke savings.

In retrieval-failure accounts of forgetting, memory traces persist yet gradually become decoupled from behavior. How might that happen in long-term sensitization? We have previously found that the transcriptional response to sensitization includes not only changes likely to promote memory but also changes likely to limit its behavioral expression. Specifically, we have found that sensitization training produces a robust and very long-lasting increase in the mRNA encoding FMRFa (Conte et al., 2017; Patel et al., 2018). In *Aplysia* FMRFamide is an inhibitory neuromodulator. It inhibits the VC neurons that help encode sensitization memory, depresses synapses, and decreases the strength of the T-SWR response (Abrams et al., 1984; Fioravante et al., 2006). This suggests the possibility that sensitization training not only builds the memory trace but also promotes inhibitory processes that can impair retrieval, a form of active forgetting (Davis and Zhong, 2017). If this is correct, it should be possible to manipulate the forgetting of sensitization memory by altering FMRFa signaling; we are now working to test this hypothesis.

Retrieval-failure accounts of forgetting propose that the memory trace is enduring, but that accessibility to retrieval is highly dynamic. Learning is thought to produce an initially accessible memory, time and new learning then decrease accessibility, but a variety of experiences can increase accessibility for short or long durations (context re-exposure, brief re-training, reminders, etc.). These different dynamics suggest that storage and retrieval are organized at different levels of neuronal function. Multiple lines of research indicate that encoding a long-term memory activates at least two distinct storage mechanisms: changes in synaptic efficacy and changes in connectivity. For example, in *Aplysia*

long-term sensitization produces morphologic changes related to increased synaptic strength (increased active zone size and vesicle complement) and additional changes related to increased connectivity, such as new synaptic varicosities and active zones (Bailey and Chen, 1988). These decay at different rates, with changes in synaptic strength decaying over time while changes in connectivity endure (Bailey and Chen, 1989). In the mammalian visual system, changes in connectivity have specifically been shown to persist beyond forgetting (Linkenhoker et al., 2005; Hofer et al., 2009). One intriguing possibility, then, is that the memory trace is represented by enduring changes in connectivity while retrieval is made possible by more labile changes in synaptic strength. One testable prediction of this dual-process account of memory expression is that a generalized increase in activity could re-potentiate synapses and re-express a seemingly forgotten sensitization memory.

One critical question is how forgetting is related to other forms of memory disruption, such as the amnesia induced by disruptions of consolidation or reconsolidation. At the behavioral level, there are strong similarities, including the fact that memories which seem lost because of an amnesic intervention can often be re-expressed through a reminder or a brief retraining (Riccio and Richardson, 1984). On the other hand, current evidence indicates induced amnesia may be because of storage failure (Haubrich et al., 2020). Amnesic agents tend to degrade the physical correlates of memory. In contrast to our findings with forgetting, memory recovery after an amnesic agent tends to refresh the physical correlates of memory (Haubrich et al., 2020). Thus, forgetting and memory impairment through disrupted consolidation may be quite distinct. One difficulty with interpreting these results is that it is not yet possible to demarcate which brain correlates of memory are related to storage and which are related to accessibility. Still, it seems possible that induced amnesia and forgetting are mechanistically distinct. It would be useful to directly compare these processes in the same system.

Grand debates in science are never settled by one study. Our results strongly implicate retrieval failure as the mechanism of forgetting for sensitization in *Aplysia*, but a definitive account of forgetting will require additional studies across organisms and learning paradigms. The strategy of monitoring neuronal changes across multiple memory states seems likely to be fruitful for better resolving how and why forgetting occurs.

References

- Abrams TW, Castellucci VF, Camardo JS, Kandel ER, Lloyd PE (1984) Two endogenous neuropeptides modulate the gill and siphon withdrawal reflex in *Aplysia* by presynaptic facilitation involving cAMP-dependent closure of a serotonin-sensitive potassium channel. *Proc Natl Acad Sci USA* 81:7956–7960.
- Alberini CM, Ghirardi M, Metz R, Kandel ER (1994) C/EBP is an immediate-early gene required for the consolidation of long-term facilitation in *Aplysia*. *Cell* 76:1099–1114.
- Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 7:55–65.

- Ally BA, Hussey EP, Ko PC, Molitor RJ (2013) Pattern separation and pattern completion in Alzheimer's disease: evidence of rapid forgetting in amnesic mild cognitive impairment. *Hippocampus* 23:1246–1258.
- Antzoulatos EG, Wainwright ML, Cleary LJ, Byrne JH (2006) Long-term sensitization training primes *Aplysia* for further learning. *Learn Mem* 13:422–425.
- Bahrnick HP (1984) Semantic memory content in permastore: fifty years of memory for Spanish learned in school. *J Exp Psychol Gen* 113:1–29.
- Bailey CH, Chen M (1988) Long-term memory in *Aplysia* modulates the total number of varicosities of single identified sensory neurons. *Proc Natl Acad Sci USA* 85:2373–2377.
- Bailey CH, Chen M (1989) Time course of structural changes at identified sensory neuron synapses during long-term sensitization in *Aplysia*. *J Neurosci* 9:1774–1780.
- Bailey CH, Montarolo P, Chen M, Kandel ER, Schacher S (1992) Inhibitors of protein and RNA synthesis block structural changes that accompany long-term heterosynaptic plasticity in *Aplysia*. *Neuron* 9:749–758.
- Barron HC, Vogels TP, Behrens TE, Ramaswami M (2017) Inhibitory engrams in perception and memory. *Proc Natl Acad Sci USA* 114:201701812.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57:289–300.
- Bonnick K, Bayas K, Belchenko D, Cyriac A, Dove M, Lass J, McBride B, Calin-Jageman IE, Calin-Jageman RJ (2012) Transcriptional changes following long-term sensitization training and in vivo serotonin exposure in *Aplysia californica*. *PLoS One* 7:e47378.
- Buonomano DV, Cleary LJ, Byrne JH (1992) Inhibitory neuron produces heterosynaptic inhibition of the sensory-to-motor neuron synapse in *Aplysia*. *Brain Res* 577:147–150.
- Calin-Jageman RJ, Cumming G (2019a) Estimation for better inference in neuroscience. *eNeuro* 6:ENEURO.0205-19.2019.
- Calin-Jageman RJ, Cumming G (2019b) The new statistics for better science: ask how much, how uncertain, and what else is known. *Am Stat* 73:271–280.
- Castellucci VF, Blumenfeld H, Goelet P, Kandel ER (1989) Inhibitor of protein synthesis blocks long-term behavioral sensitization in the isolated gill-withdrawal reflex of *Aplysia*. *J Neurobiol* 20:1–9.
- Cleary LJ, Byrne JH (1993) Identification and characterization of a multifunction neuron contributing to defensive arousal in *Aplysia*. *J Neurophysiol* 70:1767–1776.
- Conte C, Herdegen S, Kamal S, Patel J, Patel U, Perez L, Rivota M, Calin-Jageman RJ, Calin-Jageman IE (2017) Transcriptional correlates of memory maintenance following long-term sensitization of *Aplysia californica*. *Learn Mem* 24:502–515.
- Dash PK, Hochner B, Kandel ER (1990) Injection of the cAMP-responsive element into the nucleus of *Aplysia* sensory neurons blocks long-term facilitation. *Nature* 345:718–721.
- Davis RL, Zhong Y (2017) The biology of forgetting—a perspective. *Neuron* 95:490–503.
- Ebbinghaus HE (1885) *Memory: a contribution to experimental psychology*. New York: Dover Publications.
- Fioravante D, Smolen PD, Byrne JH (2006) The 5-HT- and FMRFa-activated signaling pathways interact at the level of the Erk MAPK cascade: potential inhibitory constraints on memory formation. *Neurosci Lett* 396:235–240.
- Fitzgerald Z, Mohamed A, Ricci M, Thayer Z, Miller L (2013) Accelerated long-term forgetting: a newly identified memory impairment in epilepsy. *J Clin Neurosci* 20:1486–1491.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.
- Goelet P, Castellucci VF, Schacher S, Kandel ER (1986) The long and the short of long-term memory—a molecular framework. *Nature* 322:419–422.
- Hardt O, Nader K, Nadel L (2013) Decay happens: the role of active forgetting in memory. *Trends Cogn Sci* 17:111–120.
- Haubrich J, Bernabo M, Baker AG, Nader K (2020) Impairments to consolidation, reconsolidation, and long-term memory maintenance lead to memory erasure. *Annu Rev Neurosci* 43:297–314.
- Hedges LV (1981) Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Behav Stat* 6:107–128.
- Herdegen S, Conte C, Kamal S, Calin-Jageman RJ, Calin-Jageman IE (2014a) Immediate and persistent transcriptional correlates of long-term sensitization training at different CNS loci in *Aplysia californica*. *PLoS One* 9:e114481.
- Herdegen S, Holmes G, Cyriac A, Calin-Jageman IE, Calin-Jageman RJ (2014b) Characterization of the rapid transcriptional response to long-term sensitization training in *Aplysia californica*. *Neurobiol Learn Mem* 116:27–35.
- Hofer SB, Mrcic-Flogel TD, Bonhoeffer T, Hübener M (2009) Experience leaves a lasting structural trace in cortical circuits. *Nature* 457:313–317.
- Holmes G, Herdegen S, Schuon J, Cyriac A, Lass J, Conte C, Calin-Jageman IE, Calin-Jageman RJ (2014) Transcriptional analysis of a whole-body form of long-term habituation in *Aplysia californica*. *Learn Mem* 22:11–23.
- Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *J Comput Graph Stat* 5:299–314.
- Kuhl D, Kennedy TE, Barzilai A, Kandel ER (1992) Long-term sensitization training in *Aplysia* leads to an increase in the expression of BiP, the major protein chaperon of the ER. *J Cell Biol* 119:1069–1076.
- Langaas M, Lindqvist BH, Ferkingstad E (2005) Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J R Stat Soc Series B Stat Methodol* 67:555–572.
- Linkenhoker BA, von der Ohe CG, Knudsen EI (2005) Anatomical traces of juvenile learning in the auditory system of adult barn owls. *Nat Neurosci* 8:93–98.
- Liu QR, Hattar S, Endo S, MacPhee K, Zhang H, Cleary LJ, Byrne JH, Eskin A (1997) A developmental gene (Tolloid/BMP-1) is regulated in *Aplysia* neurons by treatments that induce long-term sensitization. *J Neurosci* 17:755–764.
- Mackey SL, Glanzman DL, Small S, Dyke aM, Kandel ER, Hawkins RD (1987) Tail shock produces inhibition as well as sensitization of the siphon-withdrawal reflex of *Aplysia*: possible behavioral role for presynaptic inhibition mediated by the peptide Phe-Met-Arg-Phe-NH₂. *Proc Natl Acad Sci USA* 84:8730–8734.
- Mary A, Schreiner S, Peigneux P (2013) Accelerated long-term forgetting in aging and intra-sleep awakenings. *Front Psychol* 4:1–11.
- McCarthy DJ, Smyth GK (2009) Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25:765–771.
- Menges S, Riepe JR, Philips GT (2015) Latent memory facilitates re-learning through molecular signaling mechanisms that are distinct from original learning. *Neurobiol Learn Mem* 123:35–42.
- Mizuno K, Jeffries AR, Abel T, Peter Giese K (2020) Long-lasting transcription in hippocampal area CA1 after contextual fear conditioning. *Neurobiol Learn Mem* 172:107250.
- Montarolo PG, Goelet P, Castellucci VF, Morgan J, Kandel ER, Schacher S (1986) A critical period for macromolecular synthesis in long-term heterosynaptic facilitation in *Aplysia*. *Science* 234:1249–1254.
- Moroz LL, Edwards JR, Puthanveetil SV, Kohn AB, Ha T, Heyland A, Knudsen B, Sahni A, Yu F, Liu L, Jezzini S, Lovell P, Iannuccilli W, Chen M, Nguyen T, Sheng H, Shaw R, Kalachikov S, Panchin YV, Farmerie W, et al. (2006) Neuronal transcriptome of *Aplysia*: neuronal compartments and circuitry. *Cell* 127:1453–1467.
- Nelson TO (1985) Ebbinghaus's contribution to the measurement of retention. Savings during relearning. *J Exp Psychol Learn Mem Cogn* 11:472–479.
- Patel U, Perez L, Farrell S, Steck D, Jacob A, Rosiles T, Krause E, Nguyen M, Calin-Jageman RJ, Calin-Jageman IE (2018)

- Transcriptional changes before and after forgetting of a long-term sensitization memory in *Aplysia californica*. *Neurobiol Learn Mem* 155:474–485.
- Pavlidis P, Li Q, Noble WS (2003) The effect of replication on gene expression microarray experiments. *Bioinformatics* 19:1620–1627.
- Pearce K, Cai D, Roberts AC, Glanzman DL (2017) Role of protein synthesis and DNA methylation in the consolidation and maintenance of long-term memory in *Aplysia*. *Elife* 6:e18299.
- Perez L, Patel U, Rivota M, Calin-Jageman IE, Calin-Jageman RJ (2018) Savings memory is accompanied by transcriptional changes that persist beyond the decay of recall. *Learn Mem* 25:1–5.
- Phillips GT, Tzvetkova EI, Marinesco S, Carew TJ (2006) Latent memory for sensitization in *Aplysia*. *Learn Mem* 13:224–229.
- Revelle W (2018) psych: procedures for personality and psychological research. Northwestern University, Evanston, Illinois. Available at <https://CRAN.R-project.org/package=psych>.
- Riccio DC, Richardson R (1984) The status of memory following experimentally induced amnesias: gone, but not forgotten. *Psychobiology* 12:59–72.
- Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 23:2700–2707.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* 43:e47.
- Schaefer M, Picciotto MR, Kreiner T, Kaldany RR, Taussig R, Scheller RH (1985) *Aplysia* neurons express a gene encoding multiple FMRFamide neuropeptides. *Cell* 41:457–467.
- Scholz K, Byrne J (1987) Long-term sensitization in *Aplysia*: biophysical correlates in tail sensory neurons. *Science* 235:685–687.
- Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:1–25.
- Smyth GK (2005) Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pp 397–420. New York: Springer.
- Smyth GK, Speed T (2003) Normalization of cDNA microarray data. *Methods* 31:265–273.
- Sutton MA, Masters SE, Bagnall MW, Carew TJ (2001) Molecular mechanisms underlying a unique intermediate phase of memory in *Aplysia*. *Neuron* 31:143–154.
- Tröster AI, Butters N, Salmon DP, Cullum CM, Jacobs D, Brandt J, White RF (1993) The diagnostic utility of savings scores: differentiating Alzheimer's and Huntington's diseases with the logical memory and visual reproduction tests. *J Clin Exp Neuropsychol* 15:773–788.
- Tsai CA, Hsueh H, Chen JJ (2003) Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* 59:1071–1081.
- Walters ET, Erickson MT (1986) Directional control and the functional organization of defensive responses in *Aplysia*. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 159:339–351.
- Walters ET, Bodnarova M, Billy AJ, Dulin MF, Díaz-Ríos M, Miller MW, Moroz LL (2004) Somatotopic organization and functional properties of mechanosensory neurons expressing sensorin-A mRNA in *Aplysia californica*. *J Comp Neurol* 471:219–240.
- Wixted JT (2004) The psychology and neuroscience of forgetting. *Annu Rev Psychol* 55:235–269.
- Yang X, Scheid S, Lottac C (2019) OrderedList: similarities of ordered gene lists. R package version. Available at <http://compdiag.molgen.mpg.de/software/OrderedList.shtml>.
- Zahurak M, Parmigiani G, Yu W, Scharpf RB, Berman D, Schaeffer E, Shabbeer S, Cope L (2007) Pre-processing Agilent microarray data. *BMC Bioinformatics* 8:142.