Failure to classically condition planarian flatworms

Zachary S. Kelso[1¶*], Madeleine C. Snyder[1¶], Samuel J. Gershman[1]

1 Department of Psychology and Center for Brain Science, Harvard University, Cambridge, Massachusetts, United States of America

¶These authors contributed equally to this work.

* Corresponding author: zkelso@fas.harvard.edu

Abstract

Planarian flatworms represent one of the most evolutionarily informative nervous systems for an account of ancient bilaterian brains. Likewise, the unparalleled regenerative ability of planarians makes possible certain investigations of neural development, memory, and behavior that are simply impossible with other model organisms. Despite these facts, learning and memory are today underexplored in planarians, likely due in part to the shadow of controversial 20th-century experiments on the transfer of memories between individual flatworms. Here, we attempted to replicate and extend the classical conditioning experiments in planarians that were the basis of the later memory transfer work. We failed to find evidence for classical conditioning in any of our procedural variations and obtained similar results using computer vision methods to avoid subjectivity in manual video annotation. Our results cast doubt on the suitability of planarian flatworms for studying primitive learning processes and the molecular basis of memory using classical conditioning.

1 Introduction

Evolutionary and comparative neuroscience suggest that planarian nervous systems represent the most basal, Urbilaterian-like brains of any contemporary organism [1–3] and can serve as a worthy model for the precursor vertebrate brain [4]. Consequently, planarians offer a window into the evolutionary origins of brain functions, including cognitive functions like learning, memory, and decision making. The twentieth-century discovery that planarians appear to be capable of learning and memory was a watershed in this research program. Even more remarkably, the apparent retention of memory following decapitation and head regeneration, and the purported transferability of memories between individuals via RNA injections and cannibalism, suggested that planarian brains may shed light on unconventional mechanisms of memory storage. However, these claims have been mired in criticism and controversy, and contemporary research on planarian learning and memory is limited. The goal of this paper is to revisit basic questions about planarian learning, in an effort to resolve some of this controversy.

Historically, planarian learning and memory has been tested with various forms of classical and operant conditioning [5,6]. The classic experiment was conducted by James McConnell and Robert Thompson in 1955 [7], after which McConnell became the leading force in the planarian learning and memory transfer experiments [8–11]. These classical conditioning procedures employed light and shock as the conditioned and unconditioned stimuli, respectively, and measured learning by quantifying contraction responses to the onset of light following repeated stimulus pairings. Allan Jacobson extended these experiments with additional controls and improved procedures [12–15], for which Jacobson remarked that "[i]t appears reasonable to assert that by the strictest criteria, classical conditioning has been conclusively demonstrated in planarians" [14]. Nonetheless, several would-be replicators failed to classically condition planarians. Though a review by Corning and Kelly [6] counted 36 inductions of classical conditioning using the original light-shock paradigm as measured by conditioned response acquisition, these were weighed against seven published failures which included the mostly unsuccessful efforts of biochemist and Nobel laureate Melvin Calvin [16,17]. Other forays into planarian learning also turned up negative results [18]. By end of the 1970s, most planarian psychologists, including McConnell, had moved on to other topics or other model organisms [19]. The procedural ambiguities of how to reliably train and elicit

conditioned behavior from planarians were never clarified.

Modern attempts to classically condition or otherwise train planarians have been few in the years since the initial movement [20–26], even as planarians became a fixture of regenerative biology [27–32] and the subject of some neuroanatomical and pharmacological interest [33–36]. Research into planarians' cognitive capacities has been renewed by recent investigations into non-synaptic forms of memory [37–42], which has returned attention to the McConnell-era transfer experiments and apparent ability of planarians to remember learned information even after decapitation and whole-brain regeneration [43,44]. Nonetheless, the foundational classical conditioning experiments have not been replicated. This is surprising given the volume of the historical literature, the utility of planarians as a model of evolutionarily ancient brains, and the disruptive implications of regenerative memory retention to the neuroscience of memory.

We therefore sought to replicate the original planarian conditioning experiments in alignment with Allan Jacobson's protocol, including critical control conditions [12,13]. We extended this paradigm to test the efficacy of alternative stimuli and timing in evoking a conditioned responses with training. In all cases, we were were unable to find evidence for classical conditioning in planarians. To address the possibility that these null results arose from the subjectivity of manual scoring (with high interscorer variability), we created a computer vision pipeline to analyze worm behavior across several days of training. The results of the computer vision analyses were likewise null.

2 Results

In what follows, we describe a battery of stimulus pairings designed to replicate the initial protocols from the 1960s and leverage recent findings regarding planarian sensory capacities [12,25,45–47] to test other kinds of stimulus pairings. While our search space for a classical conditioning paradigm in planarians is far from exhaustive, our selection of assays show that salient stimuli across sensory domains fail to elicit increased conditioned response rates with repeated stimulus pairings. In addition, we found no difference in conditioned response acquisition across two species (*Girardia dorotocephala* and *Schmidtea mediterranea*) and three *G. dorotocephala* lineages (see Methods for

more details).

2.1 Classical light-shock conditioning paradigm fails to produce conditioned response in planarians

We recreated the light-shock conditioning paradigm, which consisted of a brief bout of light habituation followed by two days of paired exposure to light and shock, remaining as faithful to the original protocol as possible with modern equipment (see Methods and Fig. 1A-B). The paired conditioning group received repeated pairs of light followed by shock; the pseudoconditioning group received the same numbers of lights and shocks, but in a temporally randomized arrangement. This procedure controls for non-associative sensitization; if pseudoconditioned worms acquire a response to light, they may be due to a general increase in responsiveness rather than a stimulus-specific conditioned response.

In line with the original protocol, worm responses during the stimulus exposure were hand-scored (interscorer agreement: 91.92%). We found no significant increase in the conditioned response across days, but rather, a significant decrease in the conditioned response across days (two-sided per-worm slope test: $t(11) = -3.86, p = .003, n = 12$ worms; Page's $\mathrm{L} = 491.0, p = .998$; Fig. 1C). There was no significant difference in conditioned responses between the paired conditioning and unpaired pseudoconditioning conditions (ANOVA main effect of condition: $\mathrm{F}(1, 22) = 2.88, p = .104$). Our results were markedly different from the stark increase in conditioned responses observed in the original Jacobson study (Fig. 1D).

2.2 Multi-modal stimulus variants of classical conditioning fail to evoke a conditioned response

Because traditional light-shock methods failed to produce learning, we conducted further experiments using stimuli inspired by modern research on planarian sensory abilities related to UV light and classical studies using mechanical vibration. Thus, we trained with three new stimulus pairings: (1) white light and high-amplitude UV light, (2) blue light and high-amplitude UV, (3) low-amplitude ultraviolet (UV) light and shock, and (4) a water pulse and high-amplitude UV light. We suspected that these sets of stimuli might be more evolutionarily and ethologically relevant to

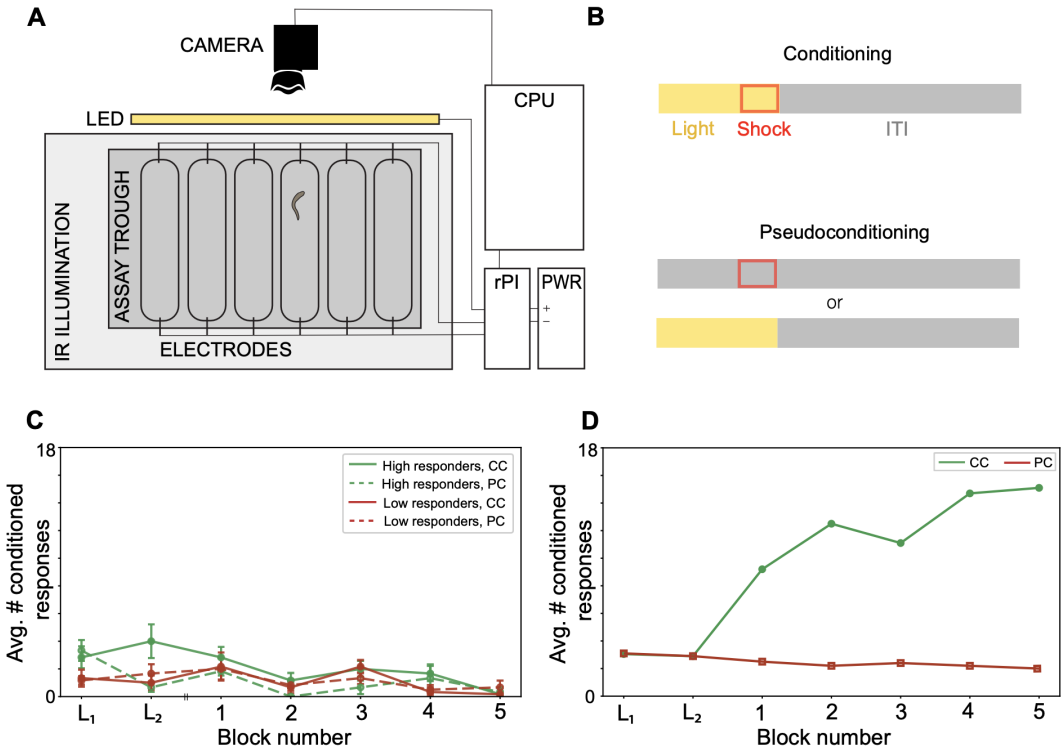

Fig 1. A. Experimental apparatus. A clear plastic trough with affixed electrodes was placed under a camera and stimulus light panel, with infrared background illumination from below. The camera, light and shock stimuli, and background illumination were controlled by a Raspberry Pi microprocessor connected to a computer. **B.** Presentation of stimuli. In the classical conditioning (CC) scenario, three seconds of light coterminate with one second of shock. In pseudoconditioning (PC), either light or shock occurs alone in random order. **C.** Experimental results. Error bars denote standard error of the mean (SEM) **D.** Adaptation of Jacobson's original data for comparison. Note that no error bars were specified in the original figure.

planarians than traditional stimuli.

Planarians have UV-sensitive extraocular opsins distributed across their bodies that mediate a rapid contraction in response to high-amplitude UV light exposure [46,48]. Low-amplitude and slowly modulating UV light evoke a more subtle and complex response, and are suitable as conditioned stimuli [36]. We first attempted to use high-amplitude UV light as the unconditioned stimulus because planarians do not habituate to phasic, high-amplitude UV exposure. No significant upward trend was observed for the UV-shock paradigm across blocks (Fig. 2A; mean slope $= -0.07 \pm 0.07$ contractions/block, 95% CI $[-0.22, +0.08]$, $t(17) = -0.95$, $p = 0.357$; equivalence testing confirmed the slope was negligible within our pre-specified threshold of ± 1 contraction/block, $p < .001$; Bayesian analysis yielded $\mathrm{BF}_{01} = 1.96$; $n = 18$ worms). We then used low-amplitude UV light as the conditioned stimulus, and shock as the unconditioned stimulus. Again, we found no acquisition of the conditioned response during training (Fig. 2A; mean slope $= -0.67 \pm 0.32$ contractions/block, 95% CI $[-1.36, +0.03]$, $t(11) = -2.11$, $p = .059$; equivalence testing was inconclusive, $p = .157$; $\mathrm{BF}_{01} = 0.38$; $n = 12$ worms; Fig. 2B). Finally, we used a water pulse stimulus as the conditioned stimulus paired with high-amplitude UV light to mimic sudden environmental disturbance and subsequent radiation exposure. This paradigm also failed to yield a robust conditioned response during training, but a small positive slope was detected (Fig. 2B; mean slope $= +0.46 \pm 0.20$ contractions/block, 95% CI $[-0.10, +1.02]$, $t(4) = +2.28$, two-sided $p = .085$; equivalence testing confirmed the slope was within the pre-specified threshold of ± 1 contraction/block, $p = .028$; $\mathrm{BF}_{10} = 3.48$; $n = 5$ worms). This statistically significant upward trend in the pulse-UV condition is likely confounded by movement or sensitization due to the water pulse stimulus itself, which in our paradigm was not dissociable from the conditioned response. As this was the most ethological paradigm, it may be worth investigating this further if the confound of movement due to the pulse can be removed.

2.3 Worm species and origin have no observable effect on learning

Literature from the 1960s claimed that worm origin, species, and feeding interval (among other factors) are important for the success of planarian learning [49]. To test whether worm provenance and species changed our outcomes, we collected specimens from three locations across the United

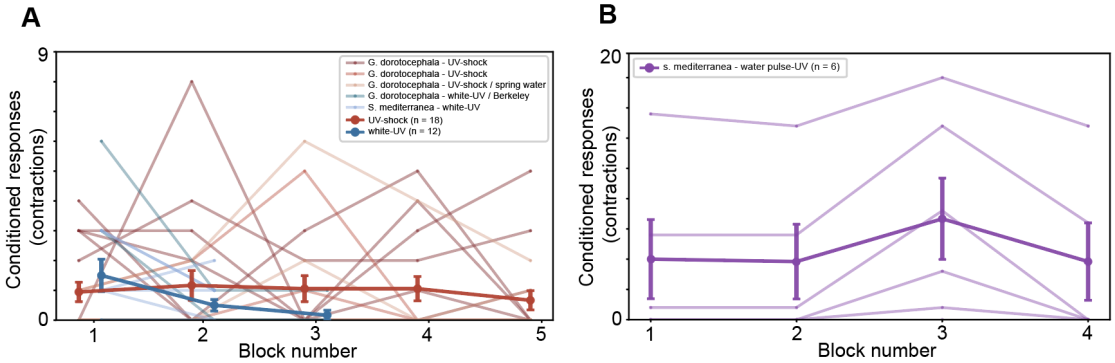

Fig 2. A. Conditioned response acquisition in response to alternative conditioning paradigms (UV light, white light, shock), using multiple sources of worms. **B.** Conditioned response acquisition in response to water pulse (CS) and UV light (US). Error bars denote standard error of the mean (SEM).

States, including the location (to the best of our knowledge) where McConnell procured his population of planarians. Commercially available *G. dorotocephala* from North Carolina, wild-caught Michigan *G. dorotocephala*, wild-caught California *G. dorotocephala* and laboratory *S. mediterranea* worms were used across our previously described experiments and showed indistinguishable levels of responding. While different species demonstrated notably different general activity levels, we did not find that a particular species or provenance had an effect on the acquisition of conditioned responses (Fig. 2A; Fig. S1). Notably, laboratory *S. mediterranea* are from the Mediterranean [50], whereas *G. dorotocephala* are native to North America, and their evolutionary distance is presumably vast, possibly as much 100 million years. [51]. The failure of American (Michigan, California, and Carolina strains) or Mediterranean species to demonstrate learning of conditioned responses suggests that this failure is not due to a species-specific idiosyncrasy.

2.4 Individual differences in sensitivity of unconditioned response to light do not affect the acquisition of the conditioned response

To investigate whether individual variability in response sensitivity to the white light stimulus may have obscured evidence of learning [52], we assayed six groups of six worms each with repeated presentations of the conditioned stimulus alone, and sorted them into low and high responders (Fig. S2). The groups had a significantly different probability of reacting to the conditioned stimulus

(white light) presentation (independent samples t-test, t = 5.58, p ¡ 0.001). We exposed both types of responders to the original light-shock conditioning paradigm and found no significant upward trend in turns or contractions for either the conditioning group (where we found a significant decrease in turns and contractions; two-sided per-worm slope $t(11) = -3.86, p = .003$) or pseudoconditioning group ($t(11) = -1.82, p = .097$), and slopes did not significantly differ between high and low responders in either the conditioning (Welch's $t(10) = -0.27, p = .794, d = -0.16$) or pseudoconditioning condition (Welch's $t(10) = +0.51, p = .619, d = +0.30$). To minimize the risk of latent inhibition arising from stimulus pre-exposure during the responsiveness assay, we spaced the assay and the conditioning paradigm 2–3 days apart, during which the worms had only exposure to ambient light. Across the groups with high enough response to discern from zero, we found no significant habituation to the conditioned stimulus (but see [6], Table VIII). The mean response rate increased significantly from Day 1 (11.8% ± 2.0% SEM) to Day 2 (17.5% ± 2.5% SEM; paired t(35) = 2.39, p = 0.022, two-sided; Cohen's $d_z = 0.40$). While this increase was consistent across five of six troupes, no individual troupe reached significance alone. Slope analysis at the trial level did not reveal any significant gradual trend across trials (mean slope = 0, t(35) = 0.16, p = 0.87), suggesting the increase reflected a between-session shift rather than a continuous within-session change (Fig. S3).

2.5 Behavioral scores are not consistent across trained observers

Having found no evidence of learning as measured through increases in conditioned response (contraction) rate, we endeavored to determine the extent to which different scorers could agree on what types of behavior count as turns and contractions. Three scorers scored a sample of 203 videos of our footage (based on our reproduction of the James et al. protocol) and timestamps of turns and contraction events were analyzed (see Methods). While the total sums of turns and contractions for each video essentially the same for each scorer (Fig. S4), an analysis of timestamps reveals severe disagreement. The primary result is that unanimous agreement of contraction events (all scorers marked a given event in a given video as a contraction) is only 28.15% (67/238). The agreement for turn events is even lower at 15.14% (106/700). These results echo those of Arlene Hartry [18], who similarly examined unanimous contraction agreement between three observers of

planarian conditioning footage and found that so "few responses fell into this category" that they were statistically irrelevant. The low agreement of multiple scorers is in marked contrast to conventional classical conditioning procedures, where it is understood that conditioned responses should be relatively obvious to the trained eye.

To further examine interscorer agreement at the timestamp level, we computed F1 metrics for each pairwise combination of the three scorers' annotations for each video. The F1 metric is the harmonic mean of precision and recall measures (Methods). In the context of our data, precision is the number of the Test Scorer's events (turns or contractions) that the Reference Scorer also marked. Recall is the proportion of Reference events that the Test Scorer correctly identified out of all of the Reference's events. These proportions describe how much the Test Scorer is able to capture a Reference Scorer's data while minimizing the detection of non-reference events. If a pair of scorers agreed perfectly, an F1 score of 1 would be expected, and severe disagreement would yield a 0. We also quantified event confusion based on the proportion of unpaired reference events that have a temporally proximate test event of the opposite type in the test events. These instances occur when the Test and Reference scorer recognize the same event as having occurred, but categorize it differently.

Table 1 shows the recall, precision, F1, and confusion rates for each scorer pair and event type. Broadly, these measures indicate quite poor agreement about when events occur across scorers: on average, a given scorer detects about half of the turns (52.3%) or contractions (59.7%) that another scorer marks, and produces an equal amount that their reference fails to detect (55.5% for turns, 57.8% for contractions). To compare contraction agreement in response to the conditioned stimulus versus the unconditioned stimulus, Scorer 1 and Scorer 2 additionally examined 198 clips from the same videoset of the shock stimuli. For shock-induced contractions, Scorer 1 and Scorer 2 agreed on 86.78% of trials (172/198). The discrepancy between high agreement about shock-induced contractions but low agreement about (supposedly) light-induced contractions does not inspire confidence that the CS-period contractions are specific, learned responses to light and not just ambient, noisy movement. The sum of these results suggest that manual assessment of planarian behavior is not trustworthy for our experimental footage. That is, when multiple scorers observe and describe the same experimental footage, their reports are not reliably similar, even if the raw

counts of turns and contractions are stable across scorers.

Behavior	Reference	Test	Recall	Precision	F1	Confusion Rate
Turns	Scorer 1	Scorer 2	0.540	0.745	0.626	0.029
Turns	Scorer 2	Scorer 3	0.397	0.640	0.490	0.009
Turns	Scorer 3	Scorer 1	0.632	0.284	0.392	0.000
Turns	**Average**	–	**0.523**	**0.556**	**0.503**	**0.013**
Contractions	Scorer 1	Scorer 2	0.827	0.602	0.697	0.000
Contractions	Scorer 2	Scorer 3	0.492	0.556	0.522	0.031
Contractions	Scorer 3	Scorer 1	0.473	0.576	0.519	0.178
Contractions	**Average**	–	**0.597**	**0.578**	**0.579**	**0.070**

Table 1. Pairwise comparison of scorer agreement across behaviors as measured by recall, precision, and F1 scores of event timestamps of 203-216 sample training videos. Note that when a reference and test scorer are flipped, the recall and precision values likewise swap places, but the F1 score remains the same.

2.6 Feature-based quantification of the conditioned response fails to provide evidence for learning

A recent study reported success conditioning planarians as measured by an increase of conditioned responses with training using a modified protocol adapted from fear conditioning experiments in rodents [25]. We attempted to reproduce this protocol in our laboratory, using both manual annotation and a novel computer vision pipeline (Fig. 3A) to quantify the acquisition of the conditioned response. From the hand-scored data we found that neither conditioned turns nor contractions showed a significant upward trend over training days in the conditioning group (turns: per-worm slope test, $t(47) = -0.78, p = 0.78$, one-sided; Page's L $= 1195.0, p = 0.60$; contractions: $t(47) = -0.29, p = 0.61$; Page's L $= 1204.5, p = 0.41; n = 48$ worms) or the pseudoconditioning group (turns: $t(11) = -1.11, p = 0.86$; contractions: $t(11) = 0.35, p = 0.37; n = 12$ worms). Linear mixed models (LMMs) confirmed no significant Day $\times$ Condition interaction for either turns ($\beta = -0.18, z = -0.21, p = 0.83$) or contractions ($\beta = 0.12, z = 0.36, p = 0.72$), indicating that the conditioning group did not acquire responses at a different rate than controls. The two groups did differ in overall turn frequency (LMM: $\beta = -7.65, z = -3.12, p = 0.002$), with the conditioning group responding more than controls, but this difference was present from the outset and did not change with training (Fig. 3B).

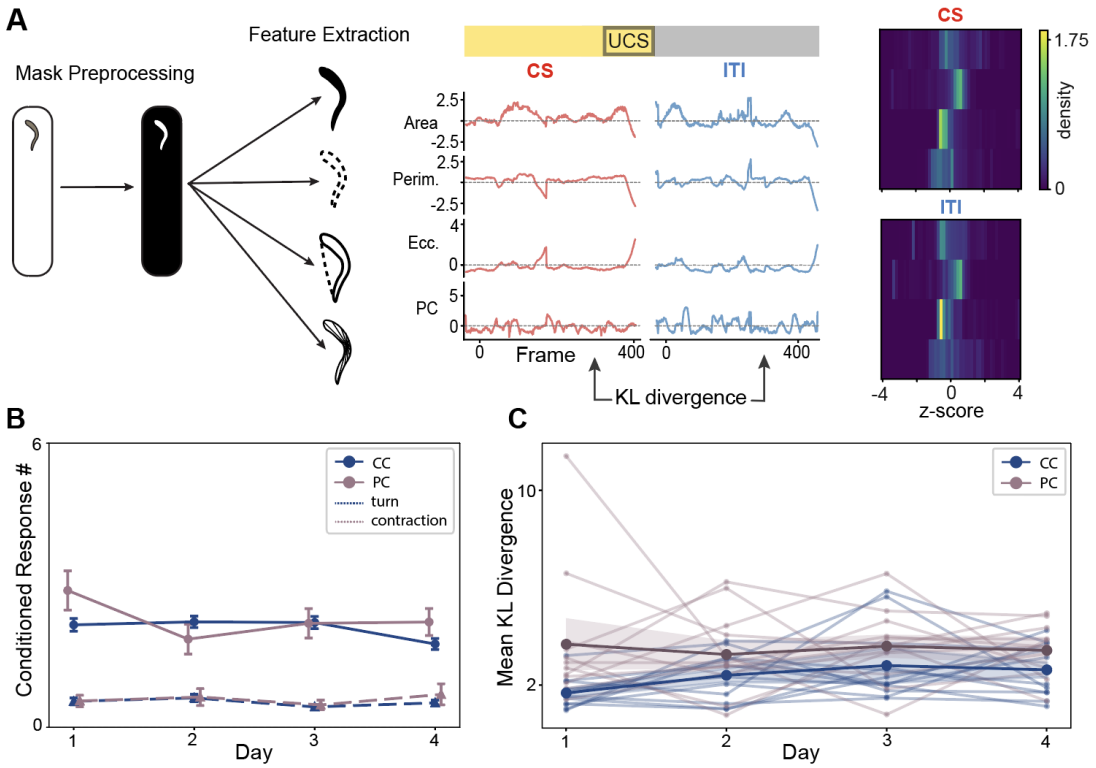

Fig 3. A. Feature extraction pipeline for quantifying conditioned response acquisition across stimulus and inter-trial intervals between experimental and control groups. Four of the ten features (Area, Perimeter, Eccentricity, and Principle Component 1 shown. Red and blue traces show a section of an example worm's z-scored features during the CS-on and ITI intervals. Heatmaps show the density of z-scored feature values across the CS-on and ITI intervals for that same worm (Area, Perimeter, Eccentricity, and PC1 from top to bottom). **B.** Hand-scored conditioned responses for experimental and control groups trained with the blue light-shock classical conditioning protocol from James et al. 2025. across different strains: *G. dorotocephala* (n = 30), *G. dorotcephala* (Michigan wild-type, n = 6), and *S. mediterranea* (n = 6). Dashed lines indicate contractions and solid lines indicate turns. **C.** KL divergence over trials for experimental and control groups of *G. dorotocephala* worms trained with the blue light-shock classical conditioning protocol from James et al. 2025 (n = 12 worms per condition). Error bars denote standard error of the mean (SEM).

Because it seemed likely that the variability in subjective manual scoring could be obscuring legitimate differences between experimental and control groups, we built a custom feature-extraction pipeline using Segment-Anything (Meta Inc.) and methods for shape analysis previously applied to planarian behavior [45,53]. Motivated by recent work on classical conditioning in rodents [54], we quantified behavioral responses by calculating the Kullback–Leibler (KL) divergence between these feature distributions sampled during stimulus-on and inter-trial intervals (Fig. 3A). Intuitively, the KL divergence measures how much the appearance of the cue changed the distribution of body shape features (e.g. the area or perimeter of the worm). If behavioral changes are acquired over the course of conditioning, the KL divergence should differ between the experimental and control groups over the course of training, and should show an upward trend only for the experimental group.

We fit a linear mixed model predicting log-transformed KL divergence from Day (continuous), Condition (conditioning vs. pseudoconditioning), and their interaction, with random intercepts for individual subjects. The interaction term was not significant ($\beta = -0.08, z = -0.95, p = 0.342$), indicating no evidence that the conditioning group's trajectory differed from that of the pseudoconditioning group. This was corroborated by a direct comparison of per-subject slopes between groups (conditioning: $+0.33 \pm 0.13$; pseudoconditioning: -0.04 ± 0.31; Welch's $t = 1.11, p = 0.286$). The two groups did differ significantly in overall KL divergence ($\beta = 0.35, z = 3.61, p < 0.001$), though notably the pseudoconditioning group showed a higher mean KL divergence, which is opposite to what a conditioning effect would predict (Fig. 3C).

3 Discussion

In the present study, we were unable to find compelling evidence for classical conditioning in planarians using an array of conditioning paradigms targeting three sensory modalities, two species of worms (with multiple different strains), and a combination of manual and automatic scoring methods. While it is possible that prolonged conditioning would eventually produce learning in these paradigm variants, no protocol yielded a magnitude of conditioned response acquisition on par with what has been previously reported (e.g., [15]). Thus, we suspect that the conditions for

eliciting robust and reproducible conditioned behavior in the planarian species we have tested are highly specific. The key variables and the necessary parameters to enable conditioning in planarians, if such truly exist, remain elusive. Because of our failure to demonstrate conditioned behavior, we were not able to address questions of molecular transfer of memory between planarians in any way.

Our finding of low interscorer agreement is concerning for the interpretation of the entire literature on planarian conditioning, almost all of which relies on manual scoring. If different scorers observing the same footage can only agree on the occurrence of conditioned responses half of the time, then those conditioned responses cannot form the basis for strong claims about learning and memory. The corollary is that weak claims about learning and memory cannot be used to make strong claims about memory transfer. Memory transfer itself is a potentially paradigm-shifting phenomenon that demands incontrovertible empirical proof of existence if it is to be accepted by the scientific community. If all planarian experiments obtain conditioned responses as unreliable as ours, then planarians seem a poor organism with which to investigate these questions.

It is difficult to reconcile our null results with the 20th century planarian learning literature. On the one hand, several laboratories (not just the laboratories of McConnell and his students) reported success in training planarians. Strikingly, several editions of the *Worm Runner's Digest* contain letters from students and laypeople describing their efforts in conditioning planarians with an approximately 4:1 ratio of success to failure [17]. Over fifty scientific papers were published reporting successful conditioning of planarians between 1959 and 1971 [6], including in top journals like *Nature*. Other demonstrations of apparent learning include instrumental and maze paradigms (e.g. [55–57]). In the modern age, even nematodes sporting a quaint 302 neurons are capable of aversive learning [58–60], in contrast to planarians' ten-thousand [61]. Nonetheless, we have failed to show classical conditioning in planarians despite several variations in procedure and species. Ultimately, we cannot explain why our results are in such contrast to those that came before.

We close with a quote from a 1964 paper by Bennet and Calvin [16], who attempted to train planarians on a variety of tasks and were mostly unsuccessful:

> "Until more adequate and more useful methods and descriptions of methods to train planarians are available, this animal appears to have little utility for proposed studies of the possible biochemical basis of memory."

This assessment appears, at least in our hands, to hold true today.

Going forward, we recommend continuing to explore alternative stimulus combinations informed by planarian ethology and ecology. We further suggest that researchers methodically document their procedures and their data. Importantly, footage demonstrating learned behavior in planarians should be made available in its entirety, including every trial of training from start to finish.

Acknowledgments

We extend our gratitude to the laboratories of Peter Reddien and Michael Levin for supplying worms and providing technical guidance, and to the Harvard NeuroTechnology Core for aid in building apparatuses. We additionally thank Reeva and Daniel Kimble for discussions of their experiences training planarians in the laboratory of James McConnell in the 1950s and 1960s, and for generously providing us with access to the complete *Worm Runner's Digest*.

Data and Code Availability

Code to reproduce the main and supplementary figures in this manuscript is available here: `https://github.com/madeleinesnyder/PlanariaPipe/tree/main` and the mp4 files used for manual scoring and binary masks used for the feature extraction pipeline are available on the Harvard Dataverse (`https://doi.org/10.7910/DVN/NSTJGZ`). The .bin files used to create the masks using our extraction pipeline are available upon request due to storage limitations.

4 Materials and methods

4.1 Planarian acquisition and care

Planarians were housed in 6- and 11-cup rectangular Pyrex dishes with air holes carved into the lids. The dishes were filled to a level of 2-3 inches with Montjuic water (1.6 mM NaCl, 1.0 mM $CaCl_2$, 1.0 mM $MgSO_4$, 0.1 mM $MgCl_2$, 0.1 mM KCl, 1.2 mM $NaHCO_3$ in deionized water). Housing dishes were cleaned biweekly on Tuesdays and Fridays. Cleaning consisted of corralling planarians into one corner using a water jet created by a transfer pipette or spray bottle containing fresh Montjuic water, draining the water from the dish, and wiping down the floor and walls with

unbleached paper towels. The corner where the planarians had been concentrated was not wiped down, leaving an untouched triangular region with an area of 2-3 square inches. The unwiped corner was alternated between opposing sides with each cleaning so that a continuous zone of slime was always present in the trough. For experiments where individual data were acquired, planarians were individually isolated into six-well plates (Falcon, item no. 353046) containing approximately 10 mL of Montjuic water four days prior to the start of training. Otherwise, cohorts of experimental worms were isolated to petri dishes (Greiner Bio-One, item no. 633181) for several days before the start of experimental training.

Planarians were fed with freshly hatched *Artemia* brine shrimp nauplii (Brine Shrimp Direct, item no. BSEA8Z), hatched in salt water according to vendor instructions. For feeding, the hatchlings were rinsed with Montjuic water and transferred to the planarian dishes with a transfer pipette. Confirmation that the planarians ate the shrimp came in the form of pursuit of the shrimp, visible pharyngeal contractions, and the tendency for planarians to take on the color of their food (orange-red with brine shrimp). Planarians were fed once or twice per week, and in the case of experiments, starved for seven days prior to testing.

The planarians used in these studies varied in source. The Jacobson-style experiments used two populations of *Girardia dorotocephala*. The first, a strain raised in the laboratory for years, came from from Michael Levin's group at Tufts. The second was purchased from Carolina Biological Supply (item no. 132954), which are wild-caught from various water sources in North Carolina. Miscellaneous experiments using alternative stimuli used the Carolina *G. dorotocephala* as well as *S. mediterranea* worms originating from the laboratory of Peter Reddien. Experiments modeled after the James et al. study used the Carolina *G. dorotocephala* worms and *S. mediterranea*. Experiments were also run with a population of worms descended from a *G. dorotocephala* individual acquired from the Huron River in Ann Arbor, Michigan, where James McConnell is suspected to have sourced worms [62]. This planarian was found June 12, 2025, on the bottom of a rock in the shallow pool at the base of the Argo Dam on the outflow side (coordinates: 42°17'24.7"N, 83°44'42.6"W). Its species was determined visually by its prominent auricles and brown, unspotted coloration. Once in the laboratory, this individual worm was amplified into a colony by repeated cycles of bisection and regeneration. In a similar fashion, two worms of the

same species were acquired from a stream on the University of California, Berkeley campus, which were used in a White-UV experiment.

4.2 Assay chamber, stimuli, and data acquisition

We designed an assay chamber resembling side-by-side duplicates of McConnell's original chamber [63]. Six separate roundbottomed, ovoid lanes were milled from a transparent block of polycarbonate and header pins were inserted at the ends of each lane. Sets of ipsilateral pins were connected by a copper wire to transmit the electrical shock stimulus to all lanes simultaneously (Figure 1A). Two alligator clips attached to the pin-spanning copper wires connected each side of the trough to a DC power supply via an H-bridge module that together produced a 10-12V, 60 Hz square-wave AC shock stimulus. Experiments were recorded at 10Hz with an overhead camera (Teledyne FLIR GS3-U3-23S6C-C (2048px x 2048 px) or GS3-U3-41C6M-C (4096px x 3000px) suspended approximately one foot above the trough. A long-pass filter (750 nm) was placed over the camera lens to eliminate saturation effects from the light stimulus. The trough was placed on a white acrylic panel which diffused another panel of infrared (IR) LED lights (850nm, Waveform Lighting flexstrips) below, so the worms were clearly visible from the overhead camera. The white light stimulus was provided by white light LED strips (6500K) glued to an acrylic panel, suspended above the trough on the same plane as the camera. Where noted, a 100-Watt incandescent bulb (Fluker's Incandescent Basking Spotlight Bulb), intense blue LED bulb (450-460nm HIGROW LED Aquarium Blue Light Blub, 36W), or 100-Watt LED UV panel (395nm Glostars 100W LED, IP66) were used. Small IR LEDs were directly coupled to the shock and light stimuli and visible in the camera frame to serve as indicators of stimulus onset.

The water pulses (Montjuic solution, contiguous with the assay chamber) were delivered via a linear actuator (AMGSH 12V; 6in) which pushed a reservoir of solution through a six-headed manifold to all chambers simultaneously. This chamber resembled our light-shock chamber, but was fitted with a removable roof and fine plastic mesh (McMaster Carr Item No. 9318T45, opening size 0.0048") that sealed the ends of the chamber to allow excess solution to drain while keeping the planarians inside. Custom software written in Python controlled the camera, conditioned stimulus, and unconditioned stimulus via a Raspberry Pi Pico (RP2020) microcontroller. Data was collected

onto a Linux computer (Ubuntu 20.04.6 LTS, 64 bit) using a custom Python (version 3.8.10) recording script (Figure 1A). The whole of these apparatuses were contained in a vertically rectangular box made of acrylic panels affixed to metal frame to isolate the experimental equipment and reduce ambient illumination.

4.3 Jacobson et al. classical conditioning paradigm

In Allan Jacobson and colleagues' experiments [12,13,15], planarians were rapidly and repeatedly presented with overlapping light and shock that resulted in the worms contracting to light even in the absence of shock. We attempted first to reproduce these experiments in their timing and stimulus parameters. Planarians of the species *Girardia dorotocephala* (known as *Dugesia dorotocephala* in Jacobson's time) were purchased from Carolina Biological Supply (item no. 132954) and subjected to 250 trials of 3 seconds of light, the final second of which included a 10 V, 60 Hz square-wave AC shock. Trials occurred in ten blocks of 25 trials, each trial separated by 30 seconds, with ten minute rest periods between blocks that were themselves spread over two days. Day 1 contained two consecutive light-only blocks in accordance with Jacobson's paradigm. These were identical to the light-shock blocks, save that the shock did not occur. Presentation of the CS alone without any paired US had been reported to diminish the effectiveness of conditioning as far back as 1959 [64], but because the Jacobson group gave these light-only trials, we did likewise. The two Day 1 light-only blocks were followed by two light-shock blocks. On day 2, worms received three light-shock blocks. In the case of the pseudoconditioning control, light-shock blocks were replaced with light-or-shock blocks. These consisted of randomly alternating light or shock presentation, with the constraint that no more than four consecutive light stimuli and no more than two consecutive shock stimuli were permitted.

Pseudoconditioning is nowadays often implemented by independent Poisson processes for different stimuli, but random alternation is the method for Jacobson's own pseudoconditioning control, so as with the light-only blocks, we followed suit. The pseudoconditioning was implemented by generating random lists of 25 alternating L and S symbols, then filtering for those that violated the consecutivity restraints. A random candidate LS sequence was selected at the start of each pseudoconditioning block and each trial in that block corresponded to a symbol in the LS sequence.

For example, a pseudoconditioning block that selected the sequence "LLS..." would present two light trials and then a shock trial. The training footage was scored on a trial-by-trial basis and condition responses were marked when a worm performed a "discernible longitudinal contraction of the body" [12] in the two second period between light onset and shock onset.

Before experiments began on any given day, a separate cohort of approximately twenty nonexperimental were allowed to "slime" the trough. After twenty minutes, this slime crew was removed and experimental worms were added and allowed to acclimate for ten minutes before experiments began. After a cohort of worms was done for the day, the trough was rinsed with DI water and scrubbed with unbleached paper towels before the next slime crew and experimental cohort were added.

Because preliminary attempts to replicate the Jacobson-style conditioning showed low and unchanging conditioned response rates to light, we attempted to train a subset of worms that were already reactive to light prior to any training. To do this, we showed a subset of worms two blocks of 20 trial of light-only trials following the same timing characteristics as described above. Responses to light (turns or contractions) were scored for each trial. The twelve highest-responding worms were split into forward conditioning and pseudoconditioning groups and called "high responders"; the next highest-responding twelve worms were split into the "low responders" groups. Figure S2 shows a box-and-whisker plot of these two groups' pre-training responses to light.

4.4 James et al. classical conditioning paradigm

Because a group at the University of Tasmania had recently reported success in conditioning planarians using vastly different temporal parameters than our experiments, we attempted to reproduce their procedures [25]. In these experiments, adapted from fear-conditioning in rodents, planarians were starved for seven days and subjected to six trials per day for four days, with trials split evenly across one morning and one afternoon block starting at approximately 10:00 and 14:00 respectively. Each trial consisted of 30 seconds of blue light exposure that coterminated with 3 seconds of shock, followed by 30 seconds of rest. The control condition consisted of an approximation of pseudoconditioning, with light-only and shock-only trials randomly dispersed throughout blocks, with at least one of each trial type occurring in each block. The day before

training trials began, worms were permitted 10 minutes to explore the trough before being returned to their home dishes. On any training day, worms were given ten minutes to acclimate to the trough before trials began. Other experimental details remained the same as the Jacobson-style experiments. These experiments bore similar results to our previous attempts at classically conditioning planarians, with no discernible increase in contraction probability to the onset of light or change in number of contractions throughout the 30 second light window. Data from these experiments were assessed by blinded manual scoring and a computer vision pipeline, described below.

We note two methodological departures from the procedures described by James at al. First, their worms appear to have been trained and tested in circular dishes, whereas ours were trained in linear lanes similarly to James McConnell's apparatuses [63] because this enabled us to train multiple worms in one sitting. Second, while the James et al. group notes that their planarians are *G. dorotocephala* ordered from a biological supply house, the company that sells them informs us that these worms are collected from ponds in Australia (Southern Biological, personal communication, July 11, 2025). Whether these *G. dorotocephala* worms are invaders from North America, a truly endemic Australian variant, or a different species altogether is unknown. The circular training/test environment and different worm strain could explain the discrepancy between our results and those of James et al., though further work is needed to resolve this.

4.5 Miscellaneous conditioning paradigms

Because repeated attempts to condition worms with a white light CS and an electric shock UCS (White-Shock) failed, we sought to use other sets of stimuli that might be more salient to planarians. These stimuli were various combinations of ultraviolet light (UV) and mechanical stimulation (Pulse, as in a pulse of water current through a narrow lane) in addition to the classic stimuli. We also varied species and the type of water that worms were housed in (Carolina Biological Supply, item no. 132450) as opposed to Montjuic water. We did not methodically cover the space of these combinations; rather, we sought to rapidly test a subsample for any indication of planarian learning, then hone in on promising candidate paradigms.

Our rationale for choosing ultraviolet light as a new stimulus was motivated by the fact that

while the shock is certainly aversive, we are unaware of any ethologically relevant situation where planarians would need to respond to electrical shocks. Furthermore, the repeated full-body shock of a submerged organism may cause side effects that could affect worms' behavior in a way that does not reflect learned behavior. In contrast, planarians have dedicated sensory machinery to detect and respond to ultraviolet light, which includes a full-body contractions to sudden onset of intense UV [46,47,65]. These morphological and behavioral characteristics are easily reconciled with planarians' benthic lifestyle, unlike shock. Likewise, though planarians lack the complex lateral line system fish, they nonetheless are subject to mechanical disturbances in water. Indeed, vibration has been used as a conditioned stimulus in planarians before [15,66]. We elected to use pulses of water through the sealed lanes of a trough closely resembling what was used in the light-shock experiments. The lanes were fixed to a removable roof and the lane ends were replaced with mesh. Montjuic solution flowing gently through the sealed lane provided a constant current, and injection into the inlet caused the current to accelerate briefly.

1. **White-UV**: The same white light from the the White-Shock paradigm was paired with an ultraviolet pulse instead of shock. Two *G. dorotocephala* worms from collected from Berkeley, California were subjected to 3 seconds of white light coterminating with 1 second of UV light. Worms were in blocks of 25 trials (2 minute ITI) trained across three blocks, the first block on Day 1 and the second two blocks on Day 2. A ten minute rest period separated neighboring blocks. Contractions to the onset of white light were scored for each trial. An additional six *S. mediterranea* worms were trained identically, save that they only received two blocks. These experiments took place between 9:30 and 14:30.

2. **UV-Shock**: The white light from White-Shock was swapped with a UV light instead and reduced in strength so as not to trigger baseline contractions. Eighteen *G. dorotocephala* worms were subjected to five blocks of 25 trials of 3-sec UV with 1 second coterminating shock using an ITI of 2 minutes. Blocks one and two occurred on Day 1, two and three on Day 2, and the fifth on Day 3. Worms were allowed 10-20 minutes to acclimate to their trough lanes before experiments began and given a ten minute rest period between blocks. The shock characteristics were the same 10-12 v, 60 Hz square wave AC described in our

Jacobson-style protocol. The ITI was 2 minutes. Of the eighteen worms, six were raised in Carolina Spring Water (Carolina Biological Supply, item no. 132450) for several weeks before experiments began. Worms were scored for contractions to the onset of the UV light. These experiments occurred between 10:00 and 16:00.

3. **Pulse-UV**: Six *S. mediterranea* worms received four blocks spread evenly across two days of 25 trials each, ten minutes of rest between same-day blocks. After ten minutes of acclimation to the flow chamber, worms were subjected to a brief 65 millisecond pulse of current followed one second later by a 2 second flash of UV light. The ITI was 60 seconds. Worms were scored for contractions to the water pulse. These experiments took place between 8:00 and 16:00.

Both mechanical disturbance (water pulses) and ultraviolet light (UV pulses) elicit contractions. Therefore, neither water pulses nor UV light are neutral stimuli in the conventional sense of the term. However, as argued by Jacobson [14], even if conditioned and unconditioned stimuli elicit the same response (contraction), learning can still be assessed by increases (or lack thereof) in response rates between controls and experimental worms. In any case, because we observed no increase in responding among worms trained on any of these paradigms, the point is moot and pseudoconditioning controls were not necessary.

4.6 Manual scoring and interscorer agreement

Learning in the classical planarian conditioning assay has historically been measured as the number of full-body contractions in response to the onset of the conditioned light stimulus over the course of training, scored by human observers. Following suit, we had one to three scorers manually score each video for conditioned responses occurring after the onset of the conditioned stimulus and before the onset of the unconditioned stimulus. For Jacobson-style experiments, scorers simply watched complete overhead footage of the experiment as it took place. Because the CS scoring period was only two seconds per trial, each scorer's response was essentially binary (one contraction or no contraction). By comparing Scorer 1 and Scorer 2's binary data on a trial-by-trial and worm-by-worm basis, we were able to count the proportions of trials in which each scorer agreed positively (contraction and contraction), agreed negatively (no contraction and no contraction), and

disagreed (contraction and no contraction, or vice versa). The proportion of agreement trials divided by the total number of trials is the interscorer agreement for this experiment.

Footage from the James-style experiments was scored differently using a custom-made computer interface called the Blind Scoring Interface (BSI, Fig. S5). This interface enabled minimally trained scorers to count turns and contractions for each trial of the experimental footage blindly. Three scorers counted conditioned responses across the same subset of 203 trials split between forward conditioning (144) and pseudoconditioning groups (59). Scorer 2 scored an additional thirteen videos (216 total), and Scorer 1 scored all footage with the blind interface. Video files representing one worm and one trial were given codenames composing of two random adjectives and one random noun; when a user scored a video, only this codename was seen, so no information about experimental group (forward or pseudoconditioning) was available to them, nor was there any indication of how much previous training the worm had undergone. The videos were created such that each trial containing light presentation was segmented into its own 27 second video spanning the onset of the light up to the frame before the onset of the shock stimulus. Timestamps of turns and contractions were recorded from each video using the blind interface by pressing a "Turn" or "Contraction" button on the frame which the behavior occurred. The video title, as well as the sum and list of timestamps for turns and contractions, were saved to a CSV from the interface. Data for interscorer reliability and agreement measures for were obtained by comparing the timestamps of turn and contraction events between the three scorers.

Because each scorer scored the same set of videos and noted timestamps of turns and contractions, we were able to compare when exactly each turn and contraction took place. This was accomplished by comparing two scorers' timestamp lists of turns and contractions per video (one video is equivalent one trial). For each trial, every possible pair of same-type timestamps was tabulated and the difference between those timestamps calculated. Pairs were sorted by increasing difference and pairs with the least difference were considered matches up to a threshold of ten frames (1 second) for turns and fifteen frames (1.5 seconds) for contractions. Any timestamps remaining when all pair combinations had been exhausted were considered orphans. This was done for same-type events, turns with turns and contractions with contractions, until each type had a list of orphan timestamps. Next, each orphan timestamp was compared to orphan timestamps of the

opposite type and, if within an event type's difference threshold, paired with the opposite-type orphan. Thus, for two scorers' assessment of a given video, we were able to calculate the number of matching timestamps, the number of unmatched timestamps, and which of those unmatched timestamps could be matched with an opposite-type event. We performed this procedure for each video and for each scorer pair and used those measures to calculate precision, recall and F1 scores.

The F1 metric is the harmonic mean of precision and recall. When treating one scorer's timestamps as reality (reference) and testing another scorer against it (test), precision is the proportion of true positives recovered by the test dataset out of all test positives. Recall is the proportion of true test positives relative to the entirety of the reference positives. In this context, precision is the proportion of a Test scorer's events that can be matched to the Reference scorer. Likewise, recall is the proportion of Reference events that the Test Scorer can match out of all of the Reference's events. These proportions describe how much the Test Scorer marks the same events as the Reference Scorer while also not marking events the referent does not have. The formulae for precision, recall, and F1 scores are given. (TP = true positive, FP = false positives, TN = true negative, FN = false negative.)

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

There are two possible variants of F1 score to calculate for scorer pairs. One option was to calculate an F1 score for each video and then average across all video-specific F1 scores to produce a scorer pair's final F1 score. However, this method risks inflating F1 scores because any video where scorers agree nothing happened (no timestamps given and therefore no possible disagreement) would raise the average. Instead, we summed the TP, FP, and FN values across all videos, then calculated precision/recall only once to produced a final F1 score. The F1 metrics calculated this way are more reflective of agreement about the events of interest.

The measures of unanimous turns and contractions across all three scorers were calculated similarly. For each video, triplets of same-type events were produced and sorted by their range (maximum frame value minus minimum frame value). The range threshold for turns was 10 and 15 for contractions. For each video, some number of unanimous contractions was counted. Because of

partial agreement (e.g., Scorer 1 and 2 agreed on a contraction but Scorer 3 disagreed), we used Venn Diagram notation to calculate the total number of distinct contraction events across all scorers. The sum of all unanimous contractions per video was divided by this number to yield the unanimity rate.

The above methods described agreement about events occurring within the window of the conditioned stimulus, i.e. conditioned responses. To score agreement about contractions induced by shock (the unconditioned stimulus) between Scorers 1 and 2, 198 individual worm trials were sampled from the classical conditioning group of the the footage from our attempted James et al. replication. The trials were observed non-blindly for contractions at the onset of the shock stimulus. Blind scoring was deemed unnecessary for shock stimuli because training day or classical/pseudoconditioning status should have no bearing on unconditioned responses. The proportion of the number of trials where Scorer 1 and Scorer 2 disagreed divided by the number of all scored videos formed the shock agreement metric.

4.7 Computer vision pipeline

We used 10 morphological features (area, perimeter, area-perimeter ratio, eccentricity, hull area, centroid coordinates, relative concavity area, and two principle components from the 100-dimensional vector of the perimeter shape) to characterize the pose of individual planarians at each frame. To extract these features we tailored the Segment-Anything computer vision pipeline (Meta, Inc) to produce a timeseries of binary masks of the worms from the same overhead experimental footage used to manually score the worms' behavior. First, each video was spatially cropped such that each video contained a single worm. Then, we labeled two points for the first frame of each video to indicate the worm, and two points to indicate the electrodes (static occlusions in the environment). We used the Segment-Anything pipeline to propagate these points through each video and obtain a timeseries of binary masks. We manually inspected each video, and if the mask extraction quality was suboptimal, split the video before the quality drop and labeled new starting points to propagate through the next video segment. This process was repeated until the entire timeseries of masks was of sufficient quality. Next, we calculated the mask area (sum of all black pixels in the mask), perimeter (using the findContours function from the

OpenCV python library), area-perimeter ratio, circularity ($(4 * \pi * A)/(P * P)$ where A is the area and P is the perimeter), hull area (using convexHull and contourArea from OpenCV python library), centroid coordinates (using moments from OpenCV), and relative concavity area ($(H - A)/H$) where H is hull area. To complement these features, for each frame we extracted the first two principle components of a 100-dimensional vector of distances from the centroid of the mask to evenly-spaced points on the perimeter of the mask. This analysis was published in [45]. The areas timeseries was used to identify outlier timepoints (> 3.5 standard deviations), which were excluded from all feature vectors.

We quantified how worm behavior differed between conditioned-stimulus (CS-on) and inter-trial (CS-off) periods by computing the Kullback–Leibler (KL) divergence between these feature distributions [67]. We first extracted the session-specific windows of time where the conditioned stimulus was on (CS-on) and off (CS-off) and approximated each condition as a multivariate Gaussian with diagonal covariance. We estimated means and unbiased variances of each feature in each condition (ignoring NaNs and adding a small constant for numerical stability), and computed the closed-form 1D Gaussian KL divergence between the CS-on and CS-off Gaussians. We obtained the total divergence KL(CS-on || CS-off) for a given worm during a training block by summing these per-feature KL values. Thus, individual worms' divergence scores could be plotted to observe any potential change in divergence over the course of the experiment's eight training blocks.

4.8 Statistical tests

Jacobson et al. replication: To test for acquisition of the conditioned response over training days, we fit a linear slope to each worm's daily response rate and tested whether the mean slope differed from zero (one-sample t-test, one-sided). We confirmed this with Page's L test for monotone increasing trends across ordered time points. Group differences (conditioning vs. pseudoconditioning) were assessed with a two-way ANOVA. For automated behavioral analyses, we fit linear mixed models (LMMs) with Day, Condition, and their interaction as fixed effects and individual subject as a random intercept, using log-transformed KL divergence as the dependent variable. The Day $\times$ Condition interaction tested whether behavioral change over time differed between groups. All analyses used Python (SciPy, statsmodels).

James et al. replication: To test for acquisition of the conditioned response from the hand-scored data, we computed a linear slope of each worm's daily response count across training days and tested whether the mean slope was significantly greater than zero using a one-sample t-test (one-sided, $\alpha = 0.05$). We supplemented this with Page's L test, and fit LMMs with Day (continuous), Condition (conditioning vs. pseudoconditioning), and their interaction as fixed effects and individual subject as a random intercept, estimated via maximum likelihood. The Day $\times$ Condition interaction served as the primary test of the conditioning hypothesis, assessing whether the rate of behavioral change over training days differed between groups. LMMs were fit separately for turns and contractions.

James et al. scoring with the computer vision pipeline: To test for changes in KL divergence values between conditioning and pseudoconditioning groups over training, we log-transformed the KL divergence values to improve normality and analyzed with a LMM predicting log KL from Day, Condition, and their interaction, with random intercepts for individual subjects. Per-subject slopes were also compared between conditions using Welch's t-test, and overall group differences in mean KL divergence were assessed with Welch's t-test and Cohen's d. All LMMs were fit using statsmodels (Python); non-parametric and slope-based tests used SciPy.

Miscellaneous conditioning paradigms: For each paradigm (UV-shock, white light–UV, and pulse-UV), we computed a linear slope of each worm's contraction count across training blocks and tested whether the mean slope differed from zero using a one-sample t-test (two-sided, $\alpha = 0.05$; one-sided where a directional hypothesis was specified). We supplemented the frequentist test with two additional analyses. First, we performed a two one-sided tests (TOST) equivalence procedure to determine whether the observed slope was negligibly small, using a pre-specified smallest effect size of interest (SESOI) of ± 1 contraction per block. Second, we computed a Bayesian t-test to quantify the relative evidence for the null hypothesis ($BF_{01} > 1$ favors no trend) versus the alternative ($BF_{10} > 1$ favors a trend). Where sufficient data permitted (≥ 3 blocks common to all troupes within a paradigm, $geq2$ worms with complete data), we also applied Page's L test. Worms were pooled across troupes within each paradigm; slopes from troupes contributing different numbers of blocks were weighted equally at the worm level.

Individual sensitivity: To test if different sensitivity to the conditioned stimulus affected acquisition

of the conditioned response, we computed each worm's linear slope across conditioning blocks and tested whether the mean slope was significantly greater than zero (one-sample t-test, one-sided). Slopes were compared between high and low responders within each condition using Welch's t-test with Cohen's d. To assess whether response rates changed across the light-only assessment sessions, we compared per-worm Day 1 and Day 2 means using a paired t-test (one-sided for increase) and the Wilcoxon signed-rank test as a non-parametric alternative, with Cohen's d (paired). To distinguish a gradual within-session trend from a between-session shift, we additionally fit per-worm linear slopes across individual trials and used a one-sample t-test.

References

1. Bailly X, Reichert H, Hartenstein V. The urbilaterian brain revisited: novel insights into old questions from new flatworm clades. Development genes and evolution. 2013;223(3):149-57.

2. Holland LZ, Carvalho JE, Escriva H, Laudet V, Schubert M, Shimeld SM, et al. Evolution of bilaterian central nervous systems: a single origin? EvoDevo. 2013;4(1):27.

3. Holland L. Invertebrate origins of vertebrate nervous systems. In: Evolutionary Neuroscience. Elsevier; 2020. p. 51-73.

4. Sarnat HB, Netsky MG. When does a ganglion become a brain? Evolutionary origin of the central nervous system. Seminars in Pediatric Neurology. 2002;9(4):240-53. Development and Dysgenesis. Available from: https://www.sciencedirect.com/science/article/pii/S1071909102800110 doi:https://doi.org/10.1053/spen.2002.32502.

5. Corning W, Riccio D. The planarian controversy. Molecular approaches to learning and memory. 1970:107-50.

6. Corning W, Kelly S. Platyhelminthes: the turbellarians. In: Invertebrate Learning: Volume 1 Protozoans Through Annelids. Springer; 1973. p. 171-224.

7. Thompson R, McConnell J. Classical conditioning in the planarian, Dugesia dorotocephala. Journal of Comparative and Physiological Psychology. 1955;48(1):65-8. Available from: https://doi.apa.org/doi/10.1037/h0041147. doi:10.1037/h0041147.

8. McConnell JV, Jacobson AL, Kimble DP. The effects of regeneration upon retention of a conditioned response in the planarian. Journal of Comparative and Physiological Psychology. 1959;52(1):1-5. Available from: https://doi.apa.org/doi/10.1037/h0048028. doi:10.1037/h0048028.

9. MCCONNELL J, JACOBSON R, HUMPHRIES B. The effects of ingestion of conditioned planaria on the response level of naive planaria: a pilot study (or:" You are what you eat?"). Worm Runner's Digest 3/1, 41. MCCONNELL, JV (1962): Memory transfer through cannibalism in planarians. J Neuropsychiat. 1961;3.

10. McConnell JV. Memory transfer through cannibalism in planarians. J Neuropsychiat. 1962;3:542-8.

11. Zelman A, Kabat L, Jacobson R, McConnell J. Transfer of training through injection of "conditioned" RNA into untrained planarians. Worm Runner's Digest. 1963;5:14-21.

12. Jacobson AL, Fried C, Horowitz SD. Planarians and memory: I. Transfer of learning by injection of ribonucleic acid. Nature. 1966;209(5023):599-601.

13. Jacobson AL, Fried C, Horowitz SD. Planarians and memory: II. The Influence of Prior Extinction on the Ribonucleic Acid Transfer Effect. Nature. 1966;209(5023):599-601.

14. Jacobson AL. Classical conditioning and the planarian. In: Chemistry of Learning: Invertebrate Research. Springer; 1967. p. 195-216.

15. Jacobson AL, Horowitz SD, Fried C. Classical conditioning, pseudoconditioning, or sensitization in the planarian. Journal of Comparative and Physiological Psychology. 1967;64(1):73-9. Available from: https://doi.apa.org/doi/10.1037/h0024808. doi:10.1037/h0024808.

16. Bennett E, Calvin M. Failure to train planarians reliably. Neurosciences Research Program Bulletin. 1964;2(4):3-24.

17. McConnell J. Failure to interpret planarian data correctly: a reply to Bennet and Calvin; 1965. What goes here?

18. HARTRY AL. Learning and Memory Transfer in Planarians. The Claremont Graduate University; 1970.

19. Gaito J. Molecular psychobiology of memory: Its appearance, contributions, and decline. Physiological Psychology. 1976 Dec;4(4):476-84. Available from: http://link.springer.com/10.3758/BF03326600. doi:10.3758/BF03326600.

20. Peretti P, Zinthefer M, Ferhandez H. Influence of Learning Trials on Memory Retention of Planaria: Dugesia Tigrina. Pakistan Journal of Psychology. 1999;30(1):35.

21. Abbott SM, Wong GK. The conditioning and memory retention of planaria (Dugesia tigrina) for directional preferences. BIOS. 2008 Dec;79(4):160-70. Available from: http://www.bioone.org/doi/abs/10.1893/0005-3155-79.4.160. doi:10.1893/0005-3155-79.4.160.

22. Prados J, Alvarez B, Howarth J, Stewart K, Gibson CL, Hutchinson CV, et al. Cue competition effects in the planarian. Animal Cognition. 2013 Mar;16(2):177-86. Available from: http://link.springer.com/10.1007/s10071-012-0561-3. doi:10.1007/s10071-012-0561-3.

23. Samuel K, Suviseshamuthu ES, Fichera ME. Addiction-Related Memory Transfer and Retention in Planaria. bioRxiv. 2021. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.09.12.459965. doi:10.1101/2021.09.12.459965.

24. Shimojo K, Shimojo E, Shimojo S. New evidence for preservation of conditioned behavior based on UV light sensitivity in dissected tail halves of planarians. Journal of Vision. 2021;21(9):2326-6.

25. James FLJ, Ahamed SK, Yang C, Thinley J, Vicario CM, Canales JJ, et al. Adapting rodent cued threat conditioning to planarians: Memory acquisition, consolidation, and reconsolidation. Behavioural Processes. 2025;235:105321. doi:10.1016/j.beproc.2025.105321.

26. Forde F. Probing the Limits of Memory: Can a Learned Response Persist Through Decapitation and Regeneration in Planaria? [Ph.D. Thesis]. Wellington, New Zealand: Open Access Te Herenga Waka-Victoria University of Wellington; 2025.

27. Elliott SA, Alvarado AS. Planarians and the history of animal regeneration: paradigm shifts and key concepts in biology. Planarian regeneration: Methods and protocols. 2018:207-39.

28. Beane WS, Morokuma J, Lemire JM, Levin M. Bioelectric signaling regulates head and organ size during planarian regeneration. Development. 2013;140(2):313-22.

29. Lobo D, Beane WS, Levin M. Modeling Planarian Regeneration: A Primer for Reverse-Engineering the Worm. PLoS Computational Biology. 2012 Apr;8(4):e1002481. Available from: https://dx.plos.org/10.1371/journal.pcbi.1002481. doi:10.1371/journal.pcbi.1002481.

30. Reddien PW, Alvarado AS. Fundamentals of planarian regeneration. Annu Rev Cell Dev Biol. 2004;20:725-57.

31. Reddien PW. The cellular and molecular basis for planarian regeneration. Cell. 2018;175(2):327-45.

32. Newmark PA, Alvarado AS. Schmidtea happens: Re-establishing the planarian as a model for studying the mechanisms of regeneration. In: Current topics in developmental biology. vol. 147. Elsevier; 2022. p. 307-44.

33. Kusayama T, Watanabe S. Reinforcing effects of methamphetamine in planarians. Neuroreport. 2000;11(11):2511-3.

34. Buttarelli FR, Pellicano C, Pontieri FE. Neuropharmacology and behavior in planarians: translations to mammals. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology. 2008;147(4):399-408.

35. Cho M, Nayak SU, Jennings T, Tallarida CS, Rawls SM. Predator odor produces anxiety-like behavioral phenotype in planarians that is counteracted by fluoxetine. Physiology & behavior. 2019;206:181-4.

36. Wyss LS, Bray SR, Wang B. Cellular diversity and developmental hierarchy in the planarian nervous system. Current opinion in genetics & development. 2022;76:101960.

37. Gershman SJ. The molecular memory code and synaptic plasticity: A synthesis. Biosystems. 2023 Feb;224:104825. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0303264722002064. doi:10.1016/j.biosystems.2022.104825.

38. Abraham WC, Jones OD, Glanzman DL. Is plasticity of synapses the mechanism of long-term memory storage? npj Science of Learning. 2019 Jul;4(1):9. Available from: https://www.nature.com/articles/s41539-019-0048-y. doi:10.1038/s41539-019-0048-y.

39. Langille JJ, Gallistel CR. Locating the engram: Should we look for plastic synapses or information-storing molecules? Neurobiology of Learning and Memory. 2020 Mar;169:107164. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1074742720300083. doi:10.1016/j.nlm.2020.107164.

40. Gallistel CR. The coding question. Trends in Cognitive Sciences. 2017 Jul;21(7):498-508. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1364661317300852. doi:10.1016/j.tics.2017.04.012.

41. Gallistel CR. The physical basis of memory. Cognition. 2021 Aug;213:104533. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0010027720303528. doi:10.1016/j.cognition.2020.104533.

42. Akhlaghpour H. An RNA-based theory of natural universal computation. Journal of Theoretical Biology. 2022 Mar;537:110984. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022519321004045. doi:10.1016/j.jtbi.2021.110984.

43. Shomrat T, Levin M. An automated training paradigm reveals long-term memory in planaria and its persistence through head regeneration. Journal of Experimental Biology. 2013 Jan;jeb.087809. doi:10.1242/jeb.087809.

44. Deochand N, Costello MS, Deochand ME. Behavioral research with planaria. Perspectives on Behavior Science. 2018 Nov;41(2):447-64. Available from: http://link.springer.com/10.1007/s40614-018-00176-w. doi:10.1007/s40614-018-00176-w.

45. Bray SR, Wyss LS, Chai C, Lozada ME, Wang B. Adaptive robustness through incoherent signaling mechanisms in a regenerative brain. Cell Reports. 2024 Aug;43(8):114580. Available from: https://doi.org/10.1016/j.celrep.2024.114580. doi:10.1016/j.celrep.2024.114580.

46. Shettigar N, Chakravarthy A, Umashankar S, Lakshmanan V, Palakodeti D, Gulyani A. Discovery of a body-wide photosensory array that matures in an adult-like animal and mediates eye–brain-independent movement and arousal. Proceedings of the National Academy of Sciences. 2021;118(20):e2021426118.

47. Shettigar N, Joshi A, Dalmeida R, Gopalkrishna R, Chakravarthy A, Patnaik S, et al. Hierarchies in light sensing and dynamic interactions between ocular and extraocular sensory networks in a flatworm. Science advances. 2017;3(7):e1603025.

48. Le D, Sabry Z, Chandra A, Kristan III WB, Collins EMS, Kristan Jr WB. Planarian fragments behave as whole animals. Current Biology. 2021;31(22):5111-7. doi:10.1016/j.cub.2021.09.056.

49. McConnell JV. Specific Factors Influencing Planarian Behavior. In: Corning WC, Ratner SC, editors. Chemistry of Learning. Boston, MA: Springer US; 1967. p. 217-33.

50. Lázaro EM, Harrath AH, Stocchino GA, Pala M, Baguñà J, Riutort M. Schmidtea mediterranea phylogeography: an old species surviving on a few Mediterranean islands? BMC evolutionary biology. 2011;11(1):274.

51. Sluys R, Riutort M. Planarian diversity and phylogeny. Planarian regeneration: methods and protocols. 2018:1-56.

52. McConnell JV. On the turning of worms: A reply to James and Halas. The Psychological Record. 1964;14(1):13.

53. Bray S, Wyss L, Wang B. Planameterization: Software to segment and parameterize videos of planarian behavior. GitHub repository. 2023. https://github.com/samuelbray32/planameterization.

54. Harris JA, Gallistel CR. Information, certainty, and learning. eLife. 2026;13:RP102155.

55. Best JB, Rubinstein I. Maze learning and associated behavior in planaria. Journal of comparative and physiological psychology. 1962;55(4):560.

56. Best JB. Protopsychology. Scientific American. 1963 Feb;208(2):54-62. Available from: https://www.scientificamerican.com/article/protopsychology. doi:10.1038/scientificamerican0263-54.

57. Best JB. Behavior of Planaria in Instrumental Learning Paradigms. National Aeronautics and Space Administration; 1963. NASA-CR-80380. Accession No. 67N13007. Available from: https://ntrs.nasa.gov/citations/19670003678.

58. Ardiel EL, Rankin CH. An elegant mind: learning and memory in Caenorhabditis elegans. Learning & memory. 2010;17(4):191-201.

59. Moore RS, Kaletsky R, Lesnik C, Cota V, Blackman E, Parsons LR, et al. The role of the Cer1 transposon in horizontal transfer of transgenerational memory. Cell. 2021 Sep;184(18):4697-712.e18. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867421008813. doi:10.1016/j.cell.2021.07.022.

60. Nishijima S, Maruyama IN. Appetitive olfactory learning and long-tearm associative memory in Caenorhabditis elegans. Frontiers in Behavioral Neuroscience. 2017 May;11:80. Available

from: http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00080/full. doi:10.3389/fnbeh.2017.00080.

61. Brown DD, Pearson BJ. A brain unfixed: unlimited neurogenesis and regeneration of the adult planarian nervous system. Frontiers in neuroscience. 2017;11:289.

62. McConnell JV. On the procuring and the care of planarians. Manual of Psychological Experimentation on Planarians, second ed Planarian Press, Ann Arbor, Michigan. 1967:128.

63. McConnell JV, Cornwell PR, Clay M. An Apparatus for Conditioning Planaria. The American Journal of Psychology. 1960 Dec;73(4):618. Available from: https://www.jstor.org/stable/1419955?origin=crossref. doi:10.2307/1419955.

64. Lubow RE, Moore A. Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. Journal of comparative and physiological psychology. 1959;52(4):415.

65. Wyss LS, Bray SR, Wang B. Neuropeptide-mediated temporal sensory filtering in a primordial nervous system. bioRxiv. 2024.

66. Griffard C. Classical conditioning of the planarian Phagocata gracilis to water flow. Journal of Comparative and Physiological Psychology. 1963;56(3):597.

67. Gallistel CR, Latham PE. Bringing Bayes and Shannon to the study of behavioural and neurobiological timing and associative learning. Timing & time perception. 2022;11(1-4):29-89.