

## RESEARCH ARTICLE

### Classical conditioning through auditory stimuli in *Drosophila*: methods and models

Gil Menda<sup>1</sup>, Haim Y. Bar<sup>2</sup>, Ben J. Arthur<sup>1</sup>, Patricia K. Rivlin<sup>1</sup>, Robert A. Wyttenbach<sup>1</sup>,  
 Robert L. Strawderman<sup>2,3</sup> and Ronald R. Hoy<sup>1,\*</sup>

<sup>1</sup>Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853, USA, <sup>2</sup>Department of Statistical Science, Cornell University, Ithaca, NY 14853, USA and <sup>3</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA

\*Author for correspondence (rrh3@cornell.edu)

Accepted 6 June 2011

#### SUMMARY

The role of sound in *Drosophila melanogaster* courtship, along with its perception *via* the antennae, is well established, as is the ability of this fly to learn in classical conditioning protocols. Here, we demonstrate that a neutral acoustic stimulus paired with a sucrose reward can be used to condition the proboscis-extension reflex, part of normal feeding behavior. This appetitive conditioning produces results comparable to those obtained with chemical stimuli in aversive conditioning protocols. We applied a logistic model with general estimating equations to predict the dynamics of learning, which successfully predicts the outcome of training and provides a quantitative estimate of the rate of learning. Use of acoustic stimuli with appetitive conditioning provides both an alternative to models most commonly used in studies of learning and memory in *Drosophila* and a means of testing hearing in both sexes, independently of courtship responsiveness.

Supplementary material available online at <http://jeb.biologists.org/cgi/content/full/214/17/2864/DC1>

Key words: proboscis-extension reflex (PER), generalized estimating equations (GEE), logistic regression, learning, Johnston's organ.

#### INTRODUCTION

Over half a century of research has established the role of sound in *Drosophila* courtship. A male extends one wing and vibrates it at a nearby female, producing a stereotyped pattern of pulse and sine song. Throughout the genus *Drosophila*, these wingbeat songs are species specific in their pattern and harmonic content (Hoikkala and Lumme, 1987; Hoy et al., 1988); *D. melanogaster* is no exception (Bennet-Clark and Ewing, 1967; Bennet-Clark and Ewing, 1969). Females detect this acoustic signal with their antennae (consisting of the arista and Johnston's organ), as has been shown by recordings of sound-evoked field potentials in Johnston's organ and from the antennal nerve (Ewing, 1978; Eberl et al., 2000; Tauber and Eberl, 2003).

Research into learning in *Drosophila* has been going on for about as long as study of courtship. In this time, most work has used aversive stimuli to test learning of chemical cues (e.g. Quinn et al., 1974; Dudai et al., 1976; Tully and Quinn, 1985; Pitman et al., 2009). For example, an odor is presented along with an electrical shock *via* the substrate; after a number of such trials, flies learn to associate the odor with the shock and make avoidance responses when presented with the (previously neutral) odor alone. In terms of classical conditioning, shock is an unconditional stimulus (US), avoidance is an unconditional response (UR) and odor is a conditional stimulus (CS). If associative learning occurs, the UR is evoked by the CS alone after a series of CS–US pairings. Some studies have used appetitive rather than aversive conditioning, with sucrose as the US and proboscis extension as the UR (e.g. Tempel et al., 1983; Chabaud et al., 2006). When a fly steps in sugar water, its tarsal chemoreceptors trigger a feeding reflex that extends the proboscis, through which it sucks the fluid (Dethier, 1976). This

proboscis-extension reflex (PER; Fig. 1D) is a fixed act, common to many insects, that has been used in studies of olfactory and taste learning in a variety of insects in addition to *Drosophila* (e.g. Nelson, 1971; Bitterman et al., 1983; Daly and Smith, 2000).

Much of the interest in learning and courtship in *Drosophila* is due to its status as a model organism in which mutants can be easily screened. Although many auditory mutants have been identified (Caldwell and Eberl, 2002), there is a limitation in that 'the only known acoustic behavior of fruit flies is their response to courtship songs' (Inagaki et al., 2010). In fact, all current methods of screening for auditory mutants are based on courtship, by testing either the receptivity of females or the tendency of males to court one another when stimulated with pulse song (Eberl et al., 1997; Inagaki et al., 2010). A new method of behaviorally testing hearing in both sexes, independent of courtship, could advance the study of hearing in *Drosophila*.

In the present study, we employed an appetitive conditioning protocol using sugar water as a reward, a non-courtship sound as a neutral stimulus and proboscis extension as an indicator of learning. We also recorded from Johnston's organ to verify that our stimuli were audible. To our knowledge, no prior studies of learning in *Drosophila* have used acoustic stimuli and only a few have used appetitive conditioning.

#### MATERIALS AND METHODS

##### Training experiments

##### Protocol

Training and testing were done according to the timeline in Fig. 2. Each fly was given six training trials. In 'paired' trials, flies were rewarded with 5 s access to sucrose (1 mol l<sup>-1</sup> solution) 5 s after

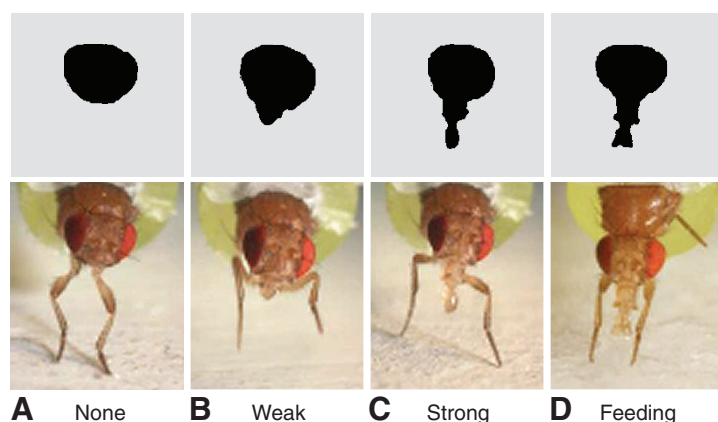


Fig. 1. Proboscis extension. When the foreleg tarsi touch sucrose solution, the proboscis reflexively extends to feed as shown in D. With lower levels of excitation, the proboscis may extend only partially (B,C) or not at all (A). Responses to sound were rated as none (A), weak (B, <50% of full extension) or strong (C,  $\geq$ 50% of full extension).

the onset of a 10 s sound stimulus. In 'unpaired' trials, sucrose was presented 30 s after the end of the sound, while sucrose was never presented in 'no-reward' trials. In all three types of trial, PER strength was rated during the first 5 s of sound stimulation, before the onset of any reward, and flies were presented with water 120 s after the sound stimulus to prevent dehydration and to wash off any sucrose remaining on the tarsi. Water presentation was isolated from sound stimulation by 2–3 min before and after, making it unlikely to affect training. The six training trials were followed by two trials testing for retention, one at 15 and one at 25 min after training, in which only the sound was presented and PER strength was rated.

#### Apparatus

The training device (Fig. 3) was built on a rotating 16 cm diameter kymograph drum (Bird Kymograph no. 70-060; Phipps & Bird Inc., Richmond, VA, USA), based on published designs (Vargo et al., 1983; Holliday and Hirsch, 1986; Brigui et al., 1990). The drum rotated fully in 5 min, presenting sound, sucrose and water on the schedule shown in Fig. 2. The sound stimulus was activated when a magnet on the drum moved past a reed switch 1 cm away from the drum. Closing the switch made no audible sound and did not transmit vibration to the drum. Sucrose solution and water were delivered from two different 8×160 cm strips of filter paper (Whatman no. 2300 916) mounted on the drum, fed by reservoirs on top of the drum. Three flies were tested at a time, each one loaded into a pipette tip and placed with its foreleg tarsi in contact with the drum. The three fly heads filled the frame of a video camera for later analysis.

#### Stimuli

The conditional stimulus was a 10 s, 400 Hz tone (DynaScan Corp. 3011 function generator; Irvine, CA, USA) broadcast through a 16 cm paper cone woofer and directed at the flies through a plastic funnel, with the 24 mm opening of the funnel 20 mm from the flies, placing flies in the near field. To avoid transmission of vibration to the rotating drum, the speaker was held by a stand not attached to the drum. Intensities were calibrated in dB SPL (re. 20  $\mu$ Pa) with a Brüel & Kjær type 2209 sound level meter with a type 4138 1/8 in microphone in the location that would be occupied by the central fly. Three intensity levels were used, 65 dB SPL (quiet – just above auditory threshold), 85 dB SPL (moderate – near the natural level of courtship song) (Bennet-Clark, 1971) and 108 dB SPL (loud). All three training conditions, paired, unpaired and no reward, were tested at each of the three intensities.

#### Response

Proboscis extension is an unconditional response when the fly's tarsi contact sucrose solution (Nelson, 1971; Médioni and Vaysse, 1975; McKenna et al., 1989). All training and test sessions were recorded on video for later frame-by-frame analysis. PER strength was categorized as no response, weak response or strong response (Fig. 1) (Chabaud et al., 2006). Rating was carried out by two observers, at least one of whom did not know the training protocol in use.

#### Subjects

Three-day-old virgin female *D. melanogaster* (Canton-S-5 strain) (McKenna et al., 1989) were taken from our laboratory cultures.

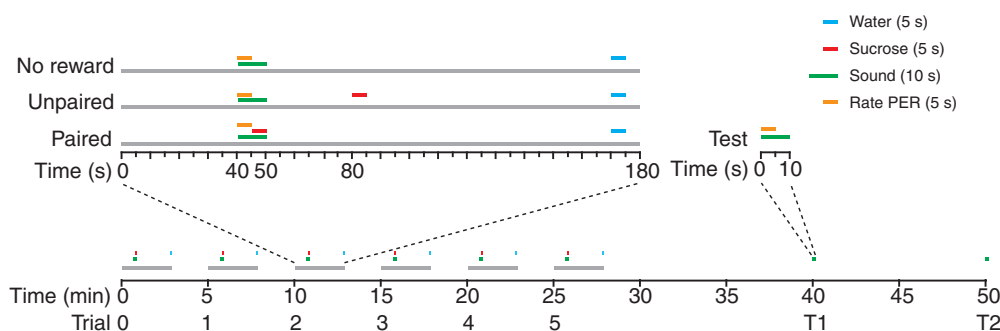


Fig. 2. Training schedule. There were three experimental groups: paired, unpaired and no reward. Each group experienced six training trials (0–5) with a 5 min inter-trial interval, followed by two retention tests (T1 and T2) occurring 15 and 25 min after training. In the paired condition, a 10 s sound stimulus overlapped with a 5 s sucrose reward; in the unpaired condition, presentation of sucrose was delayed by 30 s; in the no-reward condition, sucrose was never given. In all cases water was given 120 s after sound offset. PER, proboscis-extension reflex.

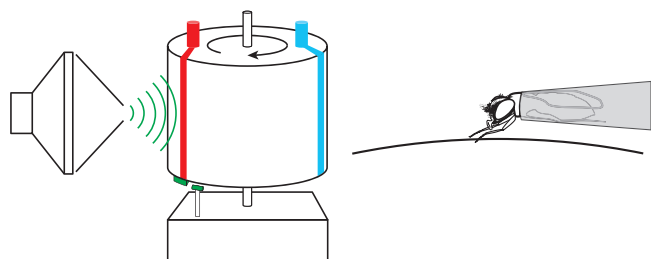


Fig. 3. Training apparatus. Flies were mounted in plastic pipette tips and held with their foreleg tarsi touching the surface of a rotating kymograph drum. Rotation of the drum controlled the presentation of sound (green), sucrose (red) and water (blue) according to the schedule shown in Fig. 2.

They were starved for 24 h before testing to increase their motivation to feed.

### Statistical analysis

In Pavlovian conditioning, learning is an increase in response over time when a stimulus predicts a reward (paired trials), compared with a constant response over time when stimulus and reward are uncorrelated (unpaired trials). To test whether learning occurs, we used a parametric statistical model in which differences between paired and unpaired groups and the change in response within each group over time are integrated into a logistic regression model:

$$\text{Test: } \log \frac{P(t)}{1 - P(t)} = \alpha + \beta_T t \quad \text{Control: } \log \frac{P(t)}{1 - P(t)} = \alpha + \beta_C t, \quad (1)$$

where  $P(t)$  is the probability that a fly responds at trial  $t$  ( $t=0$  being the initial trial). These equations describe sigmoidal growth curves for probability of response as a function of trial number. The rate of growth is reflected by the regression coefficients  $\beta_T$  (for the test, paired condition) and  $\beta_C$  (for the control, unpaired condition), while the baseline response is reflected by the intercept,  $\alpha$ .

Consequently, we can ascertain whether learning occurs using just three statistical tests. First, the log-odds of a response at  $t=0$  are equal to  $\alpha$  for both groups. Thus, we test whether the initial response rate is the same for the paired and unpaired groups, using a test for the difference between two proportions or a related procedure such as Fisher's exact test. We call this test  $T_0$  and its null hypothesis  $H_{0(0)}$ . Next, any change in response rate over time should be monotonic, as the log-odds are linear functions of  $t$  with slopes  $\beta_T$  and  $\beta_C$  for the paired and unpaired groups, respectively. In particular, if the stimulus induces learning, there should be an increase in the response rate of the paired group (i.e. the regression coefficient  $\beta_T$  is positive) but not of the control group (i.e.  $\beta_C$  is zero or less). Thus, our second ( $T_1$ ) and third ( $T_2$ ) statistical tests concern the hypotheses:

$$H_{0(1)} : \beta_C \leq 0 \text{ vs } H_{A(1)} : \beta_C > 0 \text{ and } H_{0(2)} : \beta_T \leq 0 \text{ vs } H_{A(2)} : \beta_T > 0. \quad (2)$$

Rejection of  $H_{0(2)}$  along with failure to reject  $H_{0(0)}$  and  $H_{0(1)}$  would be evidence of Pavlovian learning.

Because three tests are performed, the significance level for each test must be adjusted to ensure that the overall experimental Type I error is controlled at a level of 0.05. Using the conservative Bonferroni procedure, hypotheses  $H_{0(0)}$ ,  $H_{0(1)}$  and  $H_{0(2)}$  are tested at the reduced level  $0.05/3=0.0167$ . To account for the fact that responses across trials within each fly are dependent on each other, estimates of the regression coefficients (i.e. learning rates), their

standard errors, and  $P$ -values used in tests of  $H_{0(1)}$  and  $H_{0(2)}$  are obtained using the technique of generalized estimating equations (GEE) (Zeger and Liang, 1986; Verbeke and Molenberghs, 2001). Using GEE rather than likelihood-based methods ensures that the standard error estimates for regression coefficients reflect the dependence of the responses recorded for each fly across the series of trials, thereby avoiding inappropriate assessments of statistical significance.

### Physiological recording

As an independent test of stimulus audibility, we recorded from Johnston's organ. Recording and stimulus calibration were done as previously (Arthur et al., 2010). A fly was dorsally tethered with paraffin and positioned such that the plane of the arista was parallel to sound waves emanating from a speaker. A tungsten electrode was inserted in the second antennal segment (Johnston's organ) and the Bayesian QUEST procedure (Watson and Pelli, 1983) was used to adaptively quantify thresholds to a 400 Hz stimulus. Thresholds to periodic oscillations at the fundamental and second harmonic of the stimulus frequency were tracked in parallel (King-Smith et al., 1994), with a stimulus deemed above threshold if the response was greater during the 0.5 s stimulus presentation than during an equal amount of time immediately preceding it. The customary Weibull function was used as the assumed neurometric function with the threshold criterion set to 75%, the slope ( $\beta$ ) to 0.05 based on pilot data, the probability of failure at infinite intensity ( $\delta$ ) to 0.01, and the probability of success at negative infinity ( $\gamma$ ) to 0.5. To facilitate validation of the estimated slope parameter, each fly was used in  $\geq 200$  trials, with a uniform random intensity within the 55–95% range of the current estimated threshold for each trial (supplementary material Fig. S1). The final threshold estimate was taken to be the mean of the posterior density function when its standard deviation dropped to 3 dB. Immediately after Johnston's organ recordings were concluded, control recordings were conducted in the contralateral eye of each fly to check for stimulus artifact due to coupling with the speaker. Both antennae were removed for control recordings because attenuated potentials from antennae may be recorded throughout the head, as we and others have observed in mosquitoes (Wishart et al., 1962; Arthur et al., 2010). In all cases the reference electrode was placed in the thorax through a coxal stump following removal of a leg.

### Sound intensity and calibration

Because antennae respond to the particle-velocity component of sound (Bennet-Clark, 1971), intensities ought to be calibrated with velocity microphones and presented in dB SPVL (re.  $50 \text{ nm s}^{-1}$ ). However, measurement of particle velocity is difficult in confined spaces; indeed, both velocity and pressure can vary considerably over small distances in echoic environments. Most studies of *Drosophila* hearing and courtship take place in small and echoic chambers (e.g. Eberl et al., 1997; Inagaki et al., 2010), with calibration microphones often placed near but outside the chamber. Some use pressure microphones, some use velocity microphones, and some give no calibration detail at all. As a result, it is difficult to compare intensities across studies.

Calibration of stimuli for learning was done as described above with a pressure-sensitive microphone at the central fly position. We subsequently mapped the sound field between the funnel exit and the kymograph drum with both a pressure microphone and a pressure-gradient (particle-velocity) microphone as detailed below. Intensity was highest at the center of this space, dropping off at the edges and near the drum, with velocity varying more than pressure.

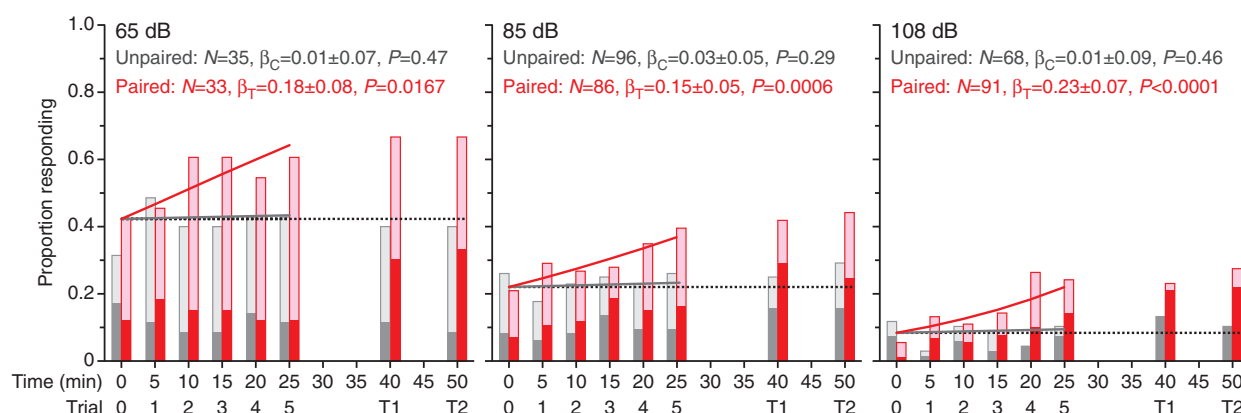


Fig. 4. Response to sound during training. Bars show the proportion responding in the paired (red) and unpaired (gray) groups. Dark portions indicate a strong PER; light portions indicate a weak PER. The no-reward group (not shown) was indistinguishable from the unpaired group. Curves show the response probabilities for each group predicted by our GEE model; dashed lines are the no-learning rates (equal to the initial response rate). For each intensity and test condition, we report the number of flies ( $N$ ), the value and standard error of the GEE regression slope parameter ( $\beta_T$  for the test paired condition;  $\beta_C$  for the control unpaired condition) and the  $P$ -value for the regression parameter. The last two points (T1 and T2) are tests of retention in which only the sound was presented.

At our moderate intensity level, for example, at the surface of the drum we measured 85 dB SPL and 95 dB SPVL at the location of the central fly, and 82 dB SPL and 90 dB SPVL at the edge of the funnel; moving the microphones away from the drum and toward the center of the funnel exit, we measured 88 dB SPL and 109 dB SPVL. The nominal values we report came from the center near the drum; intensities experienced by the three flies would vary somewhat by position and possibly by trial. While it is unlikely that intensities were precisely 65, 85 or 108 dB SPL for each fly in each trial, it is certain that the 108 dB stimulus was louder than the 85 dB stimulus, which was louder than the 65 dB stimulus.

Without the constraint of putting flies in a training apparatus, physiological recording took place in a far cleaner acoustic environment. In an anechoic far field, dB SPL and dB SPVL are numerically equal (Bennet-Clark, 1971) because a pressure of  $20\mu\text{Pa}$  (0 dB SPL) corresponds to a velocity of  $50\text{ nm s}^{-1}$  (0 dB SPVL). We calibrated in dB SPVL using a pressure-gradient microphone (Knowles NR-23158) as described previously (Arthur et al., 2010) and in dB SPL using a pressure microphone (Brüel & Kjær type 4138). The results confirm that our recording location was in fact anechoic and far-field at 400 Hz.

## RESULTS

The results of the training study are shown in Fig. 4. For the purpose of analysis, weak and strong PER responses were lumped to create simple response and no-response categories. Results in the no-reward condition were indistinguishable from those in the unpaired condition, so our analysis considered only the paired and unpaired trials. It is clear that flies in the unpaired condition did not learn to associate sound with sucrose: the proportion responding stayed at the initial level throughout. Flies in the paired condition showed an increase in response probability as trials progressed, indicating that they associated the sound with the sucrose reward. Statistically, these observations are borne out at each intensity level by the three tests described above. The initial response probability did not differ between paired and unpaired groups (Fisher's exact test:  $P=0.45$  at 65 dB,  $P=0.49$  at 85 dB,  $P=0.24$  at 108 dB), regression coefficients of unpaired trials are not significantly greater than zero (GEE Z-test:  $P=0.43$  at 65 dB,  $P=0.13$  at 85 dB,  $P=0.30$  at 108 dB), and regression coefficients of paired trials are positive (GEE Z-test:

$P=0.0167$  at 65 dB,  $P=0.0006$  at 85 dB,  $P<0.0001$  at 108 dB). Thus, the results meet the standards of learning as described by our logistic regression model. In the two test trials, responses remained high, showing that the association between sound and sucrose was retained 15 and 25 min after the end of training.

To test whether learning required the auditory modality, we conducted a set of paired trials at 108 dB with 14 flies in which the antennae were ablated. These flies did not show an increase in response as trials progressed (GEE Z-test:  $\beta_T=-0.18\pm0.12$ , mean  $\pm$  s.e.m.,  $P=0.11$ ; data not shown).

It is clear from Fig. 4 that stimulus intensity affected the proportion of flies responding in all trials, in both paired and unpaired groups. As sound intensity increased, the proportion responding decreased (chi-square test:  $P<0.0001$ ). However, the slope of the logistic regression ( $\beta_T$ ) did not vary significantly with intensity.

Recordings from Johnston's organ (Fig. 5) show that all stimuli were audible. The response consists of a phasic negative deflection of the baseline followed by a tonic positive deflection. In addition, the field potential oscillates at the frequency of the stimulus and its harmonics. Our Bayesian threshold search was based on the oscillatory potential; *post hoc* analysis of the baseline shift showed its threshold to be higher than that of the oscillatory potential. In one fly, the baseline shift was not phasic but was a tonic negative deflection (supplementary material Fig. S2) as in *Aedes* mosquitoes (Cator et al., 2009; Arthur et al., 2010). Across all 11 flies tested, thresholds with a 75% criterion were  $65\pm1$  dB (mean  $\pm$  s.e.m.) for the stimulus frequency ( $F_0$ ) and  $64\pm1$  dB for twice the stimulus frequency ( $F_1$ ) with a 0.5 s stimulus. These levels are approximately those used as the least intense CS in training experiments. However, these may be conservative estimates of threshold. With longer stimuli, such as the 10 s used in our learning experiments, thresholds could be lower if flies integrate over long periods.

## DISCUSSION

### Appetitive conditioning with acoustic stimuli

Our results and analysis show that female *Drosophila* learned to associate a tone with a sucrose reward. While *Drosophila* has long been a model system for the study of learning and memory, most previous work has used chemical stimuli as the CS and aversive stimuli, generally electrical shock, as the US (e.g. Quinn et al., 1974;



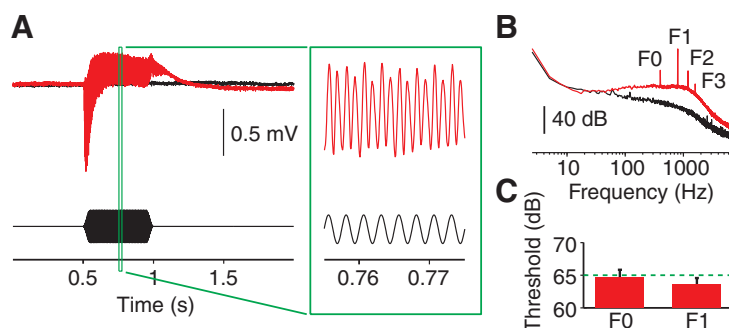


Fig. 5. Field potentials in Johnston's organ. (A) Sample response to 400 Hz in a female fly (averaged over 85 repetitions). Control recording from the eye is overlaid in black; stimulus is shown below. Recordings were made over a wide range of intensities; this example is for 107 dB, where the high signal-to-noise ratio made the response clearly visible. (B) Spectral analysis of the sample response in A. Labels F1 to F3 indicate harmonics of the 400 Hz fundamental frequency, F0. (C) Response thresholds (75% criterion) averaged across 11 females. Thresholds are shown separately for components of the response at the stimulus frequency (F0) and twice the stimulus frequency (F1). The dashed line indicates the lowest intensity used in the learning experiments.

Dudai et al., 1976; Tully and Quinn, 1985; Pitman et al., 2009). In contrast, we used an acoustic CS and a rewarding US. The prior study most similar to ours was by Chabaud and colleagues, who examined appetitive conditioning with an olfactory CS (Chabaud et al., 2006). Their wild-type flies started from a baseline of 25–40% PER, increasing to 60–75% PER after five trials, similar to the baseline and change we found in the 65 dB group.

Sound is neutral in the context of feeding but salient in the context of courtship. We chose 400 Hz as the CS in the hope of finding an auditory stimulus that was neutral but audible. This frequency is sufficiently far from that of the sinusoidal component of courtship song (160 Hz) (Wheeler et al., 1988) that it is unlikely to be perceived as courtship, but likely to be within the audible range for insects that hear with their antennae. Recordings from Johnston's organ verified that the CS was above threshold.

Given our assumption that 400 Hz is neutral, neither attractive nor aversive, the effect of intensity on proboscis extension is puzzling. While the 65 dB group started from a baseline of 40% PER, the 108 dB group started at only 10% PER. Many animals, including insects, respond to loud sounds with an acoustic startle response that, while not eliciting escape behavior, freezes or disrupts ongoing behavior (Eaton, 1984; Hoy, 1989). This may account for the low PER at 108 dB. Indeed, informal analysis of video collected before each trial found flies in the 108 dB group extending their proboscises, in the absence of stimuli, to an extent approximately equal to that found during the 65 dB trials. Thus, it appears that proboscis extension was reduced by loud sound. Despite its effect on background responsiveness, intensity had no effect on the rate of learning ( $\beta_T$  in Fig. 4), contradicting the usual expectation that rate increases with CS intensity (Davey, 1981). Taken together, these observations suggest that sound intensity can be kept near threshold in learning studies, avoiding startle without slowing the rate of learning. If a lower background response is desired, it can be achieved by increasing sound intensity.

#### Analysis of learning

Because our statistical tests for learning differ from those commonly used in *Drosophila* learning research, we offer an explanation and comparison with other methods. A logistic model with GEE estimation, hereafter referred to as LGEE, is generally applicable to associative learning. It has been used in several other learning studies; our method is most similar to that recommended by Hartz and colleagues (Hartz et al., 2001) and used by Shafir and colleagues (Shafir et al., 2005). Use of LGEE is motivated by two statistical considerations. First, the logistic model captures the dynamic, sequential nature of learning in a simple model that permits meaningful tests for time trends within groups as well as for differences in learning rates between groups. Indeed, logistic regression is the most commonly used statistical tool for modeling

trends in proportions (Collett, 2002). Second, LGEE uses all of the data collected from each subject on each trial while accounting for the dependence of responses within a subject across trials.

Although established as a learning model in other research communities, the use of LGEE with *Drosophila* is novel. Many studies simply report the difference between post-training and pre-training responses and compare it between paired and unpaired groups. While this is a valid way to determine whether learning occurs, it provides no information about the rate of learning. Chabaud and colleagues (Chabaud et al., 2006) improved on this by assessing learning as follows. (1) For each trial, employ a chi-square or related test (such as Fisher's exact) to check whether the proportion of responses in the paired group differs from that in the unpaired group. (2) For each group, apply Cochran's *Q*-test to determine whether the proportion responding is constant over time. Thus for six training trials, a total of eight statistical tests are required, six to compare responses between groups in each trial and two to test for constant response in each group. We henceforth refer to this set of procedures as CHIQ.

The CHIQ procedure enforces no preconceived notion of learning, in the form of either trends over time or the direction of differences between paired and unpaired groups. Thus, there is ambiguity in how one detects learning. Must each of the chi-square tests be statistically significant? If not, which of them must be significant in order to conclude that learning has occurred? The role of Cochran's *Q*-test is unclear when used with the full set of between-group tests. It is not even clear that the test is valid in this context. As originally designed, Cochran's *Q*-test was not meant to evaluate the equality of proportions across trials within a group unless more stringent assumptions are imposed on the probability of response for all subjects in all pairs of trials (Bhappkar, 1973; Somes and Bhappkar, 1977). A further drawback of CHIQ is the need to adjust significance levels for the larger number of tests. For an overall type I error of 0.05, each of the eight tests that would be required for our data would have to be done at a level of  $0.05/8=0.00625$ , as opposed to the 0.0167 required for LGEE. This, combined with the fact that none of the tests used in CHIQ are directional, greatly increases its conservatism in assessing learning. The associated price is a loss of statistical power, possibly severe, for detecting learning when learning in fact exists.

LGEE equates learning with a monotonic increase in the proportion responding over a series of trials in the paired group but not in the unpaired group, while CHIQ detects any kind of change in response profiles, either between or within groups. In theory, this gives CHIQ flexibility, trading loss of statistical power for greater robustness. For example, the LGEE assumption of monotonicity could be violated by satiation, although this is not a concern with a relatively small number of trials and a well-designed experiment. In principle, however, we argue that true learning must manifest

itself, at least initially, in non-decreasing (if not monotonically increasing) responses, and the primary goal should be detection of that directional trend. For this purpose, LGEE is powerful and robust. Furthermore, it would be easy to extend the logistic model to allow both non-decreasing and non-monotonic trends, although as with CHIQ, adding such flexibility complicates the definition of learning.

### Auditory responses

Recordings from Johnston's organ confirm that the auditory stimuli in learning tests were above sensory threshold. Our threshold of 64–65 dB SPVL in females is close to the 72 dB SPL threshold found behaviorally in males (Eberl et al., 1997). The oscillatory component of the evoked field potential was qualitatively similar to that recorded in mosquitoes (e.g. Tischner, 1953; Wishart et al., 1962). Most baseline shifts were phasic, as described for *Culex* mosquitoes by Warren and colleagues (Warren et al., 2009), but one fly showed a sustained deflection such as we found in *Aedes* mosquitoes (Cator et al., 2009; Arthur et al., 2010). Atypical recordings from one of 11 flies might be dismissed as damage during preparation, but this fly had a normal threshold and its responses were consistent and similar to those found in other species. We suspect that the nature of the baseline shift varies with electrode placement in a heterogeneous population of scolopidia (Kamikouchi et al., 2006; Kamikouchi et al., 2009); future studies could systematically vary placement to test this.

Recording from Johnston's organ in *Drosophila* is not difficult, and we suggest that future studies using auditory stimuli in associative learning assays be combined with recordings to narrow down the anatomic location of deficits. Similarly, if auditory mutants are used in associative learning assays, it is important to test them with other learning assays to control for possible pleiotropic effects on learning.

### CONCLUSIONS

Aside from the nice historical coincidence that Pavlov's original model works for flies as well as for dogs, there are good scientific reasons to develop diverse learning models for *Drosophila*. For example, the learning mutants *dunce* and *rutabaga* fail to learn odors in a shock-avoidance protocol but learn normally in an appetitive task (Tempel et al., 1983), while aversive and appetitive learning of odors take place through different neural and biochemical pathways in *Drosophila* larvae (Honjo and Furukubo-Tokunaga, 2009). Learning mutants may act differently not only in different protocols but also with different CS modalities. In general, a greater variety of tests would allow a finer parsing of the array of learning mutations.

In principle, our methods could be adapted to test hearing in both sexes independently of courtship, facilitating the discovery of new auditory mutations. While screening for auditory mutations in males is well established (Eberl et al., 1997), there is no known way to screen auditory mutations in females beyond testing for courtship receptivity (Inagaki et al., 2010). Given that factors other than hearing are likely to affect courtship in both sexes, a test of hearing that does not rely on courtship could be valuable; our method seems to be the only such test available at present.

Finally, we urge that researchers in *Drosophila* learning adopt the logistic regression model with generalized estimating equations. It is a more statistically valid means of analysis than those commonly used today. At the same time, it allows different learning mutants to be quantitatively compared not only in their overall level of learning but also in their rate of learning, which may differ between mutants and could eventually provide insight into mechanisms.

### ACKNOWLEDGEMENTS

We thank Dustin Rubinstein for thoughtful comments and discussion, Margaret Nelson for advice on the setup, Kristin Gawera for drawing the fly in Fig. 3, and two anonymous reviewers for helpful comments on an earlier version of the manuscript. This research was supported by NIH award 5R01DC103-36 to R.R.H. Deposited in PMC for release after 12 months.

### REFERENCES

- Arthur, B. J., Wytenbach, R. A., Harrington, L. C. and Hoy, R. R. (2010). Neural responses to one- and two-tone stimuli in the hearing organ of the dengue vector mosquito. *J. Exp. Biol.* **213**, 1376–1385.
- Bennet-Clark, H. C. (1971). Acoustics of insect song. *Nature* **234**, 255–259.
- Bennet-Clark, H. C. and Ewing, A. W. (1967). Stimuli provided by courtship of male *Drosophila melanogaster*. *Nature* **215**, 669–671.
- Bennet-Clark, H. C. and Ewing, A. W. (1969). Pulse interval as a critical parameter in the courtship song of *Drosophila melanogaster*. *Anim. Behav.* **17**, 755–759.
- Bhappkar, V. P. (1973). On the comparison of proportions in matched samples. *Sankhya - Indian J. Stat. A* **35**, 341–356.
- Bitterman, M. E., Menzel, R., Fietz, A. and Schäfer, S. (1983). Classical conditioning of proboscis extension in honeybees (*Apis mellifera*). *J. Comp. Psychol.* **97**, 107–119.
- Brigui, N., le Bourg, E. and Médioni, J. (1990). Conditioned suppression of the proboscis-extension response in young, middle-aged, and old *Drosophila melanogaster* flies: acquisition and extinction. *J. Comp. Psychol.* **104**, 289–296.
- Caldwell, J. C. and Eberl, D. F. (2002). Towards a molecular understanding of *Drosophila* hearing. *J. Neurobiol.* **53**, 172–189.
- Cator, L. J., Arthur, B. J., Harrington, L. C. and Hoy, R. R. (2009). Harmonic convergence in the love songs of the dengue vector mosquito. *Science* **323**, 1077–1079.
- Chabaud, M.-A., Devaud, J.-M., Pham-Delègue, M.-H., Preat, T. and Kaiser, L. (2006). Olfactory conditioning of proboscis activity in *Drosophila melanogaster*. *J. Comp. Physiol. A* **192**, 1335–1348.
- Collett, D. (2002). *Modelling Binary Data* (2nd edn). Boca Raton, FL: Chapman & Hall/CRC.
- Daly, K. C. and Smith, B. H. (2000). Associative olfactory learning in the moth *Manduca sexta*. *J. Exp. Biol.* **203**, 2025–2038.
- Davey, G. (1981). *Animal Learning and Conditioning*. New York: Palgrave Macmillan.
- Dethier, V. G. (1976). *The Hungry Fly: A Physiological Study of the Behavior Associated with Feeding*. New York: Harvard University Press.
- Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G. and Benzer, S. (1976). *dunce*, a mutant of *Drosophila* deficient in learning. *Proc. Natl. Acad. Sci. USA* **73**, 1684–1688.
- Eaton, R. C. (1984). *Neural Mechanisms of Startle Behavior*. New York: Springer.
- Eberl, D. F., Duyk, G. M. and Perrimon, N. (1997). A genetic screen for mutations that disrupt an auditory response in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **94**, 14837–14842.
- Eberl, D. F., Hardy, R. W. and Kernan, M. J. (2000). Genetically similar transduction mechanisms for touch and hearing in *Drosophila*. *J. Neurosci.* **20**, 5981–5988.
- Ewing, A. W. (1978). The antenna of *Drosophila* as a 'love song' receptor. *Physiol. Entomol.* **3**, 33–36.
- Hartz, S. M., Ben-Shahar, Y. and Tyler, M. (2001). Logistic growth curve analysis in associative learning data. *Anim. Cogn.* **3**, 185–189.
- Hoikkala, A. and Lumme, J. (1987). The genetic basis of evolution of the male courtship sounds in the *Drosophila virilis* group. *Evolution* **41**, 827–845.
- Holliday, M. and Hirsch, J. (1986). Excitatory conditioning of individual *Drosophila melanogaster*. *J. Exp. Psychol. Anim. Behav. Process.* **12**, 131–142.
- Honjo, K. and Furukubo-Tokunaga, K. (2009). Distinctive neuronal networks and biochemical pathways for appetitive and aversive memory in *Drosophila* larvae. *J. Neurosci.* **29**, 852–862.
- Hoy, R. R. (1989). Startle, categorical response, and attention in acoustic behavior of insects. *Annu. Rev. Neurosci.* **12**, 355–375.
- Hoy, R., Hoikkala, A. and Kaneshiro, K. (1988). Hawaiian courtship songs: evolutionary innovation in communication signals of *Drosophila*. *Science* **240**, 217–219.
- Inagaki, H. K., Kamikouchi, A. and Ito, K. (2010). Protocol for quantifying sound-sensing ability of *Drosophila melanogaster*. *Nat. Protoc.* **5**, 26–30.
- Kamikouchi, A., Shimada, T. and Ito, K. (2006). Comprehensive classification of the auditory sensory projections in the brain of the fruit fly *Drosophila melanogaster*. *J. Comp. Neurol.* **499**, 317–356.
- Kamikouchi, A., Inagaki, H. K., Effertz, T., Hendrich, O., Fiala, A., Gopfert, M. C. and Ito, K. (2009). The neural basis of *Drosophila* gravity-sensing and hearing. *Nature* **458**, 165–171.
- King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C. and Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res.* **34**, 885–912.
- McKenna, M., Monte, P., Helfand, S. L., Woodard, C. and Carlson, J. (1989). A simple chemosensory response in *Drosophila* and the isolation of *acj* mutants in which it is affected. *Proc. Natl. Acad. Sci. USA* **86**, 8118–8122.
- Médioni, J. and Vayssie, G. (1975). Conditional suppression of a reflex in *Drosophila melanogaster*: acquisition and extinction. *C. R. Seances. Soc. Biol. Fil.* **169**, 1386–1391.
- Nelson, M. C. (1971). Classical conditioning in the blowfly (*Phormia regina*): associative and excitatory factors. *J. Comp. Physiol. Psychol.* **77**, 353–368.
- Pitman, J., DasGupta, S., Krashes, M., Leung, B., Perrat, P. and Waddell, S. (2009). There are many ways to train a fly. *Fly* **3**, 3–9.
- Quinn, W. G., Harris, W. A. and Benzer, S. (1974). Conditioned behavior in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **71**, 708–712.
- Shafir, S., Menda, G. and Smith, B. H. (2005). Caste-specific differences in risk sensitivity in honeybees, *Apis mellifera*. *Anim. Behav.* **69**, 859–868.

- Somes, G. W. and Bhapkar, V. P.** (1977). A simulation study on a Wald statistic and Cochran's Q statistic for stratified samples. *Biometrics* **33**, 643-651.
- Tauber, E. and Eberl, D. F.** (2003). Acoustic communication in *Drosophila*. *Behav. Processes* **64**, 197-210.
- Tempel, B. L., Bonini, N., Dawson, D. R. and Quinn, W. G.** (1983). Reward learning in normal and mutant *Drosophila*. *Proc. Natl. Acad. Sci. USA* **80**, 1482-1486.
- Tischner, H.** (1953). Über den Gehörsinn von Stechmücken. *Acustica* **3**, 335-343.
- Tully, T. and Quinn, W. G.** (1985). Classical conditioning and retention in normal and mutant *Drosophila melanogaster*. *J. Comp. Physiol. A* **157**, 263-277.
- Vargo, M., Holliday, M. and Hirsch, J.** (1983). Automated stimulus presentation for the proboscis extension reflex in diptera. *Behav. Res. Methods Instrum.* **15**, 1-4.
- Verbeke, G. and Molenberghs, G.** (2001). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Warren, B., Gibson, G. and Russell, I. J.** (2009). Sex recognition through midflight mating duets in *Culex* mosquitoes is mediated by acoustic distortion. *Curr. Biol.* **19**, 485-491.
- Watson, A. B. and Pelli, D. G.** (1983). QUEST: a Bayesian adaptive psychometric method. *Percept. Psychophys.* **33**, 113-120.
- Wheeler, D. A., Fields, W. L. and Hall, J. C.** (1988). Spectral analysis of *Drosophila* courtship songs: *D. melanogaster*, *D. simulans*, and their interspecific hybrid. *Behav. Genet.* **18**, 675-703.
- Wishart, G., van Sickle, G. R. and Riordan, D. F.** (1962). Orientation of the males of *Aedes aegypti* (L.) (Diptera: Culicidae) to sound. *Can. Entomol.* **94**, 613-626.
- Zeger, S. L. and Liang, K. Y.** (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.