One-to-one mapping between deep network units and real neurons uncovers a visual population code for social behavior

Benjamin R. Cowley[1], Adam J. Calhoun[1], Nivedita Rangarajan[1], Jonathan W. Pillow[1], and Mala Murthy*[1]

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
*Lead contact; Authors for correspondence: pillow@princeton.edu and mmurthy@princeton.edu

Abstract

The rich variety of behaviors observed in animals arises through the complex interplay between sensory processing and motor control [1, 2, 3, 4, 5]. To understand these sensorimotor transformations, it is useful to build models that predict not only neural responses to sensory input [6, 7, 8, 9, 10] but also how each neuron causally contributes to behavior [11, 12]. Here we demonstrate a novel modeling approach to identify a one-to-one mapping between internal units in a deep neural network and real neurons by predicting the behavioral changes arising from systematic perturbations of more than a dozen neuron types. A key ingredient we introduce is "knockout training", which involves perturbing the network during training to match the perturbations of the real neurons during behavioral experiments. We apply this approach to model the sensorimotor transformation of *Drosophila melanogaster* males during a complex, visually-guided social behavior [13, 14, 15, 16]. Contrary to prevailing views [17, 18, 19], our model suggests that visual projection neurons at the interface between the eye and brain form a distributed population code that collectively sculpts social behavior. Overall, our framework consolidates behavioral effects elicited from various neural perturbations into a single, unified model, providing a detailed map from stimulus to neuron to behavior.

Main

To understand how the brain performs a sensorimotor transformation, an emerging and popular approach is to first train a deep neural network (DNN) model on a behavioral task performed by an animal (e.g., recognize an object in an image) and then compare the animal's neural activity to the internal activations of the DNN [6, 7, 8, 10, 20, 21, 22]. A shortcoming of this approach is that the DNN fails to predict how an individual neuron *causally* contributes to behavior, making it difficult to interpret a neuron's role in the sensorimotor transformation. Here we overcome this drawback by perturbing the internal units of a DNN model while predicting the behavior of animals whose neurons have also been perturbed, a method we call 'knockout training'. This approach places a strong constraint on the model: Each model unit must contribute to behavior in a way that matches the corresponding real neuron's causal contribution to behavior. An added benefit is that the model infers neural activity from (perturbed) behavior alone. This is especially useful when studying social behaviors, as it is challenging (or impossible in many systems) to record neural activity during natural, social interactions. Here we use this approach to investigate the sensorimotor transformation of *Drosophila* fruit flies during natural courtship.

Training a deep neural network to model transformations from vision to behavior

The *Drosophila* visual system contains a bottleneck between the eyes and the central brain in the form of visual projection neurons (Fig. 1**a**). The primary cell types of this bottleneck are the ∼40 Lobula Columnar or LC neuron types that receive input from the optic lobes (in the lobula or lobula plate) and send axons to a set of optic glomeruli that are read out by the central brain [18, 23]. The fact that neurons of the same LC type innervate the same optic glomerulus [18, 24] supports the hypothesis that each LC type is its own "feature detector" to extract visual information and modulate sensory-driven behavior. This has motivated recent studies to focus on a specific LC neuron type, one at a time, in order to uncover which feature that LC type detects [16, 17, 25, 26, 27, 28, 29, 30, 31]. On the other hand, the glomeruli of the olfactory system, which closely resemble optic glomeruli (e.g., each olfactory glomerulus receives input from only one olfactory receptor neuron) [32, 33], support a different picture: The olfactory glomeruli form a distributed population code that is read out in a seemingly random way by downstream neurons in the mushroom bodies [34, 35]. Whether LC neurons function as a set of labeled lines or a distributed population code remains unresolved.

To address this question, we designed a novel deep neural network (DNN) modeling approach for identifying the functional roles of individual LC neuron types using behavioral data from genetically altered flies (Fig. 1**a**, bottom diagram). Our approach relies on a DNN model with three components: (1) a front-end convolutional vision network that reflects processing in the optic lobe; (2) a "bottleneck" layer of LC units (where each model LC unit represents the summed activity of neurons of the same LC type); and (3) a decision network with dense connections that maps LC responses to behavior, reflecting downstream processing in the central brain and ventral nerve cord. We enforced the bottleneck layer to have the same number of units as LC neuron types we manipulated, al-

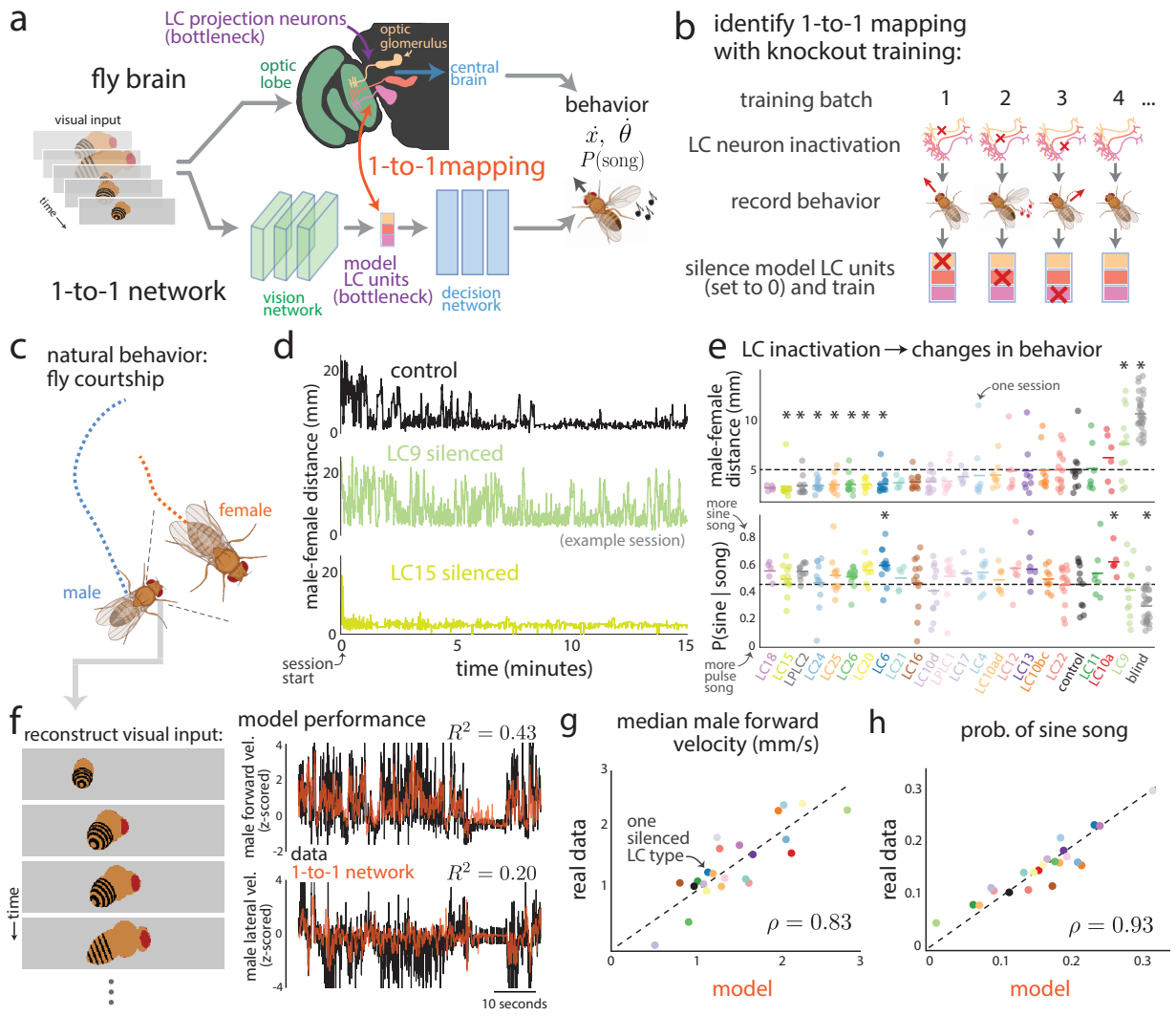


Figure 1: Identifying a one-to-one mapping between real neurons and internal units of a deep neural network with knockout training. a. Modeling the transformation from vision to behavior in male flies. The model (termed the 1-to-1 network) consists of a vision network with convolutional filters (green), a decision network with dense connections (blue), and a bottleneck of model LC units that matches the bottleneck of LC neurons found in the fly's visual system (purple). We seek a one-to-one mapping in which each model LC unit directly maps to one LC neuron type (e.g., yellow to yellow). The model takes as input a sequence of images and outputs its prediction of the male's movement, including forward, lateral, and angular velocity, as well the male's song produced by wing vibration, including sine, pulse-fast (Pfast), and pulse-slow (Pslow) song. **b**. We designed a new procedure called knockout training to fit the 1-to-1 network. Data for each training batch are the visual inputs and male behaviors involving the bilateral genetic silencing of a particular LC neuron type ('LC neuron inactivation', red Xs). For each batch, we silenced (or "knocked out") the model LC unit (i.e., set its value to 0, red Xs) that corresponds to the silenced LC neuron type for that batch. Thus, we perturbed the model in a way similar to how the male fly was perturbed. **c**. We recorded female and male behavior during natural courtship; joint positions were extracted using pose tracking [36], and male song was extracted using audio segmentation [37] (see Methods). **d**. Example sessions in which LC-silenced male flies (using a set of sparse, specific genetic driver lines from [18]) produced noticeable changes in behavior relative to control flies. Control flies began each session far apart (i.e., at larger male-female distances) and eventually decreased their distance (top row). In contrast, male flies with LC9 neurons silenced maintained a large distance from the female (middle row), while male flies with LC15 neurons silenced strongly pursued the female from the beginning of the session (bottom row, showing small distances throughout the session). (continued on next page...)

Figure 1: (...continued from previous page) **e**. Changes in male-to-female distance following silencing of each LC type in males (top row, median distance for first 5 minutes of session) and changes in the proportion of song that was sine versus pulse (bottom row, taken over entire session)—sine is typically sung at close distances to the female, whereas pulse is produced at larger distances and faster male speeds [15]. Each dot denotes one courtship session (one male-female pair). Short lines denote means across sessions for each LC type; horizontal dashed line denotes mean of control sessions. Asterisks denote signficant deviation from control; $p < 0.05$, permutation test. **f**. The 1-to-1 network takes as input a sequence of 10 images (left inset, grayscale 64 x 256 pixel images—color images are shown for clarity) corresponding to the 10 most recent time frames ($\sim$300 ms) of the male's visual experience (further increasing the number of time frames did not improve prediction). Each image is a reconstruction of what the male fly observed based on the male and female joint positions of that time frame (e.g., **c**). Based on visual inputs alone, the 1-to-1 network is able to predict the male fly's behavior, including forward velocity (right, top row) and lateral velocity (right, bottom row). Reported R^2's are computed from heldout control behavior and not from these example traces. **g** and **h**. The 1-to-1 network also accurately predicts changes in behavior across flies with different silenced LC neuron types. Correlations ρ were significant ($p < 0.002$, permutation test).

lowing us to identify a one-to-one mapping between model LC units and LC neuron types. We collected training data to fit the model by genetically inactivating 22 individual LC types one at a time in male flies while recording the fly's movements and song production during courtship (see Methods). We then devised a fitting procedure called 'knockout training', which involves training the model using the entire dataset of perturbed and unperturbed behavior. Critically, when training the model on data from a fly with a particular silenced LC type, we set to 0 (i.e., knock out) the activity of the corresponding model LC unit (Fig. 1**b**). Thus, the resulting model is trained to capture the behavioral repertoire of each genetically altered fly when the corresponding model LC unit is silenced, thereby aligning the model LC units to the real LC neurons. We refer to the resulting model as the 1-to-1 network.

Before fitting the model with courtship data (Fig. 1**c**), we quantified the extent to which silencing (bilaterally) each LC neuron type changes behavior of the male fly. Only one LC type, LC10a, has so far been implicated in male courtship behavior [26, 16]; however, we found that the silencing of other LC types also affected courtship behavior. For example, silencing LC9 neurons resulted in failures to initiate chasing, as measured by large male-female distances over time (Fig. 1**d**, middle panel, and Fig. 1**e**, top panel). On the other hand, silencing LC15 neurons resulted in stronger and more persistent chasing, as measured by small male-female distances over time (Fig. 1**d**, bottom panel, and Fig. 1**e**, top panel). Silencing different LC neuron types also led to different proportions of song type (sine or pulse song, Fig. 1**e**, bottom row). Overall, we found that almost every LC neuron type, when silenced, resulted in some change in behavior (Fig. 1**e** and Ext. Data Fig. 1); these changes were small relative to blind flies (Fig. 1**e**, 'blind'). This suggests no single LC type is the sole contributor to the male's courtship behavior.

Using this perturbation data, we performed knockout training to fit the parameters of the 1-to-1 network. The model inputs were pixel images that reflected the visual input to the male fly (i.e., a fictive female fly changing her size, position, and rotation) (Fig. 1**f**, left panel; see Methods); the model outputs comprised the male movements (forward, lateral, and angular velocity) and song production, which included sine song and two forms of pulse song (Pfast and Pslow as identified by [37]). The 1-to-1 network reliably predicted these behavioral variables in held-out data (Fig. 1**f**, right panels, and Ext. Data Fig. 2**a**). Importantly, the 1-to-1 network also predicted the differences in behavior observed across silenced LC types, such as changes in the male's forward velocity (Fig. 1**g**), his sine song production (Fig. 1**h**), and other behavioral outputs (Ext. Data Fig 2**b**). We further confirmed that the 1-to-1 network predicted the behavior of males with individual LC types silenced better than a model with the same architecture but without knockout training (Ext. Data Fig. 2**c**), indicating that knockout training effectively captured the differences in behavior across silenced LC types.

Comparing real and model neural activity

Our next challenge was to determine if the knockout training procedure, which relied solely on behavioral data, enabled the model LC units in the 1-to-1 network to mirror the response properties of real LC neurons. To address this question, we compared model LC activity with activity of real LC neurons by imaging calcium dynamics recorded in head-fixed, behaving male flies expressing GCaMP6f (Fig. 2**a**, see Methods). We targeted 5 different

LC neuron types (Fig. 2**b**, LC6, LC11, LC12, LC15, and LC17); we chose these LC types because silencing each one led to noticeable changes in behavior (Fig. 1**e** and Ext. Data Fig. 1) and their corresponding model LC units strongly responded to visual input of a moving fictive female (Ext. Data Fig 3). Since the 1-to-1 network can take any visual stimulus as input (i.e., an image-computable model), we presented both synthetic and naturalistic stimulus sequences. We first presented a moving or looming spot (Fig. 2**c**-**e**, see Methods) commonly used to characterize LC responses in previous studies [38, 39]. For example, LC11 is known to be tuned to the speed of a small moving object [39, 40], also observed in our recorded responses (Fig. 2**c**, top panel). We fed these stimulus sequences into the 1-to-1 network and took the responses of the corresponding model LC units as the model's predictions (Fig. 2**a**, black box). Despite the fact that the 1-to-1 network had never been trained on neural data, we found that the model LC responses (Fig. 2**c**-**e**, bottom row) qualitatively matched their corresponding real LC responses (Fig. 2**c**-**e**, top row). We found these matches surprising as these synthetic stimulus properties (moving spots) represent a small portion of the full repertoire of female motion during natural courtship on which the model was trained.

Given these matches, we then tested the 1-to-1 network's predictions on more naturalistic stimulus sequences (i.e., a fictive female varying her position, size, and rotation; see Supplementary Video 1)—LC responses to such naturalistic stimuli have rarely been collected before. We found that the recorded LC neurons responded to many of these naturalistic stimulus sequences and that each LC type produced unique responses (Fig. 2**f**, color traces, and Ext. Data Fig. 3). Despite the fact that the tethered male fly from which we recorded real LC responses was not in a courtship state, which can modulate LC responses [16], we found good matches between real LC responses and their corresponding model LC responses (Fig. 2**f**, black traces vs color traces, and Ext. Data Fig. 3). Quantitatively, the 1-to-1 network reliably predicted real LC responses across stimuli and LC neuron types with an average noise-corrected R^2 of $\sim$0.25 (Fig. 2**g**). The 1-to-1 network better predicted real responses versus other networks with the same architecture but different training schemes, including dropout training, training without the knockout procedure, and no training at all (Fig. 2**h**). When we gave the 1-to-1 network access to neural data (e.g., to train a linear mapping between all 22 model LC units and one real LC neuron type), we found that held-out prediction improved to a noise-corrected R^2 of $\sim$0.6 (Ext. Data Fig. 4). To our knowledge, our model is the first highly-predictive, image-computable encoding model of LC neurons.

Encoding of visual stimulus features by the model LC population

Given that the responses of model LC units can accurately reproduce those of real LC neurons, we next investigated how the population activity of 22 model LC units encodes visual stimuli. Previous studies found that LC neurons (mostly recorded in non-behaving flies) sparsely encode simple visual stimuli, such as moving dots, bars, and gratings [38, 39]. Visual stimuli of a moving spot that loosely approximates the statistics of the female during courtship have only been tested on LC10 [16, 26], a subset whose neurons are male-specific [42]; it is not clear if the other LC neuron types would respond sparsely, if at all, to this type of naturalistic stimulus. Here, we presented visual stimuli of a fictive female with naturalistic movement to test if the model LC responses are indeed locally sparse [43], a defining property of a "labeled line" code. Instead of sparse responses, we found that the majority of model LC units responded to the naturalistic stimuli (Fig. 3**a**). Despite these large response variations, almost no model LC units linearly encoded any single visual parameter of female size, position, or rotation (Fig. 3**b**, low R^2's). On the other hand, a linear mapping of all model LC units could encode each visual parameter (Fig. 3**b**, 'all', high R^2). This suggests that rather than one parameter being entirely encoded by a single LC (consistent with a "labeled line" encoding scheme), the model relies on multiple LC neuron types to encode a parameter (consistent with a distributed encoding scheme). That the real LC neurons we recorded all varied their activity to the same stimulus sequence (Ext. Data Fig. 3, all 5 LC types responded to 'vary female position' and 'natural sequence') is consistent with this conclusion.

We further investigated the 'tuning' of each model LC unit by systematically varying all three primary parameters of the female visual input and passing these input sequences into the model—we assembled the responses for each model LC unit into a 3-dimensional 'tuning curve' (Fig. 3**c** and Ext. Data Fig. 5). While some model LC units were largely driven by a single parameter, such as model LC4 tuned to female position (Fig. 3**d**, top row) and model LC17 tuned to female size (Fig. 3**d**, middle row), other model LC units were tuned to interactions between two or more parameters (Fig. 3**d**, bottom row, model LC18 encodes position for large female sizes but

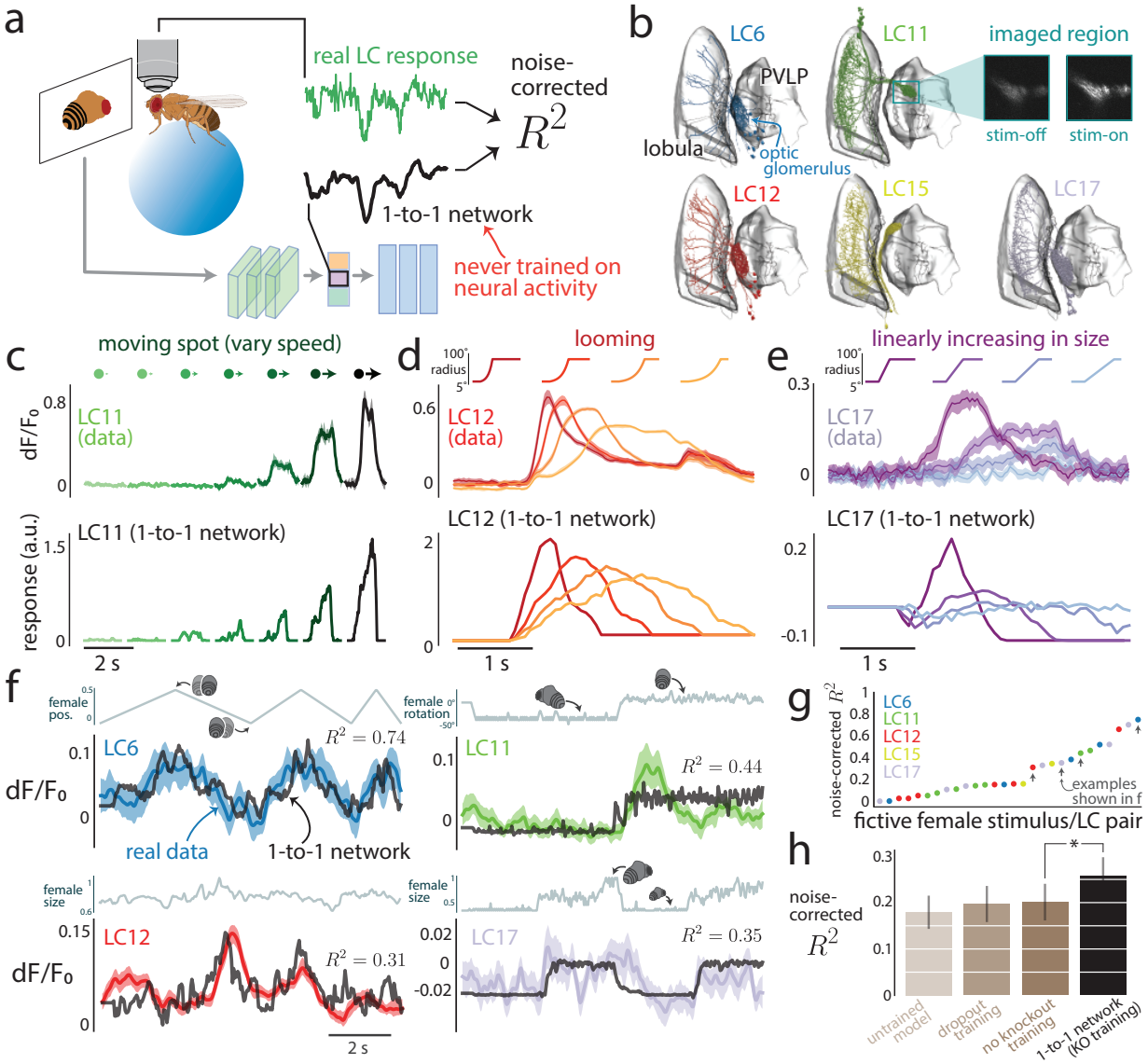


Figure 2: Model LC responses from the 1-to-1 network match real LC neural responses. a. We recorded LC responses using calcium imaging while a head-fixed male fly (walking on a moveable, air-supported ball) viewed dynamic stimulus sequences of a moving spot or a fictive female fly on a projection screen (see Methods). We feed a visual stimulus sequence as input into the 1-to-1 network and asked if the predicted responses (black trace) for a given model LC unit matched the real response of the same LC neuron (green trace) by computing the noise-corrected R^2 between the two over time (see Methods). A noise-corrected R^2 of 1 indicates that a model perfectly predicts repeat-averaged responses after taking into account the amount of variability across repeats. The 1-to-1 network never had access to any of the real LC responses and was only trained on behavioral data (Fig. 1**b**). **b**. We targeted LC6, LC11, LC12, LC15, and LC17, each of which has neurons whose dendrites span the lobula and converge onto the same optic glomerulus (images taken from Janelia hemibrain software [41]). We image the region around the glomerulus (top right) and take as the response the summed calcium dynamics across the glomerulus ('stim-off'/'stim-on': activity before/during visual stimulus was presented). **c**-**e**. Real responses for LC11, LC12, and LC17 (top row) to three different visual stimulus sequences in which a spot moved to the right at different speeds (**c**), loomed at different speeds (**d**), and linearly increased in radius at different speeds (**e**). Model LC responses to these stimulus sequences (bottom row) largely matched the real LC responses. Because the model was only trained on images of a fictive female, for model input we replaced the spot with a fictive female facing away from the male with the same size and location of the spot for each frame (see Methods).
(continued on next page...)

Figure 2: (...continued from previous page) **f**. We also presented various stimulus sequences of a fictive female fly (top traces denote the values of a visual parameter such as female size or position) and collected both real LC responses (color traces) and model LC responses from the 1-to-1 network (black traces)—each example here has real and model responses from a different LC neuron. Shaded regions denote 90% bootstrapped confidence intervals of the mean across repeats. **g**. Prediction performance (noise-corrected R^2 as defined in **a**) of the 1-to-1 network for all stimulus/LC pairs. Each dot represents the prediction of responses of one of 5 possible LC types to one of 9 possible fictive female stimulus sequences (i.e., 45 stimulus/LC pairs). Only LC responses with high SNR across repeats were considered (28 out of the 45 possible stimulus/LC pairs; see Methods). Arrows denote example stimulus sequences in **f**. **g**. Average noise-corrected R^2 across all stimulus sequences and LC types for different models (see Methods). The average of the 1-to-1 network (i.e., the average across dots in **g**) was significantly higher than each average of the other models (asterisk; $p < 0.01$, paired permutation test). Error bars indicate 90% bootstrapped confidence intervals of the mean across stimulus sequence/LC pairs.

not for small female sizes). To quantify these interactions, we decomposed the response variance [44] of each model LC unit into four components: the variance solely due to either female size, position, or rotation and the remaining response variance (corresponding to interactions among any of the visual parameters, see Methods) (Fig. 3**e**). Most model LC units encoded changes in female position (Fig. 3**e**, orange bars), half encoded female size (Fig. 3**e**, blue bars), and female rotation was weakly encoded (Fig. 3**e**, green bars are small for all model LC units). However, almost all model LC units encoded some nonlinear interaction among the three visual parameters (Fig. 3**e**, black bars, $\sim$20% of the response variance for each model LC unit). The presence of such interactions matched our finding that all three visual parameters were needed to predict the male's behavior (Ext. Data Fig. 6). These interactions would not have been discovered had we only varied one visual parameter at a time—motivating the use of more complex stimuli to probe LC function.

The analyses presented so far ignore changes to the visual input over time, but LC neurons do respond to dynamically-changing stimuli [38, 39]. Thus, we considered dynamic stimulus sequences in which the fictive female exhibits variation in one visual parameter while the other two parameters remain fixed (Fig. 3**f**, dashed lines, see Supplementary Video 2). For example, we varied the female's size over time with different speeds (Fig. 3**f**, top row, dashed lines). We found that some model LC units perfectly encoded female size (Fig. 3**f**, top left panel, LPLC1), some model LC units encoded a time-delayed version of size (Fig. 3**f**, top row, middle panel, LC9), while other model LC units encoded the speed at which female size changed (Fig. 3**f**, top row, right plot, LC13). Similar relationships were present for other stimulus sequences and model LC units (Fig. 3**f**, bottom two rows). Across the three stimulus sequences, the model LC units encoded many different temporal features of the visual input (Fig. 3**g**); we note that the model was predictive of real LC responses for these same stimulus sequences (Fig. 2**f** and **g**). The model's predictions were consistent with findings from previous studies, such as LC11 encoding the position of a small moving spot [30, 31] (Fig. 3**g**, LC11 has high R^2 for 'position' in 'vary female position') and LPLC2 encoding loom [27] (Fig. 3**g**, LPLC2 has high R^2 for 'size' in 'vary female size'). Indeed, model units LPLC1, LC18, LPLC2, LC17, and LC12 all strongly encoded female size (Fig. 3**g**, topmost row), matching recent findings that these LC neurons respond to looming and moving objects of various sizes [38, 39]. That the model's predictions were consistent with these relationships, typically characterized under specific contexts, was unexpected, given that the model only had access to the stimuli and behavior during courtship. This suggests that LC neurons have similar tuning across different behavioral contexts, and it follows that this tuning is reused for different behaviors.

Taken together, we found that almost all of the model LC units encode some aspect of female size, position, rotation, or motion during courtship. In addition, many model LC units encoded information about the same parameter (the average signal correlation across model LC units during natural stimulus sequences was $R^2 = 0.22$). We conclude that the model LC units form a distributed, combinatorial code for visual stimuli: Each visual stimulus feature is encoded by multiple LCs (i.e., Fig. 3**g**, each row has multiple red squares) while each LC encodes multiple visual stimulus features (i.e., Fig. 3**g**, each column has multiple red squares). This result suggests that LC neural responses may be read out as a population to drive male courtship behavior—we therefore next asked whether a majority of the model LC units in the population were necessary and sufficient to perform different actions during courtship.

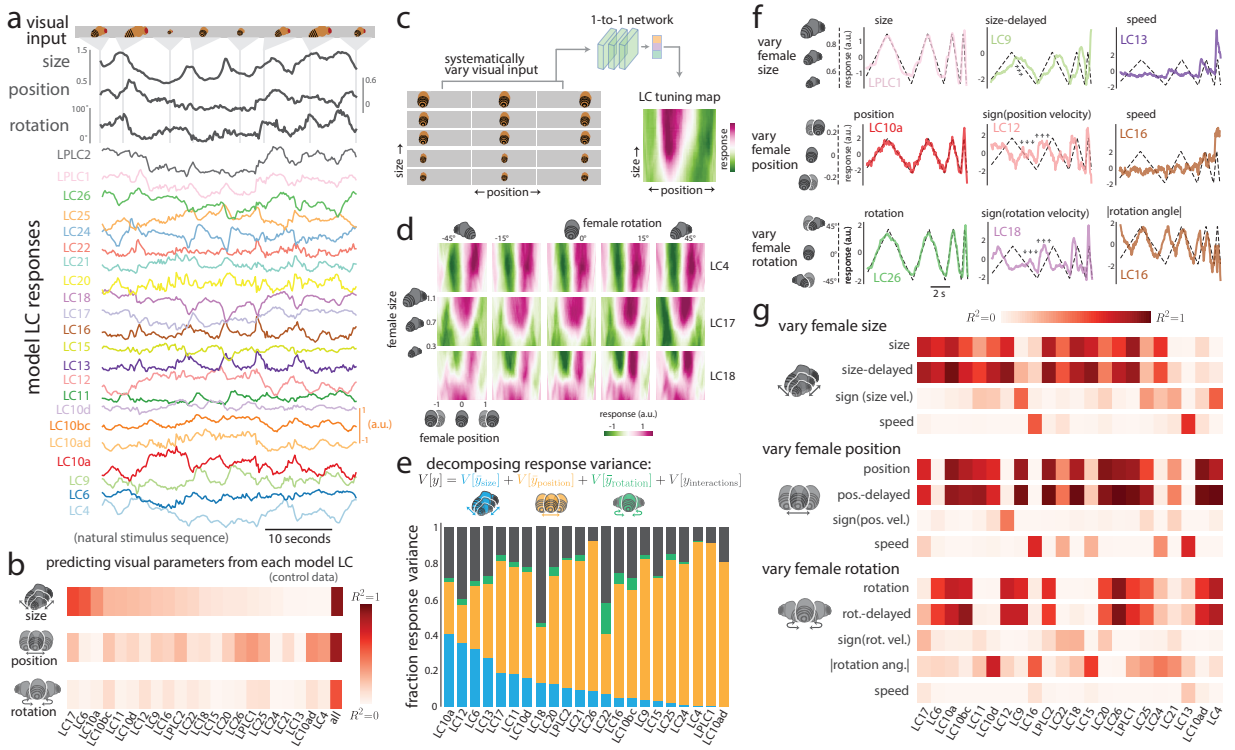


Figure 3: Model LC units combinatorially encode the features of female motion during courtship. a. Model LC responses to a natural stimulus sequence in which a fictive female changes her size, position, and rotation. **b**. Cross-validated R^2 between each primary visual parameter and model LC responses for stimulus sequences observed during natural courtship (assessed on heldout stimuli from control flies). Columns are sorted based on female size (first row). The end column of each row ('all') is the cross-validated R^2 between a linear combination (identified via ridge regression) between all model LC neurons and a single visual parameter. Because female position and rotation are circular variables, we converted each variable x to a 2-d vector $[\cos(x), \sin(x)]$ and took the average R^2 across both variables. **c**. To characterize the stimulus preferences of each model LC unit, we systematically varied the visual parameters of female size, position, and rotation. Because the 1-to-1 network takes in as input a sequence of 10 frames, for simplicity we repeated the same image of the fictive female for all 10 frames for a given set of parameter values (see Methods). For each model LC unit, we computed a heatmap of the LC responses to visualize the unit's tuning preferences. **d**. LC tuning curves as heatmaps for three example model LC units. Female size and position varies within each 2-d heatmap; female rotation varies across columns. **e**. We used variance decomposition (see Methods) to decompose the response variance of each model LC unit (i.e., the total variance of responses to the systematically-varied stimulus sequences in **c** and **d**) into a component solely due to either female size (blue), position (orange), and rotation (green) as well as interactions between these three visual parameters (black). A large fraction of response variance for a given visual parameter indicates that a model LC unit strongly changes its response to variations in this parameter relative to variations in other parameters. Because the 1-to-1 network is deterministic, all response variance can be attributed to variations of the visual parameters (i.e., no component of the variability can be attributed to 'noise' across repeats of the same stimulus). **f**. Example model LC responses to dynamic stimulus sequences in which the fictive female varied either her size, position, or rotation angle over time while the other two parameters remained fixed (dashed lines; dashed y-axis values correspond to plots in the first column only). Different model LC units appear either to directly encode a visual parameter (e.g., LPLC1 encodes 'size') or encode features derived from the parameter (e.g., LC13 encodes 'speed', the speed at which female size changes). Responses for all model LC units are in Ext. Data Fig. 7. **g**. R^2 between model responses and visual parameter features for the three different stimulus sequences in **f**. Columns have the same ordering as that of **b**.

Model LC units form a population code for behavior

During courtship, male flies chase, orient towards, and sing to females. Each of these behaviors could be modulated by separate, individual LC neuron types, as has been proposed [17, 18]. However, our results showing that model LC units combinatorially encode the motion of the female (Fig. 3) suggest instead that a combination of LC types modulate male behaviors toward the female. Here, we directly tested this hypothesis by systematically inactivating model LC units in all different combinations (or alone)—experiments not easily performed in a real fly—and then examined which model LC units were necessary and sufficient to modulate behavior (Fig. 4**a**).

We started by asking which model LC unit, when silenced, maintained the best performance in predicting the behavior of control flies. For predicting male forward velocity, for example, silencing LC10d had a negligible effect on prediction performance (Fig. 4**b**, 'LC10d' vs 'none'), leading us to conclude that LC10d does not contribute meaningfully to male forward velocity. Next, in a greedy and cumulative manner, we repeatedly inactivated the model LC unit that maintained the best performance while keeping all previously-chosen LCs inactivated (no retraining was performed); eventually prediction performance had to decrease (Fig. 4**b**, rightmost dots) because of the bottleneck imposed by the model LC units. Thus, the inactivated model LC units that led to the largest drops in performance were the strongest contributors to behavior (Fig. 4**b**, rightmost dots). We performed this procedure for all 6 behavioral outputs (Fig. 4**c** and Ext. Data Fig. 8). We observed that different model LC units contributed to different behaviors (Fig. 4**c**, different column orderings across rows) and that many model LC units contributed to multiple behavioral outputs (Fig. 4**c**, multiple red squares per column). These results suggest that model LC units do not employ a "labeled line" coding scheme in which each courtship behavior is determined by the responses of one or two LC types (different across behaviors). Rather, the results suggest a distributed population coding scheme in which almost all model LC units contribute in some way to all courtship behaviors examined here.

Because natural behavior is an aggregate over many different behavioral contexts, it may be the case that LC neuron types more resemble a 'labeled line' code for a simpler context. This motivated us to consider the simple, time-varying stimulus sequences for which we recorded responses of a subset of real LC neurons (Fig. 2**f**) and that we used to analyze model LC tuning properties (Fig. 3**f** and **g**). Through systematic inactivation, we again identified the model LC units that were both necessary and sufficient to produce the model's output to these more synthetic stimuli, each representing a different context. For example, we found that when we varied female size only (Fig. 4**d**, top panel, dashed line), inactivating 12 different model LC units (Fig. 4**d**, middle panel, squares with red Xs, identified via cumulative inactivation, see Methods) resulted in no change in forward velocity (Fig. 4**d**, middle panel, green trace overlays black trace, and Ext. Data Fig. 9). This suggests that the *other*, activated model LC units (Fig. 4**d**, middle panel, squares without a red X) were sufficient. We then inactivated these "sufficient" model LC units (keeping all other model LC units activated) and found a large behavioral deficit (Fig. 4**d**, bottom panel, red trace does not overlay black trace), indicating that these inactivated model LC units were also necessary. We show examples for the same type of analysis for the male's angular velocity in response to varying female position (Fig. 4**e**, left column) and for the production of Pslow song in response to varying female rotation (Fig. 4**e**, right column). For these examples, all model LC units except two were necessary and sufficient to fully shape the model's output (Fig. 4**d** and **e**, bottom row, union of red Xs across columns; only LC15 and LC21 did not contribute to behavior).

Across all behavioral outputs, even for these simple stimulus sequences, we found that combinations of multiple model LC units were necessary to drive behavior. This is supported by the observations that multiple LC units contributed to the same behavior (Fig. 4**f**, multiple red squares per row) and that most model LC units contributed to multiple behaviors (Fig. 4**f**, multiple red squares per column); indeed, no single model LC unit was necessary and sufficient for these behaviors (Ext. Data Fig. 10). Nonetheless, the results of the 1-to-1 network are consistent with previous findings [3, 18, 27, 29, 30, 31, 45, 46]. For example, sexually dimorphic LC10a neurons are known to be involved in tracking the female during courtship chasing [16, 26]; this is also true in our model (Fig. 4**c**, LC10a contributes to forward, lateral, and angular velocity). However, our model extends this prior work to uncover a major role for LC10a in driving song choice (Fig. 4**c**, LC10a contributes to Pfast song but not sine song). Overall, our results support the notion that the model LC units form a distributed, population code.

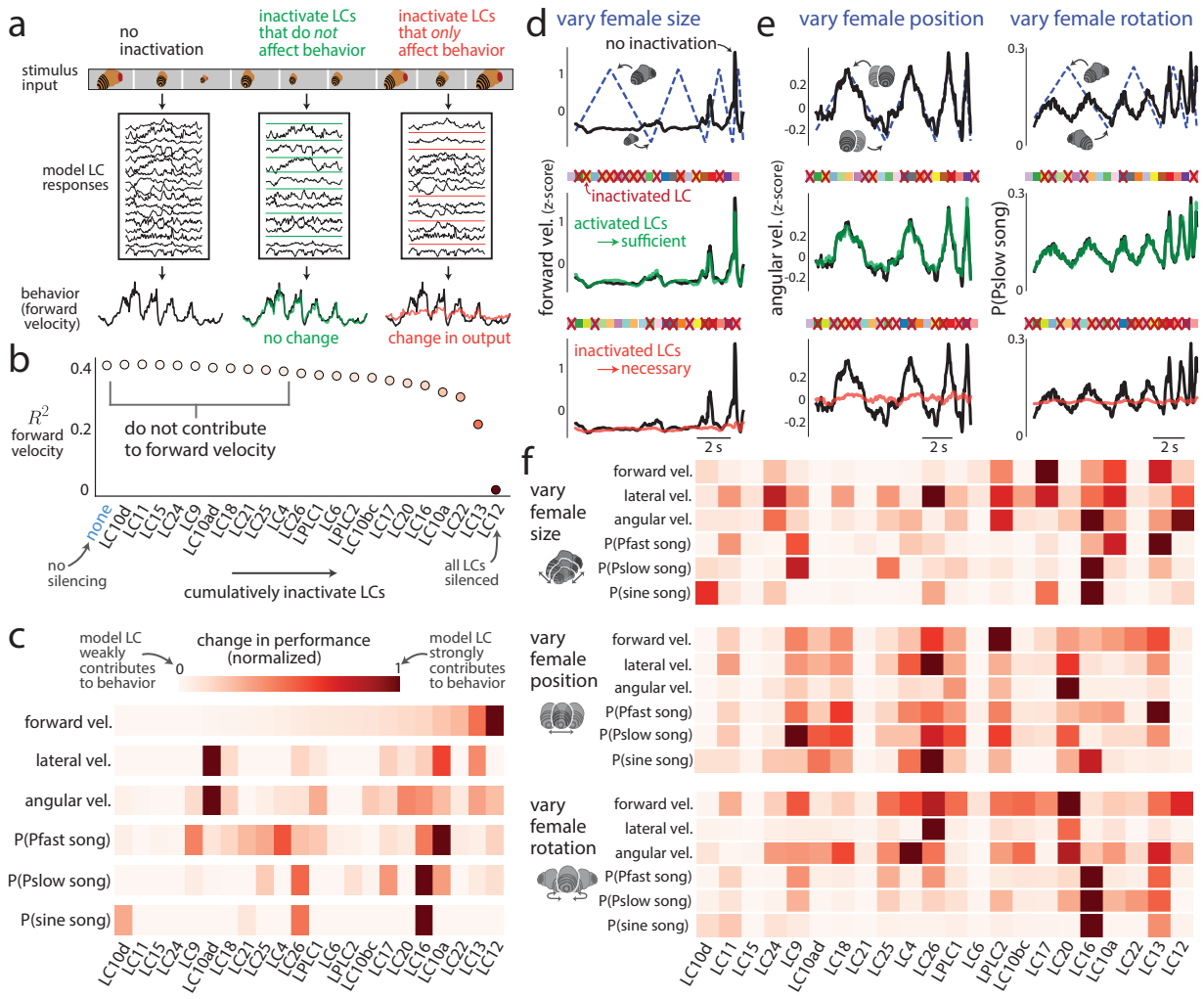


Figure 4: Model LC units form a combinatorial code for behavior. a. Illustration of our approach to investigate which model LC units contribute to which behavioral outputs for any stimulus sequence. We can assess if a group of model LC units are sufficient and necessary for behavior if we inactivate all model LC units not in that group (middle panel, 'sufficient') or inactivate only that group of model LC units (right panel, 'necessary'). **b**. We identify which model LC units contribute to forward velocity via cumulative inactivation for stimulus sequences observed during natural courtship (see Methods). At each iteration, we inactivate the model LC unit that, once inactivated, maintains the best prediction of forward velocity (assessed by R^2 between predicted and real, held-out behavior of control flies) while still inactivating all previously-inactivated model LC units (i.e., cumulative inactivation in a greedy manner). The inactivated model LC units that lead to the largest decreases in R^2 (e.g., LC12 and LC13) contribute the most to forward velocity. **c**. Results for cumulative inactivation for all 6 behavioral outputs; forward velocity (top row) is the same as in **b**. Normalized change in performance is the difference in R^2 between no silencing ('none') and silencing K model LC units, normalized by the R^2 of no silencing. A normalized change in performance close to 1 indicates that the model LC unit strongly contributes to the behavior. Columns of each row are ordered based on the ordering of forward velocity (top row). **d**. We considered the model's predicted behavior for simple stimulus sequences (e.g., dashed line in top panel: varying the female's size over time while her position and rotation remained fixed). Inactivating a chosen group of model LC units (red Xs denote inactivation; model LC units are represented by squares whose colors match the LC colors in Fig. 3**a** and whose ordering matches that in **c**) led to almost no change in the model's output (middle panel, green trace overlays black trace), indicating that the remaining activated model LC units (squares with no red X) are sufficient to produce the model's output. (continued on next page...)

Figure 4: (...continued from previous page) Inactivating these "sufficient" model LC units (bottom panel, red Xs are swapped from squares in middle panel) led to a large behavioral deficit (bottom panel, red traces do not match black traces), indicating that these model LC units are also necessary. Thus, this group of model LC units are both necessary and sufficient for the 1-to-1 network's prediction of forward velocity for a stimulus sequence in which the female's size is varied. **e**. Other example inactivations to assess necessity and sufficiency for two other behaviors (angular velocity and Pslow song) and stimulus sequences (varying female position and rotation). Same format is in **d**. **f**. Results of cumulative inactivation for the three simple stimulus sequences in **d** and **e**. Same format, color legend, and ordering of columns as in **c**.

LC synaptic connectivity reveals a visual population code in *Drosophila*

We aggregated results both from how the model LC neurons encode visual input (Fig. 3) and contribute to behavior (Fig. 4) and outline these relationships (Fig. 5**a**). The picture is complicated: Many model LC units encode multiple visual features of the female (Fig. 5**a**, left connections) and contribute to multiple behavioral outputs (Fig. 5**a**, right connections). Given our model's ability to predict both the changes to behavior during genetic silencing (Fig. 1**g** and **h**) and responses of real LC neurons (Fig. 2), we propose that the LC neurons in the fly visual system also form this complex population code. Such a code would need a rich scaffolding of downstream connections in order for downstream areas to read out and integrate information from multiple LC types. To test this, we analyzed a connectomic resource called the hemibrain [23] in which 39 LC neuron types in total have been identified. We found that LC neurons formed direct connections with neurons innervating a variety of downstream areas; for example, a single LC4 neuron connected with neurons innervating 9 different neuropils (Fig. 5**b**). We computed the synaptic connectivity matrix for LC neuron types and their downstream partners (Fig. 5**c** and Ext. Data Fig. 11**a**), where each entry was either 1 (denoted by a blue square in Fig. 5**c**) if at least 3 synapses were identified between the neurons of an LC type and downstream neurons of a given type, else 0. We found evidence in support of a population code: Similar to the example LC4 neuron (Fig. 5**b**), neurons of each LC type innervated neurons from multiple downstream areas, suggesting each LC type contributes to a variety of behaviors (Fig. 5**c**, multiple blue squares per row). In addition, many post-synaptic partners received direct input from multiple LC neuron types (Fig. 5**c**, multiple blue squares per column)—in fact, 54% of downstream neurons received input from 2 or more LC types; 34% from 3 or more LC types (Ext. Data Fig. 11**b**), suggesting that a downstream neuron integrates information from multiple LC types. An additional observation not yet reported is that many LCs directly connect with other LCs in the lobula (Fig. 5**c**, 'LC + LPLC' have many blue squares in its columns); such recurrence muddles the idea that each LC is an independent feature detector (although these lateral connections may implement a winner-takes-all mechanism such as divisive normalization [47, 48] or may sparsify the code [49]).

Discussion

Why would the fly visual system, with its highly-structured set of glomeruli, use a distributed population code of LC neurons? Such a code can bring computational advantages, such as noise robustness [50, 51, 52, 53], flexibility to accommodate multiple tasks [20, 54, 55, 56, 57], and coding efficiency [5, 51, 58, 59]. It may be the case that in an early point in evolutionary history, the optic glomeruli mostly formed a 'labeled line' code to allow for quick reflexes (e.g., escape from a looming predator). However, as the visual and motor systems adapted to changing environmental presures, these systems may have updated their "software" (reusing LC neurons for different behaviors) while keeping the same "hardware" (a fixed number of glomeruli) to allow for a richer repertoire of complex behaviors (e.g., courtship). A key to uncovering this population code was our modeling framework with knockout training, which is general and broadly-applicable. We used this framework to identify a one-to-one mapping between internal units of a deep neural network and real neurons—further revealing the relationships between stimulus, neuron, and behavior.

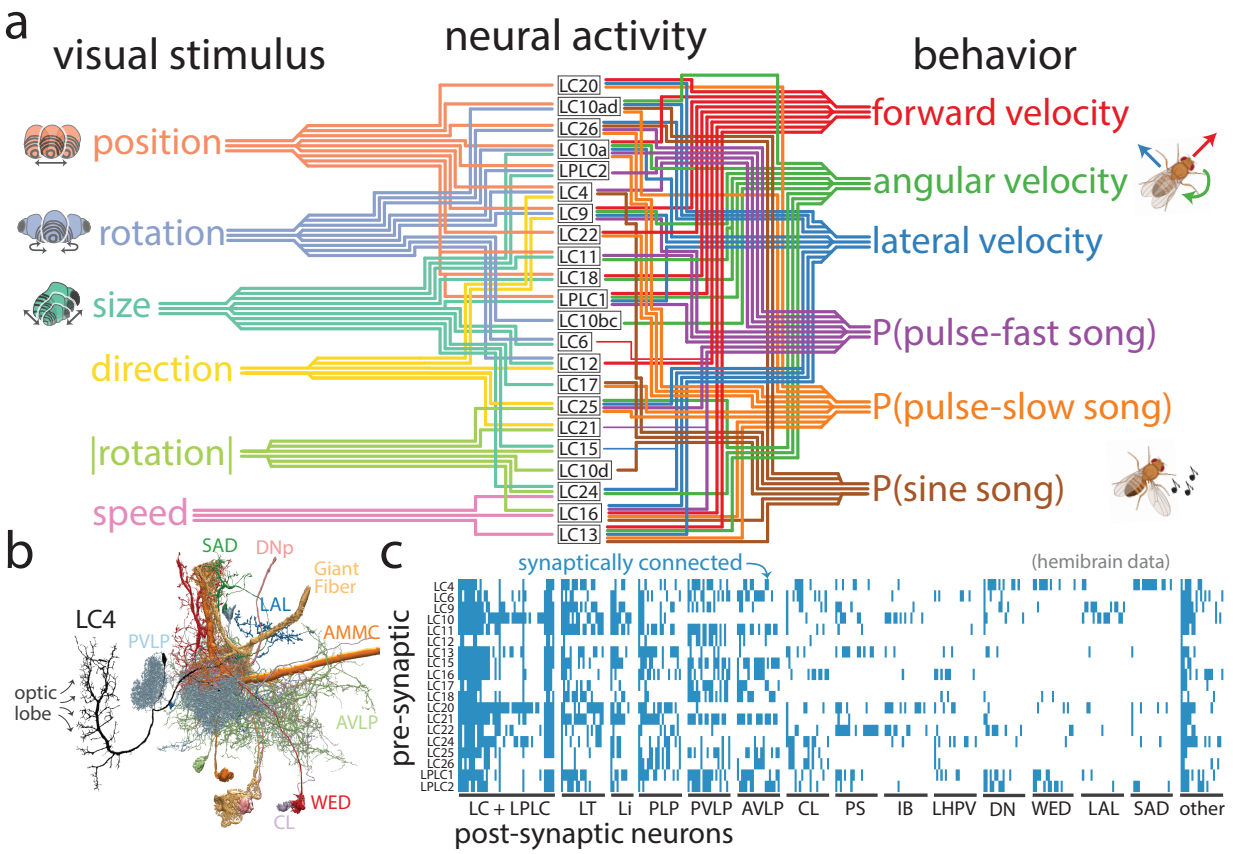


Figure 5: The role LC neurons play in the sensorimotor transformation of the male fly during courtship. a. Summary of our findings. Each line denotes a relationship between a model LC unit and either a visual feature ($R^2 > 0.30$ in Fig. 3**b** or **g**) or a behavioral variable (a change in performance greater than 30% in Fig. 4**c** or **f**; weaker relationships with a change greater than 10% but less than 30% denoted with thin lines). For illustrative purposes, the 9 strongest connections are displayed for each feature or behavioral variable. A lack of connection does not rule out a relationship, as relationships may exist in other contexts. Even at these high criteria (i.e., cutoffs at 0.3), many model LC units encode more than one visual feature and contribute to more than one behavioral variable. **b**. A single LC4 neuron (black) has dendrites in the lobula of the optic lobe and synaptically connects to downstream neurons in 9 different neuropils. Image taken from Janelia hemibrain software [41]. **c**. Synaptic connectivity matrix for pre-synaptic LC or LPLC neurons (rows) and post-synaptic, downstream neurons (columns). Each blue square indicates at least 3 direct synaptic connections were identified between neurons of a given LC type (e.g., LC4) and post-synaptic neurons of a given type (e.g., PVLP). Columns are grouped into neuron types/brain areas based on Janelia's hemibrain connectome dataset [23, 41]; each group was limited to 20 neuron types—see Ext. Data Fig. 11 for the synaptic connectivity matrix of all downstream neuron types.

Methods

Flies

For all experiments, we used 4-7 day old virgin flies harvested from density-controlled bottles seeded with 8 males and 8 females. Fly bottles were kept at 25°C and 60% relative humidity. Virgined flies were then housed individually across all experiments. Female virgined flies were individually housed and kept in behavioral incubators under a 12-12 hr light-dark cycling. Prior to recording with a female, males were painted with a small spot of opaque ultraviolet-cured glue (Norland Optical and Electronic Adhesives) on the dorsal mesothorax to facilitate identification during tracking. Data from CS-blind flies were collected in a previous study [15]. UAS-TNT-C was obtained from the Bloomington stock center. All LC lines and BDP control were generously provided by Michael Reiser.

Courtship experiments

Behavioral chambers were constructed as previously described [15, 60]. Each recording chamber had a floor lined with white plastic mesh and equipped with 16 microphones. Video was recorded from above the chamber at a 60 Hz frame rate; features for behavioral tracking were extracted from the video and downsampled to 30 Hz for later analysis. Flies were introduced gently into the chamber using an aspirator. Recordings were timed to be within 150 minutes of the behavioral incubator lights switching on to catch the morning activity peak. Recordings were stopped either after 30 minutes or after copulation, whichever came sooner. All flies were used; we did not use any criteria (e.g., if males sang during the first five minutes of the experiment or not) to drop fly sessions from analyses. In total, behavior was recorded and analyzed from 219 male flies; the number of flies per condition were as follows:

LC type	LC4	LC6	LC9	LC10a	LC10ad	LC10bc	LC10d	LC11	LC12	LC13	LC15	LC16	LC17
number of flies	9	14	12	6	8	9	11	7	6	8	11	12	5

LC type	LC18	LC20	LC21	LC22	LC24	LC25	LC26	LPLC1	LPLC2	control	CSTul-blind	total
number of flies	5	9	5	14	11	11	13	11	11	11	32	219

Joint positions for the male and female for every frame were tracked with a deep neural network trained for multi-animal pose tracking called SLEAP [36]. We used the default values for the parameters. We estimated the presence of sine, Pfast, and Pslow song for every frame using a song segmenter on the audio signals recorded from the chamber's microphones according to a previous study [61, 37].

From the tracked joint positions and song recordings, we extracted the following 6 behavioral variables of the male fly that represented his moment-to-moment behavior. 1) Forward velocity was the difference between the male's current position minus his position one frame in his past; this difference in position was projected onto his heading direction (i.e., the vector from the male's body position to his head position). 2) Lateral velocity was the same difference in position as computed for forward velocity except this difference was projected onto the direction orthogonal to his heading direction; rightward movements were taken as positive. 3) Angular velocity was the angle between the male's current heading direction and the male's heading direction one frame in the past; rightward turns were taken as positive, and angles were reported in visual degrees. 4) Probability of sine song was computed as a binary variable for each frame, where a value of 1 was reported if sine song was present during that frame, else 0 was reported. 5) Probability of Pfast and 6) probability of Pslow were computed in the same manner as that for the probability of sine song. These six behavioral output variables described the male's movement (forward, lateral, and angular velocity) as well as his song production (probability of sine, Pfast, and Pslow song).

Visual input reconstruction

To best mimic how a male fly transforms his retina's visual input into behavior, we desired an image-computable model (i.e., one that takes as input an image rather than abstract variables determined by the experimenter, such as female size or male-to-female distance). We approximately reconstructed the male's visual input based on the joint positions of both the male and female fly during courtship, as described in the following process. For each frame, we created a 64-pixel $\times$ 256-pixel grayscale image with a white background. Given the female rotation, size, and

location (see below), we placed an image patch of a grayscale fictive female (composed of ellipses that represented the head, eyes, thorax, and tail of the female; no wings were included) occluding the white background. Because male flies perceive roughly 160 visual degrees on either side [62], we removed from the image the 40 visual degrees directly behind the male, leading to images with 64×228 pixels. Example input images are shown in Figure 1**f**, where the approximated female flies were colored and on gray background for illustrative purposes. Example videos of input image sequences are present in Supplementary Video 1 and Supplementary Video 2.

We computed the female's rotation, size, and location in the following way. For female rotation, we first rotated the coordinate axes of the absolute positions of both the male and female to align with the male's heading direction (i.e., a change of reference so the male is facing the same direction for every frame). We defined the rotation angle as the angle between the direction of the male head to female body and the direction of the female's heading in this rotated space. A rotation angle of 0° indicates the female is facing away from the male, $\pm 180°$ indicates the female is facing towards the male, and -90°/+90° indicates the female is facing to the left/right of the male. We preprocessed a set of 360 image patches (25×25 pixels) that depicted the female rotated by a different visual degree. Given the computed rotation angle, we accessed the image patch corresponding to that rotation angle. For female size, we treated the female fly as a sphere (whose diameter matched the average size of a female fly, $\sim$4 mm) and computed as size the visual angle between the two vectors of the male's head position to the two outermost points on the sphere projected onto the horizontal viewing axis of the male; this angle was normalized so that a size of 1 corresponded to 180 visual degrees. This size determined the width and height of the selected image patch to be placed into the 64×228-pixel image. Here, size indicates the size of the image patch, not the actual size of the fictive female (which may vary because a female facing away is smaller than a female facing to the left or right). For female position, we computed the visual angle between the male's heading direction and the direction between the male's head and the female's body position. We normalized this angle such that a position of 0 is directly in front of the male, a position of either -1 or 1 is directly behind the male fly, and a position of -0.5/+0.5 is immediately to the left/right. We then used this position to place the image patch (with its chosen rotation and size) at a certain pixel location along the horizontal axis of the image. Because the male and female flies did not have room to fly in the experimental chamber, we assumed that only the female's lateral position (and not vertical position) could change.

Description of 1-to-1 network

We designed our 1-to-1 network to predict the male fly's behavior (i.e., movement and song production) from his visual input. Although the male can use other sensory modalities such as olfaction or proprioception to detect the female, we chose to focus only on visual inputs because 1) the male relies primarily on his visual feedback for courtship [63, 15], and 2) we wanted the model to have a representation solely based on vision to match the representations of visual LC neurons. The model took as input the images of the 10 most previous time frames (corresponding to $\sim$300 ms)—longer input sequences did not lead to an improvement in predicting behavior. Each grayscale image was 64×228 pixels (with values between 0 and 255) depicting a fictive female fly on a white background (see "Visual input reconstruction" above). Before being fed into the network, the input was first re-centered by subtracting 255 from each pixel intensity to ensure the background pixels had values of 0. The model output 6 behavioral variables of the male fly: forward velocity, lateral velocity, angular velocity, probability of sine song, probability of Pfast, and probability of Pslow (see "Courtship experiments" above).

Our model was a deep neural network (DNN) that comprised three components: a vision network, an LC bottleneck, and a decision network (Fig. 1**a**). The first 3 layers of the vision network were convolutional, and each layer had 32 filters with kernel size (3×3) and a downsampling stride of 2. Each convolution was followed by a batchnorm operation [64] and a relu activation function. The fourth and final dense layer of the vision network linearly mapped the output activity of the third convolutional layer to 16 embedding variables. The vision network processed each of the 10 input images separately; in other words, the vision network's weights were shared across time frames. For simplicity, the vision network's input was the entire image (i.e., the entire visual field); thus, we did not include two retinae as in the fly. We found that incorporating two retinae into the model, while more biologically-plausible, made it more difficult to interpret the tuning of each LC neuron type. For example, for a two-retinae model, it is difficult to determine if differences in tuning for two model units of the same LC type but in different retinae are true differences in real LC types or instead differences due to overfitting between the two

retinal vision networks. The 1-to-1 network avoids this discrepancy through the simplifying assumption that each LC type has a similar tuning across both retinae.

The next component of the DNN model was the LC bottleneck, which received 10 16-dimensional embedding vectors corresponding to the past 10 time frames. The LC bottleneck processed this input with a dense layer of 22 filters followed by batchnorm and relus—allowing each LC to integrate information over time. This was then followed by another dense layer of 22 filters (termed the "LC bottleneck"), corresponding to the 22 LC types we considered in our silencing experiments. This second dense layer was important for knockout training, described below. Each model LC unit represented the summed activity of all neurons of the same LC type. We allowed model LC units to have positive and negative responses to account for the fact that we did not incorporate two retinae into the model; thus, the model could encode, for example, summed LC activity of the left optic lobe as positive responses and summed LC activity of the right optic lobe as negative responses, although this was not enforced. For two perturbations (LC10ad and LC10bc), we only had genetic lines to silence two LC neuron types together. For simplicity, we corresponded each of these with its own model LC unit, which represented the summed activity of all neurons from both types (e.g., LC10a and LC10d for LC10ad). The decision network took as input the 22 LC bottleneck units and comprised 3 dense layers, where each layer had 128 filters followed by batchnorm and relus. The decision network predicted the movement output variables (forward velocity, lateral velocity, and angular velocity) each with a linear mapping and the song production variables (probability of sine, Pfast, and Pslow song) each with a linear mapping followed by a sigmoid activation function. The hyperparameters of number of layers and filters were chosen based on prediction of a heldout validation dataset.

Knockout training

We sought a one-to-one mapping between the model's 22 LC units in its bottleneck and the 22 LC neuron types in our silencing experiments (Fig. 1**a**). To identify this mapping, we devised knockout training. We first describe the high-level training procedure and then give details about the optimization. We initialized the 1-to-1 network by training on control data only. This presumably sculpted the model's LC bottleneck representation to match the representation of the real LC neurons (but not in a one-to-one way). We performed two steps to identify the one-to-one mapping. First, for each silenced LC type, we searched for a linear combination of the model LC bottleneck variables that, if subtracted from the model LC activity (i.e., "knocked out"), would lead to better prediction of male flies with that silenced LC type than if we did no subtraction. We used stochastic gradient descent (with learning rate 1e1 and momentum 0.7) to identify these linear combinations for all LCs, which gave us an initial estimate for the one-to-one mapping. We found that this initialization process was helpful for the following knockout procedure to succeed. For knockout training, we trained this initialized 1-to-1 network on all silenced and control data (Fig. 1**b**). For each silenced LC type, we "knocked out" (i.e., set to 0) the corresponding model LC unit (no model units were silenced for control data). This is similar to dropout training [65] except that hidden units were purposefully—not randomly—chosen. The intuition behind knockout training is that the remaining, non-knocked-out model LC units must encode enough information to predict the silenced behavior; any extra information not helpful for prediction (i.e., information encoded by the knocked-out model LC) will not be encoded into the non-knocked-out LC units (as the back-propagated error would not contain this information). For example, let us assume that LPLC1 solely encodes female size and contributes strongly to forward velocity. To predict LPLC1-silenced behavior (which would not rely on female size), the other model LC units would need only to encode other aspects of the female stimulus (e.g., position or rotation). In fact, any other model LC unit encoding female size would hurt prediction because LPLC1-silenced behavior does not depend on it. Another view of knockout training is that we constrain the model not only to predict behavior but also to predict behavior with certain constraints on what internal representations the model may use. These constraints are set by the perturbations (e.g., genetic silencing) we use in our experiments.

The optimization details are as follows. The model was trained end-to-end using stochastic gradient descent with learning rate 1e-3 and momentum 0.7. Each training batch had 150 samples, where each sample was a sequence of 10 images and 6 output values. Each batch was balanced across LC types and types of song (sine song, pulse song, or no song). This data balancing equalized the amount of training data for each silenced LC type (as different LC types had a different number of flies) and the amount of song observed (as different flies sang different amounts of song). The model treated different flies for the same silenced LC type as the same to capture overall trends of

an "average" fly. The training data was augmented by randomly flipping some input from left to right (flipping each image in the image sequence and correspondingly changing the sign of the lateral and angular velocity), which helped to account for the fact that the flies tended to prefer to walk along the edge of the chamber in either a clockwise or counter-clockwise manner. We z-scored the movement behavioral variables (forward, lateral, and angular velocity) based on the mean and standard deviation of the control data in order to have similarly-sized gradients from each output variable. The loss functions were mean squared error for forward, lateral, and angular velocity and binary cross-entropy for the probabilities of sine, Pfast, and Pslow song. The model instantiation and optimization was coded in Keras [66] on top of Tensorflow [67]; we used the default random initialization parameters to initialize weights. The overall training dataset had 219 flies (including silenced and control) and $\sim$9.3 million training samples (non-augmented) in total. We took the model with the best prediction performance (on heldout validation data) after 200,000 training batches. We evaluated the model's predictions of behavioral outputs for each silenced LC type with 3,000 heldout frames randomly chosen across flies with the same silenced LC.

Two-photon functional imaging

We recorded LC responses of a head-fixed male fly using a custom-built two-photon microscope with a 40x objective and a two-photon laser (Coherent) tuned to 920 nm for imaging of GCaMP6f. A 562 nm dichroic split the emission light into red and green channels, which were then passed through a red 545-604 nm and green 485-555 nm bandpass filter respectively. We recorded the imaging data from the green channel with a single plane at 50 Hz. Before head-fixation, the male's cuticle above the brain was surgically removed, and the brain was perfused with an extracellular saline composition. The male's temperature was controlled at 30°C by flowing saline through a Peltier device and measured via a water bath with a thermistor (BioscienceTools TC2-80-150). We targeted LC neuron types LC6, LC11, LC12, LC15, and LC17 (Fig. 2**b**) for their proximity to the surface (and thus better imaging signal), changes to behavior when silenced (Fig. 1**e** and Ext. Data Fig. 1), and their corresponding model LC responses to stimuli (Fig. 3**b** and **g**). The genotype for a particular target (e.g., LC6) was +/w-; LC6 AD/ UAS GCaMP6f; LC6 DBD/+.

Each awake, head-fixed male fly rested its legs on a freely-moveable, air-supported ball and viewed a translucent projection screen placed in the right visual hemifield (matching our recording location in the right retina). The flat screen was slanted 40 visual degrees from the heading direction of the fly and perpendicular to the axis along the direction between the fly's head and the center of the screen (with a distance of 9 cm between the two). An LED-projector (DLP Lightcrafter LC3000-G2-PRO) with a Semrock FF01-468/SP-25-STR filter projected stimulus sequences onto the back of the screen at a frame rate of 180 fps. A neutral density filter of optical density 1.3 was added to the output of the projector to reduce light intensity. The stimulus sequences (described below) comprised a moving spot and a fictive female that varied her size, position, and rotation.

We recorded a number of sessions for each targeted LC: LC6 (5 flies), LC11 (5 flies), LC12 (6 flies), LC15 (4 flies), and LC17 (5 flies). We computed dF/F_0 at targeted ROIs using a baseline ROI for F_0 that had no discernible response and was far away from targeted ROIs. For each LC and stimulus sequence, we concatenated repeats across flies. To remove effects due to adaptation and differences among flies, we de-trended responses by taking the z-score across time for each repeat; we then scaled and re-centered these z-scores by the standard deviation and mean of the original repeats (i.e., the original and de-noised repeats had the same overall mean and standard deviation). We sought a way to test whether an LC was responsive to a stimulus sequence or not. To do this, we computed a metric akin to a signal-to-noise ratio for each combination of LC type and stimulus sequence in the following way. For a single run, we split the repeats into two separate groups (same number of repeats per group) and computed the repeat-averaged response for each group. We then computed the R^2 between the two repeat-averaged responses by computing the Pearson correlation over time and squaring it. We performed 50 runs with random split groups of repeats to establish a distribution of R^2's. We compared this distribution to a null distribution of R^2's that retained the timecourses of the responses but none of the time-varying relationships among repeats. To compute this null distribution, we sampled 50 runs of split groups (same number of repeats as the actual split groups) from the set of repeats for all stimulus sequences; in addition, the responses for each repeat were randomly flipped in time (i.e., reversed) or in sign, breaking any possible co-variation across time among repeats. For each combination of LC type and stimulus, we computed the sensitivity d' [68] between the actual R^2 distri-

bution and the null R^2 distribution. We designated a threshold $d' > 1$ to indicate that an LC was responsive for a given stimulus sequence (i.e., we had a reliable estimate of the repeat-averaged response). After this procedure, a total of 28 combinations of stimulus sequence and LC type out of a possible 45 combinations remained (Fig. 2**g**).

We considered two types of stimulus sequences: a moving spot and a moving fictive female. The moving spot (black on isoluminant gray background) had three different stimulus sequences (Fig. 3**e**). The first stimulus sequence was a spot with fixed diameter of 20° that moved from the left to right with a velocity chosen from candidate velocities {1, 2, 5, 10, 20, 40, 80} °/s; each sequence lasted 2 seconds. The second stimulus sequence was a spot that loomed from a starting diameter of 80° to a final diameter of 180° according to the formula $\theta(t) = -2\tan^{-1}(-r/v \cdot 1/t)$, where r/v is the radius-to-speed ratio with units in ms and t is the time (in ms) until the object reaches its maximum diameter (i.e., $t = t_{\text{final}}$ - t_{current}) [46]. A larger r/v establishes a slower object loom. We presented different loom speed ratios chosen from candidate $r/v \in \{10, 20, 40, 80\}$ ms. Once a diameter of 180° was reached, the diameter remained constant. The third stimulus sequence was a spot that linearly increased its size from a starting diameter of 10° according to the formula $\theta = 10 + v \cdot t$, where v is the angular velocity (in °/s) and t is the time from stimulus onset (in seconds). The final diameter of the enlarging spot for each velocity (30°, 50°, 90°, or 90°) was determined based on the chosen angular velocity $v \in \{10, 20, 40, 80\}$ °/s. Once a diameter of 90° was reached, the diameter remained constant.

The second type of stimulus sequence was a fictive female varying her size, position, and rotation. The fictive female was generated in the same manner as that for the input of the 1-to-1 network (see Methods section "Visual input reconstruction"). We considered three kinds of fictive female stimulus sequences; we first describe them at a high level and then separately in more detail. The first kind consisted of sequences in which the female varied only one visual parameter (e.g., size) while the other two parameters remained fixed (e.g., position and rotation); we varied this parameter with three different speeds (Fig. 3**f** and Fig. 4**d** and **e**). Second, we generated sequences that optimized a model output variable (e.g., maximizing or minimizing forward velocity). Third, we used a natural image sequence taken from a courtship session. Each stimulus sequence lasted for 10 seconds (300 frames).

In total, we presented 9 different stimulus sequences of the fictive female fly. Details of the fictive female sequences are as follows:

- *vary female position*: The female varied only her lateral position (with a fixed size of 0.8 and a rotation angle of 0° facing away from the male) from left-to-right (75 frames) then right-to-left (75 frames). Positions were linearly sampled in equal intervals between the range of -0.1 and 0.5. This range of positions was biased to the rightside of the visual field to account for the fact that the projection screen was oriented in the male fly's right visual hemifield. After the intial pass of left-to-right and right-to-left (150 frames total), we repeated this same pass two more times with shorter periods (100 frames and 50 frames, respectively), interpolating positions in the same manner as the initial pass.
- *vary female size*: The same generation procedure as for 'vary female position' except that instead of position, we varied female size from 0.4 to 0.9 (sampled in equal intervals) with a fixed position of 0.25 and a rotation angle of 0° facing away from the male. We note that the recorded LCs weakly responded to this stimulus (Ext. Data Fig. 3); we suspect this is because female size needed to be scaled larger when projected onto the screen.
- *vary female rotation*: The same generation procedure as for 'vary female position' except that instead of position, we varied the female rotation angle from -180° to 180° (sampled in equal intervals) with a fixed position of 0.25 and a fixed size of 0.8.
- *optimize for forward velocity*: We optimized a 10-second stimulus sequence in which female size, position, and rotation were chosen to maximize the model output of forward velocity for 5 seconds and then minimize male forward velocity for 5 seconds. In a greedy manner, the next image in the sequence was chosen from candidate images to maximize the objective. We confirmed that this approach did yield large variations in the model's output. To ensure smooth transitions, the candidate images were images "nearby" in parameter space (i.e., if the current size was 0.8, we would only consider candidate images with sizes in the range of 0.75 to 0.85). Images were not allowed to be the same in consecutive frames and had to have a female size greater than 0.3 and a female position between -0.1 and 0.5.

- *optimize for lateral velocity*: The same generation procedure as for 'optimize for forward velocity' except that we optimized for the model output of lateral velocity. In this case, maximizing lateral velocity is akin to asking the model to output the action of moving to the right while facing forward.
- *optimize for angular velocity*: The same generation procedure as for 'optimize for forward velocity' except that we optimized for the model output of angular velocity. In this case, maximizing angular velocity is akin to asking the model to output the action of turning to the right while pivoting.
- *optimize for forward velocity with fixed position*: The same generation procedure as for 'optimize for forward velocity' except that we limited female position p to be within the tight range of $0.225 < p < 0.275$. This ensured that most changes of the female stemmed from changes in either female size or rotation, not position.
- *optimize for lateral velocity with multiple transitions*: The same generation procedure as for 'optimize for lateral velocity' except that we had four optimization periods: maximize for 2.5 seconds, minimize for 2.5 seconds, maximize for 2.5 seconds, and minimize for 2.5 seconds.
- *natural stimulus sequence*: A 10-second stimulus sequence taken from a real courtship session. This sequence was chosen to ensure large variation in the visual parameters and that the female fly was mostly in the right visual field between positions -0.1 and 0.5 (as stimuli were presented to the right visual hemifield of the male fly).

For each recording session, we presented the stimuli in the following way. For the moving spot stimuli, each stimulus sequence was preceded by 400 ms of a blank, isoluminant gray screen. For the fictive female stimuli, a stimulus sequence of the same kind (e.g., 'vary female size') was presented in three consecutive repeats for a total of 30 seconds; this stimulus block was preceded by 400 ms of a blank, isoluminant gray screen. All stimulus sequences (both moving spot and the fictive female) were presented one time each in a random ordering. Another round (with the same ordering) was presented if time allowed; usually, we presented 3 to 4 stimulus rounds before an experiment concluded. This typically provided 9 or more repeats per stimulus sequence per fly.

Predicting real neural responses

To obtain the model predictions for the moving spot stimuli (Fig. 2**c**-**e**, bottom row), we generated a fictive female facing away from the male whose size and position matched that of the moving spot. This was done to prevent any artifacts from presenting a stimulus (e.g., a high-contrast moving spot) on which the model had not been trained (i.e., the model only observed a fictive female). For moving spot with varying speed (Fig. 2**c**), the fictive female translated from left to right (i.e., same as the stimuli presented to the male fly); however, because the 1-to-1 network did not have an adaptation mechanism (i.e., model responses could remain constant and not return to 0), after the first translation we also translated the fictive female from the right to the left for 30 frames. This ensured the female would start and stop at the same position and that the model responses would begin and end at the same value. To obtain model predictions for the fictive female stimuli (Fig. 2**f**-**h**), we input the same stimululus sequences presented to the fly except that we changed the grayscale background to white (to match the training images).

To evaluate the extent to which the 1-to-1 model predicted the repeat-averaged LC responses for each stimulus sequence of the moving fictive female, we sought a prediction performance metric (e.g., R^2) that accounted for the fact that our estimates of the repeat-averaged responses were noisy. Any metric not accounting for this noise would undervalue the true prediction performance (i.e., the prediction performance between a model and a repeat-averaged response with an infinite number of repeats). To measure prediction performance, we chose a noise-corrected R^2 metric recently proposed [69] that precisely accounts for noise across repeats and is unbiased at estimating the ground-truth normalized R^2. A noise-corrected $R^2 = 1$ indicates that our model perfectly predicts the ground-truth repeat-averaged responses up to the amount of noise across repeats.

We computed this noise-corrected R^2 between the 1-to-1 network and real responses for each LC and stimulus sequence (Fig. 2**g**) for which the LC was responsive (i.e., $d' > 1$). Importantly, the 1-to-1 network never had access to any neural data in its training; instead, for a given LC type, we used the model LC unit in its bottleneck

as the 1-to-1 network's predicted response. For comparison, we also used other models to predict the real LC responses (Fig. 2**h**). These models all had the same architecture (e.g., number of layers and filters) but were trained differently. They include an untrained network (i.e., random initial weights), a network trained with dropout in its LC bottleneck layer (i.e., for every training sample, a random model LC unit was knocked out) [65], and a network with no knockout training (i.e., no model LC unit was set to 0 during training). The trained networks had the same optimization procedure, hyperparameters, and the same training data as that for the 1-to-1 network. The fact that the untrained model was predictive (Fig. 2**h**, leftmost bar above 0) indicates that our chosen architecture for the 1-to-1 network, including spatial and temporal convolutions, was conducive for detecting prominent changes of the visual stimulus (e.g., a looming object changing many pixels over space and time) that LC neurons also detect; this motivates the use of more complex stimuli, such as natural stimulus sequences, when comparing model predictions to real LC responses.

Analyzing model LC responses to visual input

To better understand how each model LC unit responds to the visual input, we systematically varied the visual parameters, input these stimulus sequences into the 1-to-1 network, and formed heatmaps out of the model LC responses (Fig. 3**c**). For each input stimulus sequence, each of its 10 images was a repeat of the same image of a fictive female with a given size, lateral position, and rotation angle (i.e., the fictive female did not move over time for each 10-frame input sequence). Across stimulus sequences, we varied female size (50 values linearly interpolated between 0.3 to 1.1), lateral position (50 values linearly interpolated between -1 to 1), and rotation angle (50 values linearly interpolated between -180 and 180 visual degrees), resulting in $50 \times 50 \times 50 = 125,000$ different stimulus sequences (i.e., all possible combinations). To understand the extent to which each visual parameter contributed to a model LC neuron's response, we decomposed the total response variance into different components [44]. The first three components represent the variance of the marginal response to each of the 3 visual parameters (which we had independently varied). We computed these marginalized variances by 1) taking the mean response for each value of a given visual parameter by averaging over all values of the other two parameters and 2) taking the variance of this mean response over values of the marginalized parameter. Any remaining variance (subtracting the three marginalized variances from the total response variance) represents response variance arising from interactions among the three visual parameters (e.g., the model LC response depends on female size but only if the female is in the center and faces away from the male). Because the 1-to-1 network was deterministic, no response variance was attributed to 'noise' across repeats (unlike trial-to-trial variability observed in responses of real neurons).

While analyzing the model LC responses to a large bank of static stimuli is helpful to understand LC tuning (Fig. 3**d** and **e**), we may miss important relationships between the features of the visual input and model LC responses without considering dynamics (e.g., the speed at which female size changes). To account for these other temporal features, we devised three stimulus sequences that varied in time for roughly 10 seconds each (Fig. 3**f**); these stimuli were similar to a subset of stimuli we presented to real male flies (see Methods section "Two-photon functional imaging"). For each stimulus sequence, we varied one visual parameter while the other two remained fixed at nominal values chosen based on natural sequence statistics. The first 2.5 seconds of each stimulus were the following:

1. *vary female size*: linearly increase from 0.5 to 0.9 with position $= 0$ and rotation $= 0°$
2. *vary female position*: linearly increase from -0.25 to 0.25 with size $= 0.8$ and rotation $= 0°$
3. *vary female rotation*: linearly increase from -45° to 45° with size $= 0.8$ and position $= 0$

The next 2.5 seconds were the same as the first 2.5 seconds except reversed in time (e.g., if the female increased in size the first 2.5 seconds, then the female decreased in size at the same speed for the next 2.5 seconds). Thus, the first 5 seconds was one period in which the female increased and decreased one parameter. The stimulus sequence contained 4 repeats of this period with different lengths (i.e., different speeds): 5, 3.33, 1.66, and 0.66 seconds (corresponding to 150, 100, 50, and 10 time frames, respectively). We passed these stimulus sequences as input into the 1-to-1 network (i.e., for each time frame, the 10 most recent images were passed into the model) and collected the model LC responses over time.

Analyzing how model LCs contribute to behavior

Because the 1-to-1 network identifies a one-to-one mapping, the model predicts not only the response of an LC neuron but also how that LC neuron causally relates to behavior. We wondered to what extent each model LC unit causally contributed to each behavioral output variable. We designed an ablation approach (termed the CumuLative Inactivation Procedure or CLIP) to identify which model LCs contributed the most to each behavioral output. The first step in CLIP is to inactivate each model LC unit individually by setting a model LC's activity value for all time frames to a constant value (chosen to be the mean activity value across all time frames). We then test to what extent the 1-to-1 network with the inactivated model LC predicts the behavioral output of heldout behavior from control flies. We choose the model LC unit that, once inactivated, leads to the *least* drop in prediction performance (i.e., the model LC unit that contributes the least to the behavioral output). We then iteratively repeat this step, keeping all previously-inactivated model LC units still inactivated. In this way, we greedily ablate model LC units until only one model LC remains. After performing CLIP, we obtain an ordering of model LC units from weakest to strongest contribution of a particular behavioral output (Fig. 4**b** and **c**). We then use this ordering (and prediction performance) to infer which model LC units contribute to which behavioral outputs. We performed CLIP to predict heldout behavior from control flies (Fig. 4**c**) as well as the model output to simple stimulus sequences (Fig. 4**f**). For the simple stimulus sequences (where we did not have real behavioral data), we used the model output when no silencing occurred as ground truth behavior.

Connectome analysis

To identify the output synaptic connections of the LC neurons, we used Janelia's Hemibrain FlyEM dataset [23]. We downloaded the summarized synaptic connection matrix at `https://storage.cloud.google.com/hemibrain/v1.2/exported-traced-adjacencies-v1.2.tar.gz` of the v1.2 version obtained with neuPrint+ [41]. This matrix contained the number of synaptic connections between $\sim$20,000 neurons in a female fly's brain. From this matrix, we identified which downstream neurons had direct synaptic connections to read out the LC neurons (Fig. 5**c** and Ext. Data Fig. 11). We summed the number of synaptic connections across all neurons with the same neuron type. We denoted a connection (Fig. 5**c**, blue square) if at least 3 synaptic connections existed between an LC/LPLC neuron type and a downstream neuron type.

Statistical analysis

Unless otherwise stated, all statistical hypothesis testing was conducted with permutation tests, which do not assume any parametric form of the underlying probability distributions of the sample. All tests were two-sided and non-paired, unless otherwise noted. Each test was performed with 1,000 runs, where $p < 0.001$ indicates the highest signficance achievable given the number of runs performed. When comparing changes in behavior due to genetic silencing versus control flies (Fig. 1**e**), we did not correct p-values for multiple comparisons, as we did not make any claim about a specific change being signficant. Still, 9 out of the 44 tests were significant, which is larger than the number of tests expected to be falsely significant (i.e., rejecting the null hypothesis when the null hypothesis is true) for a given α-level (for $\alpha = 0.05$, 2.2 tests were expected to be falsely significant). Paired permutation tests were performed when comparing prediction performance between models (Fig. 2**h**) for which the sample ordering was permuted in the same way for both samples. Error bars in Fig. 2 were 90% bootstrapped confidence intervals of the means, computed by randomly sampling with replacement. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those of previous studies [e.g., 16, 18, 38, 39]. Experimenters were not blinded to the conditions of the experiments during data collection and analysis.

Acknowledgments.

We thank R. Pang and M. Aragon for comments on the manuscript. This work was supported by a C.V. Starr Fellowship to B.R.C., a Simons Collaboration on the Global Brain Postdoctoral Fellowship to A.J.C., the Simons Collaboration on the Global Brain and NIH BRAIN Initiative (R01 NS104899) to M.M. and J.W.P., and an HHMI Faculty Scholar Award and NINDS R35 Research Program Award to M.M.

Data availability.

Data are available at `https://doi.org/10.34770/rmry-cs38`.

Code availability.

The code for extracting fly body positions (SLEAP) is available at `https://sleap.ai/`. Song segmentation was performed with code found at `https://github.com/murthylab/MurthyLab_FlySongSegmenter`. Model weights, example stimuli, and code are available at `https://github.com/murthylab/one2one-mapping`.

Competing interests

The authors declare no competing interests.

Author contributions.

B.R.C., A.J.C., N.R., J.W.P., and M.M. conceived of and designed the study. A.J.C. designed and performed the silencing experiments. A.J.C. and N.R. designed and performed the imaging experiments. B.R.C. analyzed the imaging data. B.R.C. and A.J.C. designed the model; B.R.C. trained and analyzed the model. B.R.C. analyzed the connectome data. B.R.C., J.W.P., and M.M. wrote the manuscript with input from A.J.C. and N.R.

References

[1] Christopher A Buneo, Murray R Jarvis, Aaron P Batista, and Richard A Andersen. Direct visuomotor transformations for reaching. *Nature*, 416(6881):632–636, 2002.

[2] Arthur R Houweling and Michael Brecht. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 451(7174):65–68, 2008.

[3] Gwyneth Card and Michael H Dickinson. Visually mediated motor planning in the escape response of drosophila. *Current Biology*, 18(17):1300–1307, 2008.

[4] Alexander Borst, Juergen Haag, and Dierk F Reiff. Fly motion vision. *Annual review of neuroscience*, 33: 49–70, 2010.

[5] Alexandre Pouget, Peter Dayan, and Richard Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–132, 2000.

[6] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365, 2016.

[7] David Sussillo, Mark M Churchland, Matthew T Kaufman, and Krishna V Shenoy. A neural network that finds a naturalistic solution for the production of muscle activity. *Nature neuroscience*, 18(7):1025–1033, 2015.

[8] Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. A deep learning framework for neuroscience. *Nature neuroscience*, 22(11):1761–1770, 2019.

[9] Daniel A Butts. Data-driven approaches to understanding visual neuron activity. *Annual review of vision science*, 5:451–477, 2019.

[10] Omer Mano, Matthew S Creamer, Bara A Badwan, and Damon A Clark. Predicting individual neuron responses with anatomically constrained task optimization. *Current Biology*, 31(18):4062–4075, 2021.

[11] Hendrikje Nienborg and Bruce Cumming. Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? *Current opinion in neurobiology*, 20(3):376–381, 2010.

[12] Xaq Pitkow, Sheng Liu, Dora E Angelaki, Gregory C DeAngelis, and Alexandre Pouget. How can single sensory neurons predict behavior? *Neuron*, 87(2):411–423, 2015.

[13] Arthur W Ewing. Functional aspects of drosophila courtship. *Biological Reviews*, 58(2):275–292, 1983.

[14] Petra Stockinger, Duda Kvitsiani, Shay Rotkopf, László Tirián, and Barry J Dickson. Neural circuitry that governs drosophila male courtship behavior. *Cell*, 121(5):795–807, 2005.

[15] Philip Coen, Jan Clemens, Andrew J Weinstein, Diego A Pacheco, Yi Deng, and Mala Murthy. Dynamic sensory cues shape song structure in drosophila. *Nature*, 507(7491):233–237, 2014.

[16] Tom Hindmarsh Sten, Rufei Li, Adriane Otopalik, and Vanessa Ruta. Sexual arousal gates visual processing during drosophila courtship. *Nature*, 595(7868):549–553, 2021.

[17] Mehmet Keleş and Mark A Frye. Visual behavior: The eyes have it. *Elife*, 6:e24896, 2017.

[18] Ming Wu, Aljoscha Nern, W Ryan Williamson, Mai M Morimoto, Michael B Reiser, Gwyneth M Card, and Gerald M Rubin. Visual projection neurons in the drosophila lobula link feature detection to distinct behavioral programs. *Elife*, 5:e21022, 2016.

[19] Han SJ Cheong, Igor Siwanowicz, and Gwyneth M Card. Multi-regional circuits underlying visually guided decision-making in drosophila. *Current Opinion in Neurobiology*, 65:77–87, 2020.

[20] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474):78–84, 2013.

[21] Jonathan A Michaels, Stefan Schaffelhofer, Andres Agudelo-Toro, and Hansjörg Scherberger. A modular neural network model of grasp movement generation. *bioRxiv*, page 742189, 2020.

[22] Baohua Zhou, Zifan Li, Sunnie Kim, John Lafferty, and Damon A Clark. Shallow neural networks trained to detect collisions recover features of visual loom-selective neurons. *Elife*, 11:e72067, 2022.

[23] Louis K Scheffer, C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J Hayworth, Gary B Huang, Kazunori Shinomiya, Jeremy Maitlin-Shepard, Stuart Berg, et al. A connectome and analysis of the adult drosophila central brain. *Elife*, 9:e57443, 2020.

[24] Hideo Otsuna and Kei Ito. Systematic analysis of the visual projection neurons of drosophila melanogaster. i. lobula-specific pathways. *Journal of Comparative Neurology*, 497(6):928–958, 2006.

[25] Salil S Bidaye, Meghan Laturney, Amy K Chang, Yuejiang Liu, Till Bockemühl, Ansgar Büschges, and Kristin Scott. Two brain pathways initiate distinct forward walking programs in drosophila. *Neuron*, 108(3): 469–485, 2020.

[26] Inês MA Ribeiro, Michael Drews, Armin Bahl, Christian Machacek, Alexander Borst, and Barry J Dickson. Visual projection neurons mediating directed courtship in drosophila. *Cell*, 174(3):607–621, 2018.

[27] Jan M Ache, Jason Polsky, Shada Alghailani, Ruchi Parekh, Patrick Breads, Martin Y Peek, Davi D Bock, Catherine R von Reyn, and Gwyneth M Card. Neural basis for looming size and velocity encoding in the drosophila giant fiber escape pathway. *Current Biology*, 29(6):1073–1081, 2019.

[28] Nathan C Klapoetke, Aljoscha Nern, Martin Y Peek, Edward M Rogers, Patrick Breads, Gerald M Rubin, Michael B Reiser, and Gwyneth M Card. Ultra-selective looming detection from radial motion opponency. *Nature*, 551(7679):237–241, 2017.

[29] Rajyashree Sen, Ming Wu, Kristin Branson, Alice Robie, Gerald M Rubin, and Barry J Dickson. Moonwalker descending neurons mediate visually evoked retreat in drosophila. *Current Biology*, 27(5):766–771, 2017.

[30] Ryosuke Tanaka and Damon A Clark. Object-displacement-sensitive visual neurons drive freezing in drosophila. *Current Biology*, 30(13):2532–2550, 2020.

[31] Mehmet F Keleş and Mark A Frye. Object-detecting neurons in drosophila. *Current Biology*, 27(5):680–687, 2017.

[32] Laiyong Mu, Kei Ito, Jonathan P Bacon, and Nicholas J Strausfeld. Optic glomeruli and their inputs in drosophila share an organizational ground pattern with the antennal lobes. *Journal of Neuroscience*, 32(18): 6061–6071, 2012.

[33] Rachel I Wilson. Early olfactory processing in drosophila: mechanisms and principles. *Annual review of neuroscience*, 36:217, 2013.

[34] Mala Murthy, Ila Fiete, and Gilles Laurent. Testing odor response stereotypy in the drosophila mushroom body. *Neuron*, 59(6):1009–1023, 2008.

[35] Sophie JC Caron, Vanessa Ruta, LF Abbott, and Richard Axel. Random convergence of olfactory inputs in the drosophila mushroom body. *Nature*, 497(7447):113–117, 2013.

[36] Talmo D Pereira, Nathaniel Tabris, Arie Matsliah, David M Turner, Junyu Li, Shruthi Ravindranath, Eleni S Papadoyannis, Edna Normand, David S Deutsch, Z Yan Wang, et al. Sleap: A deep learning system for multi-animal pose tracking. *Nature methods*, 19(4):486–495, 2022.

[37] Jan Clemens, Philip Coen, Frederic A Roemschied, Talmo D Pereira, David Mazumder, Diego E Aldarondo, Diego A Pacheco, and Mala Murthy. Discovery of a new song mode in drosophila reveals hidden structure in the sensory and neural drivers of behavior. *Current biology*, 28(15):2400–2412, 2018.

[38] Carola Städele, Mehmet F Keleş, Jean-Michel Mongeau, and Mark A Frye. Non-canonical receptive field properties and neuromodulation of feature-detecting neurons in flies. *Current Biology*, 30(13):2508–2519, 2020.

[39] Nathan C Klapoetke, Aljoscha Nern, Edward M Rogers, Gerald M Rubin, Michael B Reiser, and Gwyneth M Card. A functionally ordered visual feature map in the drosophila brain. *Neuron*, 2022.

[40] Mehmet F Keleş, Ben J Hardcastle, Carola Städele, Qi Xiao, and Mark A Frye. Inhibitory interactions and columnar inputs to an object motion detector in drosophila. *Cell reports*, 30(7):2115–2124, 2020.

[41] Jody Clements, Tom Dolafi, Lowell Umayam, Nicole L Neubarth, Stuart Berg, Louis K Scheffer, and Stephen M Plaza. neuprint: analysis tools for em connectomics. *bioRxiv*, 2020.

[42] Hung-Hsiang Yu, Takeshi Awasaki, Mark David Schroeder, Fuhui Long, Jacob S Yang, Yisheng He, Peng Ding, Jui-Chun Kao, Gloria Yueh-Yi Wu, Hanchuan Peng, et al. Clonal development and organization of the adult drosophila central brain. *Current Biology*, 23(8):633–643, 2013.

[43] John T Wixted, Larry R Squire, Yoonhee Jang, Megan H Papesh, Stephen D Goldinger, Joel R Kuhn, Kris A Smith, David M Treiman, and Peter N Steinmetz. Sparse and distributed coding of episodic memory in neurons of the human hippocampus. *Proceedings of the National Academy of Sciences*, 111(26):9621–9626, 2014.

[44] Wieland Brendel, Ranulfo Romo, and Christian K Machens. Demixed principal component analysis. *Advances in neural information processing systems*, 24, 2011.

[45] Clara H Ferreira and Marta A Moita. Behavioral and neuronal underpinnings of safety in numbers in fruit flies. *Nature communications*, 11(1):1–10, 2020.

[46] Catherine R Von Reyn, Aljoscha Nern, W Ryan Williamson, Patrick Breads, Ming Wu, Shigehiro Namiki, and Gwyneth M Card. Feature integration drives probabilistic behavior in the drosophila escape response. *Neuron*, 94(6):1190–1204, 2017.

[47] Shawn R Olsen, Vikas Bhandawat, and Rachel I Wilson. Divisive normalization in olfactory population codes. *Neuron*, 66(2):287–299, 2010.

[48] Matteo Carandini and David J Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012.

[49] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[50] Michael N Shadlen and William T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896, 1998.

[51] Vikas Bhandawat, Shawn R Olsen, Nathan W Gouwens, Michelle L Schlief, and Rachel I Wilson. Sensory processing in the drosophila antennal lobe increases reliability and separability of ensemble odor representations. *Nature neuroscience*, 10(11):1474–1482, 2007.

[52] Kerry MM Walker, Jennifer K Bizley, Andrew J King, and Jan WH Schnupp. Multiplexed and robust representations of sound features in auditory cortex. *Journal of Neuroscience*, 31(41):14565–14576, 2011.

[53] Jeffrey W Johnston, Stephanie E Palmer, and David J Freedman. Nonlinear mixed selectivity supports reliable neural computation. *PLOS computational biology*, 16(2):e1007544, 2020.

[54] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012.

[55] Robert AA Campbell, Kyle S Honegger, Hongtao Qin, Wanhe Li, Ebru Demir, and Glenn C Turner. Imaging a population code for odor identity in the drosophila mushroom body. *Journal of Neuroscience*, 33(25): 10568–10581, 2013.

[56] Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590, 2013.

[57] Alexandria H Jaeger, Molly Stanley, Zachary F Weiss, Pierre-Yves Musso, Rachel CW Chan, Han Zhang, Damian Feldman-Kiss, and Michael D Gordon. A complex peripheral code for salt taste in drosophila. *Elife*, 7: e37167, 2018.

[58] Rodrigo Quian Quiroga and Stefano Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3):173–185, 2009.

[59] Shinichiro Kira, Houman Safaai, Ari S Morcos, Stefano Panzeri, and Christopher D Harvey. A distributed and efficient population code of mixed selectivity neurons for flexible navigation decisions. *bioRxiv*, 2022.

[60] Adam J Calhoun, Jonathan W Pillow, and Mala Murthy. Unsupervised identification of the internal states that shape natural behavior. *Nature neuroscience*, 22(12):2040–2049, 2019.

[61] Benjamin J Arthur, Tomoko Sunayama-Morita, Philip Coen, Mala Murthy, and David L Stern. Multi-channel acoustic recording and automated analysis of drosophila courtship songs. *BMC biology*, 11(1):1–11, 2013.

[62] Justin P Kumar. Building an ommatidium one cell at a time. *Developmental Dynamics*, 241(1):136–149, 2012.

[63] Sweta Agrawal, Steve Safarik, and Michael Dickinson. The relative roles of vision and chemosensation in mate recognition of drosophila melanogaster. *Journal of Experimental Biology*, 217(15):2796–2805, 2014.

[64] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[65] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1): 1929–1958, 2014.

[66] François Chollet et al. Keras. `https://keras.io`, 2015.

[67] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

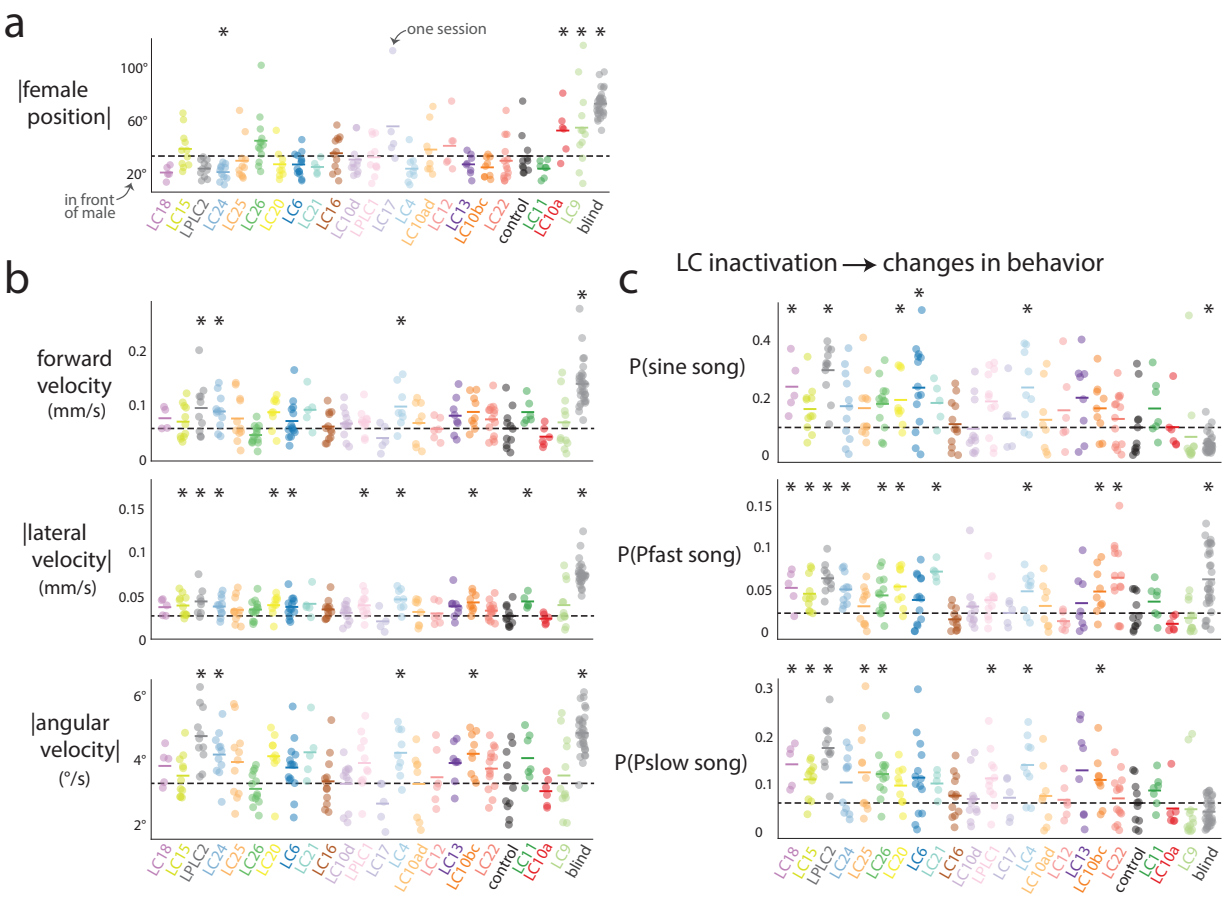
[68] Michael J Hautus, Neil A Macmillan, and C Douglas Creelman. *Detection theory: A user's guide*. Routledge, 2021.

[69] Dean A Pospisil and Wyeth Bair. The unbiased estimation of the fraction of variance explained by a model. *PLoS computational biology*, 17(8):e1009212, 2021.

[70] Mai M Morimoto, Aljoscha Nern, Arthur Zhao, Edward M Rogers, Allan M Wong, Mathew D Isaacson, Davi D Bock, Gerald M Rubin, and Michael B Reiser. Spatial readout of visual looming in the central brain of drosophila. *Elife*, 9:e57685, 2020.

[71] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.

Extended Data Figures



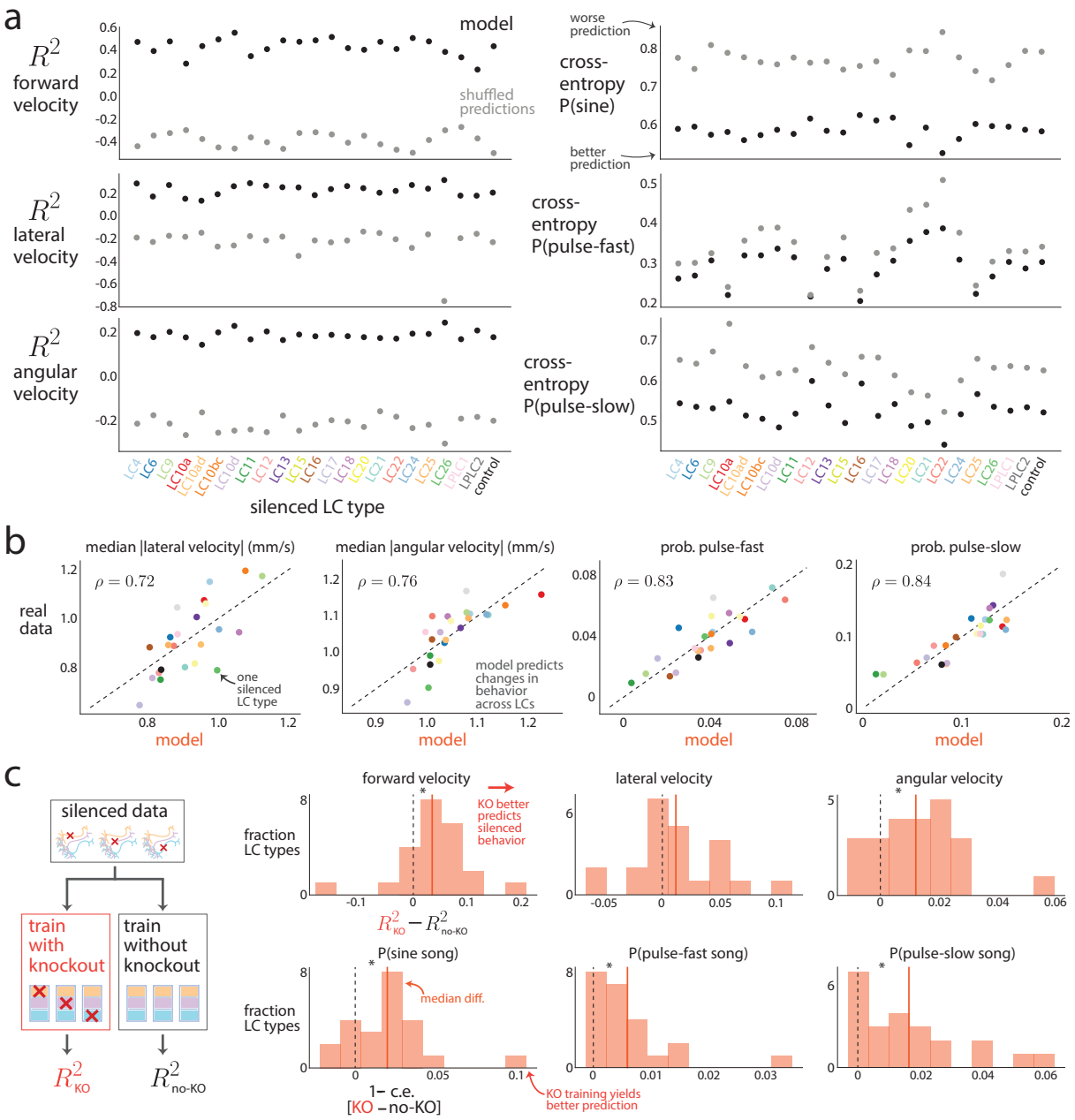
Extended Data Figure 1: We observed different changes to behavior when silencing different LC neuron types of the male's visual system.

a. We observed differences when considering the male's visual input, including the position of the female relative to the male's heading direction (absolute value of position, where 0° is directly in front of the male).

b. We observed changes in the male's movement statistics, including forward velocity, lateral velocity, and angular velocity. The absolute value was taken for lateral and angular velocity, as we were interested in changes away from the male's heading direction (e.g., a large turn to the right or left indicated a large deviance).

c. We observed changes in the male's song production, including the probability of sine, Pfast, and Pslow song.

All plots have the same format, including the column ordering, as in Fig. 1**e**. Each dot denotes the mean for one courtship session (one male-female pair); each line indicates the mean across flies for that condition. The black dashed line is the mean of the control flies. Asterisks denote significant deviation from control flies ($p < 0.05$, permutation test). We did not correct p-values for multiple comparisons, as we wondered how many significant changes there were—not any specific change. We observed 42 significant changes (not including blind flies) out of 154 comparisons, which is larger than the ∼8 tests expected to be falsely significant for an $\alpha = 0.05$. We also note that these metrics were not exhaustive, and there were likely behavioral changes between control and silenced flies not observed here. This was a primary motivation to model the behavior of flies with different silenced LCs.



Extended Data Figure 2: Model predictions of behavior. We tested the ability of the 1-to-1 network to predict behavior in multiple ways.

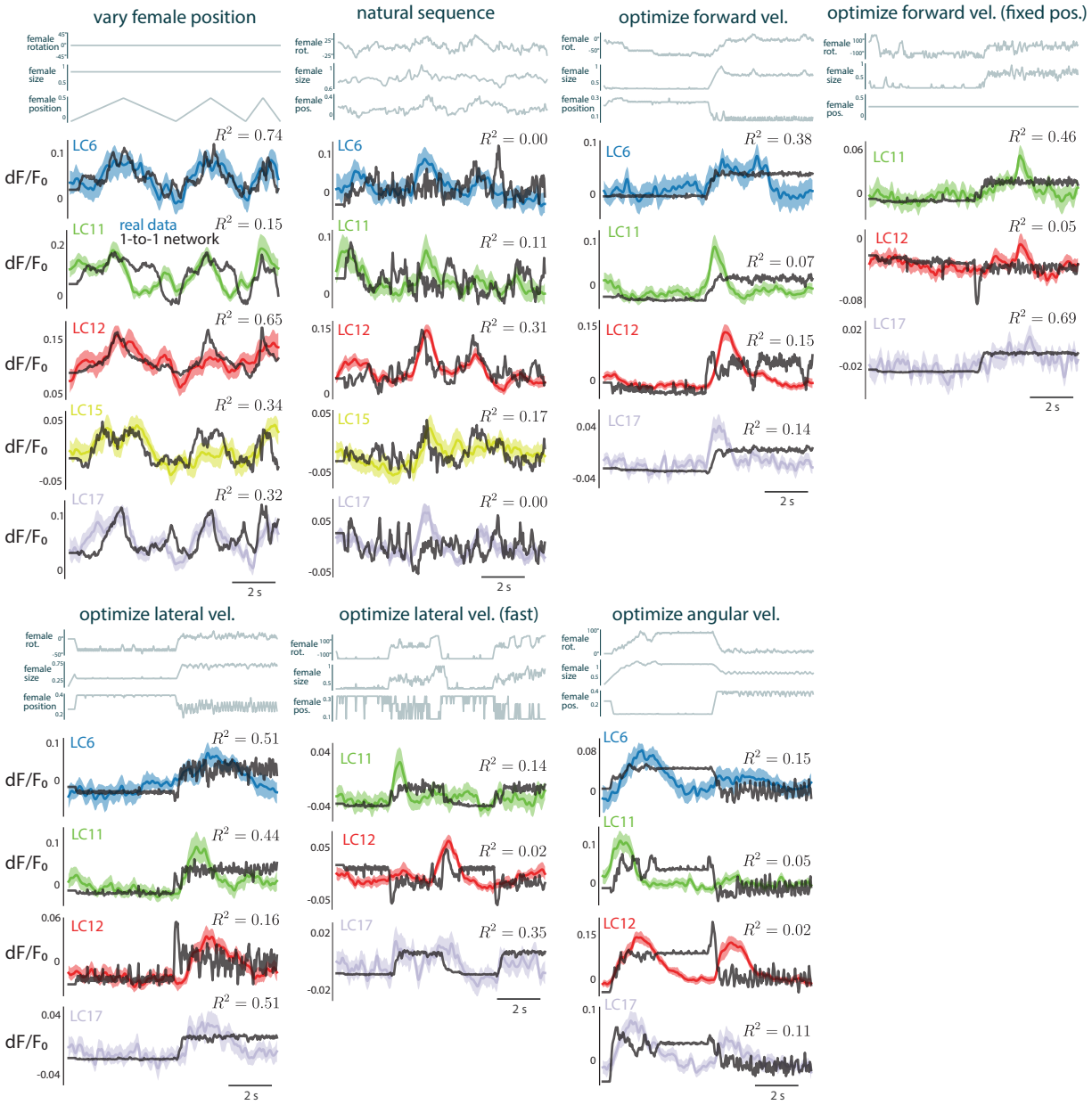
a. The model predicted the male fly's movement and song production (heldout data) across many of the LCs and control flies. Each black dot corresponds to the prediction performance (R^2 for movement variables, binary cross-entropy for the song variables) for heldout frames across flies with the same silenced LC neuron type. Because the values of binary cross-entropy can be difficult to interpret, we re-computed prediction performance after shuffling the heldout frames but keeping the predictions fixed (gray dots). Overall, the 1-to-1 network predicted movement and song production better than expected by chance. We note that R^2 values for movement (left column, $R^2 \approx 0.4$ for forward velocity, $R^2 \approx 0.2$ for lateral and angular velocity) are not close to 1 because we are predicting rapid changes to movement variables frame-to-frame (with a frame rate of 30 Hz). Example traces are in Fig. 1**f**.
(continued on next page...)

Extended Data Figure 2: (...continued from previous page)

Smoothing and removing outliers would likely improve R^2; however, this would still fail to account for the fact that a male fly's moment-to-moment decision is stochastic—in other words, the male responds differently to repeats of the same stimulus sequence. To take this stochasticity into account, one would need to present identical repeats of the same visual stimulus sequence. This is not possible for our natural courtship experiments, where a male fly's visual experience is determined by his behavior. However, this may be possible in experiments using virtual reality, where the experimenter has greater control over a male fly's visual input.

b. Changes in behavior predicted by the model (x-axis) versus observed in real data (y-axis). All reported correlations ρ were significant ($p < 0.001$, permutation test).

c. A key assumption of knockout training is that there are noticeable differences between silenced and control behavior (else the 1-to-1 network would not be able to differentiate behavior resulting from silencing different LCs). To test this, we compared our model with and without knockout training. For each LC neuron type, we computed the R^2 on heldout silenced behavior for our model with knockout training (R^2_{KO}) and without knockout training ($R^2_{\text{no-KO}}$). If $R^2_{\text{KO}} - R^2_{\text{no-KO}}$ is greater than 0, it indicates that by giving our model information about which LC neuron type was silenced, our model can better predict the changes in behavior. The 1-to-1 network had significantly better prediction than that of a network with no knockout training (orange lines to the right of black dashed lines, $p < 0.05$ denoted by asterisks, one-tailed paired permutation test) for all behavioral variables except lateral velocity ($p = 0.17$).



Extended Data Figure 3: Real LC responses and predicted responses to stimulus sequences of a moving fictive female.

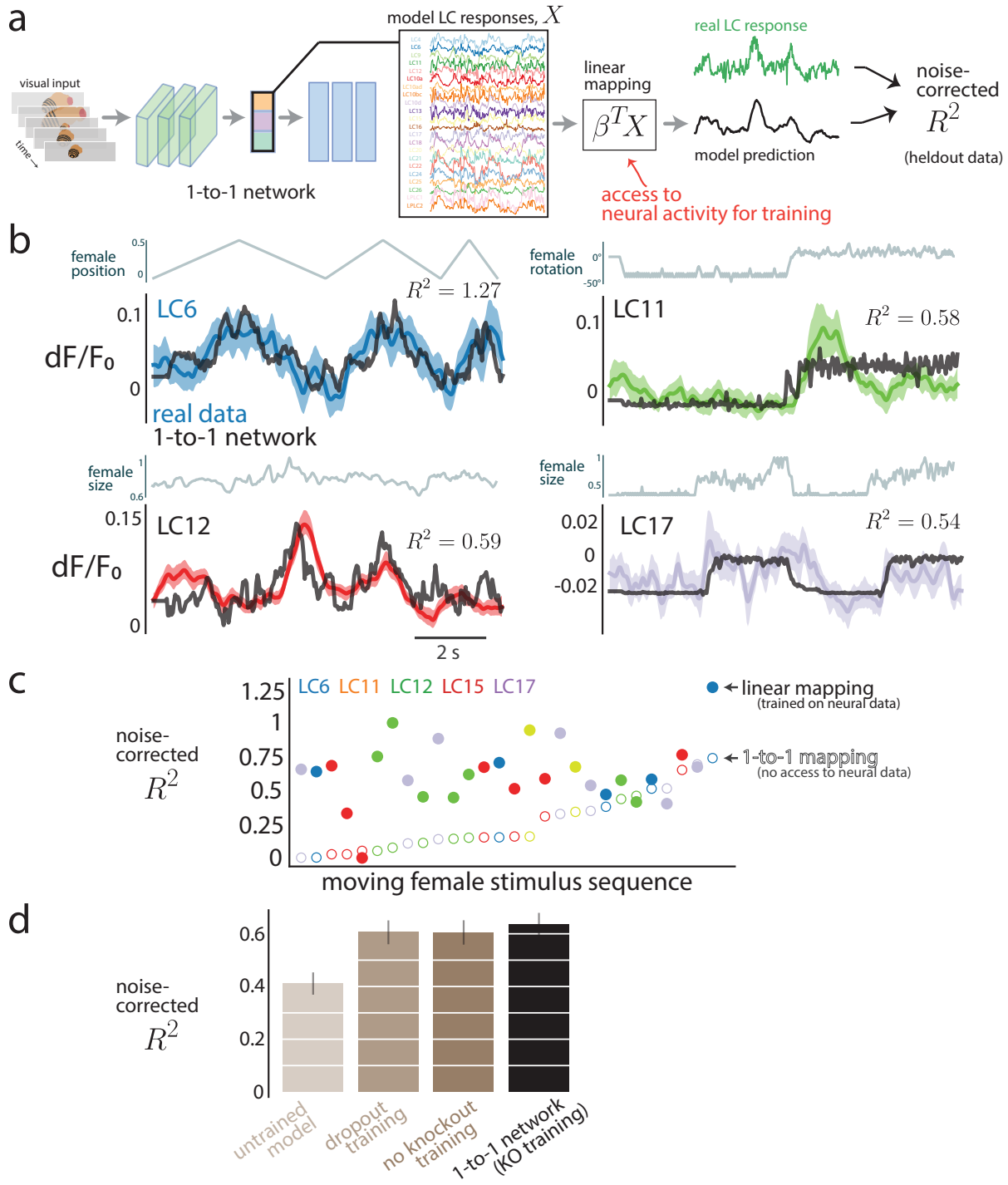
We considered 9 different stimulus sequences in which a female varied her rotation, size, and position (three top traces for each stimulus sequence, see Methods for stimulus descriptions). We found that the 1-to-1 network's predictions (black traces) largely predicted the responses of the real LC neurons (color traces), despite the fact that the 1-to-1 network was never given access to neural data. The reported noise-corrected R^2's are the same as in Figure 2**g**.

We only considered stimulus sequences for which the real LCs responses reliably varied across time for the stimulus sequence (i.e., their signal-to-noise ratio across repeats surpassed a threshold, see Methods). For some LCs, we detected a reliable response to only one or a few stimuli (e.g., LC15 only responded to 'vary female position' and 'natural sequence').
(continued on next page...)

Extended Data Figure 3: (...continued from previous page)

We noticed that none of the LC neurons responded to stimulus sequences in which the fictive female varied either her size ('vary female size') or rotation ('vary female rotation'). This was surpising given that previous studies [31, 38, 70] and the 1-to-1 network (Fig. 3**e**, **f**, and **g**) indicate that many LCs are responsive to the size of the object. Although we varied the female size within a realistic range observed during courtship, we suspect that this range was not large enough when we presented the fictive female on the projection screen to real flies. Thus, from the male's perspective, the fictive female likely appeared to be far away and not changing in size.

We noticed two salient differences between the real LC responses and the 1-to-1 network's predictions. First, the real LC responses had adaptation effects (e.g., the large transients for the LC neurons at stimulus onset of 'optimize angular vel.'). Because the 1-to-1 network did not have an adaption mechanism, the model could not predict these transients. Second, because the LC responses were recorded with calcium imaging (and averaged over many repeats), the real LC responses tended to be smoother than the 1-to-1 network's predictions. Incorporating an adaption mechanism and accounting for response smoothness would likely increase prediction performance; however, we do not foresee these improvements largely changing the conclusions inferred from our model, namely the relationship between stimulus features, LC responses, and behavior.



Extended Data Figure 4: A linear mapping from model LC units to real LC neurons yields high prediction performance. The 1-to-1 network, with no access to neural data during training, was predictive of real LC responses to stimulus sequences in which a fictive female varied her size, position, and rotation (Fig. 2**f**-**h**). Here, we ask if we were to give our 1-to-1 network access to neural data for training, to what extent would the model's prediction of real LC responses improve.

(continued on next page...)

Extended Data Figure 4: (...continued from previous page)

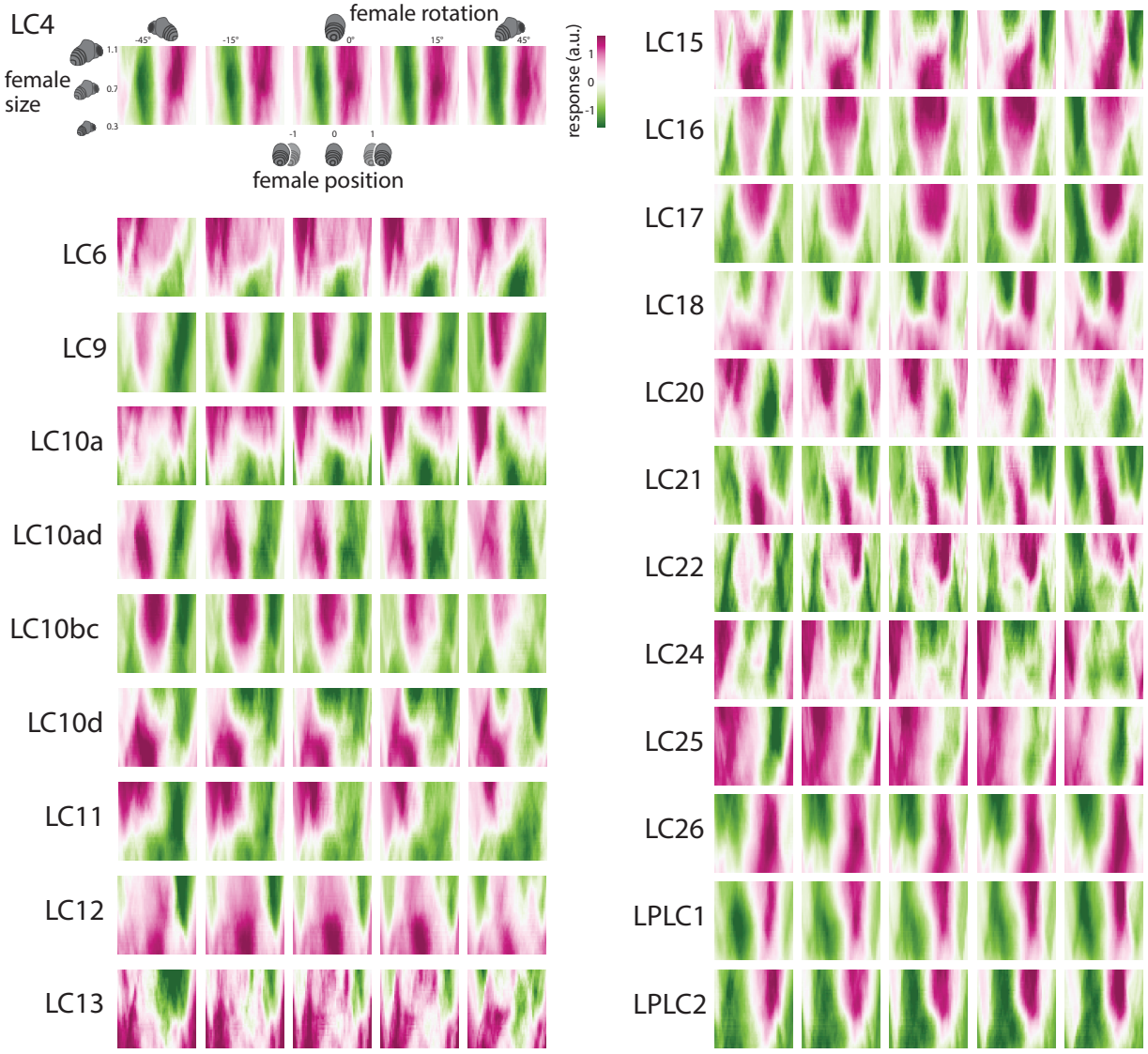
a. Basic setup. We input a stimulus sequence into the 1-to-1 network (fully trained with knockout training) and collect responses from all model LC units, denoted as $X \in \mathcal{R}^{K \times T}$ for K model LC units (here, $K = 22$) and the T timepoints of the stimulus sequence. We then define a linear mapping $\beta \in \mathcal{R}^{K}$ to map the K model LC responses to the real LC response. We use real LC responses to train β. Specifically, for each of the 4 cross-validation folds, we train β on 75% of the real LC responses (randomly selected) using ridge regression. We then predict the responses for the remaining heldout timepoints. We concatenate the predictions across the 4 folds and then compute the noise-corrected R^2 in the same way as in Figure 2**f**-**h**. Thus, the reported cross-validated noise-corrected R^2's indicate to what extent the 1-to-1 network, given neural data on which to train, can predict heldout real LC responses. Another view is that in this setting, the 1-to-1 network is a task-driven model trained on behavioral data with an internal representation (the model LC bottleneck) that reflects the activity of real LC neurons up to a linear transformation [6].

b. Example stimulus sequences (one parameter shown for each, top row) with real LC responses (color traces) and predicted responses (black traces). Same format as in Figure 2**f**. Note that a noise-corrected R^2 above 1, as we observed for LC6 ($R^2 = 1.27$), is possible due to the unbiasing terms in computing an unbiased R^2 [69]; the true noise-corrected R^2 is below but likely close to 1.

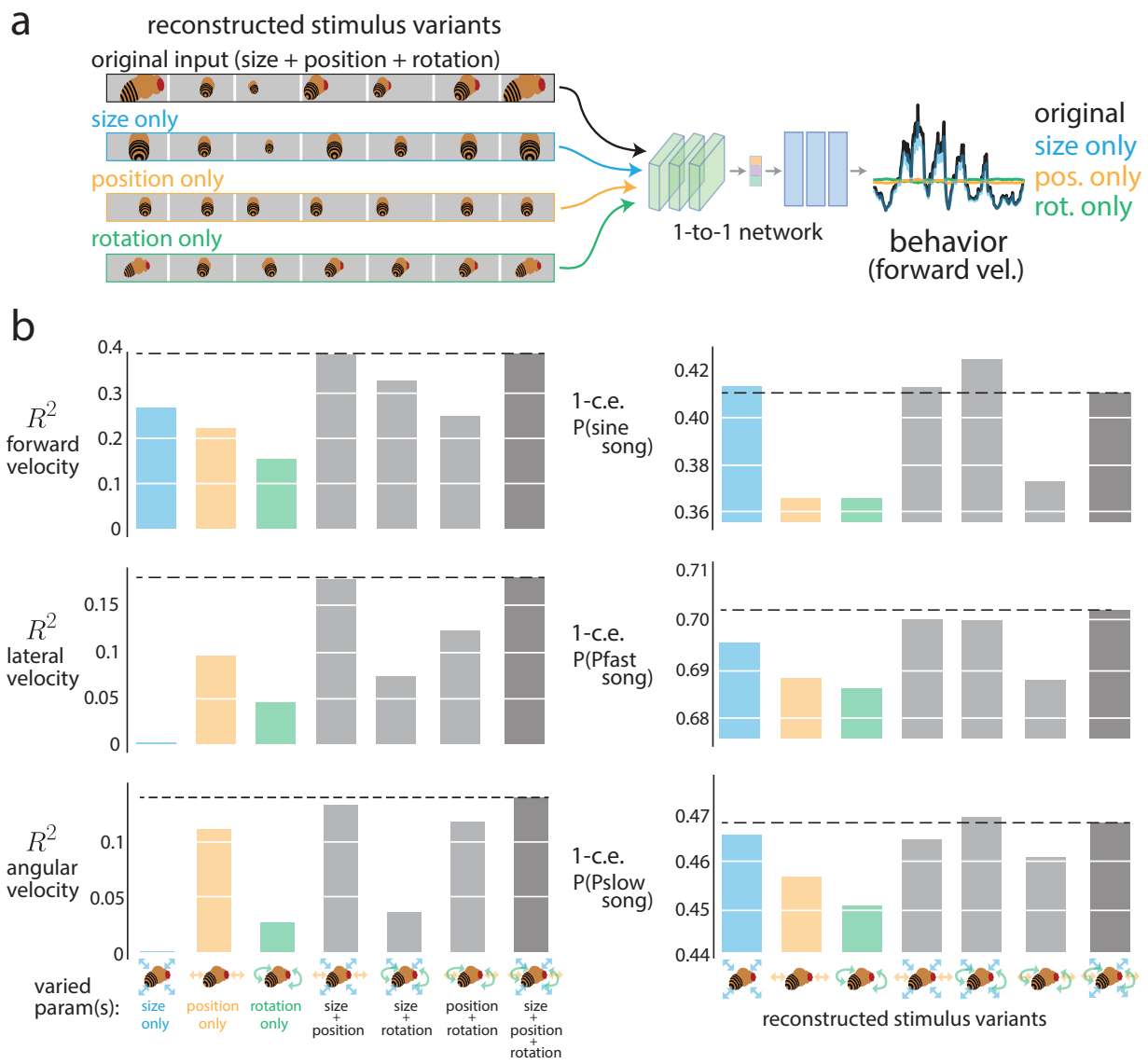
c. Prediction performance of the linear mapping (filled dots) versus the 1-to-1 network with its one-to-one mapping (unfilled dots). As expected, the prediction performance of the linear mapping tends to be larger for each pair of stimulus sequence and LC neuron type than for each pair of the one-to-one mapping (filled dots are above unfilled dots). Same format as in Figure 2**g**; unfilled dots are the same as the dots in Figure 2**g**.

d. Prediction performance using the linear mapping for different networks. For each network, we trained a new linear mapping between the model LC responses and the real LC responses. Overall, prediction performance greatly increased: The 1-to-1 network with the linear mapping had a noise-corrected R^2 at $\sim$65%, an additive increase of $\sim$40% over that of the 1-to-1 network with the one-to-one mapping ($\sim$25%, Fig. 2**h**). We also found that, for the linear mapping, the performance of the 1-to-1 network was similar to those of the other trained models (black bar close to bars for dropout training and no knockout training). This was not unexpected and indicates that all 3 networks (trained on the same data) have similar internal representations (up to a linear transformation) at the layer of their LC bottlenecks. However, the 1-to-1 network's representation is better aligned along its coordinate axes (i.e., for which each model LC unit corresponds to one axis) than those of the other networks (Fig. 2**h**) when comparing those axes to the LC neurons. The untrained model was predictive of LC responses (bar for 'untrained model' above 0), indicating that this model's convolutional filters, even with randomized weights, could detect large changes of the visual stimulus (e.g., a fictive female moving back and forth, see Methods). That a linear combination of random features is often predictive in a regression setting is a well-studied phenomenon in machine learning [71].

These results indicate that by simply training a network on courtship behavioral data (i.e., a task-driven approach), we have identified a highly-predictive image-computable model of LC neurons. To our knowledge, ours is the first image-computable model of LC neurons proposed.



Extended Data Figure 5: Model LC tuning heat maps. Each "pixel" in the heatmap corresponds to the response of the model LC unit to one input stimulus sequence in which a static fictive female fly had a given size, position, and rotation (i.e., all 10 images of the input sequence were the same). We then varied female size, position, and rotation across stimulus sequences. Same format as in Figure 3**c** and **d**).



Extended Data Figure 6: Assessing which visual stimulus parameters of the fictive female predict behavior.

a. We tested to what extent female size, position, rotation, or a combination of the three predicted male behavior by only reconstructing certain parameters of the male's visual input (while keeping the other parameters fixed). For example, when reconstructing the male's visual input, we allowed only the female size to vary, keeping the female's lateral position and rotation angle fixed ('size only', blue). For each reconstructed stimulus variant, we trained our model with control data only (i.e., no knockout training) and used the trained model to predict heldout control data. Thus, each training and testing had the same data except that the visual input was manipulated in some way. We tested all possible combinations of visual parameters: The reconstructed stimulus variants were {size}, {position}, {rotation}, {size, position}, {size, rotation}, {position, rotation}, {size, position, rotation}.

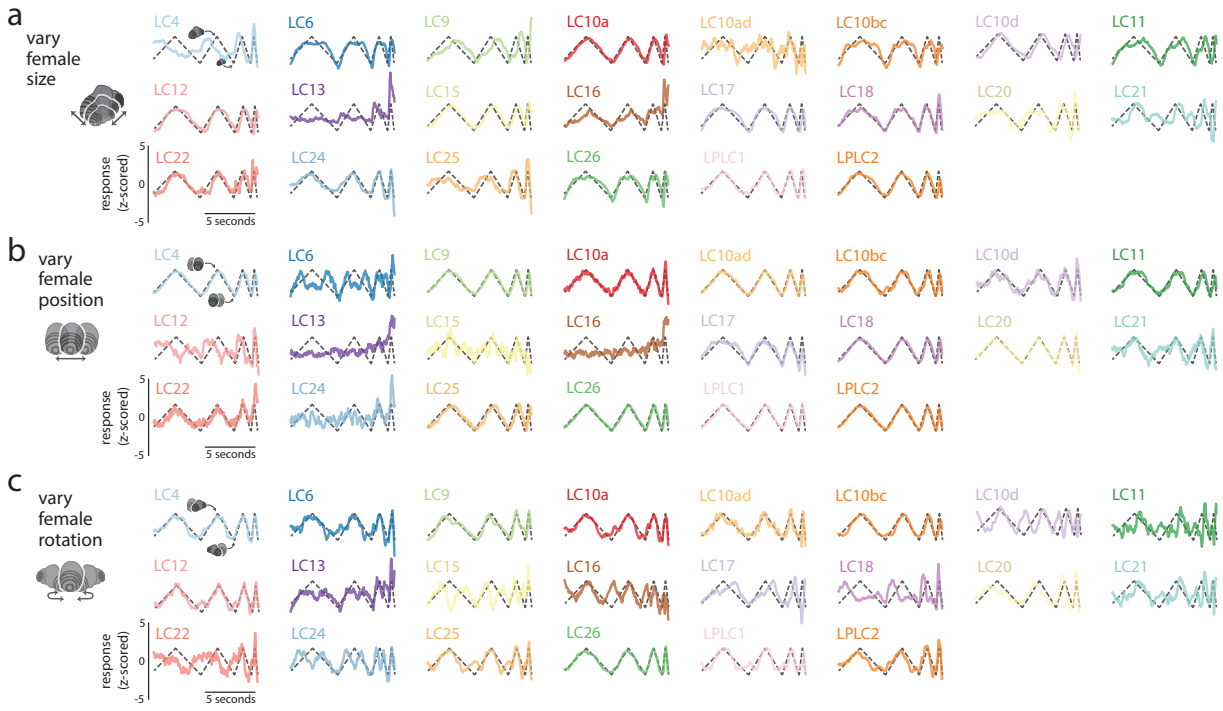
In this illustration, the reconstructed stimulus variant of 'size only' (blue), in which only the female's size varies, entirely predicts the behavioral output (e.g., forward velocity) of the 'original' variant (blue trace overlays black trace), but 'position only' and 'rotation only' predict no behavior (orange and green traces are flat). This indicates that size alone predicts forward velocity in this example.
(continued on next page...)

Extended Data Figure 6: (...continued from previous page)

b. Each variant's prediction performance (on heldout behavior of control flies) for the male fly's movement (left column, measured in explained variance or R^2) and song production (right column, measured in $1 -$ binary cross-entropy or '1-c.e.' of the probability of either sine, Pfast, or Pslow song). Giving the model access to all visual parameters ('size + pos. + rot.') leads to the best prediction (black dashed lines) except for 'size + rot.' for sine and Pslow song. Thus, a bar close to a black dashed line indicates that the corresponding combination of visual parameters predicts the behavioral output as well as using all three visual parameters.

We had two findings. First, we found that female size was most predictive of forward velocity (top left panel, blue bar above yellow and green bars) and that female position was most predictive of lateral and angular velocity (bottom two left panels, yellow bar above blue and green bars). This complements our finding that size and position are the most heavily encoded in the model LC population (Fig. 3**e**). Still, female rotation provided some prediction for all three movement-related outputs (left column, green bars above 0). Of all three parameters, female size was most predictive of song production (right column, blue bars above yellow and green bars).

Second, female size and position together was the strongest predictor for movement (left column, 'size + position' bars are close to black dashed line), while female size and rotation together tended to be the strongest predictor for song production (right column, 'position + rotation' bars are close or even above black dashed line). These results suggest that the male fly integrates information from all three visual parameters to guide his behavior. This motivated us to analyze how each LC encodes these three visual parameters (Fig. 3).



Extended Data Figure 7: All model LC responses to dynamic stimulus sequences in which only one visual parameter of the fictive female varied.

Same stimulus sequences and format as in Figure 3**f**; these responses were used to compute the R^2's in Fig. 3**g**).

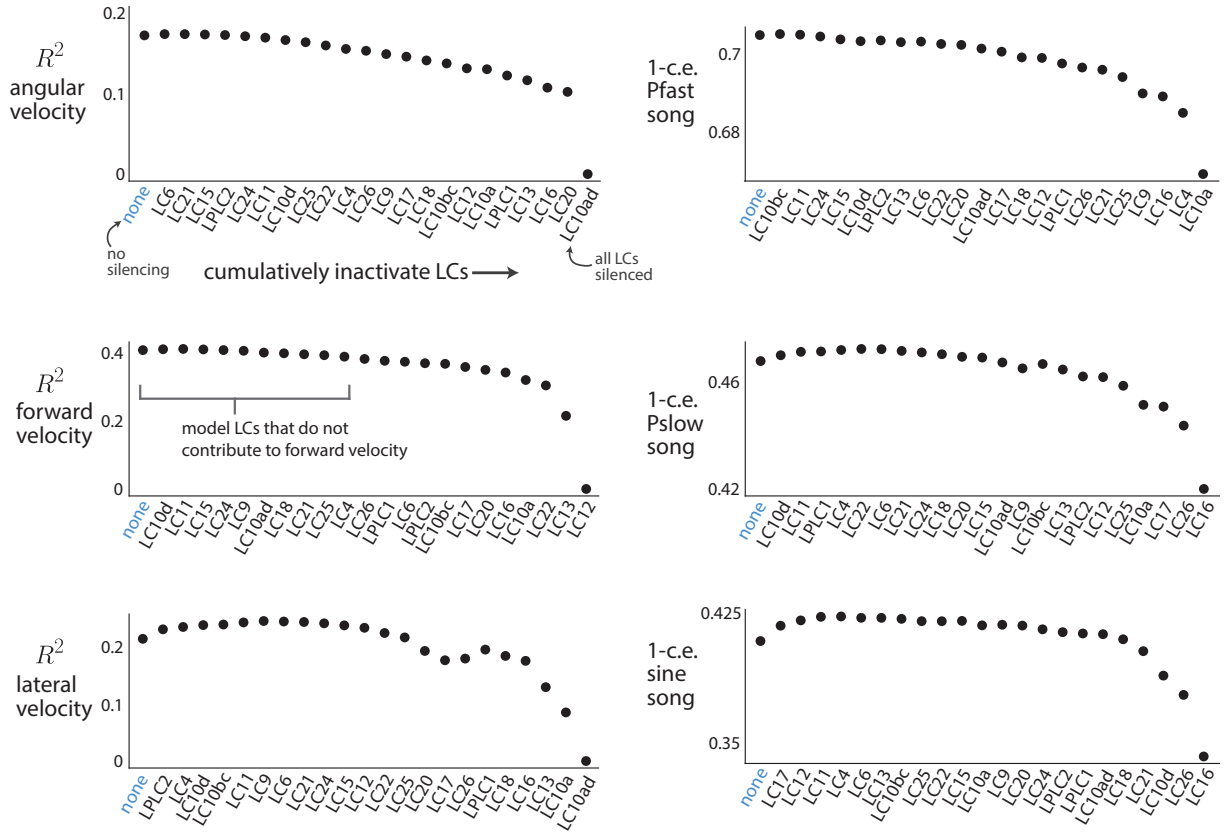
Stimulus sequences include the following:

a. Varying female size while the female stays in the middle facing away from the male.

b. Varying female position while the female has a fixed, relatively large size while facing away from the male.

c. Varying female rotation while the female has a fixed, relatively large size and stays in the middle.

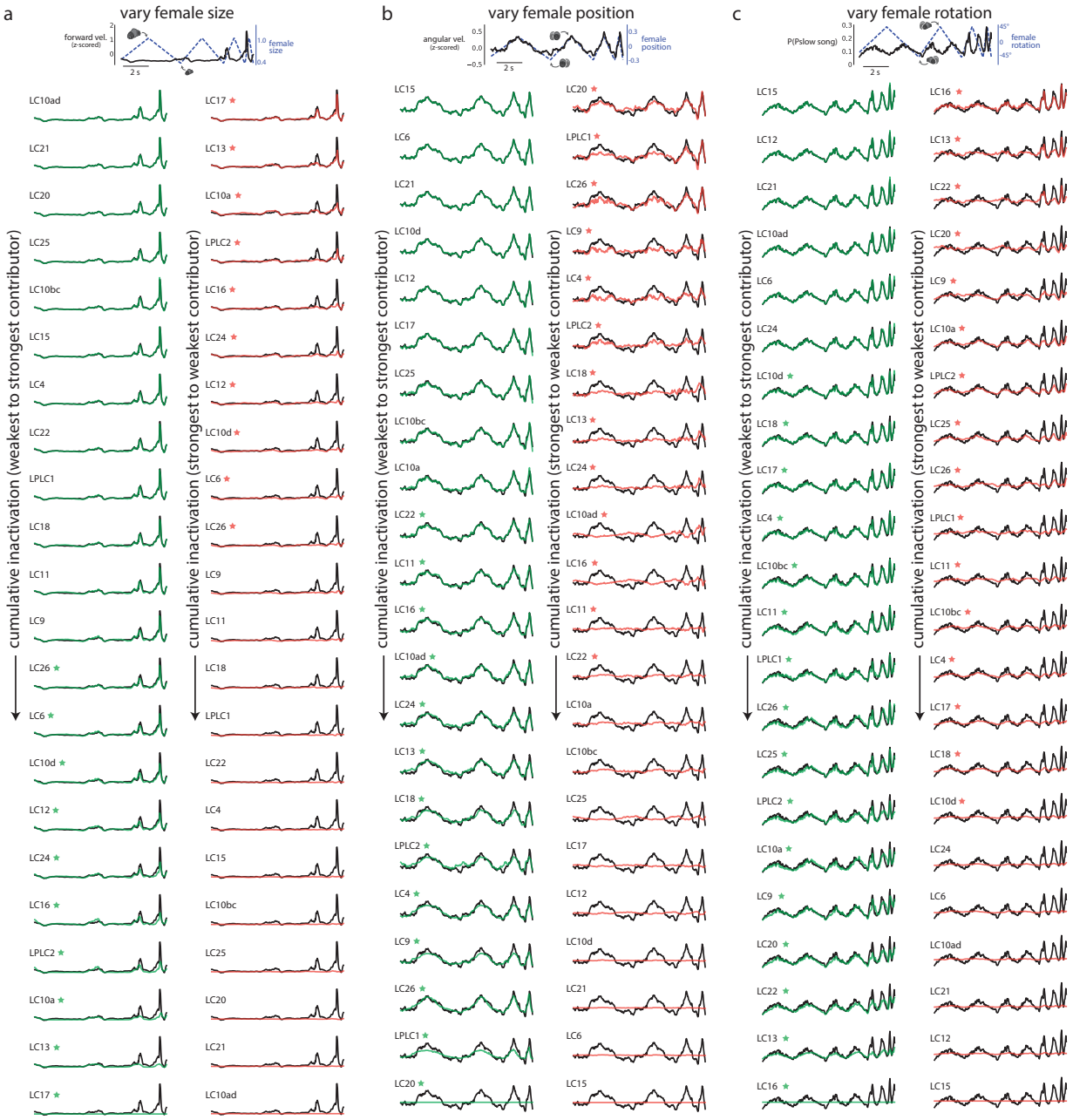
Each trace's sign was flipped to have a positive correlation with the varying visual parameter of the corresponding stimulus sequence.



Extended Data Figure 8: Cumulative inactivation procedure for each of the 6 behavioral output variables during natural courtship.

We inactivated model LC units in a cumulative, greedy manner and observed to what extent the responses for the remaining model LC units predict heldout behavioral data from control flies. Same format as in Figure 4**b**. For each plot, model LC units on the left contribute the least to the given behavioral output; model LC units on the right contribute the most. We found that when inactivating some model LC units, performance actually *increased* (e.g., LC17 for sine song, bottom right); this is likely due to overfitting and is one of the reasons dropout is useful for training deep neural networks [65]. However, while inactivating one model LC unit for one behavioral output may increase performance (e.g., LC17 for sine song), inactivating this model LC unit will likely hurt prediction for another behavioral output (e.g., LC17 is the third strongest contributor for Pslow song, middle plot in right column).

The red squares of the heatmaps in Figure 4**c** (which condense the information plotted here) correspond to the differences between the performance value (R^2 or 1-c.e.) for each model LC unit and no inactivation ('none'), divided by the maximum difference (i.e., the difference between the value for the rightmost model LC and that for 'none').



Extended Data Figure 9: Cumulative inactivation for dynamic stimulus sequences to identify necessary and sufficient model LCs.

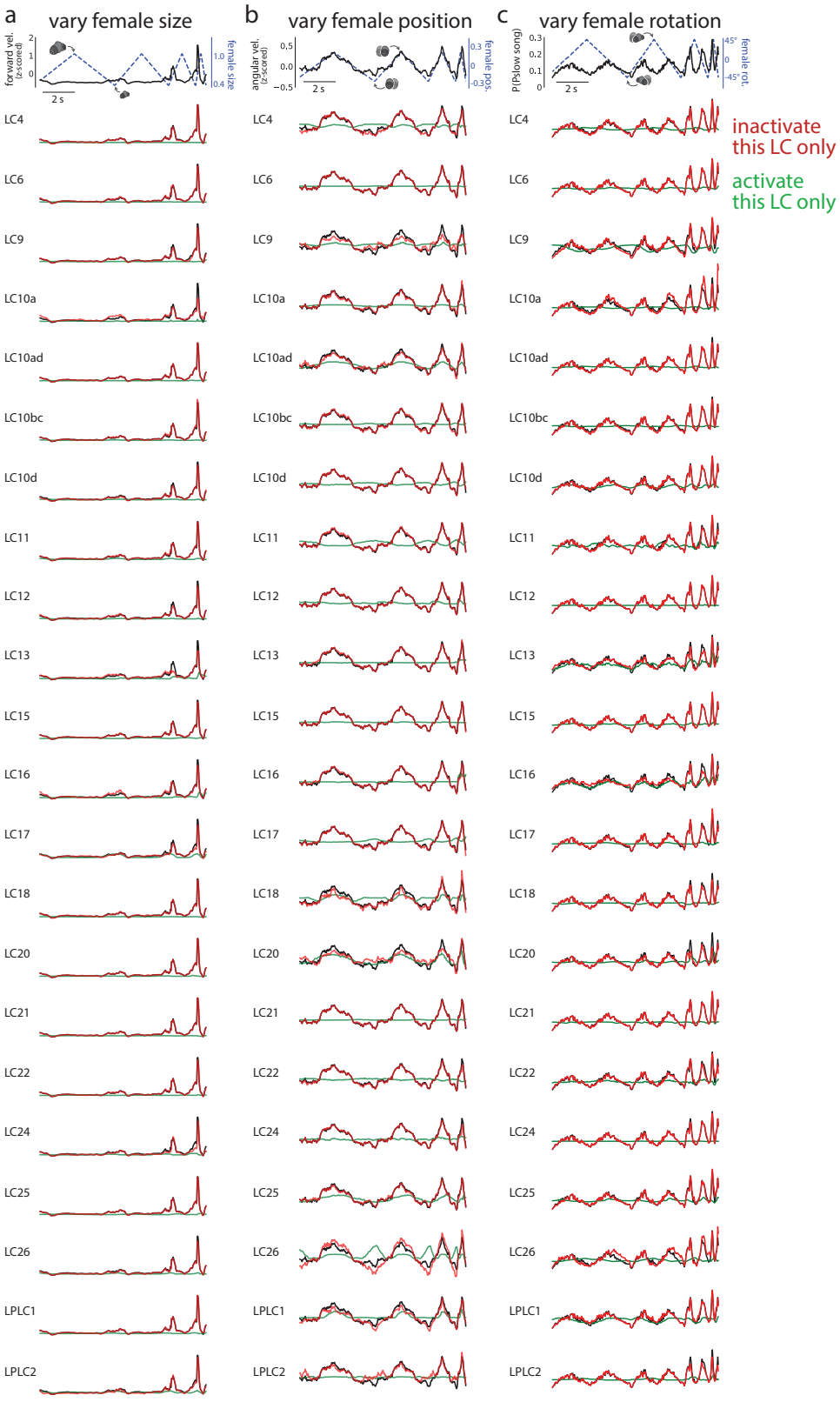
We wanted to identify which model LC neurons were necessary and sufficient for behavior in response to simple, dynamic stimulus sequences (e.g., varying female size while holding female position and rotation fixed). Here, we show the results for cumulatively inactivating each model LC unit one-by-one; the ordering was determined in a greedy manner by choosing the next model LC unit that, once inactivated, led to the least change in behavior (same process as in Fig. 4**b**). We used these results to determine which model LC units were necessary and sufficient for these behaviors as reported in Figure 4**d** and **e**.
(continued on next page...)

Extended Data Figure 9: (...continued from previous page)

a. Cumulative inactivation of model LC units for forward velocity in response to a stimulus sequence in which the fictive female only varies her size (same stimulus sequence as in Fig. 4**d**). We observed that the model LC units from LC17 to LC26 (second column, red stars) needed to be inactivated to fully eliminate forward velocity (i.e., necessity); silencing the other model LC units led to no change in behavior (left column, model LC units with green stars are sufficient).

b. Results for angular velocity in response to a stimulus sequence in which the fictive female only varies her position (same stimulus sequence as in Fig. 4 **e**, left column). We observed that the model LC units from LC20 to LC22 were necessary and sufficient (stars).

c. Results for probability of Pslow song in response to a stimulus sequence in which the fictive female only varies her rotation (same stimulus sequence as in Fig. 4 **e**, right column). We observed that the model LC units from LC16 to LC10d were necessary and sufficient (stars).

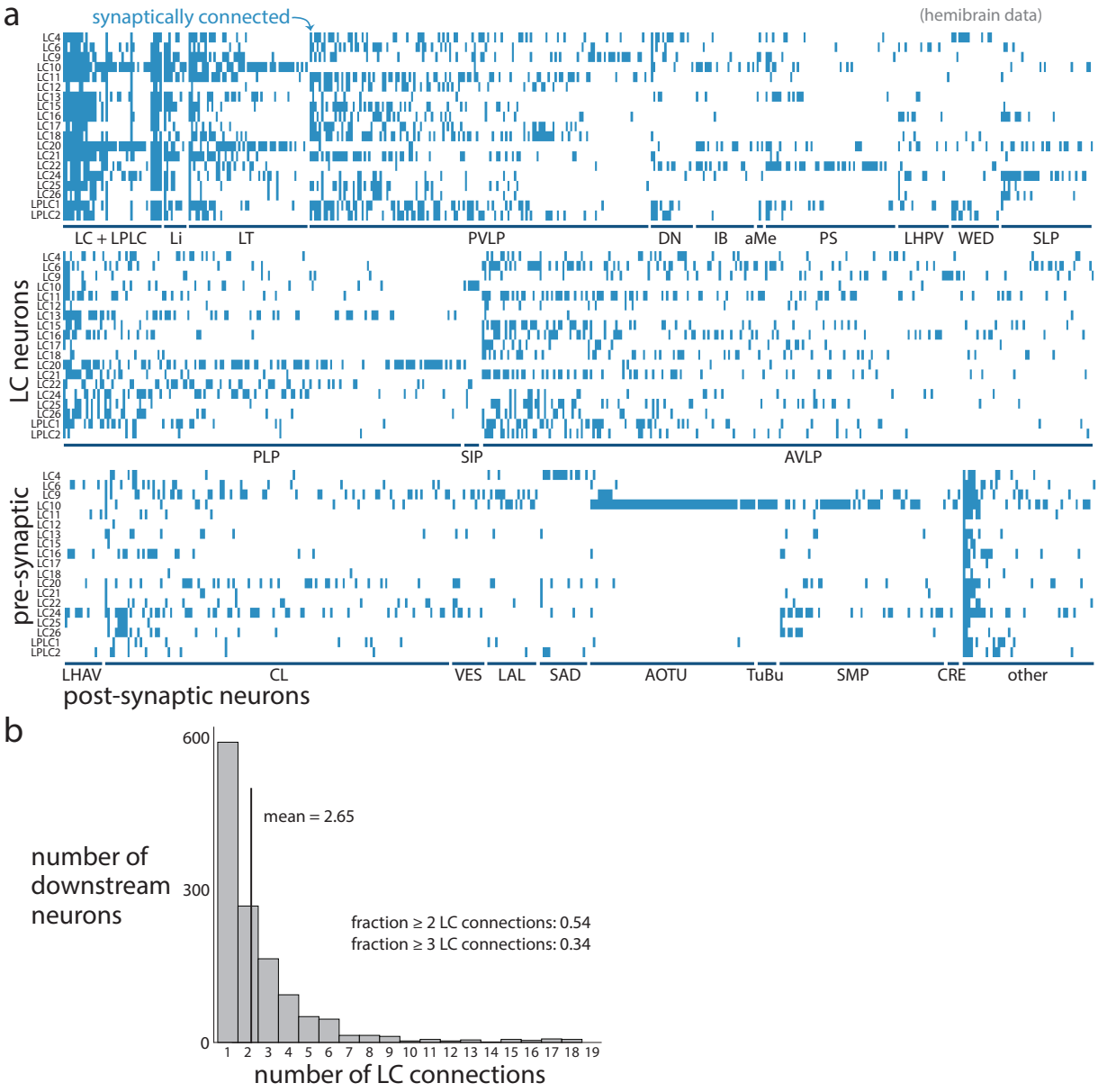


Extended Data Figure 10: "One hot" inactivation and activation of individual model LCs.

(caption continued on next page...)

Extended Data Figure 10: (...continued from previous page)

To observe how an individual model LC unit may affect behavior (top row, black traces), we either inactivated only that single model LC unit while keeping all other model LC units active (i.e., one hot inactivation; red traces) or inactivated all model LC units except that model LC unit (i.e., one hot activation; green traces). We performed this for the three different dynamic stimulus sequences (same sequences as in Fig. 4**d** and **e**). We considered the behavioral outputs of forward velocity (**a**), angular velocity (**b**), and P(Pslow song) (**c**), all of which varied with the corresponding visual parameters (top row, black traces vary with dashed lines). We observed that inactivating any single model LC unit led to almost no change in behavior (red traces overlap with black traces). Likewise, one-hot activation of any model LC unit produced little variation in behavioral output (almost all green traces are flat). We observed small but noticeable changes in behavior for model LC17 in **a**; model LC4, LC9, LC10ad, LC11, LC18, LC20, LC25, LC26, LPLC1, and LPLC2 in **b**; and model LC9, LC11, LC13, LC16, LC26, and LPLC1 in **c**. Overall, these results indicate that no single model LC unit solely contributes to one behavioral output; combinations of model LC units need to be inactivated/activated in order to see appreciable changes in behavioral output.



Extended Data Figure 11: Full synaptic connectivity matrix.

a. Full synaptic connectivity matrix between all pre-synaptic LC and LPLC neuron types (rows) and all post-synaptic neuron types (columns) identified using the Janelia hemibrain [23, 41]. Same format as in Figure 5**c**. Each blue square indicates at least 3 synaptic connections between neurons of an LC type and post-synaptic neurons of a given type (acronyms are from the Janelia hemibrain). We identified 1,298 post-synaptic neuron types (i.e., 1,298 columns in total across the three rows).

(continued on next page...)

Extended Data Figure 11: (...continued from previous page)

b. If the LC neuron types form a population code, it follows that many downstream neurons would read out from 2 or more LC types. To test this, we binned the downstream neuron types by how many different LC neuron types were pre-synaptically connected (i.e., the number of blue squares of a column in **a**). The mean of the distribution was $\mu = 2.65$ with a median of 2. We found that 54% of downstream neuron types read out 2 or more LC types; 34% read out 3 or more LC types. When we took the union of connections across neurons per downstream area (i.e., the x-axis labels in **a**), we found that all but 1 downstream area read out from 3 or more LC neuron types (TuBu neurons only read out from LC10) with a median of 11 LC/LPLC connections per downstream area. These results suggest that many downstream neurons integrate information from 2 or more LC neuron types, and each downstream area has access to visual information from many LC types. These results further support the conclusion that LC neuron types form a population code.