

Sensory and Motor Systems

A Standardized Nonvisual Behavioral Event Is Broadcasted Homogeneously across Cortical Visual Areas without Modulating Visual Responses

Mahdi Ramadan,^{1,2} Eric Kenji Lee,^{1,3} Saskia de Vries,¹ Shiella Caldejon,¹ India Kato,¹ Kate Roll,¹ Fiona Griffin,¹ Thuyanh V. Nguyen,¹ Josh Larkin,¹ Paul Rhoads,¹ Kyla Mace,¹ Ali Kriedberg,¹ Robert Howard,¹ Nathan Berbesque,¹ and  Jérôme Lecoq¹

<https://doi.org/10.1523/ENEURO.0491-21.2022>

¹Allen Institute for Brain Science, Seattle 98103, WA, ²Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge 02139, MA, and ³Department of Psychological and Brain Sciences, Boston University, Boston 02215, MA

Abstract

Multiple recent studies have shown that motor activity greatly impacts the activity of primary sensory areas like V1. Yet, the role of this motor related activity in sensory processing is still unclear. Here, we dissect how these behavior signals are broadcast to different layers and areas of the visual cortex. To do so, we leveraged a standardized and spontaneous behavioral fidget event in passively viewing mice. Importantly, this behavior event had no relevance to any ongoing task allowing us to compare its neuronal correlates with visually relevant behaviors (e.g., running). A large two-photon Ca^{2+} imaging database of neuronal responses uncovered four neural response types during fidgets that were consistent in their proportion and response patterns across all visual areas and layers of the visual cortex. Indeed, the layer and area identity could not be decoded above chance level based only on neuronal recordings. In contrast to running behavior, fidget evoked neural responses that were independent to visual processing. The broad availability of visually orthogonal standardized behavior signals could be a key component in how the cortex selects, learns and binds local sensory information with motor outputs. Contrary to behaviorally relevant motor outputs, irrelevant motor signals could project to separate local neural subspaces.

Significance Statement

Recent studies have shown contextual and behavioral variables to dominate brain-wide activity, yet it is unknown what role they play across the majority of cortex. To study their impact, we leveraged a large two-photon dataset collected in mice passively viewing a battery of visual stimuli. We characterized the neuronal response of neurons of the visual cortex to a standardized fidget behavior. We found that as much as 47% of excitatory neurons show significant co-activity with fidgets. Throughout all areas and layers we recorded from, those responses were distributed across three robust neural response types. Further analysis showed no interaction between fidget neural events and cells' visual stimulus responses, contrasting with feedback neural signals induced by running. These contrasting response distribution patterns suggest that behavioral neuronal correlates are broadly available but will modulate sensory responses depending on their relevance to local sensory inputs.

Received November 23, 2021; accepted August 23, 2022; First published September 7, 2022.

The authors declare no competing financial interests.

Author contributions: M.R. and J.L. designed research; S.C., I.K., K.R., F.G., T.V.N., J.L., P.R., K.M., A.K., R.H., N.B., and J.L. performed research; M.R., E.K.L., S.d.V., and J.L. analyzed data; M.R. and J.L. wrote the paper.

Introduction

Traditionally, the sensory cortex has been modeled as a feed-forward structure where low level information, e.g., pixel-wise visual inputs, are integrated with global behavior signals like motor output downstream of the visual cortex (Eliasmith et al., 2012). Many principles of deep learning were inspired by this view and fueled the modern rise of artificial networks. Indeed, initial reports of visual response modulation in V1 show weak modulation by behavior in monkeys (Wurtz and Mohler, 1976). This result was experimentally challenged in mice with the discovery of strong running modulation in V1 (Niell and Stryker, 2010) as well as across multiple sensory areas (Lecoq et al., 2014; Schneider et al., 2014; Ayaz et al., 2019; de Vries et al., 2020). In addition, it was demonstrated that nonvisual events are not mere modulators but can also directly evoke neuronal activity in V1 (Keller et al., 2012). In fact, contextual and behavioral variables have recently been shown to largely dominate brain-wide activity (Musall et al., 2019; Stringer et al., 2019; Guitchounts et al., 2020). This result brings into question the role of these events. If the brain broadcasts behavior relevant variables like motor outputs, this should allow each brain area to integrate this information into its computation (Schneider et al., 2014; Harris et al., 2019). As a result, understanding the microcircuit computation occurring across all cortical layers and cell types in this context requires a detailed physiological characterization of the neuronal correlates of motor outputs.

One approach to tackle this challenge is to monitor all potential behavior events and characterize all associated neuronal correlates (Petreanu et al., 2012). This is challenging since motor outputs, contrary to sensory stimuli, are highly variable from trial to trial and hard to standardize and control. In addition, behavioral events like running are correlated with a complex symphony of sensory-motor events, tangling together feedforward visual and visually-relevant feedback signals. Running behaviors are also inherently associated with visual motion, triggering visual predictions mismatch in head-fixed mice (Keller et al., 2012; Heindorf et al., 2018). Thus, to extract principles on the neuronal correlates of behavioral events, it is helpful to investigate a nonvisually relevant behavioral event. Here, our goal is to characterize how a standardized behavior output differently affects all areas, layers and cell types of the cortex to provide foundational knowledge for

This work was supported by the Allen Institute and the National Institutes of Health Training Program in Neural Computation and Engineering at the University of Washington Grant R90DA033461.

Acknowledgments: We thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support; Fuhui Long for support and guidance in training SVM; Eric Shea-Brown and Adrienne Fairhall for continuous support and encouragement; Amy Bernard for initial internship support; Doug Ollerenshaw for comments on this manuscript; and Ahmed Al Dulaimy for his coding expertise on the video annotator.

Correspondence should be addressed to Mahdi Ramadan at mramadan@mit.edu or Jérôme Lecoq at jerome@alleninstitute.org.

<https://doi.org/10.1523/ENEURO.0491-21.2022>

Copyright © 2022 Ramadan et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

modeling cortical computation. To achieve this goal, we leveraged the natural occurrence of fidgets in experimental mice.

Fidgets are stereotypical behavioral responses that are potentially part of a stressful state (Cooke et al., 2015). They manifest as a fast, spontaneous, startle response accompanied with stereotypical body chest movements. We used fidgets detected in behavioral videos to quantify and compare the neuronal correlates of standardized behavior events across all layers and most areas of the mouse visual cortex. To this end, we leveraged a large survey of neuronal responses recorded with *in vivo* two-photon calcium imaging in the mouse visual cortex (de Vries et al., 2020). While the brain-wide impact of behavioral events is now established, our analysis revealed that neurons in three different cortical layers and four visual areas have homogenous post-fidget neural responses. Fidget response profiles were stereotypical and equally distributed among four response types. In addition, in contrast to running behaviors, visually evoked neuronal responses showed no interaction with fidget neural events.

Materials and Methods

Transgenic mice

All animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at the Allen Institute for Brain Science. Triple transgenic mice (Ai93, tTA, Cre) were generated by first crossing Ai93 mice with Camk2a-tTA mice, which preferentially express tTA in fore-brain excitatory neurons. Double transgenic mice were then crossed with a Cre driver line to generate mice in which GCaMP6f expression is induced in the specific populations of neurons that express both Cre and tTA. A total of 28 males and 23 female mice were used, with each mouse exposed in multiple sessions to various visual stimuli.

Rorb-IRES2-Cre;Camk2a-tTA;Ai93 ($n = 10$) exhibit GCaMP6f in excitatory neurons in cortical layer 4 (dense patches) and layers 5, 6 (sparse). Cux2-CreERT2;Camk2a-tTA;Ai93 ($n = 16$) expression is regulated by the tamoxifen-inducible Cux2 promoter, induction of which results in Cre-mediated expression of GCaMP6f predominantly in superficial cortical layers 2, 3, and 4. Slc17a7-IRES2-Cre;Camk2a-tTA;Ai93 ($n = 2$) is a pan-excitatory line and shows expression throughout all cortical layers. Scnn1a-Tg3-Cre;Camk2a-tTA;Ai93 ($n = 5$) exhibit GCaMP6f in excitatory neurons in cortical layer 4 and in restricted areas within the cortex, in particular primary sensory cortices. Nr5a1-Cre;Camk2a-tTA;Ai93 ($n = 1$) exhibit GCaMP6f in excitatory neurons in cortical layer 4. Rbp4-Cre;Camk2a-tTA;Ai93 ($n = 11$) exhibit GCaMP6f in excitatory neurons in cortical layer 5. Ntsr1-Cre_GN220; Ai148 ($n = 1$) exhibit CaMP6f in excitatory corticothalamic neurons in cortical layer 6.

Animal head-implants and cortical window implantation

Transgenic mice expressing GCaMP6f were weaned and genotyped at \sim postnatal days (p)21, and surgery was performed between p37 and p63. Surgical protocols

were described in previous publications associated with the two-photon datasets (de Vries et al., 2020).

Intrinsic imaging and mapping of the visual cortex

Retinotopic mapping was used to delineate functionally defined visual area boundaries and enable targeting of the *in vivo* two-photon calcium imaging to retinotopically defined locations in primary and secondary visual areas. Retinotopic mapping protocols were described in previous publications associated with the two-photon datasets (de Vries et al., 2020).

In vivo two-photon chronic imaging

Calcium imaging was performed using a two-photon-imaging instrument, Nikon A1R MP+. The Nikon system was adapted to provide space to accommodate the behavior apparatus). Laser excitation was provided by a Ti:Sapphire laser (ChameleonVision – Coherent) at 910 nm. Precompensation was set at $\sim 10,000$ fs². Movies were recorded at 30 Hz using resonant scanners over a 400- μ m field of view.

Mice were head-fixed on top of a rotating disk and free to walk at will. The disk was covered with a layer of removable foam (Super-Resilient Foam, 86375K242, McMaster) to alleviate motion-induced artifacts during imaging sessions.

An experiment container consisted of three imaging sessions (60 min each) at a given field of view during which mice passively observed three different stimuli. The same location was targeted for imaging on all three recording days to allow repeat comparison of the same neurons across sessions. One imaging session was performed per day, for a maximum of 16 sessions for each mouse.

On the first day of imaging at a new field of view, the ISI targeting map was used to select spatial coordinates. A comparison of surface vasculature patterns was used to verify the appropriate location by imaging over a field of view of ~ 800 μ m using epi-fluorescence and blue light illumination. Once a cortical region was selected, the imaging objective was shrouded from stray light from the stimulus screen using opaque black tape. In two-photon imaging mode, the desired depth of imaging was set to record from a specific cortical depth. On subsequent imaging days, we returned to the same location by matching (1) the pattern of vessels in epi-fluorescence with (2) the pattern of vessels in two-photon imaging and (3) the pattern of cellular labeling in two-photon imaging at the previously recorded location.

Calcium imaging data were collected at the four cortical depths of 175, 275, 350, and 375 μ m. Throughout our analysis, data from the cortical depth of 175 μ m were classified as layer 2/3, 275 and 350 μ m as layer 4, and 375 μ m as layer 5.

More details on our imaging protocols are available in our previous publication (de Vries et al., 2020).

Visual stimulation

Visual stimuli were generated using custom scripts written in PsychoPy as described previously (de Vries et al., 2020).

Visual stimuli included drifting gratings, static gratings, locally sparse noise, natural scenes and natural movies. These stimuli were distributed across three ~ 60 -min imaging sessions. During session A, the drifting gratings, natural movie one and natural movie three stimuli were presented. During session B, the static gratings, natural scenes, and natural movie were presented. During session C, the locally sparse noise, natural movie one and natural movie two were presented. In each session, the different stimuli were presented in segments of 5–13 min and interleaved with each other. In addition, at least 5 min of spontaneous activity were recorded in each session.

Drifting gratings

The total stimulus duration was 31.5 min. The stimulus consisted of a full field drifting sinusoidal grating at a single spatial frequency (0.04 cycles/degree) and contrast (80%). The grating was presented at eight different directions (separated by 45°) and at five temporal frequencies (1, 2, 4, 8, 15 Hz). Each grating was presented for 2 s, followed by 1 s of mean luminance gray before the next grating. Each grating condition (direction and temporal frequency combination) was presented 15 times, in a random order. There were blank sweeps (i.e., mean luminance gray instead of grating) presented approximately once every 20 gratings. This stimulus was used to measure the direction tuning, orientation tuning, and temporal frequency tuning of the cells.

Static gratings

The total stimulus duration was 26 min. The stimulus consisted of a full field static sinusoidal grating at a single contrast (80%). The grating was presented at six different orientations (separated by 30°), five spatial frequencies (0.02, 0.04, 0.08, 0.16, 0.32 cycles/degree), and four phases (0, 0.25, 0.5, 0.75). The grating was presented for 0.25 s, with no intergrating gray period. Each grating condition (orientation, spatial frequency, and phase) was presented ~ 50 times in a random order. There were blank sweeps (i.e., mean luminance gray instead of grating) presented roughly once every 25 gratings. This stimulus was used to measure the spatial frequency tuning and the orientation tuning of the cells, providing a finer measurement of orientation than provided from the drifting grating stimulus.

Locally sparse noise

Stimulus. The total stimulus duration was 37.5 min. The locally sparse noise stimulus consisted of a 16×28 array of pixels, each 4.65° on a side. For each frame of the stimulus (which was presented for 0.25 s), a small number of pixels were white, a small number were black, and the rest were gray. The white and black spots were distributed such that no two spots were within five pixels of each other.

Natural scenes

The stimulus consisted of 118 natural images. Images 1–58 were from the Berkeley Segmentation Dataset, images 59–101 from the van Hateren Natural Image Dataset, and images 102–118 are from the McGill Calibrated Colour Image Database. The images were presented in grayscale

and were contrast normalized and resized to 1174×918 pixels. The images were presented for 0.25 s each, with no interimage gray period. Each image was presented ~ 50 times, in random order, and there were blank sweeps (i.e., mean luminance gray instead of an image) roughly once every 100 images.

Natural movie

Three different clips were used from the opening scene of the movie *Touch of Evil* (Welles, 1958). Natural Movie 1 and natural Movie 2 were both 30-s clips while natural Movie 3 was a 120-s clip. All clips had been contrast-normalized and were presented in grayscale at 30 fps. Each movie was presented 10 times in a row with no intertrial gray period.

Behavioral monitoring

During calcium imaging experiments, eye movements and animal posture were recorded. The left side of each mouse was imaged with the stimulation screen in the background to provide a detailed record of the animal response to all stimuli. The eye facing the stimulus monitor (right eye) was recorded using a custom IR imaging system. No pupillary reflex was evoked by any of these illumination LEDs.

Eye tracking video hardware includes a camera (Allied Vision, Mako G-032B with GigE interface) acquiring at a rate of 30 fps, with a 33-ms exposure time and gain between 10 and 20. These videos were illuminated using an LED (Engin Inc, LZ1-10R602) at a wavelength of 850 nm with a fixed lens in front (Thorlabs, LB-1092-B-ML).

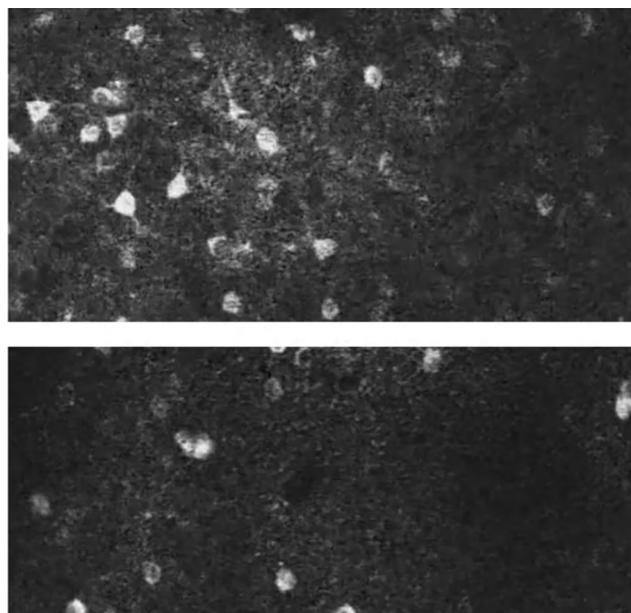
Animal behavior monitoring hardware includes a camera (Allied Vision Mako G-032B with GigE interface) with a 785-nm short pass filter (Semrock, BSP01-785R-25), and lens (Thorlabs MVL8M23, 8 mm EFL, f/1.4). The camera acquires at a rate of 30 fps, with a 33-ms exposure time and a set gain of 10. The short-pass filter is affixed to the camera to suppress any light from the eye tracking LED. Illumination for the behavior monitoring camera comes from a 740-nm LED (LED Engine Inc, LZ4, 40R308-000). A bandpass filter (747 ± 33 nm, Thorlabs, LB1092-B-ML) is affixed in front of the illumination LED to prevent visible portion of the LED spectrum from reaching the mouse eye.

Two-photon movies 512×512 pixels, 30 Hz), eye tracking (30 Hz), and a side-view full body camera (30 Hz) were recorded and continuously monitored.

Processing of two-photon calcium imaging movies

For each two-photon imaging session, the image processing pipeline performed: (1) spatial or temporal calibration, (2) motion correction, (3) image normalization to minimize confounding random variations between sessions, (4) segmentation of connected shapes, and (5) classification of soma-like shapes from remaining clutter.

The motion correction algorithm relied on phase correlation and only corrected for rigid translational errors. Each movie was partitioned into 400 consecutive frame blocks, representing 13.3 s of video. Each block was registered iteratively to its own average three times. A second



Movie 1. Example live calcium imaging of neuronal activity during fidget (fidget initiation indicated). The size of each calcium imaging frame is $400 \times 400 \mu\text{m}$. Videos are trial-averaged calcium responses during fidget [-3 (s) to $+6$ (s) relative to fidget initialization] across an imaging session. Video frames are converted to greyscale, played at $2\times$ speed and compressed as a .mp4 file. [View online]

stage of registration integrated the periodic average frames themselves into a single global average frame through six additional iterations. The global average frame served as the reference image for the final resampling of every raw frame in the video.

Fluorescence movies were processed using a segmentation algorithm to identify somatic regions of interest (ROIs) that was described previously (Niell and Stryker, 2010). Segmented ROIs were matched across imaging sessions. For each ROI, events were detected from $\Delta F/F$ by using an L0-regularized algorithm. For each neuron, we z score $\Delta F/F$ trial activity and compute the mean z-scored response of each neuron aligned to the time of fidget onset (0 s).

To determine the significance of neural activity modulation postfidget, we apply several threshold criteria to clustered neuronal activity. The threshold criteria were used in previous calcium imaging literature, and here we present two that gave very different results: one criterion where the mean $\Delta F/F$ is larger than 6%, and one criterion where the maximum $\Delta F/F$ during the postfidget period is $>5\%$.

Behavioral analysis

We use histogram of oriented gradients (HOG) descriptors as our model features because of their invariance to position, rotation, scale, and changes to lighting between mice. Additionally, HOG vectors have been shown to perform well in dynamic behavioral classification across different individuals (de Vries et al., 2020).

Side-view full-body camera (30 Hz) videos were converted to grayscale, normalized using power law compression

before processing, and manually cropped to exclude background elements including the screen and rotation disk. Cropping was done by a manual selection tool that constructed a rectangle from four clicked points. Pixels outside this rectangle were cropped out. The four points were systematically chosen in this order: upper-left as the eye of the mouse, lower-left as the closest point on the running disk forming a line to the first chosen point that is parallel to the vertical axis, lower-right as the point forming a line to the second point that is parallel to the surface of the running disk, and finally upper-right as the closest point intersecting the head-stage forming a line to the third point that is parallel to the vertical axis. HOGs were then computed for each frame using the following parameters: eight histogram orientation bins, using square cells with a height of 32 pixels and one cell per block, and then dimensionality reduced using principal component analysis (PCA; 50 dimensions). The features were then concatenated in one second blocks and fed into the model.

The training set of ~100,000 video frame labels was collected from six human annotators who used a custom XML-based tool. A radial basis function support vector machine (SVM) was trained on the training set with a cross-entropy loss function (C and γ parameters of the SVM found using a grid search).

The normalized fidget rate was computed by subtracting the baseline fidget rate during interstimulus gray stimulus presentation and dividing by the SD of the fidget rate.

To estimate fidget magnitude, an optical flow measure was computed for each grayscale cropped frame using python's OpenCV Optical flow function. We used the Gunner's Farneback algorithm using two-frame motion estimation based on polynomial expansion with a 30-pixel kernel size. For each example, the optical flow measure was integrated over the duration of all continuous frames labeled as fidget.

Fidget neuronal response analysis

Neural responses were aligned to the onset of the fidget behavior and cropped to keep 100 frames (~3 s) preceding the initiation of the fidget and 200 frames (~6 s) postfidget initiation. Neural activity was normalized on a trial-by-trial basis by subtracting the mean activity of the 100 frames (~3 s) of baseline neural activity preceding the initiation of the fidget response and dividing by the SD of activity. Across trial z-scored neural activity was then averaged to get the mean z-scored activity for each neuron.

Clustering of fidget neuronal responses

The mean z-scored activity for all neurons postfidget was passed into the k-means++ clustering algorithm with clustering evaluated using the gap statistic. The optimal number of clusters was found to be four.

Analysis of visually evoked responses in the visual cortex along with their fidget and running modulation

To compare visually evoked responses during fidget and nonfidget, we computed the mean $\Delta F/F$ trial activity during frames annotated as fidget and nonfidget. Specifically,

we only compute this metric during cells' preferred direction and temporal frequency tuning for the drifting gratings stimulus, defined as the stimulus parameters which elicit the largest mean response.

To investigate whether there is an interaction between fidget and running modulation, we computed a Fidget Cohen's d metric defined as:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s},$$

where \bar{x}_1 is the mean $\Delta F/F$ during fidget, \bar{x}_2 is the mean $\Delta F/F$ during nonfidget, and s is the norm of the SD of $\Delta F/F$ during fidget and nonfidget. We then computed the running Cohen's d metric same as above but for running versus stationary behavior. For each cell, we then plotted the fidget and running Cohen's d against each other per individual cell, and computed a covariance that indicates whether cell's that were highly modulated by fidget were also highly modulated by running.

Uniform Manifold Approximation and Projection (UMAP) and boosted decision tree classification

To search for any differences between fidget neural responses across different cortical layers and visual areas we took 200 frames (~6 s) of each neuron's normalized average postfidget activity and passed them into a nonlinear dimensionality reduction method, UMAP. In Python, we used the umap package (umap 0.4.0rc3, hyperparameters $\text{min_dist} = 0.0$ and $\text{n_neighbors} = 20$) to project the dataset into a two-dimensional embedding to elicit any differences in the data. The embedded data points were then passed into a boosted decision tree classifier. An 85–15 train-test split was used along with fivefold cross-validation with XGBoost (xgboost 1.2.0) and Scikit-Learn (sklearn 0.23.2). A grid hyperparameter search was used to optimize the classifier. To counter class imbalance in the training and test sets, each class was randomly subsampled down to the count of the least prevalent class in the set. After training, F1-scores

($F1 = \frac{TP}{TP + 0.5 * (FP + FN)}$; TP, FP, and FN are true positives, false positives, and false negatives, respectively) were estimated for each class. To find chance-level F1-score baselines, labels were randomized and the classifiers trained in the same way as before. We also verified that the similarity of neuronal responses was not because of the particular choice of dimensionality reduction method. To do this, we conducted the same analysis as described above (and now additionally including different Cre-lines), but processed the neuronal responses with UMAP at two other n_neighbors values (5 and 200), with PCA (10 components), and UMAP projected to three dimensions. As before, chance-level F1-score baselines were calculated by training the classifier on a dataset with randomized labels.

Statistical tests

Multiple comparisons were corrected using the Benjamini-Hochberg false discovery rate framework ($q < 0.05$), and all

statistical tests in the study were two-tailed, two-sample Kolmogorov–Smirnov tests.

Code accessibility

The data that support the findings of this study are available at this GitHub repository available to the public (https://github.com/mahdiramadan/Fidget_Paper), and the needed data can be downloaded from this Dryad link (https://datadryad.org/stash/share/S-QAenoJCDeDypT64bgG6Vxz6HE0bQAFPomMx6eUW_k).

Results

Fidget as a standardized behavior output

We first sought to characterize the range of behavioral events that mice displayed under head fixation. While we recorded neuronal activity *in vivo* using two-photon imaging, mice were free to run on a rotating disk while a camera captured mouse body posture (Fig. 1a–c). We observed a variety of behaviors such as whisking, grooming, mastication, flailing (uncoordinated movement), walking, running, and a startle behavior we denote as a “fidget.” Fidgets manifested as a combination of abdominal flexion (causing the abdomen to be raised above the rotating disk) and an upward force generated from the lower limbs causing lower trunk curvature and contraction (Fig. 1d). Fidgets were qualitatively stereotyped across mice in their duration, pattern of movement, and motor response magnitude. Following this observation, we sought to develop a computer vision model to automatically identify fidget events from hours of mice behavioral videos. Six human annotators first established a training dataset (20 mice, 10,000 fidgets manually annotated; see Materials and Methods). We computed the HOGs for each video frame and concatenated a feature vector from a 1-s section of frames (30 frames; Fig. 1e). HOGs are transformation invariant visual features extracted using edge-detection-like computation. HOG features are largely invariant to variation in lighting conditions and image transformations such as translation, rotation, scale. This allows us to carry out robust behavioral feature detection as others have firmly established (Dolensek et al., 2020). Another advantage of using HOG vectors is the biologically inspired emphasis on interpretable edge detection computations and has been found to be superior to Eigen-feature-based face recognition models (Turk and Pentland, 1991).

Using this feature, we trained a SVM classifier in a supervised manner using the human-annotated labels. Our final trained model had a recall performance of $74 \pm 4.2\%$ (mean \pm SD, $n=7$) and a precision performance of $78 \pm 5.3\%$ (mean \pm SD, $n=7$) for the seven 1-h-long experiments held out as a validation test set. Our trained classifier was as accurate at identifying fidgets and other mouse behaviors as human annotators. Indeed, seven pairs of annotators analyzed the same videos and their annotations were compared head-to-head. Each video was drawn from a seven-video validation test set. Head-to-head human versus human performance recall for the

seven videos was $73 \pm 5.9\%$ (mean \pm SD, $n=7$) and had a performance precision of $74 \pm 7.2\%$ (mean \pm SD, $n=7$); this was within the range of the model's performance. Having established a robust computer vision model, we automatically annotated 144 1-h experiments total (recall $p=0.51$, performance $p=0.26$).

To quantify the standardization of fidget events across mice, we integrated the optical flow magnitude of the fidget motor response (see Materials and Methods) over the duration of the fidget. 80% of all fidget events from 20 1-h experiments and across 20 unique mice fell within 30% of the maximum magnitude; this consistency reaffirmed the stereotypy of fidget events (Fig. 1f).

Occurrence of fidget across mice and visual stimuli

We next sought to establish whether the occurrence of fidgets could relate to our visual stimuli. Fidget behavior has been associated with stress and surprise responses in mice (Toth et al., 2013; Yilmaz and Meister, 2013; Cooke et al., 2015). As described in a previous publication (de Vries et al., 2020), mice passively viewed a range of both artificial (drifting and static gratings, locally sparse noise) and natural visual stimuli (natural scenes and natural movies), organized into three different recording session (sessions A, B, and C; Fig. 2a). We hypothesized that artificial visual stimuli (e.g., drifting gratings) induce a more stressful or surprising context than natural stimuli that are more ethologically familiar to the mouse (e.g., natural movies). In particular, the moving drifting gratings, through its perceived motion, evoke an innate avoidance response (Lecoq et al., 2014). In line with our prediction, the average normalized fidget rate was significantly higher during drifting gratings (Fig. 2a, $p=0.023$, two tailed t test, $n=60$) than other stimuli (Fig. 2b).

We found that fidget rate was highly variable across mice (see Extended Data Fig. 2-1), raising the possibility that various mouse Cre-lines have different stress sensitivities, thus accounting for the fidget rate variance. The absolute fidget rate did not significantly differ between mice from different Cre-lines (ANOVA, $p=0.14$, $n=144$). This result did not exclude that mice could be more sensitive to individual stimuli. To account for variability across individual mice, we normalized the change in fidget rate evoked during the session with drifting gratings (session A) by the fidget rate during session C (Fig. 2c). Similar to the absolute rate, we saw nonsignificant changes (ANOVA, $p=0.14$, $n=144$).

Previous research has shown that fidget behaviors can be learned (Cooke et al., 2015). Our passive viewing protocol included two weeks of habituation to our visual stimuli (see Materials and Methods), suggesting we could have reached a more stable state. To check whether the fidget occurrence we see is learned over the course of our two-photon experiments, we quantified the average fidget rate across mice as a function of the number of visual stimulation sessions the mouse has already seen. We found a nonsignificant change in the fidget rate with an increased number of sessions experienced (Fig. 2d), supporting the claim that we are operating in a stable

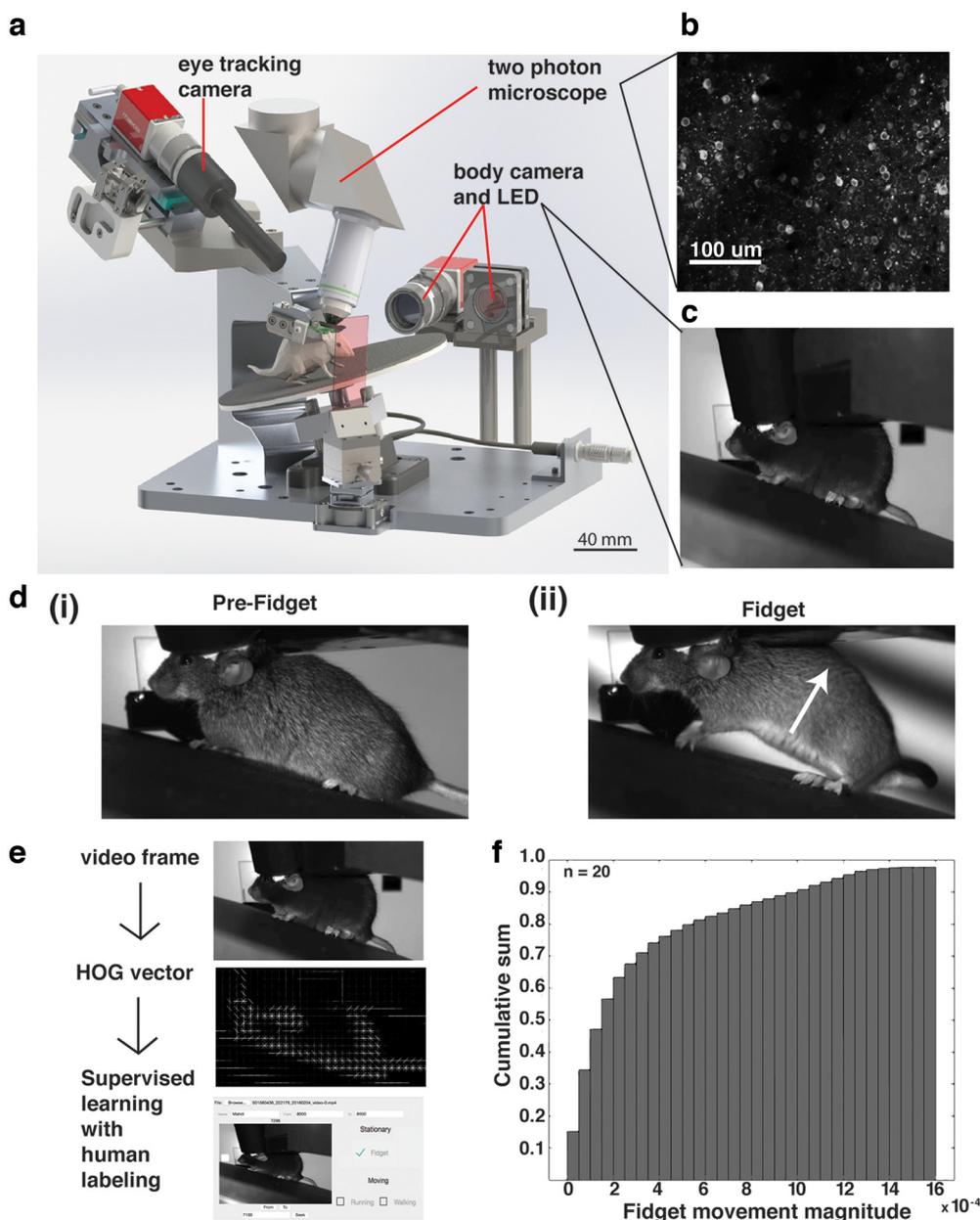


Figure 1. Detection of behavioral fidget during two-photon imaging in freely-viewing mice. **a**, Computer design of our apparatus to monitor the behavior of mice during head-fixation and two-photon imaging. **b**, Example two-photon imaging field of view ($400 \times 400 \mu\text{m}$) showcasing neurons labeled with Gcamp6f. **c**, Example video frame of mouse captured by the body camera at 30 Hz. **d**, A pair of video frames showing the progression of a prototype fidget behavior in time. **i**, First, the mouse is stationary. **ii**, Then, during the initiation of the startle response, the mouse stereotypically pushes its body up using its bottom paws while arching the back and contracting the abdomen. **e**, Computational strategy for detecting fidgets automatically. **f**, Cumulative probability sum of labeled fidget moment magnitude, showcasing the consistency of the stereotyped behavior across experiments and mice. Fidget magnitude is calculated as the sum of 2D optical flow vectors throughout the fidget duration ($n = 20$ sessions; for optical flow calculation, see Materials and Methods).

behavioral regime and that the mice are adequately habituated (ANOVA, $p = 0.21$, $n = 144$).

In summary, we saw no significant difference in the fidget rate between Cre-lines but different visual stimuli evoked different fidget rates. Importantly, there were no learning effects as mice were already habituated to the stimulus. These results allowed us to explore the neuronal correlates of fidgets across layers and areas of the visual cortex with two-photon calcium imaging.

Neuronal correlates of fidgets

We analyzed experiments where adult mice (90 ± 15 d) expressed a genetically encoded calcium sensor (GCaMP6f) under the control of specific Cre-line drivers (Rorb, Cux2, Rbp4, and Scnn1a excitatory lines; see Materials and Methods). Data were collected from four visual cortical areas (VISal, VISI, VISp, VISpm) and three different cortical layers (Layer 2/3, Layer 4, and Layer 5; 175-, 275-, and 375- μm laminar depths, respectively) based on the

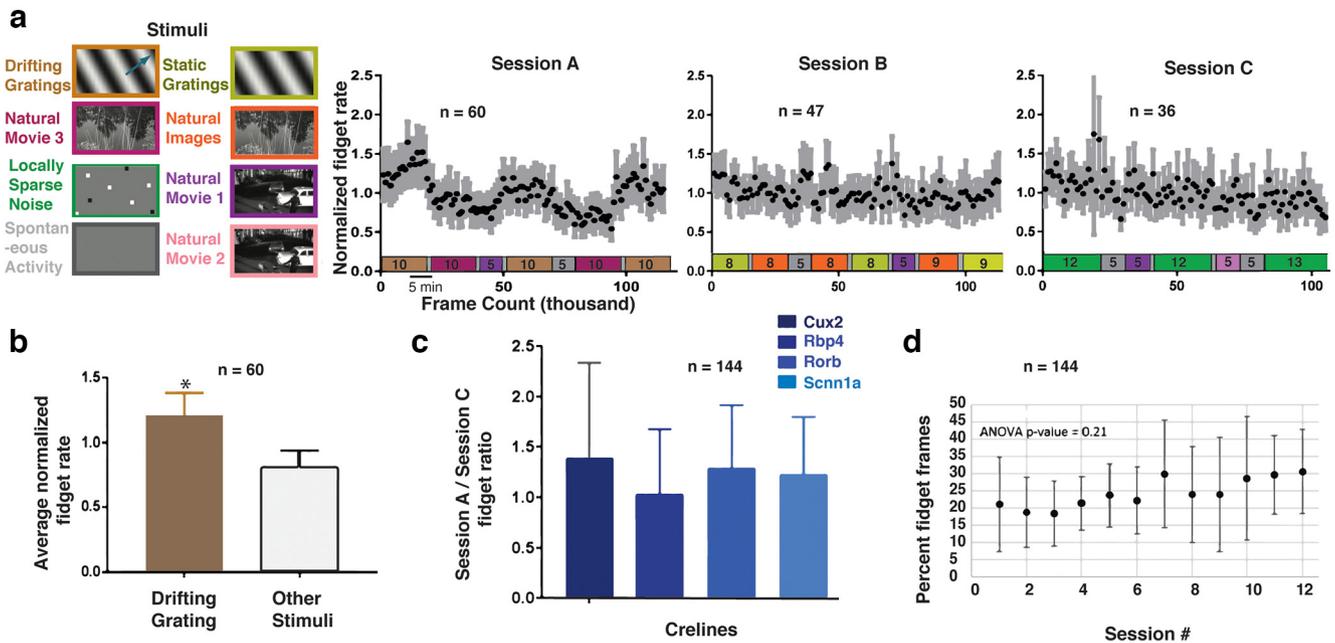


Figure 2. Fidget rate is correlated with visual stimulus type but independent of a mouse driver-line or session number. **a**, Left, Standardized experimental design of sensory visual stimuli. Six blocks of different stimuli were presented to mice and were distributed into three separate protocols identified as session A, session B, and session C (right). Normalized fidget rate (black dots) for all three session types across all mice. Gray bars indicate 95% mean confidence intervals. Color coded stimulus protocol with indicated durations (minutes) are aligned with the time axis for all three session types. *N* indicates the number of sessions. **b**, Average normalized fidget rate across all mice during the presentation of drifting grating visual stimuli in comparison to all other stimuli ($p = 0.023$). *N* indicates the number of sessions. **c**, Comparison of the fidget rate ratio between different session types across cre-lines. No significant effect was found (ANOVA, $p = 0.14$). *N* indicates the number of sessions. **d**, average percent of video frames labeled as fidget versus the number of experimental sessions a mouse has been exposed to, no significant learning effect found (ANOVA, $p = 0.21$). *N* indicates the number of sessions. Mice displayed a high variance in frequency of behaviors across sessions, but this variance was not explained by mouse cre-line (Extended Data Fig. 2-1; ANOVA, $p = 0.14$, $n = 144$). Error bars in Fig. 2b-d are standard deviation. Asterisk in **b** indicates statistical significance.

work of Niell and Stryker (2010). In total, we analyzed the activity of 20, 253 neurons imaged during 144 1-h imaging sessions. Visual responses of neurons at the retinotopic center of gaze were recorded in response to drifting gratings, flashed static gratings, locally sparse noise, natural scenes and natural movies displayed on a screen. The analysis of these visually evoked responses was published previously (Niell and Stryker, 2010). Here, we focused on the neuronal correlates associated with fidgets and their overlap with a subset of visual stimuli.

Many neurons showed a robust and prolonged response after fidget onset, in line with multiple previous studies studying the brain wide effects of motor responses (Niell and Stryker, 2010; Saleem et al., 2013; Musall et al., 2019; Steinmetz et al., 2019; Fig. 3b). We first checked that this large neural response was not because of motion artifacts caused by the fidget itself (see Extended Data Fig. 3-1). First, these calcium events were prolonged and delayed, a temporal dynamic incompatible with an immediate motion artifact. In addition, many fidgets evoked clear global events across large portions of the field of view with minimal movements of the 2p image (see Movie 1).

To investigate the structure of these seemingly global neural events, we looked at individual neuron responses across trials. Interestingly, neuronal responses fell into

four distinct types based on the direction, magnitude and durations of the response: Neutral neurons did not have activity changes from pre-fidget to post-fidget, Phasic neurons displayed a transient increase of activity right after fidget initiation, Active neurons maintained this increase throughout the post-fidget period, while Depressed neurons displayed a decrease in activity post-fidget. These response patterns were stable across trials for each neuron and displayed little deviation from the trial mean. Similar response clusters have been identified previously (Hyland, 1998; Lemke et al., 2019; Sauerbrei et al., 2020; Fig. 3c).

Neuronal fidget responses across layers and areas are uniform

Although it is now established that behaviorally related neural events are evoked in the visual cortex, we wanted to evaluate whether these neural response types were localized to a certain layer or area of visual cortex. To quantify how these response types were distributed, we first used a time series k-means clustering algorithm (see Materials and Methods). Across all 144 sessions, we found a surprisingly large portion of neurons whose activity was impacted by fidgets. 47.2% of neurons were classified as active (12%), phasic (13.9%), or depressed (21.3%), and as neutral (52.8%). The clustered activity

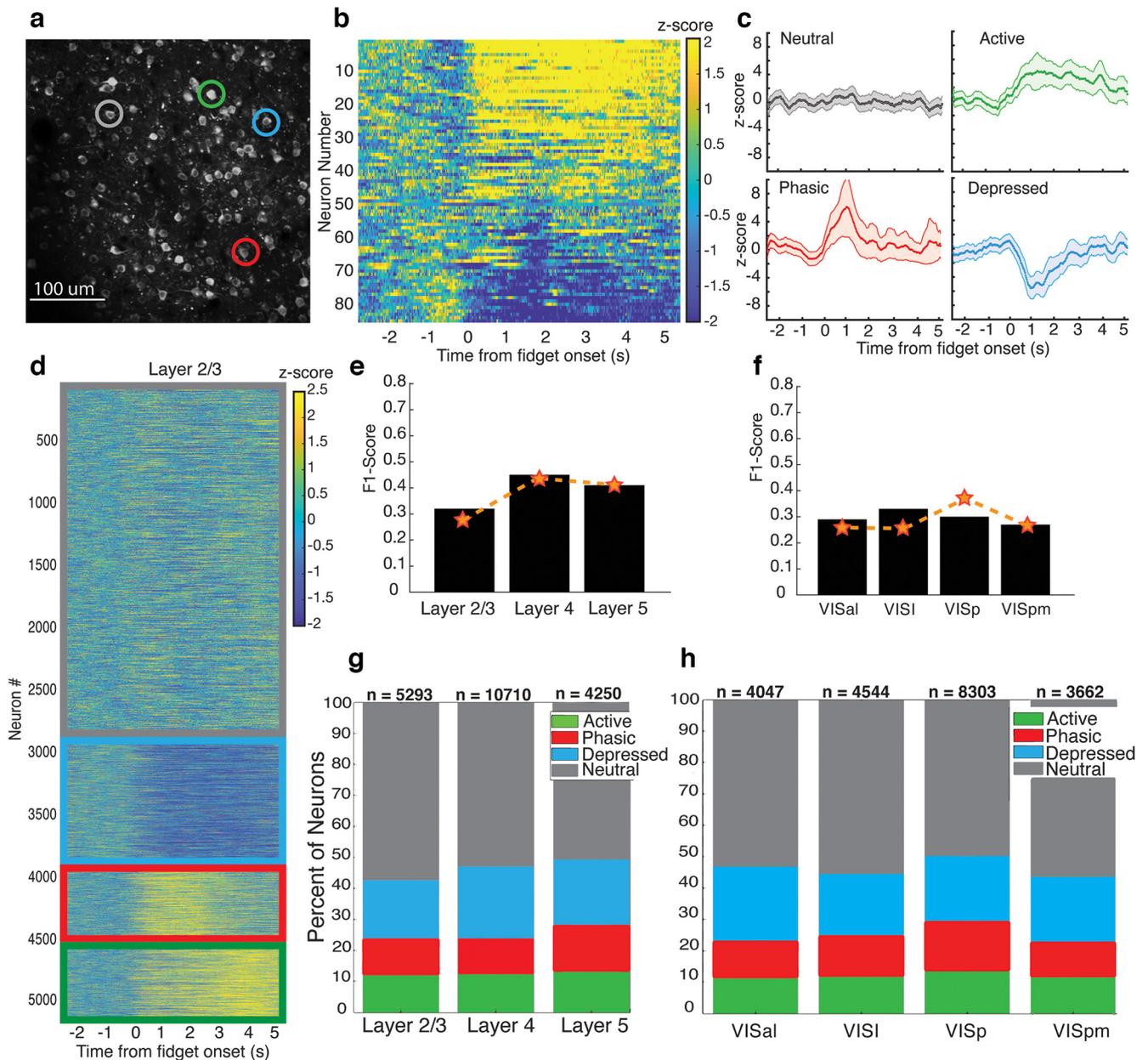


Figure 3. Neuronal fidget response types are distributed equally across visual cortical layers and areas. **a**, Example two-photon imaging field of view (Cux2, 400 × 400 μm) showcasing all neurons recorded in one session. Four unique neuronal response types are displayed by the four neurons identified by colored circles. Scale bar: 100 μm. **b**, The trial averaged z-scored activity of all neurons from one experiment, aligned to the time of fidget initiation (0 s). **c**, The activity profiles of exemplary neurons showcasing the four types of neuronal responses identified using clustering (see Materials and Methods). Top-left in gray, Neutral. Top-right in green, Active. Bottom-left in red, Phasic. Bottom-right in blue, Depressed. Colored points indicate the trial averaged z-scored activity, shaded region indicate 1 SD across trials for each time point. **d**, All layer 2/3 neurons from all experiments clustered into the four neuronal response types outlined by color coded borders (neutral, gray; depressed, blue; phasic, red; active, green). **e**, Decoding of neural response type using UMAP feature vector across cortical depths (see Materials and Methods). Results suggest that the neural response types cannot be differentiated across cortical depth (black bars: F1 score based on neural data, orange dashed line: mean F1 score based on shuffled neural data). **f**, Decoding of neural response type using UMAP feature vector across cortical areas. Results suggest that the neural response types cannot be differentiated across cortical area (black bars: F1 score based on neural data, orange dashed line: F1 score based on shuffled neural data). **g**, Percent distribution of neuronal response types per cortical layer for all cortical areas. *N* indicates the number of neurons. **h**, Percent distribution of the neuronal response types per cortical area for all cortical layers. *N* indicates the number of neurons. Fidgets did not induce additional translational motion of the cortex compared with when the mouse was at rest (Extended Data Fig. 3-1, *n* = 20 h sessions). Equal distribution of clustered neuronal response types across area and layer, and clustering homogeneity is robust to threshold criteria (Extended Data Fig. 3-2). UMAP embeddings of postfidget single neuron responses by area, layer, and mouse Cre-line (Extended Data Fig. 3-3). Lack of neuronal activity differences between Cre-line, area, and depth is not because of the method of dimensionality reduction or hyperparameter choice (Extended Data Fig. 3-4).

profiles displayed high consistency within each layer and area (Fig. 3*d*), and the clustering was not differentiated between layers or areas (Fig. 3*e,f*). Crucially, when conditioned on different layers and areas, the distribution of neural response types was surprisingly consistent, with around 47% of neurons on average being classified as active, phasic or depressed (Fig. 3*g,h*). To verify our result was not impacted by selection criteria on neurons to be clustered, we applied several different criteria derived from previous publications (Mesa et al., 2021) to our clustered neural data as a control. As expected, the percent of neurons significantly modulated in the postfidget period spans from ~8% to ~25% of all neurons depending on the strictness of the threshold criteria. Crucially, the distribution of neural response types remained consistent when conditioned on a threshold for different layers and areas (see Extended Data Fig. 3-2).

We next investigated whether layers or visual areas were differentially modulated by what could be a behaviorally relevant feedback input. To do so, we projected each 200-dimensional postfidget neutral response (each time point being a feature) into a two-dimensional subspace found using UMAP, a nonlinear dimensionality reduction technique (McInnes et al., 2018). If cortical layers or areas differed in postfidget response, we would expect to see distinct clusters of data points corresponding to each. Instead, we find that when labeled by layer or area identity, the data were mixed and could not be visually separated, this implies that postfidget responses did not differ between layer or area (see Extended Data Fig. 3-3). To validate that area or layer identity could not be distinguished based on postfidget neuronal responses with UMAP, we trained a boosted decision tree classifier based on the UMAP-projected dataset using either the layer (175, 275, or 375 μm) or area as labels (VISp, VISpm, VISal, VISl). If differences between postfidget responses for each of these layer/area classes exist, the model should achieve high classification accuracy. After training (see Materials and Methods), we found instead that the classifier had near chance-level performance for both area and layer on test data, with comparable performance to the classifier trained on randomly permuted labels (Fig. 3*e,f*). This was not because of the choice of certain parameters for UMAP (such as number of nearest neighbors) or of the choice of UMAP over other methods like PCA (see Extended Data Fig. 3-4).

Contrary to behaviors such as running, fidgets in passively viewing mice are nonrelevant to vision. We therefore hypothesized that fidget neuronal response would not affect visual responses as opposed to running modulation. To do so, we compared the response of cells to different drifting grating orientations and spatial frequencies during baseline condition with during fidget (Fig. 4*a* for examples). We found that all four subtypes of fidget responsive cell were equally responsive to visual stimuli and their sensory tuning properties to drifting gratings (direction, orientation and temporal frequency selectivity) were not significantly different (Fig. 4*b-d*, $p = 0.24, 0.30, 0.06, 0, 16$, two-sided t test, comparing the response amplitude during fidget and baseline for active, phasic, depressed, and neutral subclass, respectively). When comparing visual

stimuli during and outside of fidget behaviors, a low percent of neurons had significant visually evoked responses (5.5%, 5.0%, 2.8%, and 4.1% for active, phasic, depressed, and neutral subclass, respectively; Fig. 4*e*). This is in contrast to a visually relevant behavior such as running (Guitchounts et al., 2020), which had a larger percentage of neurons with significantly modulated visually evoked responses (12.7%, 14.2%, 11.6%, and 10.0%, for active, phasic, depressed, and neutral subclass, respectively; Fig. 4*f*), as reported previously (Niell and Stryker, 2010; de Vries et al., 2020).

Could a subset of the same cells be involved in behavior modulation in the visual cortex? To look at this, we compared the Cohen's d of fidget (see Materials and Methods) and running evoked responses across the four identified neuron response types. Cells that were running modulated ($p = 0.0009, 0.0002, 0.13, 0.0004$, two-sided t test, comparing the response amplitude during running versus stationary for active, phasic, depressed, and neutral subclass, respectively) were in fact not modulated by fidget ($p = 0.16, 0.24, 0.06, 0.30$, two-sided t test, comparing the response amplitude during running versus stationary for active, phasic, depressed, and neutral subclass, respectively; Fig. 4*g*). In fact, among cells that were significantly modulated by running, only 4% of them were also significantly modulated by fidget. In conclusion, our data support that in passively viewing conditions, a behavioral event with no visual relevance broadly impacts the visual cortex but does not modulate visual responses.

Discussion

Multiple recent studies have shown that motor activity greatly impacts the activity of not only the motor cortex but primary sensory areas like V1 (Niell and Stryker, 2010; Drew et al., 2019; Musall et al., 2019; Stringer et al., 2019; Guitchounts et al., 2020). Given the rapid advance in brain-wide neuronal recording (Kim et al., 2016; Jun et al., 2017), there has been increasing concern related to the importance of monitoring behavioral activity to account for spontaneous brain-wide neuronal activity (Steinmetz et al., 2019). Using a large two-photon Ca^{2+} imaging dataset collected in mice passively viewing a battery of standardized visual stimuli, we characterized the neuronal response of neurons of the visual cortex to fidgets, a single standardized motor output analogous to a startle response. We found that 47% of neurons show significant co-activity with fidgets throughout all areas and layers we recorded from. Previous studies in behaving mice have shown that brain-wide activity is better accounted for by uninstructed motor outputs than task driven signals (Stringer et al., 2019). Our study confirms the importance of taking into account motor activity when analyzing neuronal data.

We here propose a complementary approach to uncovering the role of motor signals in primary sensory areas. Many behavior tasks are associated with rich behavior outputs that can only be properly captured with multiple video cameras (Musall et al., 2019). Even with appropriate monitoring, the dimensionality and variability of behavioral outputs make any interpretation more challenging. By

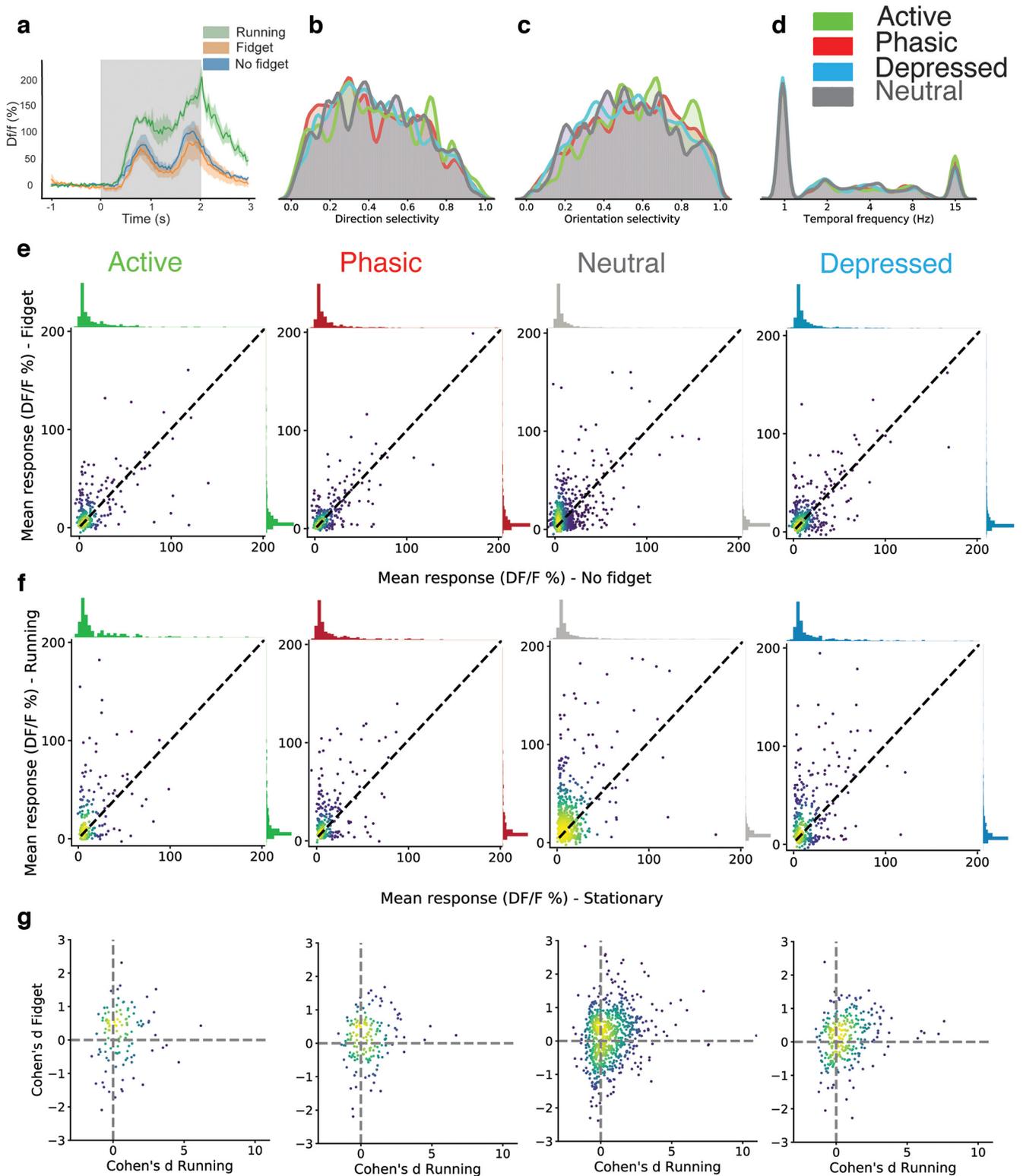


Figure 4. Contrary to running responses, neuronal fidget responses do not modulate visually evoked activity. **a**, Example visual evoked responses of a cell to drifting gratings during running, fidgets, and resting conditions, shaded error bars indicate \pm SEM. **b**, Normalized histograms of the direction selectivity of the four cells types during the drifting gratings stimulus is plotted (**c**) same as **b** but for cells' preferred drifting grating orientation (**d**) same as **b** but for cells' preferred temporal frequency of drifting gratings. **e**, The mean trial $\Delta F/F$ during fidget is plotted against nonfidget behavior for each cell during its preferred direction and temporal frequency of the drifting grating stimulus. The dotted line indicates unity. This metric is plotted separately for all four cell types. **f**, Same as **d** but for running versus stationary behaviors. **g**, The Cohen's d metric for fidget is plotted against the Cohen's d for running behavior for each cell, across all four cell types.

focusing our analysis on a single standardized behavior event, we could group together a large number of behavioral trials, akin to visually-evoked stimulus trials. Closed-loop experimental designs could prove instrumental to extend this approach, by pairing specific visual stimuli with specific tracked behavioral events. Such an approach could allow a more direct analysis of how motor output modulates visual inputs.

“Visual” fidgets were previously analyzed in relation to visual stimuli (Cooke et al., 2015). They observed an increase of behavioral visual fidget response to gratings as supported here. Contrary to this study, all of our analyses were conducted in a familiar sensory environment where all mice were habituated to stimuli. We also included all fidgets in our subsequent analysis regardless of their timing with a visual stimulus. This allowed us to compare a large number of single cell responses to either fidgets or visual stimuli, extending this previous characterization to a novel direction.

We found that excitatory neurons were responding with three distinct temporal profiles to fidgets. Remarkably the proportion and responses of neurons in each class was maintained in all layers and brain areas we looked at, and consequently we could not predict the location of our recording using the response to fidget despite a large database to train our decoder on. This result suggests that behavioral information is not only broadcasted broadly, but also homogeneously throughout the cortical mantle. This result is to contrast with running modulation which typically impacts deeper layers more strongly and is not homogeneously strong across the visual cortex (de Vries et al., 2020). Our interpretation is that local sensory inputs shape local behavioral representations depending on the causal overlap of those events. Future research could test whether a visually conditioned behavioral event transitions from a broad modulation to be more specific to certain visual areas during learning (Lillicrap et al., 2016, 2020). The broad availability of standardized behavior signals could be a key component in how the cortex selects, learns and binds local sensory information with relevant motor outputs.

References

- Ayaz A, Stäuble A, Hamada M, Wulf M-A, Saleem AB, Helmchen F (2019) Layer-specific integration of locomotion and sensory information in mouse barrel cortex. *Nat Commun* 10:1–14.
- Cooke SF, Komorowski RW, Kaplan ES, Gavornik JP, Bear MF (2015) Visual recognition memory, manifested as long-term habituation, requires synaptic plasticity in V1. *Nat Neurosci* 18:262–271.
- de Vries SEJ, et al. (2020) A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat Neurosci* 23:138–151.
- Dolensek N, Gehrlach DA, Klein AS, Gogolla N (2020) Facial expressions of emotion states and their neuronal correlates in mice. *Science* 368:89–94.
- Drew PJ, Winder AT, Zhang Q (2019) Twitches, blinks, and fidgets: important generators of ongoing neural activity. *Neuroscientist* 25:298–313.
- Eliasmith C, Stewart TC, Choo X, Bekolay T, DeWolf T, Tang Y, Tang C, Rasmussen D (2012) A large-scale model of the functioning brain. *Science* 338:1202–1205.
- Guitchoyants G, Masis J, Wolff SBE, Cox D (2020) Encoding of 3D head orienting movements in the primary visual cortex. *Neuron* 108:512–525.e4.
- Harris JA, et al. (2019) Hierarchical organization of cortical and thalamic connectivity. *Nature* 575:195–202.
- Heindorf M, Arber S, Keller GB (2018) Mouse motor cortex coordinates the behavioral response to unpredicted sensory feedback. *Neuron* 99:1040–1054.e5.
- Hyland B (1998) Neural activity related to reaching and grasping in rostral and caudal regions of rat motor cortex. *Behav Brain Res* 94:255–269.
- Jun JJ, et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551:232–236.
- Keller GB, Bonhoeffer T, Hübener M (2012) Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* 74:809–815.
- Kim TH, Zhang Y, Lecoq J, Jung JC, Li J, Zeng H, Niell CM, Schnitzer MJ (2016) Long-term optical access to an estimated one million neurons in the live mouse cortex. *Cell Reports* 17:3385–3394.
- Lecoq J, Savall J, Vučinić D, Grewe BF, Kim H, Li JZ, Kitch LJ, Schnitzer MJ (2014) Visualizing mammalian brain area interactions by dual-axis two-photon calcium imaging. *Nat Neurosci* 17:1825–1829.
- Lemke SM, Ramanathan DS, Guo L, Won SJ, Ganguly K (2019) Emergent modular neural control drives coordinated motor actions. *Nat Neurosci* 22:1122–1131.
- Lillicrap TP, Cownden D, Tweed DB, Akerman CJ (2016) Random synaptic feedback weights support error backpropagation for deep learning. *Nat Commun* 7:13276.
- Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G (2020) Backpropagation and the brain. *Nat Rev Neurosci* 21:335–346.
- McInnes L, Healy J, Melville J (2018) UMAP: uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426.
- Mesa N, Waters J, de Vries SEJ (2021) The effect of inclusion criteria on the functional properties reported in mouse visual cortex. *eNeuro* 8:ENEURO.0188-20.2021.
- Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK (2019) Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci* 22:1677–1686.
- Niell CM, Stryker MP (2010) Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65:472–479.
- Petreaanu L, Gutnisky DA, Huber D, Xu NL, O’Connor DH, Tian L, Looger L, Svoboda K (2012) Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature* 489:299–303.
- Saleem AB, Ayaz A, Jeffery KJ, Harris KD, Carandini M (2013) Integration of visual motion and locomotion in mouse visual cortex. *Nat Neurosci* 16:1864–1869.
- Sauerbrei BA, Guo JZ, Cohen JD, Mischiati M, Guo W, Kabra M, Verma N, Mensh B, Branson K, Hantman AW (2020) Cortical pattern generation during dexterous movement is input-driven. *Nature* 577:386–391.
- Schneider DM, Nelson A, Mooney R (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* 513:189–194.
- Steinmetz NA, Zátka-Haas P, Carandini M, Harris KD (2019) Distributed coding of choice, action and engagement across the mouse brain. *Nature* 576:266–273.
- Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364:eaav7893.
- Toth M, Ziegler M, Sun P, Gresack J, Risbrough V (2013) Impaired conditioned fear response and startle reactivity in epinephrine-deficient mice. *Behav Pharmacol* 24:1–9.
- Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3:71–86.
- Wurtz RH, Mohler CW (1976) Enhancement of visual responses in monkey striate cortex and frontal eye fields. *J Neurophysiol* 39:766–772.
- Yilmaz M, Meister M (2013) Rapid innate defensive responses of mice to looming visual stimuli. *Curr Biol* 23:2011–2015.