



Published in final edited form as:

*Nat Neurosci.* 2019 October ; 22(10): 1677–1686. doi:10.1038/s41593-019-0502-4.

## Single-trial neural dynamics are dominated by richly varied movements

Simon Musall<sup>1,4</sup>, Matthew T. Kaufman<sup>1,2,4</sup>, Ashley L. Juavinett<sup>1,3</sup>, Steven Gluf<sup>1</sup>, Anne K. Churchland<sup>1</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Neuroscience, Cold Spring Harbor, NY

<sup>2</sup>Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL

<sup>3</sup>Division of Biological Sciences, UC San Diego, San Diego, CA

<sup>4</sup>These authors contributed equally to this work.

### Abstract

When experts are immersed in a task, do their brains prioritize task-related activity? Most efforts to understand neural activity during well-learned tasks focus on cognitive computations and task-related movements. We wondered whether task-performing animals explore a broader movement landscape, and how this impacts neural activity. We characterized movements using video and other sensors and measured neural activity using widefield and two-photon imaging. Cortex-wide activity was dominated by movements, especially uninstructed movements not required for the task. Some uninstructed movements were aligned to trial events. Accounting for them revealed that neurons with similar trial-averaged activity often reflected utterly different combinations of cognitive and movement variables. Other movements occurred idiosyncratically, accounting for trial-by-trial fluctuations that are often considered “noise”. This held true throughout task-learning and for extracellular Neuropixels recordings that included subcortical areas. Our observations argue that animals execute expert decisions while performing richly varied, uninstructed movements that profoundly shape neural activity.

### Introduction

Cognitive functions, such as perception, attention, and decision-making, are often studied in the context of movements. This is because most cognitive processes will naturally lead to action: pondering where to go or with whom to interact both ultimately lead to movement. Laboratory studies of cognition therefore often rely on the careful quantification of

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Corresponding author: Anne K Churchland; churchland@cshl.edu.

Author contributions

S.M., M.T.K. and A.K.C. designed the experiments. S.M. and S.G. trained animals and recorded widefield data. S.M. performed surgeries. M.T.K. and S.M. acquired 2-photon data, designed the linear model and performed data analysis. A.L.J. recorded and spike-sorted Neuropixels data. A.K.C., M.T.K and S.M. wrote the paper with assistance from S.G. and A.L.J.. S.M. and M.T.K. contributed equally.

Competing financial interests

The authors declare no competing interests.

instructed movements (e.g., key press, saccade or spout lick), to track outcomes of cognitive computations. However, studying cognition in the context of movements leads to a well-known challenge: untangling neural activity that is specific to cognition from neural activity related to the instructed reporting movements.

To address this problem, researchers often isolate the time of decision from the time of movement using a delay period<sup>1,2</sup>, or prevent the subject from knowing which movement to use until late in the trial<sup>3,4</sup>. A second strategy is to account for the extent to which putative cognitive activity is modulated by metrics of the instructed movement: for instance, assessing how decision-related activity co-varies with saccade parameters<sup>5,6</sup> or orienting movements<sup>7</sup>. However, instructed movements are only a subset of possible movements that may modulate neural activity.

In the absence of a behavioral task, uninstructed movements, such as running on a wheel, can drive considerable neural activity even in sensory areas<sup>8,9</sup>. Despite this, other uninstructed movements, such as hindlimb flexions during a lick task, are usually not considered when analyzing neural data. This exposes two implicit assumptions (Fig. 1A). The first assumption is that uninstructed movements have a negligible impact on neural activity compared to task-related activity or instructed movements. The second assumption is that uninstructed movements occur infrequently and at random times, while instructed movements are task-aligned, occurring at stereotyped moments on each trial. In this case, uninstructed movements would increase neural variability trial-by-trial but their effects could be removed by trial averaging. Both assumptions are largely untested, however, in part due to the difficulty in accurately measuring multiple movement types and systematically relating them to neural activity.

Revealing the impact of uninstructed movements on neural activity is critical for behavioral experiments: movements may account for considerable trial-by-trial variance and also mimic or overshadow the cognitive processes that are commonly studied in trial-averaged data. It is also unclear whether movements may drive substantial activity in specific brain areas but have negligible effects in others. We therefore leveraged multiple sensors and dual video recordings to track a wide array of movements and measure their impact on neural activity across the dorsal cortex as well as subcortical structures. A linear encoding model demonstrated that neural variability was dominated by movements, with uninstructed movements outpacing the predictive power of instructed movements and task variables. This movement dominance persisted throughout task learning. Both instructed and uninstructed movements also accounted for a large degree of task-aligned activity that was still present after trial-averaging. We then separated out movement-related activity, thereby recovering task-related dynamics that were otherwise obscured. Taken together these results argue that during cognition, animals engage in a diverse array of uninstructed movements, with profound consequences for neural activity.

## Results

### Cortex-wide imaging during auditory and visual decision-making

We trained head-fixed mice to report the spatial position of 0.6-s long sequences of auditory click sounds or visual LED stimuli. Animals grabbed handles to initiate trials, then stimuli were presented at randomized times after the handle grab (Fig. 1B). Each stimulus was presented twice and after a 1s delay, animals reported a decision and received a water reward for licking the spout corresponding to the stimulus presentation side (Fig. 1C). Animals were trained to expert performance on either auditory or visual stimuli (Fig. 1D).

We measured cortex-wide neural dynamics using widefield calcium imaging<sup>10</sup>. Mice were transgenic, expressing the Ca<sup>2+</sup>-indicator GCaMP6f in excitatory neurons. Fluorescence was measured through the cleared, intact skull<sup>11</sup> (Fig. 1E). Imaging data were then aligned to the Allen Mouse Common Coordinate Framework v3 (CCF, Supplementary Fig. 1, 2)<sup>12</sup>.

Baseline-corrected fluorescence ( $\Delta F/F$ ) revealed clear modulation of neural activity across dorsal cortex (Fig. 1F, Video S1; average response to visual trials, 22 sessions from 11 mice). During trial initiation (handle grab), cortical activity was strongest in sensorimotor areas for hind- and forepaw ('Hold'). The first visual stimulus caused robust activation of visual areas and weaker responses in medial secondary motor cortex (M2) ('Stim 1'). Activity in anterior cortex increased during the second stimulus presentation ('Stim 2') and the delay ('Delay'). During the response period, neural activity strongly increased throughout dorsal cortex ('Response'). A comparison of neural activity across conditions confirmed that neural activity was modulated by whether the stimulus was visual or auditory (Fig. 1G) and presented on the left or the right (Fig. 1H). Differences across conditions were mainly restricted to primary and secondary visual areas and the posteromedial part of M2.

### Movements dominate cortical activity

To assess the impact of movements, we built a linear encoding model<sup>13,14</sup> that could take into account a large array of instructed and uninstructed movements alongside task-variables to predict changes in single-trial cortex-wide activity. To reduce computational cost and prevent overfitting, we used singular value decomposition (SVD) on the imaging data and ridge regression to fit the model to data components. The model's design matrix included two types of variables: discrete behavioral events and continuous (analog) variables. Discrete behavioral events included task-related features, such as the animal's choice and the time of sensory stimulus onset (Fig. 2A). Other events were movement-related, either instructed (e.g., licking) or uninstructed (e.g., whisking, Fig. 2B,C). For each event, ridge regression produced a time-varying event kernel, relating that variable to neural activity (Fig. 2C). Event kernels flexibly captured delayed or multiphasic responses (Fig. 2D). Similar approaches are often used for motion-correction in neuroimaging<sup>15</sup> or fitting complex neural responses to sensory or motor events during cognitive tasks<sup>13,14</sup>. The model included analog variables for continuous signals like pupil diameter (Fig. 2B) and the 200 highest-variance dimensions of video data from the animal's face and body<sup>16</sup>. For each analog variable, ridge regression produced a single scaling weight, making these variables useful in capturing

neural responses that scale linearly with a given variable (Fig. 2E line 1). However, analog variables cannot account for delayed or multiphasic responses (Fig. 2E lines 2-4).

To evaluate how well the model captured neural activity at different cortical locations, we computed the 10-fold cross-validated  $R^2$  ( $cvR^2$ ). Cross-validation ensures that no variance is predicted by chance, regardless of the number of model variables (shuffled control:  $cvR^2 < 10^{-6}\%$  in all recordings). The upper bound for the  $cvR^2$  is 100%, because SVD effectively removed single-pixel noise from the imaging data. This ceiling was confirmed in each recording by peer-prediction for 500 randomly-selected pixels, predicting  $\sim 100\%$  of the variance in the remaining pixels. While some areas were particularly well predicted in specific trial epochs (e.g., V1 during stimulus presentation), there was high predictive power throughout the cortex during all epochs of the trial (Fig. 3A). For all data ('Whole trial'), the model predicted  $41.2 \pm 0.9\%$  (mean  $\pm$  SEM,  $n=22$  sessions) of all variance across cortex. Explained variance was even higher at lower sampling rates, indicating that the model mostly predicted slower fluctuations in neural activity instead of frame-to-frame changes at 30 Hz (Supplementary Fig. 3).

Spatiotemporal dynamics of the event kernels sensibly matched known roles of sensory and motor cortices. For example, pixels in left primary visual cortex (V1) were highly positive in response to a rightward visual stimulus (Fig. 3B) with temporal dynamics that matched stimulus responses when averaging over visual trials (Fig. 1H). Left somatosensory (FL) and primary motor forelimb areas were activated during right handle grabs (Fig. 3C) and bi-phasic responses in the olfactory bulb (OB) were associated with nose movements (Fig. 3D). In contrast, changes in pupil size were associated with changes in numerous cortical areas, resembling previously observed effects of arousal<sup>17</sup> (Fig. 3E). To assess the contribution of video data, we performed principal component analysis (PCA) on the model's weights for all video variables (Fig. 3F). In many sessions, video weights were at least 8-dimensional, indicating that the video contained different movement patterns that accounted for separate neural responses. To reveal the diverse relationship between neural responses and video, the first six dimensions of the cortical response weights are shown for an example session, after sparsening (Fig. 3G). When projecting model weights onto video pixels (Fig. 3H), several specific areas of the animal's face were particularly important, especially around the animal's nose, eye, and jaw. This indicates that video dimensions contributed additional information to the model that was not contained in other movement variables.

We next sought to address which variables were most important for the model's success. The simplest way to gauge importance is to fit a single-variable model and ask how well it predicts the data compared to the full model (Fig. 4A, light green versus both circles combined). However, the lack of spatial precision in  $cvR^2$  maps for different single-variable models (Fig. 4B, top row) suggested that there might be overlap in the explanatory power of some model variables that contain related information (Fig. 4A, light green inside the black circle). For instance, right visual stimulation also predicted neural activity in the somatosensory mouth area and parts of anterior cortex, likely because visual stimulation is followed by animal movements in the response period. Similarly,  $cvR^2$  maps for the right handle and nose were almost identical, indicating that both are associated with other movements (Fig. 4B, top right). Single-variable models are therefore an upper bound for the

linear information within a given variable. However, the same variable may contribute little to the full model's overall predictive power if it is largely redundant with other model variables. To measure each variable's unique contribution to the model, we created reduced models in which we temporally shuffled a particular model variable. The resulting loss of predictive power ( $R^2$ ) relative to the full model provides a lower bound for the amount of unique information (non-redundant to other model variables) that the given variable contributed to the model (Fig. 4A, dark green crescent). In contrast to single-variable  $cvR^2$ ,  $R^2$  maps (Fig. 4B, bottom row) were highly spatially localized and matched brain areas that were also most modulated in their corresponding event kernels (Fig. 3). This was consistent in both visual and auditory experts (Supplementary Fig. 4). Control recordings, from animals expressing GFP instead of a GCaMP indicator, confirmed that this was not explained by hemodynamic signals or potential motion artifacts (Supplementary Fig. 5).

We then compared  $cvR^2$  and  $R^2$  values (averaged over all pixels) for all model variables. Many variables individually predicted a large amount of neural variance, and movement variables contained particularly high predictive power compared to task variables (Fig. 4C, light green). Video ('Video') and video motion energy ('Video ME') variables had the most predictive power, each explaining ~23% of all variance. Differences were even more striking for  $R^2$  values, which were particularly low for task variables (Fig. 4C, dark green). Consider the 'time' variable, which captures activity at consistent times in each trial (similar to an average over all trials). A 'time'-only model captured considerable variance (light green), but eliminating this variable had a negligible effect on the model's predictive power. This is because other task variables could capture time-varying modulation equally well. In contrast, movement variables made much larger unique contributions.

To compare the importance of all instructed or uninstructed movements (Fig. 1A, columns) relative to task variables, we repeated the analysis above on groups of variables (Fig. 4D). Both task and instructed movement groups contained similar predictive power ( $cvR^2_{\text{Task}}=17.6\pm0.6\%$ ;  $cvR^2_{\text{Instructed}}=17.3\pm0.6\%$ ) and unique contributions ( $R^2_{\text{Task}}=2.9\pm0.2\%$ ;  $R^2_{\text{Instructed}}=3.6\pm0.2\%$ ). Uninstructed movements, however, were more than twice as predictive ( $cvR^2_{\text{Uninstructed}}=38.3\pm0.9\%$ ), with a 5-fold higher unique contribution ( $R^2_{\text{Uninstructed}}=17.8\pm0.6\%$ ). To assess the spatial extent of this large difference, we created pixel-wise  $R^2$  maps for every group (Fig. 4E). Throughout cortex, unique contributions of uninstructed movements were much higher than either instructed movements or the task group. The contribution of uninstructed movements was particularly prominent in anterior somatosensory and motor cortices, but also clearly evident in posterior areas such as retrosplenial cortex and even V1.

### Controls for the importance of uninstructed movements

These results strongly suggest that uninstructed movements are critical for predicting cortical activity. We considered four possible confounds. First, the task might be too simple to engage cortex. We therefore tested animals in a more challenging auditory rate discrimination task, known to require parietal and frontal cortices<sup>18,19</sup>. Results were very similar, arguing that our results were not due to insufficient task complexity (Supplementary Fig. 6). Second, including many analog variables may have bolstered the predictive power of

uninstructed movements; indeed, the high  $R^2$  for video shows that analog variables are particularly effective in unsupervised movement detection. However, in a model that excluded all analog variables the most predictive group remained uninstructed movements, demonstrating that their importance does not hinge on using analog variables (Supplementary Fig. 7). Third, we ensured that our results were not affected by potential epileptiform activity that has been observed in the transgenic mouse line used here<sup>20</sup> (Supplementary Fig. 8 and Methods).

Lastly, uninstructed movements might be important because they simply provide additional information about instructed movements (e.g., a preparatory body movement before licking) instead of independent, self-generated actions. We therefore computed  $R^2$  for each group at every time point in the trial (Fig. 4F, Video S2). Here,  $R^2_{\text{Uninstructed}}$  was highest in the baseline period when  $R^2_{\text{Instructed}}$  was low (black and blue traces). Conversely,

$R^2_{\text{Uninstructed}}$  was reduced at times when  $R^2_{\text{Instructed}}$  was high, arguing against a tight relation between instructed and uninstructed movements. However, some uninstructed movements may still be related to instructed behavior, such as facial movements during licking or changes in posture when grabbing the handles.

### The impact of movements on single-trial and trial-averaged neural activity

We next asked whether movements were aligned to specific times in the task (Fig. 1A, rows). The task model captures all consistent neural responses to task-related events, similar to linearly combining all possible peri-event time histograms (PETHs) for different task events. All variance that is not captured by the task-only model (~83% across all recordings, Fig. 4D) must therefore be due to trial-by-trial variability (Fig. 5A). Accordingly, the task model predicted trial-averaged data almost perfectly (Fig. 5B, left) but not variability in overall single-trial activity relative to the average (Fig. 5C, left). Results for instructed movements were largely similar, confirming that they are mostly aligned with the task. In contrast, the uninstructed movement model predicted not only a large amount of trial-by-trial variability, but also trial-averaged data, indicating that it contained a high degree of task-aligned as well as task-independent information.

To determine the task-aligned and task-independent contributions for each movement variable (Fig. 1A), we computed the increase in explained variance of the task-only model when a given movement variable was added (Fig. 5D). This is conceptually similar to computing unique contributions as described above but here indicates the task-independent contribution (dark blue portion) of a given movement variable. Subtracting this task-independent component from the overall explained variance of each single-movement-variable model (complete blue circle in Fig. 5D, light green bars in Fig. 4C) yields the task-aligned contribution (light blue fill). As expected, instructed movements had a considerable task-aligned contribution, especially the lick variable (Fig. 5E, light blue bars, left). Most uninstructed movement variables also made considerable task-aligned contributions, but in addition made large task-independent contributions (Fig. 5E, right). This was even clearer after pooling variables into groups (Fig. 5F). Here, uninstructed movements contained more task-aligned ( $\text{cv}R^2_{\text{Instructed}}=11.0\pm 0.5\%$ ;  $\text{cv}R^2_{\text{Uninstructed}}=14.5\pm 0.5\%$ ) and task-independent ( $R^2_{\text{Instructed}}=6.3\pm 0.3\%$ ;  $R^2_{\text{Uninstructed}}=23.8\pm 0.7\%$ ) explained variance than instructed

movements. For both movement groups, the amount of task-aligned information was substantial: instructed and uninstructed movements captured ~63% and ~83% of all information in the task model, respectively.

These results have important implications for interpreting both trial-averaged and single-trial data (Fig. 5G). First, although a task-only model accurately predicted the PETH, it cannot account for most single-trial fluctuations (Fig. 5G, left column). Instructed movements also predicted parts of the PETH, in particular at times when instructed movements were required (black dashed lines, middle top). As with the task model, the instructed movement model failed to predict most single-trial fluctuations (middle, bottom 3 rows). In contrast, an uninstructed movement model also predicted the PETH with good accuracy and was the only one to also accurately predict neural activity on single trials (right column). This demonstrates that much of the single-trial activity that is often assumed to be noise or reflect random network fluctuations is instead related to uninstructed movements.

Importantly, because many movements are task-aligned, some apparent task features could also be explained by uninstructed movements (Fig. 5G, red dashed arrows). We therefore developed an approach to “partition” the PETH into respective contributions from the task, instructed and uninstructed movements, based on how much variance each group accounted for (Fig. 5H). First, we fitted the full model to the imaging data (gray trace), then split the full model prediction (red trace) into three components based on the weights for each group, without re-fitting. This provides the best estimate of the relative contributions of the task (green traces), instructed movements (blue traces) and uninstructed movements (black traces) to the PETH while explaining the most neural variance. In M2, PETH partitioning revealed that diverse trial-averaged dynamics, e.g., during the delay (bottom right, red dashed arrow), are best accounted for by uninstructed movements. Conversely, PETH partitioning in V1 showed that most features were reassuringly well-explained by the task as they are most likely due to visual responses (Fig. 5I). This analysis therefore allows a better interpretation of trial-averaged PETHs by using trial-by-trial data to separate task-related activity from the impact of task-aligned movements.

Taken together, these results highlight that most of the activity without trial-averaging is explained by uninstructed movements, which is critical to interpret trial-by-trial variability. Furthermore, uninstructed movements can also occur at predictable time points and significantly impact the shape of a PETH, sometimes masquerading as activity related to instructed movements or cognitive variables. This reveals a conundrum that initially seemed difficult to resolve: many PETH features can be explained by either task or movement variables alone (Fig. 5G top row). Our approach offers insight into this issue: exploiting trial-by-trial variability to partition the PETH into multiple components that jointly recreate the complete shape (Fig. 5H,I).

### Changes in task and movement contributions during task learning

To assess how the relationship between task and movement variables develops over learning, we analyzed behavioral and neural data from the 10-week training period from novice to fully-trained auditory experts. The  $cvR^2$  of both the instructed movement model (Fig. 6A, blue) and the task model (green) strongly increased over the first 20 training days, indicating

that neural activity became increasingly aligned with the temporal structure of the task. In contrast, the  $cvR^2$  of an uninstructed movement model decreased modestly with training and eventually reached a stable level (black lines).

While task and instructed movement models were increasingly predictive of neural activity in anterior regions, uninstructed movements lost predictive power in posterior regions (Fig. 6B,C). This reduction was not explained by reduced neural responsiveness to movements: model weights for uninstructed movements were constant over the entire training duration (Fig. 6D), regardless of cortical location (Fig. 6E). Instead, the decrease in predictive power was explained by two specific changes in uninstructed movements: First, the rate of uninstructed movements decreased over time (Fig. 6F, bottom), which was reflected in reduced trial-by-trial variability of the neural data (Fig. 6G). Second, uninstructed movements became increasingly task-aligned and less variable across trials (Fig. 6H, I). Task-aligned  $cvR^2$  of uninstructed movements mostly increased in anterior cortex while task-independent  $cvR^2$  was reduced throughout the cortex (Fig. 6J, right column). Importantly, uninstructed movements were consistently the strongest predictor of neural activity and task-independent contributions plateaued at high levels even after up to 50 days of training. These results suggest that animals express a richly varied array of movements during all training stages, but their movement landscape evolves as they transition from novice to expert status.

### The impact of movements on single neuron activity

Widefield imaging reflects the pooled activity over many neurons, especially those in superficial layers<sup>12</sup> and is affected by dendritic and axonal projections<sup>21</sup>. Although we observed substantial area-level specificity (Figs. 3,4B), we further extended our approach to single neurons. Using 2-photon imaging, we measured the activity of over 13,000 single neurons in 10 animals expressing either GCaMP6f (7 mice) or GCaMP6s (3 mice) at depths spanning 150–450 $\mu$ m. We recorded in 5 functionally identified areas, covering large parts of the dorsal cortex. In anterior cortex, we targeted M2 and functionally identified the anterior lateral motor cortex (ALM) and medial motor cortex (MM) based on stereotactic coordinates and their averaged neural responses to licking<sup>22</sup> (Fig. 7A, top). In posterior cortex we targeted V1 (identified by retinotopic mapping<sup>23,24</sup>), retrosplenial (RS) and primary somatosensory cortex (S1) (Fig. 7A, bottom). In each area, average population activity was modulated as expected for sensory, motor, and association areas (Fig. 7B).

We then applied our linear model to the 2-photon data, now aiming to predict single cell activity. In the single-cell data, as in the widefield data, the model had high predictive power, capturing  $37.0 \pm 1.7\%$  (mean  $\pm$  SEM,  $n=10$  mice) of the variance across all animals. Again, individual movement variables outperformed task variables (Fig. 7C, light green bars) and made larger unique contributions (dark green bars). Given the known causal role of ALM for licking<sup>25,26</sup>, one might expect that the licking variable would be particularly important for predicting single cell activity. Instead, in agreement with our widefield results, almost all movement variables made large contributions. Video variables were again far more powerful than other model variables. The same was true when analyzing cortical areas separately (Supplementary Fig. 9). When grouping variables, we again found that uninstructed



movements had much higher predictive power ( $cvR^2_{\text{Task}}=12.3\pm 1.0\%$ ;  $cvR^2_{\text{Instructed}}=10.7\pm 0.9\%$ ;  $cvR^2_{\text{Uninstructed}}=32.6\pm 1.8\%$ ) and made larger unique contributions ( $R^2_{\text{Task}}=1.8\pm 0.1\%$ ;  $R^2_{\text{Instructed}}=1.2\pm 0.3\%$ ;  $R^2_{\text{Uninstructed}}=21.2\pm 1.2\%$ ) (Fig. 7D, left). Across areas, full model performance varied from  $cvR^2=43.7\pm 1.4\%$  in ALM to  $cvR^2=31.1\pm 2.0\%$  in V1 (black dashed lines) potentially due to differences in signal-to-noise ratio across mice. Despite this, the relative predictive power of variable groups was highly consistent (Fig. 7D, right). These differences were found in almost every recorded neuron (Fig. 7E). Uninstructed movements were consistently the strongest predictors of single neuron activity (Fig. 7E, top black trace) and had the greatest unique contribution to the model (Fig. 7E, bottom black trace). The above results also held true in aggressive controls for potential imaging plane motion artifacts, performed by excluding frames with XY motion (Supplementary Fig. 10).

We again leveraged PETH partitioning to gain deeper insight into the functional tuning of single neurons. As with widefield data, PETH partitioning into different components revealed individual contributions from task, instructed and uninstructed movement variables at different trial times. This is exemplified using data averaged over all ALM neurons (Fig. 7F). Average population activity rose consistently over the trial (gray trace), which could be interpreted as a progressively increasing cognitive process. However, PETH partitioning revealed that this rise reflected the combination of several components, each with its own dynamics. Task-related contributions explained the rise during stimulus presentation, which plateaued during the delay (green trace). Uninstructed movement contributions increased during the delay (black trace) and plateaued before the response period. Lastly, instructed movement contributions increased ~350 ms before and throughout the response period. The contributions by each group were distinct in time and together created average neural activity that smoothly rose over the entire second half of the trial.

We then examined the PETHs of individual cells. To identify cells that were mostly modulated by one variable group versus another, we computed three modulation indices (MIs), specifying the extent to which each cell's PETH was best explained by the task, instructed or uninstructed movements (Fig. 7G). Each MI ranges from 0 to 1, with high MI values indicating stronger PETH modulation due to the variable group of interest. Individual cells at the extremes of the MI distributions exemplify strong modulation by task (Fig. 7H, cell 1), uninstructed movements (cell 2) or instructed movements (cell 3). Importantly, the impact of uninstructed movements could not have been inferred from the shape of the PETH alone: while cell 3 was clearly modulated by an instructed movement (licking in the response period), the PETHs of cell 1 and cell 2 exhibited similar temporal dynamics. However, PETH partitioning revealed that the activity of these cells reflected different processes. Notably, this distinction would not have been possible without measuring and analyzing uninstructed movements. PETH partitioning therefore provides additional insight for interpreting the trial average and improves isolation of neurons of interest. Aside from identifying neurons that were purely modulated by a single variable group, we also found many cells where PETH partitioning could isolate task-related dynamics that were otherwise overshadowed by the impact of movements (cell 4). As in studies of mixed selectivity, most cells were affected by a combination of different variables<sup>27</sup>. Nonetheless, PETH

partitioning could also be used to identify cells that were most modulated by a single variable of interest (Supplementary Fig. 11).

Finally, we tested whether our results were similar when considering electrophysiological recordings in subcortical areas. This was critical because calcium imaging only approximates spiking activity and includes slow indicator-related dynamics that may distort the contribution of particular movement variables. Moreover, many subcortical targets like the superior colliculus (SC) also contain cognitive and movement-related activity<sup>28</sup>, raising the question of whether they are modulated by uninstructed movements as in cortical areas. We therefore recorded brain activity using Neuropixels probes<sup>29</sup> in visual cortex (VC) and SC. Because Neuropixels probes span almost the entire depth of the mouse brain, we were also able to simultaneously record from the midbrain reticular nucleus (MRN) (Fig. 8A).

We presented visual stimuli to two head-fixed mice. As before, we used PETH partitioning to separate trial-averaged data into stimulus and movement components (Fig. 8B). Consistent with our widefield calcium imaging results from V1 (Fig. 5I), trial-averaged neural responses were well-explained by stimulus variables with minimal movement contributions. In contrast, across all neurons, trial-by-trial  $cvR^2$  was far higher for movement variables than for visual stimuli, demonstrating that moment-by-moment neural activity was mostly related to movements (Fig. 8C). This can be seen in an example neuron that exhibited slow fluctuations in firing rate during the recording session (~40 minutes, Fig. 8D), which were not explained by visual stimuli but were highly predictable with movement variables. The same was also true when comparing  $cvR^2$  and  $R^2$  for stimuli and movements in different cortical and subcortical areas (Fig. 8E). In all recorded brain areas, movements were far more important for predicting neural activity, indicating that the dominance of movements in driving neural activity is not limited to cortex but potentially a brain-wide phenomenon.

## Discussion

Our results demonstrate that cortex-wide neural activity is dominated by movements. By including a wide array of uninstructed movements, our model predicted single-trial neural activity with high accuracy. This was true for thousands of individual neurons, recorded with three different methods, spanning multiple cortical areas and depths. Both instructed and uninstructed movements also had large task-aligned contributions, affecting trial-averaged neural activity. By partitioning PETHs, we separated movement-related from task-related contributions, providing a new tool to unmask truly task-related dynamics and isolate the impact of movements on cortical computations.

Movement-related modulations of neural activity can either reflect the movement itself (due to movement generation, efference copy, or sensory feedback), or changes in internal state that correlate with movements<sup>30,31</sup>. Here, we found that movement kernels were highly specific in terms of both cortical areas and temporal dynamics (Fig. 3), indicating that they were not predominately reflective of internal state changes. State changes (e.g., arousal) might instead be best captured by analog variables like the pupil<sup>32</sup>, as indicated by its broad effects across cortex. Video variables could, in principle, also provide significant internal

state information, which may partially explain their high model contributions. However, video weights were at least 8-dimensional, indicating that they reflected more than a one-dimensional state like arousal (Fig. 3). The combined importance of internal states and specific movements highlights the need for tracking multiple movements when assessing their combined impact on neural activity. Using video recordings to track animal movements is a relatively easy way to achieve this goal, especially with new tools that are available for interpreting video data<sup>33,34</sup>, and should therefore become standard practice for physiological recordings.

The vast majority of single neuron activity was best predicted by uninstructed movements. Remarkably, this was true across cortical areas and depths as well as subcortical regions (Figs. 7D, 8E, Supplementary Fig. 10). Moreover, the amount of explained neural variance by uninstructed movements was highly similar between 2-photon and electrophysiological data, demonstrating that our results are largely independent of recording technique. However, our Neuropixels recordings were done in passive mice and included only visual stimuli as task variables. Further studies with brain-wide recordings in task-performing animals are therefore needed to assess whether stronger contributions of task-variables on neural activity might be present in task-relevant subcortical areas.

The prevalence of movement modulation may also explain why apparent task-related activity has been observed in a variety of areas<sup>12,35</sup> and highlights the importance of additional controls, like neural inactivation, to establish involvement in a given behavior<sup>36</sup>. However, in areas with established causal relevance, movements must still be considered. For instance, in ALM, which is known to play a causal role in similar tasks to the one studied here<sup>11,26</sup>, many neurons were well-explained by uninstructed movement variables (Fig. 7D,H). Our partitioning method can help to better interpret such neural activity by identifying neurons that are best explained by task variables. Moreover, for neurons with mixed selectivity for task and movement variables<sup>27</sup>, the model can separate their respective contributions and reveal obscured task-related neural dynamics (Supplementary Fig. 11).

Virtually all recorded cortical neurons exhibited mixed selectivity and were strongly modulated by movements regardless of whether or not they were task-modulated (Fig. 7E). Importantly, this does not indicate that task-related dynamics are absent from cortical circuits but rather that movements account for a much larger amount of neural variance. Therefore, even for neurons that are strongly tuned to task variables, cognitive computations are often embedded within a broader context of movement-related information.

A recent paper on movement-related activity in large neural populations likewise observed that facial movements dominate neural activity<sup>16</sup>. Our approach confirms and extends this important work in several ways. First, we show that including additional movement variables (e.g., hindlimb) further improves model performance, and we provide a tool to separate task- and movement related activity through PETH partitioning. Most importantly, we demonstrate the importance of uninstructed movements during a cognitive decision-making task. This shows that the impact of movements remains large during task performance, even in well-trained expert animals. Moreover, by studying the task and movement contributions during learning, we found that uninstructed movements increasingly overlapped with the

task as performance improved (Fig. 6A,F). This indicates that the animals consistently performed rich movement sequences that become more stereotyped during learning. This may be because movements like whisking indicate reward anticipation during the delay period and reflect task-related knowledge (Fig. 6I). Another intriguing possibility is that additional movements persist to aid cognitive function. For example, the animal might perform a stereotyped movement sequence following the stimulus that then spans the delay period and ‘outsources’ working memory demands into a sequence of motor commands<sup>37</sup>. This resembles earlier results from rats learning a complex motor task<sup>38</sup> and highlights that responses to movement can be tightly tied up with the reward signals required for learning<sup>39</sup>, and ultimately become embedded in cognitive processes.

The profound impact of uninstructed movements on neural activity suggests that movement signals play a critical role in cortical information processing. Widespread integration of movement signals might be advantageous for sensory processing<sup>9,40</sup>, canceling of self-motion<sup>41</sup>, gating of inputs<sup>42</sup> or permitting distributed associational learning<sup>43,44</sup>. All of these functions build on the idea that the brain creates internal models of the world that are based on the integration of sensory and motor information to predict future events. The creation of such models requires detailed movement information. Our results thus highlight the notion that such predictive coding is not limited to sensory areas but might be a common theme across cortex: while sensory areas integrate motion signals to predict expected future stimuli<sup>45</sup>, motor areas equally integrate sensory signals to detect deviations between intended and actual motor state<sup>46-48</sup>. The principles of comparing current and predicted feedback also extend to other brain areas like the ventral tegmental area<sup>49</sup> or the cerebellum<sup>50</sup>. This suggests that robust movement representations may be found throughout the brain.

## Methods

### Animal Subjects

All surgical and behavioral procedures conformed to the guidelines established by the National Institutes of Health and were approved by the Institutional Animal Care and Use Committee of Cold Spring Harbor Laboratory. Experiments were conducted with male mice from the ages of 8-25 weeks. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications<sup>10,12</sup>. All mouse strains were of C57BL/6J background and purchased from Jackson Laboratory. Six transgenic strains were used to create the transgenic mice used for imaging: Emx-Cre (JAX 005628), LSL-tTA (JAX 008600), CaMK2 $\alpha$ -tTA (JAX 003010), Ai93 (JAX 024103), G6s2 (JAX 024742) and H2B-eGFP (JAX 006069). Ai93 mice as described below were Ai93; Emx-Cre; LSL-tTA; CaMK2 $\alpha$ -tTA. G6s2 as described below were CaMK2 $\alpha$ -tTA; tetO GCaMP6s.

For widefield imaging we used 11 Ai93 mice (4 mice were negative for CaMK2 $\alpha$ -tTA). Two G6s2 and H2B-eGFP mice were used for widefield control experiments. For two-photon imaging, we recorded from 7 Ai93 mice (6 mice were negative for CaMK2 $\alpha$ -tTA) and 3 G6s2 mice. All trained mice were housed in groups of two or more under an inverted 12:12-h light-dark regime and trained during their active dark cycle.

Data collection and analysis were not performed blind to the conditions of the experiments. Animals were randomly assigned to cohorts that were trained with auditory stimuli versus visual stimuli. Auditory and visual trials were then randomly interleaved in both groups.

### Surgical procedures

All surgeries were performed under 1-2% isoflurane in oxygen anesthesia. After induction of anesthesia, 1.2 mg/kg of meloxicam was injected subcutaneously and lidocaine ointment was topically applied to the skin. After making a medial incision, the skin was pushed to the side and fixed in position with tissue adhesive (Vetbond, 3M). We then created an outer wall using dental cement (C&B Metabond, Parkell; Ortho-Jet, Lang Dental) while leaving as much of the skull exposed as possible. A circular headbar was attached to the dental cement. For widefield imaging, after carefully cleaning the exposed skull we applied a layer of cyanoacrylate (Zap-A-Gap CA+, Pacer technology) to clear the bone. After the cyanoacrylate was cured, cortical blood vessels were clearly visible.

For 2-photon imaging, instead of clearing the skull, we performed a circular craniotomy using a biopsy punch. For imaging anterior cortex, a 3-mm wide craniotomy was centered 1.5 mm lateral and 1.5 mm anterior to bregma. For imaging posterior cortex, a 4-mm wide craniotomy was centered 1.7 mm lateral and 2 mm posterior to bregma. We then positioned a circular coverslip of similar size over the cortex and sealed the remaining gap between the bone and glass with tissue adhesive. The coverslip window was then secured to the skull using C&B Metabond (Parkell) and the remaining exposed skull was sealed using dental cement. After surgery, animals were kept on a heating mat for recovery and a daily dose of analgesia (1.2 mg/kg meloxicam) and antibiotics (2.3 mg/kg enrofloxacin) were administered subcutaneously for 3 days or longer.

For electrophysiology experiments, we used 13-17 week old male mice. Mice were given medicated food cups (MediGel CPF, Clear H20 74-05-5022) 1-2 days prior to surgery. We performed a small circular craniotomy over visual cortex using a dental drill. Rather than a circular headbar, a boomerang-shaped custom Titanium headbar was cemented to the skull, just posterior to the eyes, near Bregma. In addition, a small ground screw was drilled into the skull over the cerebellum. The probe was mounted onto a 3D printed piece within an external casing, affixed to a custom stereotaxic adapter, and lowered into the brain as previously described<sup>51</sup>. Recordings were performed after the mouse had fully recovered from surgery (3-4 days).

### Behavior

The behavioral setup was based on an Arduino-controlled finite state machine (Bpod r0.5, Sanworks) and custom Matlab code (2015b, Mathworks) running on a linux PC. Servo motors (Turnigy TGY-306G-HV), touch sensors and visual stimuli were controlled by microcontrollers (Teensy 3.2, PJRC) running custom code. Twenty-five mice were trained on a delayed two-alternative forced choice (2AFC) spatial discrimination task. Mice initiated trials by touching either of two handles with their forepaws. Handles were mounted on servo motors and were moved out of reach between trials. After one second of holding a handle, sensory stimuli were presented. Sensory stimuli consisted of either a sequence of auditory

clicks, or repeated presentation of a visual moving bar (3 repetitions, 200 ms each). Auditory stimuli were presented from either a left or right speaker, and visual stimuli were presented on one of two small LED displays on the left or right side. The sensory stimulus was presented for 600 ms, there was a 500 ms pause with no stimulus, and then the stimulus was repeated for another 600 ms. After the second stimulus, a 1000 ms delay was imposed, then servo motors moved two lick spouts into close proximity of the animal's mouth. If the animal licked twice to the spout on the same side as the stimulus, he was rewarded with a drop of water. After one spout was contacted twice, the other spout was moved out of reach to force the animal to commit to its decision.

Animals were trained over the course of approximately 30 days. After 2-3 days of restricted water access, animals were head-fixed and received water in the setup. Water was given by presenting a sensory stimulus, subsequently moving only the correct spout close to the animal, then dispensing water automatically. After several habituation sessions, animals had to touch the handles to trigger the stimulus presentation. Once animals reliably reached for the handles, the required touch duration was gradually increased up to 1 second. Lastly, the probability for fully self-performed trials, in which both spouts were moved towards the animal after stimulus presentation, was gradually increased until animals reached stable detection performance of 80% or higher.

Each animal was trained exclusively on a single modality (6 visual animals, 5 auditory for widefield imaging; 5 visual animals, 5 auditory animals for 2-photon imaging). Only during imaging sessions were trials of the untrained modality presented as well. This allowed us to compare neural activity on trials where animals performed stimulus-guided sensorimotor transformations versus trials where animal decisions were random. To ensure that detection performance was not overly affected by presentation of the untrained modality, the trained modality was presented in 75% and the untrained modality in 25% of all trials. Additionally, 4 auditory animals (2 GFP, 2 G6s mice) were used to control for hemodynamic corrections in the widefield data (Supplementary Fig. 5). Since animal performance was not important for these controls, all recordings were done with 50% visual and 50% auditory stimuli.

### **Behavioral sensors**

We used information from several sensors in the behavioral setup to measure different aspects of animal movement. The handles detected contact with the animal's forepaws, and the lick spouts detected contact with the tongue using a grounding circuit. An additional piezo sensor (1740, Adafruit LLC) below the animal's trunk was used to detect hindlimb movements. Sensor data were normalized and thresholded at 2 standard deviations to extract hindlimb movements. Based on hindlimb events we created an event-kernel design matrix that was also included in the linear model (see below).

### **Video monitoring**

Two webcams (C920 and B920, Logitech) were used to monitor animal movements. Cameras were positioned to capture the animal's face (side view) and the body (bottom view). To target particular behavioral variables of interest, we defined subregions of the video which were then examined in more detail. These included a region surrounding the

eye, the whisker pad and the nose. From the eye region we extracted changes in pupil diameter using custom Matlab code. To analyze whisker movements, we computed the absolute temporal derivative averaged over the entire whisker pad. The resulting 1-D trace was then normalized and thresholded at 2 standard deviations to extract whisking events. Based on whisking events we created an event-kernel design matrix that was also included in the linear model (see below). The same approach was used for the nose and pupil diameter, as well as the bottom camera to capture whole-body movements.

### Widefield imaging

Widefield imaging was done using an inverted tandem-lens microscope<sup>52</sup> in combination with an sCMOS camera (Edge 5.5, PCO) running at 60 fps. The top lens had a focal length of 105 mm (DC-Nikkor, Nikon) and the bottom lens 85 mm (85M-S, Rokinon), resulting in a magnification of 1.24x. The total field of view was  $12.5 \times 10.5$  mm and the image resolution was  $640 \times 540$  pixels after 4x spatial binning (spatial resolution:  $\sim 20\mu\text{m}/\text{pixel}$ ). To capture GCaMP fluorescence, a 525 nm band-pass filter (#86-963, Edmund optics) was placed in front of the camera. Using excitation light at two different wavelengths, we isolated  $\text{Ca}^{2+}$ -dependent fluorescence and corrected for intrinsic signals (e.g., hemodynamic responses)<sup>10,12</sup>. Excitation light was projected on the cortical surface using a 495 nm long-pass dichroic mirror (T495lpxr, Chroma) placed between the two macro lenses. The excitation light was generated by a collimated blue LED (470 nm, M470L3, Thorlabs) and a collimated violet LED (405 nm, M405L3, Thorlabs) that were coupled into the same excitation path using a dichroic mirror (#87-063, Edmund optics). We alternated illumination between the two LEDs from frame to frame, resulting in one set of frames with blue and the other with violet excitation at 30 fps each. Excitation of GCaMP at 405 nm results in non-calcium dependent fluorescence<sup>53</sup>, allowing us to isolate the true calcium-dependent signal by rescaling and subtracting frames with violet illumination from the preceding frames with blue illumination. All subsequent analysis was based on this differential signal at 30 fps. To confirm accurate CCF alignment, we performed retinotopic visual mapping<sup>23,24</sup> in each animal and found high correspondence between functionally identified visual areas and the CCF (Supplementary Fig. 2).

To ensure that our correction approach effectively removed non-calcium dependent fluorescence, we performed control experiments with two GFP-expressing animals (CAG-H2B-eGFP). Here, the hemodynamic correction removed  $\sim 90\%$  of variance in the data and the model lost most of its predictive power and spatial specificity (Supplementary Fig. 5). We also imaged two G6s2 mice expressing GCaMP6s<sup>10</sup>. These animals were expected to exhibit stronger calcium-dependent fluorescence due to large signal-to-noise ratio (SNR) of the GCaMP6s indicator. Accordingly, the remaining variance after hemodynamic correction was much higher (Supplementary Fig. 5A-B) and the model predicted even more variance than in Ai93 mice (Supplementary Fig. 5C-D). The  $\beta$ -weights in modulated areas were also strongest with GCaMP6s-expressing mice and close to zero in GFP controls (Supplementary Fig. 5F&H). Together, these controls therefore provide strong evidence that our widefield results were not due to potential contributions from uncorrected hemodynamic signals.

## Two-photon imaging

Two-photon imaging was performed with a resonant-scanning two-photon microscope (Sutter Instruments, Movable Objective Microscope, configured with the “Janelia” option for collection optics), a Ti:Sapphire femtosecond pulsed laser (Ultra II, Coherent Inc.), and a 16X 0.8 NA objective (Nikon Instruments). Images were acquired at 30.9 Hz with an excitation wavelength of 930 nm. All focal planes were between 150-450  $\mu\text{m}$  below the pial surface. To label blood vessels, we subcutaneously injected 50 $\mu\text{l}$  of 1.25% Texas Red (D3328, Thermo Fisher) 15 minutes before imaging. During recording, the red labeled vasculature was used as a reference to correct for slow z-drift of the focal plane. The objective height was manually adjusted during the recording in 1-2  $\mu\text{m}$  increments as often as necessary to maintain the same focal plane. This was done every 5-10 minutes in each session.

## Preprocessing of neural data

To analyze widefield data, we used SVD to compute the 200 highest-variance dimensions. These dimensions accounted for at least 90% of the total variance in the data. Using 500 dimensions accounted for little additional variance ( $\sim 0.15\%$ ), indicating that additional dimensions were mostly capturing recording noise. SVD returns ‘spatial components’  $U$  (of size pixels x components), ‘temporal components’  $V^T$  (of size components x frames) and singular values  $S$  (of size components x components) to scale components to match the original data. To reduce computational cost, all subsequent analysis was performed on the product  $SV^T$ .  $SV^T$  was high-pass filtered above 0.1Hz using a zero-phase, second-order Butterworth filter. Results of analyses on  $SV^T$  were later multiplied with  $U$ , to recover results for the original pixel space. All widefield data was rigidly aligned to the Allen Common Coordinate Framework v3, using four anatomical landmarks: the left, center, and right points where anterior cortex meets the olfactory bulbs and the medial point at the base of retrosplenial cortex.

To analyze two-photon data we used Suite2P<sup>54</sup> with model-based background subtraction. The algorithm was used to perform rigid motion correction on the image stack, identify neurons, extract their fluorescence, and correct for neuropil contamination. F/F traces were produced using the method of Jia et al.<sup>55</sup>, skipping the final filtering step. Using these traces, we produced a matrix of size neurons x time, and treated this similarly to  $SV^T$  above. Finally, we confirmed imaging stability by examining the average firing rate of neurons over trials. If all neurons varied substantially at the beginning or end of a session, trials containing the unstable portion were discarded. Recording sessions yielded  $188.63 \pm 19.28$  neurons and contained  $378.56 \pm 9$  performed trials (mean  $\pm$  SEM).

As an image motion control, we also used the XY-translation values from the initial motion correction to ask whether motion of the imaging plane (due to imperfect registration or Z-motion) could contribute to our two-photon results. To account for this possibility, we removed all frames that were translated by more than 2 pixels in the X or Y direction and repeated the analysis (Extended Data Figure 10). In 4983 neurons, we observed negative unique contributions from the task-group, indicative of model overfitting when removing too many data frames. We therefore rejected these neurons from further analysis. For the



remaining 8787 neurons, explained variance was highly similar as in our original findings, demonstrating that our results could not be explained by motion of the imaging plane.

To compute trial-averages, imaging data were double-aligned to the time when animals initiated a trial and to the stimulus onset. After alignment, single trials consisted of 1.8 s of baseline, 0.83 s of handle touch and 3.3 s following stimulus onset. The randomized additional interval between initiation and stimulus onset (0 - 0.25 s) was discarded in each trial and the resulting trials of equal length were averaged together.

### Neuropixels recordings

To investigate single-neuron responses, a set of linearly expanding (looming) or contracting (receding) dots (40 cm/s) were presented 20 cm (dots were 2 to 20 degrees of visual space) above the mouse's head while the mouse was head-restrained but free to move on a wheel. Stimuli were high or low contrast, and either visual only or with an additional auditory looming stimulus -- white noise of increasing volume (80 dB at max volume). Eight different types of stimuli were presented: high and low contrast visual looming, high and low contrast visual receding, high and low contrast audiovisual looming, and high and low contrast audiovisual receding. Stimuli (20 repeats of each type) were 0.5 s long and randomly presented, with randomized gaps (more than 4 s) between stimuli. Results in Fig. 8B were obtained by averaging over all stimulus conditions. A video camera (Basler AG) in combination with infrared lights was used to record the pupil and face during visual stimulation. These videos were synchronized with the electrophysiology data for subsequent analysis.

Electrophysiology data was collected with SpikeGLX (Bill Karsh, <https://github.com/billkarsh/SpikeGLX>). We recorded from 384 channels spanning ~4 mm of the brain<sup>29</sup>. The data were first median subtracted across channels and time, sorted with Kilosort spike sorting software<sup>56</sup>, and manually curated using phy (<https://github.com/kwikteam/phy>). Sorted data were analyzed using custom MATLAB code.

### Linear model

The linear model was constructed by combining multiple sets of variables into a design matrix, to capture signal modulation by different task or motor events (Fig. 2C). Each variable (except for analog variables) was structured to capture a time-varying event kernel. Each variable therefore consisted of a binary vector containing a pulse at the time of the relevant event, and numerous copies of this vector, each shifted in time by one frame relative to the original. For sensory stimuli, the time-shifted copies spanned all frames from stimulus onset until the end of the trial. For motor events like licking or whisking, the time-shifted copies spanned the frames from 0.5 s before until 1 s after each event. Lastly, for some variables, the time-shifted copies spanned the whole trial. These whole-trial variables were aligned to stimulus onset and contained information about decision variables, such as animal choice or whether a given trial was rewarded. The model also contained several analog variables, such as the pupil diameter. These analog variables provided additional information on movements that we had not previously considered<sup>16,57</sup>. To capture animal movements, we used SVD to compute the 200 highest dimensions of video information in both cameras.

SVD was performed either on the raw video data ('video') or the absolute temporal derivative (motion energy; 'video ME'). SVD analysis of behavioral video was the same as for the widefield data, and we used the product  $SV^T$  of temporal components and singular values as analog variables in the linear model. This is conceptually similar to the methods available with the FaceMap toolbox by Stringer et al<sup>16</sup>. Importantly, we did not use any lagged versions of analog variables. Supplementary table 1 provides an overview of all model variables and how they were generated.

To use video variables, it was important to ensure that their explanatory power did not overlap with other model variables like licking and whisking that can also be inferred from video data. To accomplish this, we first created a reduced design matrix  $X_r$ , containing all movement variables as well as times when spouts or handles were moving or visual stimuli were presented.  $X_r$  was ordered so that the motion energy and video columns were at the end. We then performed a QR decomposition of  $X_r$ <sup>58</sup>. The QR decomposition of a matrix  $A$  is  $A = QR$ , where  $Q$  is an orthonormal matrix and  $R$  is upper triangular. Columns 1 to  $j$  of  $Q$  therefore span the same space as columns 1 to  $j$  of  $A$  for all  $j$ , but all the columns are orthogonal to one another. Finally, we replaced the motion and video columns of the full design matrix  $X$  with the corresponding columns of  $Q$ . This allowed the model to improve the fit to the data using any unique contributions of the motion and video variables, while ensuring that the weights given to other variables were not altered. Note that the QR decomposition is limited to the original space spanned by  $X$ , and therefore cannot enhance the explanatory power of the video variables - it can only reduce it.

Subsequently, we used the same approach to orthogonalize all uninstructed movement variables against instructed movements. This was done to ensure that all information that could be explained by instructed movements could not be accounted for by correlated, uninstructed movement variables. As described above, this can only reduce the explanatory power of the uninstructed movement variables.

To ensure that model variables were sufficiently decorrelated, we introduced an additional randomized delay (0-0.25 s) between the animal grabbing the handles and stimulus onset. Movement and task variables were also decorrelated due to the natural variability in response times and the number of actions.

When a design matrix has columns that are close to linearly dependent (multicollinear), model fits are not reliable. To test for this, we devised a novel method we call "cumulative subspace angles." The idea is that for each column of the design matrix, we wish to know how far it lies from the space spanned by the previous columns (note that pairwise angles do not suffice to determine multicollinearity). Our method works as follows: (1) the columns of the matrix  $X$  were normalized to unit magnitude, (2) a QR decomposition of  $X$  was performed, (3) the absolute value of the elements along the diagonal of  $R$  were examined. Each of these values is the absolute dot product of the original vector with the same vector orthogonalized relative to all previous vectors. The values range from zero to one, where zero indicates complete degeneracy and one indicates no multicollinearity at all. Over all experiments, the most collinear variable received a 0.26, indicating that it was 15° from the

space of all other variables. The average value was 0.84, corresponding to a mean angle of 57°.

To avoid overfitting, the model was fit using ridge regression. The regularization penalty was estimated separately for each column of the widefield data using marginal maximum likelihood estimation<sup>59</sup> with minor modifications that reduced numerical instability for large regularization parameters. Ridge regression was chosen over other regularization methods like a sparseness prior (using LASSO or elastic net) to allow for similar contributions from different correlated variables. This was important, because a sparseness prior would have isolated the contribution of the best model variables while completely rejecting correlated but slightly less informative variables. Instead, we sought to include contributions from all model variables that were related to the imaging data. This was particularly important for the interpretation of the model event kernels and PETH partitioning. To quantify upper and lower bounds on the explanatory power of each model variable we computed the overall explained variance ( $cvR^2$ ) and unique predictive power ( $R^2$ ) for each variable (see below).

### Variance analysis

Explained variance ( $cvR^2$ ) was obtained using 10-fold cross-validation. To compute all explained variance by individual model variables, we created reduced models where all variables apart from the specified one were shuffled in time. The explained variance by each reduced model revealed the maximum potential predictive power of the corresponding model variable.

To assess unique explained variance by individual variables, we created reduced models in which all the time points for only the specified variable were shuffled. The difference in explained variance between the full and the reduced model yielded the unique contribution  $R^2$  of that model variable. The same approach was used to compute unique contributions for groups of variables, i.e., ‘instructed movements’, ‘uninstructed movements’ or ‘task’. Here, all variables that corresponded to a given group were shuffled at once.

To compute the ‘task-aligned’ or ‘task-independent’ explained variance for each movement variable, we created a reduced ‘task-only’ model where all movement variables were shuffled in time. This task-only model was then compared to other reduced models where all movement variables but one were shuffled. The difference between the task-only model and this model yielded the task-independent contribution of that movement variable. The task-aligned contribution was computed as the difference between the total variance explained by a given variable (its  $cvR^2$ ) and its task-independent contribution.

### PETH partitioning

Reconstructed trial averages (Figs. 5 and 7) were produced by fitting the full model and averaging the reconstructed data over all trials. To split the model into the respective contributions of instructed movements, uninstructed movements and task variables, we reconstructed the data based on that variable group alone (using the weights from the full model, without re-fitting) and averaging over all trials. To evaluate the relative impact of each group variable on the perievent time histogram (PETH), we computed a modulation index (MI), defined as

$$MI_{Group} = \frac{\left(1 + \frac{\Delta_{Group} - \Delta_{nonGroup}}{\Delta_{Group} + \Delta_{nonGroup}}\right)}{2},$$

where  $\Delta_{Group}$  and  $\Delta_{nonGroup}$  denotes the sum of the absolute deviation from zero when reconstructing the PETH either based on all variables of a given group ( $\Delta_{Group}$ ) or all other variables ( $\Delta_{nonGroup}$ ). The MI ranges from 0 (variable group has no impact on PETH) to 1 (PETH is fully explained by variable group). Intermediate values denote a mixed contribution to the PETH from different group variables (Supplementary Fig. 11).

### Model-based video reconstruction

To better understand how the video related to the neural data, we analyzed the portion of the  $\beta$ -weight matrix that corresponded to the video variables. This portion of the matrix was projected back up into the original video space. The result was of size  $p \times d$ , where  $p$  is the number of video pixels (153,600) and  $d$  is the number of dimensions of the widefield data (200). We performed PCA on this matrix, reducing the number of rows. The top few ‘scores’ (projections onto the principal components) are low-dimensional representations of the widefield maps that were most strongly influenced by the video. To choose the dimensionality, we used the number of dimensions required to account for >90% of the variance (Fig. 3F). To obtain the widefield maps showing how the video was related to neural activity (Fig. 3G), we projected the scores back into widefield data pixel space and sparsened them using the varimax rotation. To determine the influence of each video pixel on the widefield (Fig. 3H), we projected the low-dimensional  $\beta$ -weights into video pixel space, took the magnitude of the  $\beta$ -weights for each pixel, and multiplied by the original standard deviation for that pixel (to reverse the Z-scoring step of PCA).

### Aberrant cortical activity in Ai93 transgenic animals

Mice with both *Emx-Cre* and *Ai93* transgenes can exhibit aberrant, epileptiform cortical activity patterns, especially when expressing *GCaMP6* during development<sup>20</sup>. To avoid this issue, we raised most of the 11 mice in our widefield data set (6 mice) on a doxycycline-containing diet (DOX), preventing *GCaMP6* expression until they were 6 weeks or older. This was also true for all *Ai93* mice in our 2-photon data set. However, 5 mice were raised on standard diet, raising the concern that aberrant activity may have affected our widefield results.

To test for the presence of epileptiform activity, we used the same comparison as Steinmetz et al.<sup>20</sup> on the cortex-wide average. A peak in cortical activity was flagged as a potential interictal event if it had a width of 60-235 ms and a prominence of 0.03 or higher. These parameters flagged nearly all cases of apparent interictal events (Supplementary Fig. 8A) and identified four out of the 11 mice in the widefield data set to exhibit potential epileptiform activity (Supplementary Fig. 8B). None of the identified mice were raised on DOX.

To ensure that epileptiform activity would not bias our results, we removed flagged events and interpolated over the resulting gaps (in low-D) with Matlab’s built-in autoregressive

modeling (fillgaps.m) and a 20-frame prediction window. The result did not show any perturbations around the former interictal events (Supplementary Fig. 8C). When comparing modeling results between DOX- and non-DOX-raised mice analyzed separately, predicted variance was highly similar in all cases (Supplementary Fig. 8D-G). This shows that our results were not due to epileptiform activity and indicated that it was safe to include all mice in the dataset with this data processing step.

This was further supported by our additional experiments with GCaMP6s-expressing animals. Consistent with previous observations<sup>20</sup>, we found no epileptiform activity in these mice. Nevertheless, our model predicted an even larger amount of variance than in Ai93 mice and produced highly-specific maps of unique contributions from single variables (Supplementary Fig. 5).

### Code availability

The Matlab code used for the data analysis in this study are available online under <http://churchlandlab.labsites.cshl.edu/code/>.

### Reporting Summary

Further information on research design is also available in the Life Sciences Reporting Summary linked to this article.

### Data availability

The data in from this study will be stored on a dedicated, backed-up repository maintained by Cold Spring Harbor Laboratory. A link to the repository can be found under <http://churchlandlab.labsites.cshl.edu/code/>.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

We thank O. Odoemene, S. Pisupati and H. Nguyen for technical assistance and scientific discussions, H. Zeng for providing Ai93 mice, J. Tucciarone and F. Marbach for breeding assistance, A. Mills and P. Shrestha for providing GFP mice, T. Harris, S. Caddick and the Allen Institute for Brain Sciences for assistance with the Neuropixels probes, and N. Steinmetz, M. Pachitariu and K. Harris for widefield analysis code. Financial support was received from the Swiss National Science foundation (SM, P2ZHP3\_161770), the Pew Charitable Trusts (AKC), the Simons Collaboration on the Global Brain (AKC, MTK), the NIH (EY R01EY022979) and the Army Research Office (ARO) under contract W911NF-16-1-0368 as part of the collaboration between US DOD, UK MOD and UK Engineering and Physical Research Council (EPSRC) under the Multidisciplinary University Research Initiative (MURI)(AKC).

### References

1. Shadlen MN & Newsome WT Motion perception: seeing and deciding. *Proc Natl Acad Sci U S A* 93, 628–633 (1996). [PubMed: 8570606]
2. Maimon G & Assad JA A cognitive signal for the proactive timing of action in macaque LIP. *Nat. Neurosci.* 9, 948–955 (2006). [PubMed: 16751764]

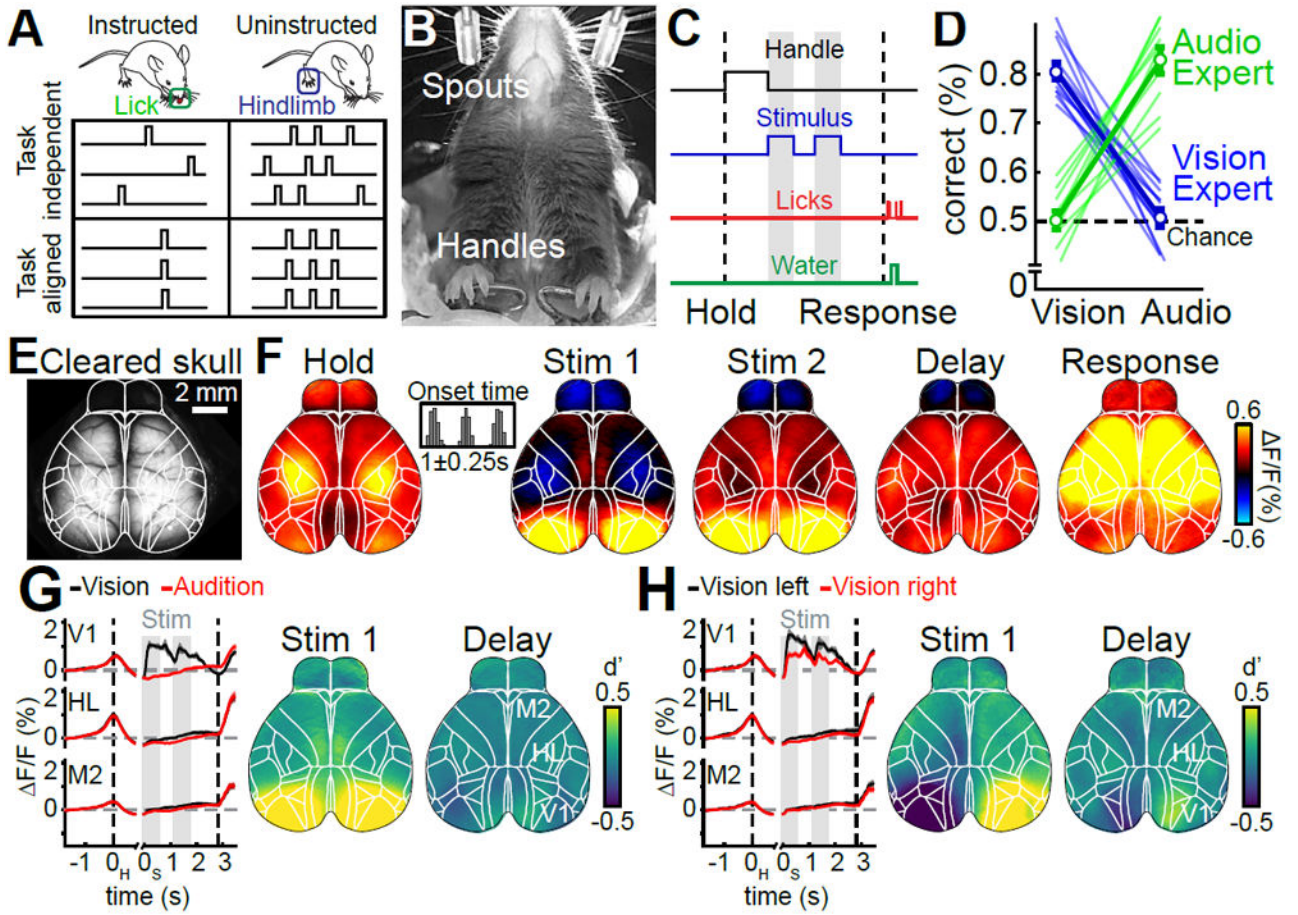
3. Horwitz GD, Batista AP & Newsome WT Representation of an abstract perceptual decision in macaque superior colliculus. *J. Neurophysiol.* 91, 2281–2296 (2004). [PubMed: 14711971]
4. Gold JI & Shadlen MN The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J. Neurosci.* 23, 632–651 (2003). [PubMed: 12533623]
5. Roitman JD & Shadlen MN Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* 22, 9475–9489 (2002). [PubMed: 12417672]
6. Churchland AK, Kiani R & Shadlen MN Decision-making with multiple alternatives. *Nat Neurosci* 11, 693–702 (2008). [PubMed: 18488024]
7. Erlich JC, Bialek M & Brody CD A Cortical Substrate for Memory-Guided Orienting in the Rat. *Neuron* 72, 330–343 (2011). [PubMed: 22017991]
8. Niell CM & Stryker MP Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* 65, 472–479 (2010). [PubMed: 20188652]
9. Saleem AB, Ayaz A, Jeffery KJ, Harris KD & Carandini M Integration of visual motion and locomotion in mouse visual cortex. *Nat. Neurosci.* 16, 1864–1869 (2013). [PubMed: 24185423]
10. Wekselblatt JB, Flister ED, Piscopo DM & Niell CM Large-scale imaging of cortical dynamics during sensory perception and behavior. *Journal of Neurophysiology* 115, 2852–2866 (2016). [PubMed: 26912600]
11. Guo ZV et al. Flow of Cortical Activity Underlying a Tactile Decision in Mice. *Neuron* 81, 179–194 (2014). [PubMed: 24361077]
12. Allen WE et al. Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron* 94, 891–907.e6 (2017). [PubMed: 28521139]
13. Runyan CA, Piasini E, Panzeri S & Harvey CD Distinct timescales of population coding across cortex. *Nature* 548, 92–96 (2017). [PubMed: 28723889]
14. Scott BB et al. Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron* 95, 385–398.e5 (2017). [PubMed: 28669543]
15. Caballero-Gaudes C & Reynolds RC Methods for cleaning the BOLD fMRI signal. *NeuroImage* 154, 128–149 (2017). [PubMed: 27956209]
16. Stringer C et al. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364, eaav7893 (2019).
17. Shimaoka D, Harris KD & Carandini M Effects of Arousal on Mouse Sensory Cortex Depend on Modality. *Cell Reports* 22, 3160–3167 (2018). [PubMed: 29562173]
18. Hanks TD et al. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520, 220–223 (2015). [PubMed: 25600270]
19. Erlich JC, Brunton BW, Duan CA, Hanks TD & Brody CD Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife Sciences* 4, e05457 (2015).
20. Steinmetz NA et al. Aberrant Cortical Activity in Multiple GCaMP6-Expressing Transgenic Mouse Lines. *eNeuro* ENEURO.0207-17.2017 (2017). doi:10.1523/ENEURO.0207-17.2017
21. Zhuang J et al. An extended retinotopic map of mouse cortex. *Elife* 6, (2017).
22. Chen T-W, Li N, Daie K & Svoboda K A Map of Anticipatory Activity in Mouse Motor Cortex. *Neuron* 94, 866–879.e4 (2017). [PubMed: 28521137]
23. Garrett ME, Nauhaus I, Marshel JH & Callaway EM Topography and Areal Organization of Mouse Visual Cortex. *J. Neurosci.* 34, 12587–12600 (2014). [PubMed: 25209296]
24. Marshel JH, Garrett ME, Nauhaus I & Callaway EM Functional Specialization of Seven Mouse Visual Cortical Areas. *Neuron* 72, 1040–1054 (2011). [PubMed: 22196338]
25. Guo ZV et al. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* 545, 181–186 (2017). [PubMed: 28467817]
26. Li N, Chen T-W, Guo ZV, Gerfen CR & Svoboda K A motor cortex circuit for motor planning and movement. *Nature* 519, 51–56 (2015). [PubMed: 25731172]
27. Raposo D, Kaufman MT & Churchland AK A category-free neural population supports evolving demands during decision-making. *Nat Neurosci* 17, 1784–1792 (2014). [PubMed: 25383902]

28. Horwitz GD & Newsome WT Separate Signals for Target Selection and Movement Specification in the Superior Colliculus. *Science* 284, 1158–1161 (1999). [PubMed: 10325224]
29. Jun JJ et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236 (2017). [PubMed: 29120427]
30. Vinck M, Batista-Brito R, Knoblich U & Cardin JA Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron* 86, 740–754 (2015). [PubMed: 25892300]
31. Polack P-O, Friedman J & Golshani P Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nature Neuroscience* 16, 1331–1339 (2013). [PubMed: 23872595]
32. Reimer J et al. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* 84, 355–362 (2014). [PubMed: 25374359]
33. Pereira TD et al. Fast animal pose estimation using deep neural networks. *Nature Methods* 16, 117 (2019). [PubMed: 30573820]
34. Mathis A et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* 21, 1281 (2018). [PubMed: 30127430]
35. Le Merre P et al. Reward-Based Learning Drives Rapid Sensory Signals in Medial Prefrontal Cortex and Dorsal Hippocampus Necessary for Goal-Directed Behavior. *Neuron* 97, 83–91.e5 (2018). [PubMed: 29249287]
36. Gilad A, Gallero-Salas Y, Groos D & Helmchen F Behavioral Strategy Determines Frontal or Posterior Location of Short-Term Memory in Neocortex. *Neuron* 99, 814–828.e7 (2018). [PubMed: 30100254]
37. Euston DR & McNaughton BL Apparent Encoding of Sequential Context in Rat Medial Prefrontal Cortex Is Accounted for by Behavioral Variability. *J. Neurosci.* 26, 13143–13155 (2006). [PubMed: 17182765]
38. Kawai R et al. Motor cortex is required for learning but not for executing a motor skill. *Neuron* 86, 800–812 (2015). [PubMed: 25892304]
39. Coddington LT & Dudman JT The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nature Neuroscience* 21, 1563 (2018). [PubMed: 30323275]
40. Ayaz A, Saleem AB, Schölvinck ML & Carandini M Locomotion controls spatial integration in mouse visual cortex. *Curr. Biol.* 23, 890–894 (2013). [PubMed: 23664971]
41. Sommer MA & Wurtz RH Brain circuits for the internal monitoring of movements. *Annu Rev Neurosci* 31, 317 (2008). [PubMed: 18558858]
42. Schmitt LI et al. Thalamic amplification of cortical connectivity sustains attentional control. *Nature* 545, 219–223 (2017). [PubMed: 28467827]
43. Wang L, Rangarajan KV, Gerfen CR & Krauzlis RJ Activation of Striatal Neurons Causes a Perceptual Decision Bias during Visual Change Detection in Mice. *Neuron* 97, 1369–1381.e5 (2018). [PubMed: 29503185]
44. Engel TA, Chaisangmongkon W, Freedman DJ & Wang X-J Choice-correlated activity fluctuations underlie learning of neuronal category representation. *Nature Communications* 6, 6454 (2015).
45. Keller GB, Bonhoeffer T & Hübener M Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* 74, 809–815 (2012). [PubMed: 22681686]
46. Wolpert DM, Ghahramani Z & Jordan MI An internal model for sensorimotor integration. *Science* 269, 1880–1882 (1995). [PubMed: 7569931]
47. Wolpert DM & Miall RC Forward Models for Physiological Motor Control. *Neural Netw* 9, 1265–1279 (1996). [PubMed: 12662535]
48. Wolpert DM & Kawato M Multiple paired forward and inverse models for motor control. *Neural Netw* 11, 1317–1329 (1998). [PubMed: 12662752]
49. Schultz W Dopamine neurons and their role in reward mechanisms. *Curr. Opin. Neurobiol.* 7, 191–197 (1997). [PubMed: 9142754]
50. Wolpert DM, Miall RC & Kawato M Internal models in the cerebellum. *Trends Cogn. Sci. (Regul. Ed.)* 2, 338–347 (1998). [PubMed: 21227230]

## Additional References

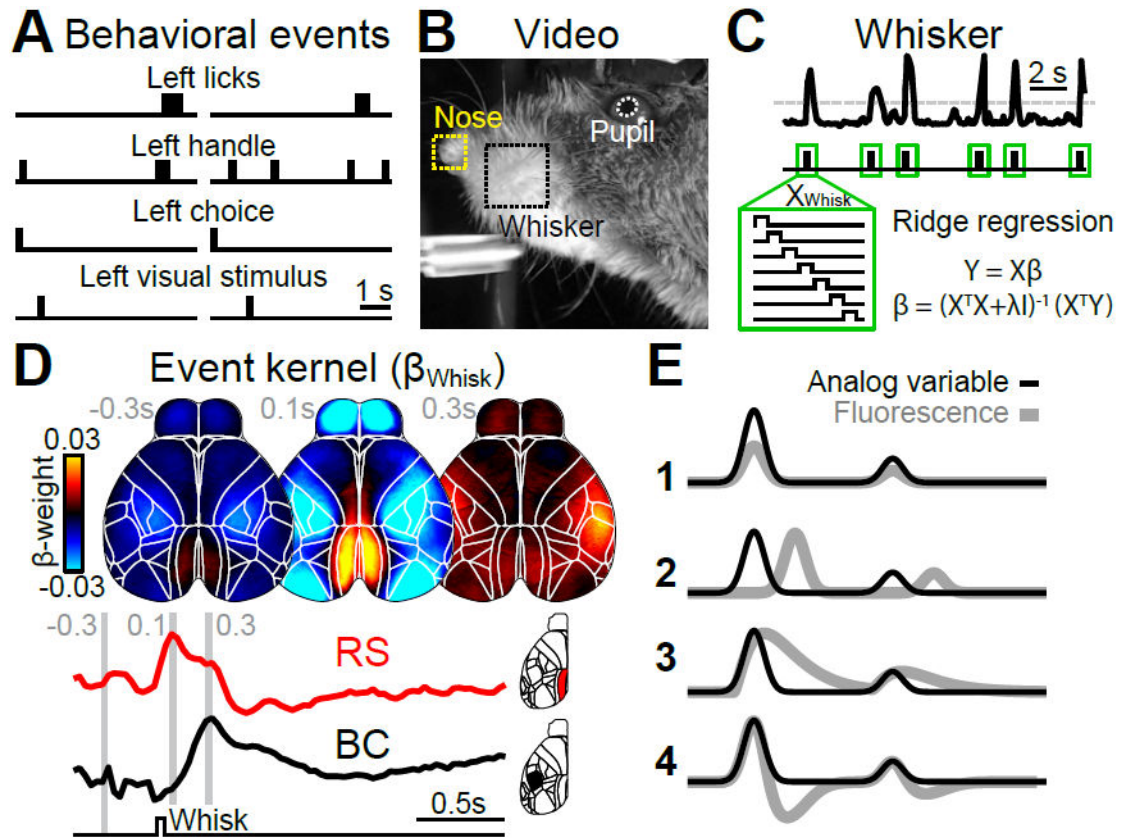
51. Juavinett AL, Bekheet G & Churchland AK Chronically-implanted Neuropixels probes enable high yield recordings in freely moving mice. *bioRxiv* 406074 (2018). doi:10.1101/406074
52. Ratzlaff EH & Grinvald A A tandem-lens epifluorescence microscope: hundred-fold brightness advantage for wide-field imaging. *J. Neurosci. Methods* 36, 127–137 (1991). [PubMed: 1905769]
53. Lerner TN et al. Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* 162, 635–647 (2015). [PubMed: 26232229]
54. Pachitariu M et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 061507 (2016). doi:10.1101/061507
55. Jia H, Rochefort NL, Chen X & Konnerth A In vivo two-photon imaging of sensory-evoked dendritic calcium signals in cortical neurons. *Nat Protoc* 6, 28–35 (2011). [PubMed: 21212780]
56. Pachitariu M, Steinmetz N, Kadir S, Carandini M & H. K. D Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* 061481 (2016). doi: 10.1101/061481
57. Powell K, Mathy A, Duguid I & Häusser M Synaptic representation of locomotion in single cerebellar granule cells. *eLife Sciences* 4, e07290 (2015).
58. Mumford JA, Poline J-B & Poldrack RA Orthogonalization of Regressors in fMRI Models. *PLOS ONE* 10, e0126255 (2015). [PubMed: 25919488]
59. Karabatsos G Marginal maximum likelihood estimation methods for the tuning parameters of ridge, power ridge, and generalized ridge regression. *Communications in Statistics - Simulation and Computation* (2017).





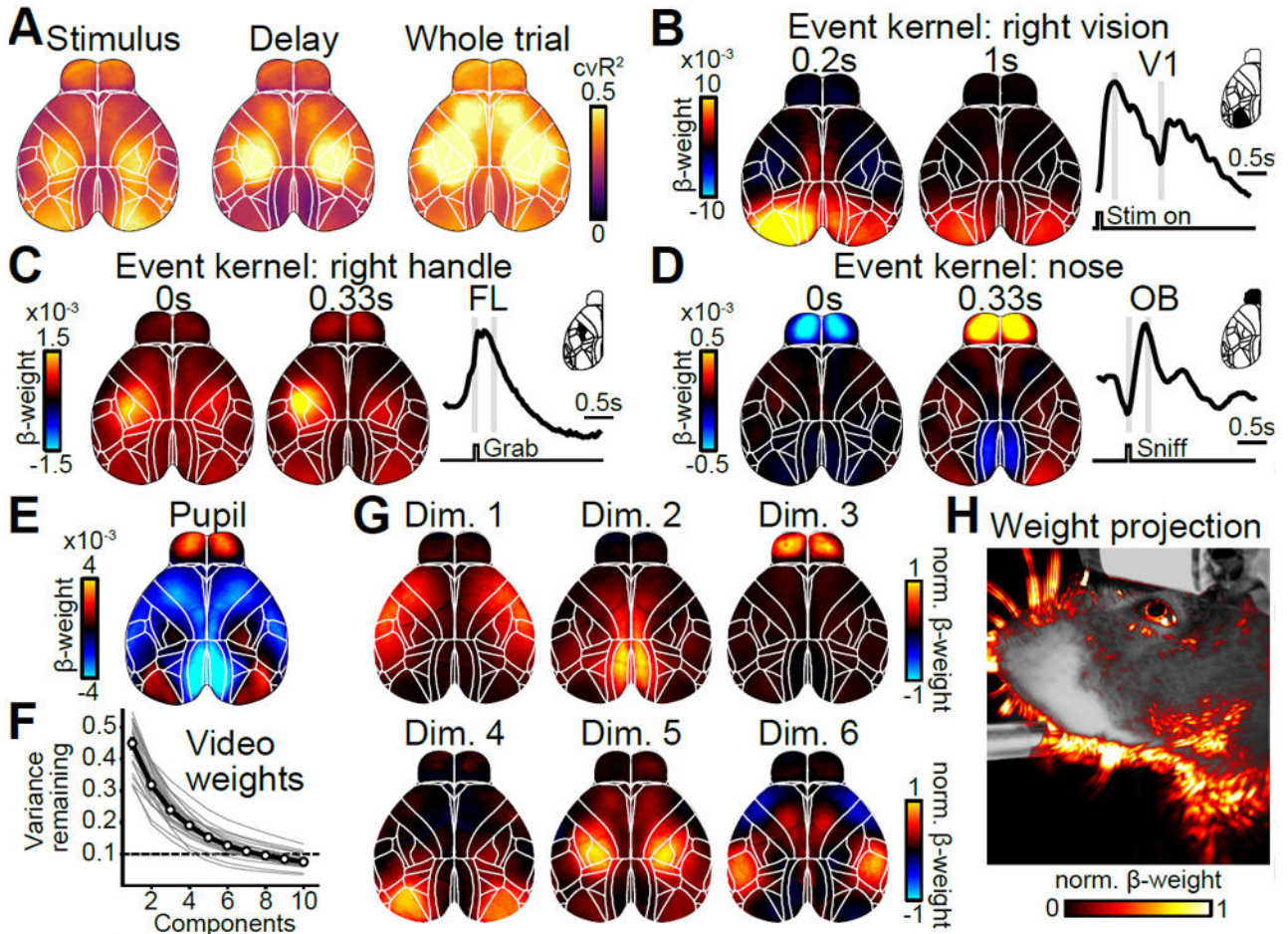
**Figure 1. Widefield calcium imaging during auditory and visual decision making.**

(A) Schematic for the two main questions addressed in this work. Uninstructed movements are exemplified as ‘hindlimb’ but numerous movements are considered throughout this work. (B) Bottom view of a mouse in the behavioral setup. (C) Single-trial timing of behavior. Mice held the handles for 1s ( $\pm 0.25$ s) to trigger the stimulus sequence. One second after stimulus end, water spouts moved towards the mouse so they could report a choice. (D) Visual experts (blue) had high performance with visual but chance performance with auditory stimuli. Auditory experts (green) showed the converse. Thin lines: animals, thick lines: mean. Error bars: mean $\pm$ SEM; n=11 mice. (E) Example image of cortical surface after skull clearing. Overlaid white lines show Allen CCF borders. (F) Cortical activity during different task episodes. Shown are responses when holding the handles (‘Hold’), visual stimulus presentation (‘Stim 1&2’), the subsequent delay (‘Delay’) and the response period (‘Response’). In each trial, stimulus onset was pseudo-randomized within a 0.25-s long time window (inset). (G) Left: Traces show average responses in primary visual cortex (V1), hindlimb somatosensory cortex (HL) and secondary motor cortex (M2) of the right hemisphere during visual (black) or auditory (red) stimulation. Trial averages are aligned to both the time of trial initiation (left dashed line) and stimulus onset (gray bars). Right dashed line indicates response period, shading indicates SEM. Right:  $d'$  between visual and auditory trials during first visual stimulus and the subsequent delay period. (H) Same as (G) but for correct visual trials on the left versus right side. (F-H) (n=22 sessions).



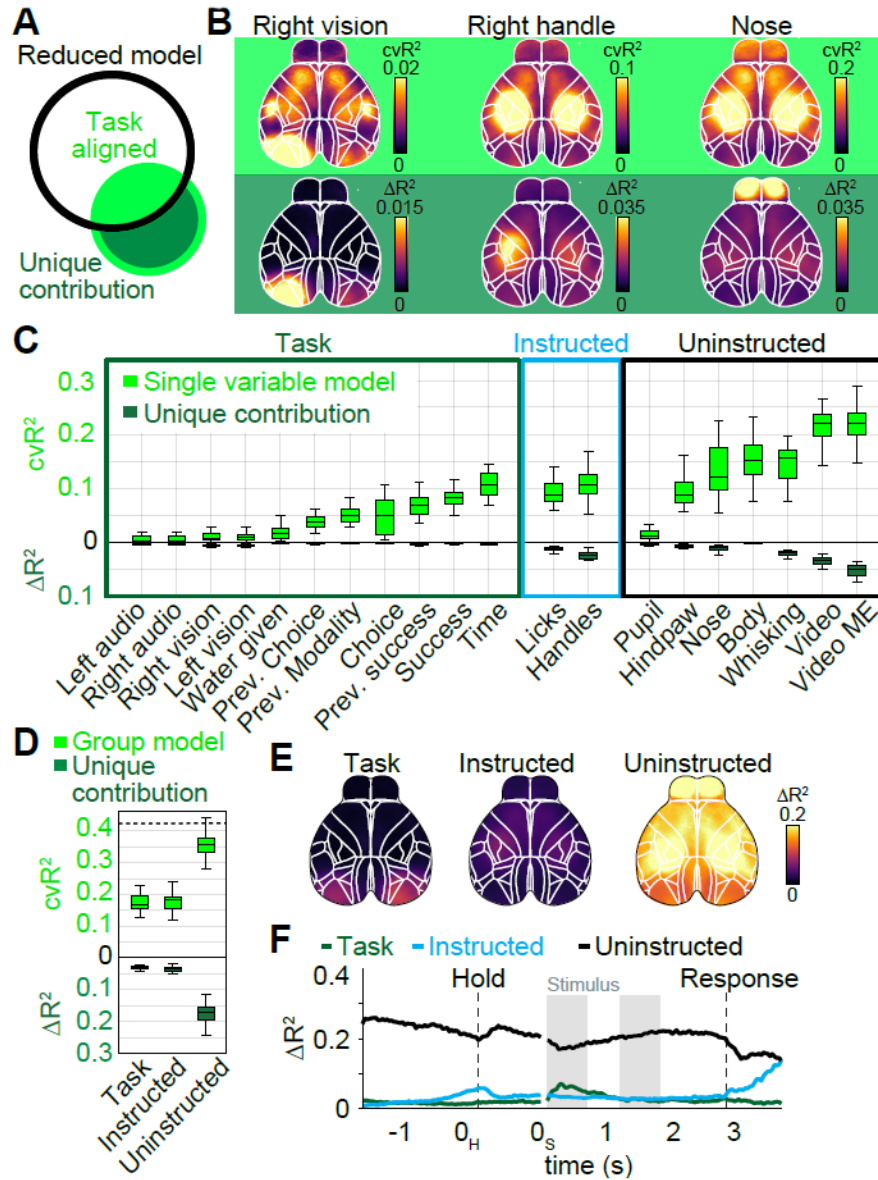
**Figure 2. A linear model to reveal behavioral correlates of cortical activity.**

(A) Two example trials, illustrating different classes of behavioral events. (B) Image of facial video data with 3 movement variables used in the model. (C) Absolute averaged motion energy in the whisker-pad over two trials, showing individual bouts of movement (top). Whisking events were inferred by thresholding (dashed line) and a time-shifted design matrix ( $X_{\text{Whisk}}$ ) was used to compute an event kernel ( $\beta_{\text{Whisk}}$ ). (D) Average  $\beta_{\text{Whisk}}$  maps  $-0.3$ ,  $0.1$  and  $0.3$  seconds relative to whisk onset ( $n=22$  sessions). Whisking caused different responses across cortex, with retrosplenial (RS) being most active  $0.1$  seconds after whisk onset (red trace) and barrel cortex (BC) after  $0.3$  seconds (black trace). (E) A schematic analog variable (black) fitted to cortical activity (gray). Analog variables are linearly scaled to fit neural data instead of assuming a fixed event response structure (1). In contrast to the event kernel traces in (D), analog variables cannot account for neural responses that are shifted in time (2) or include additional response features (3-4).



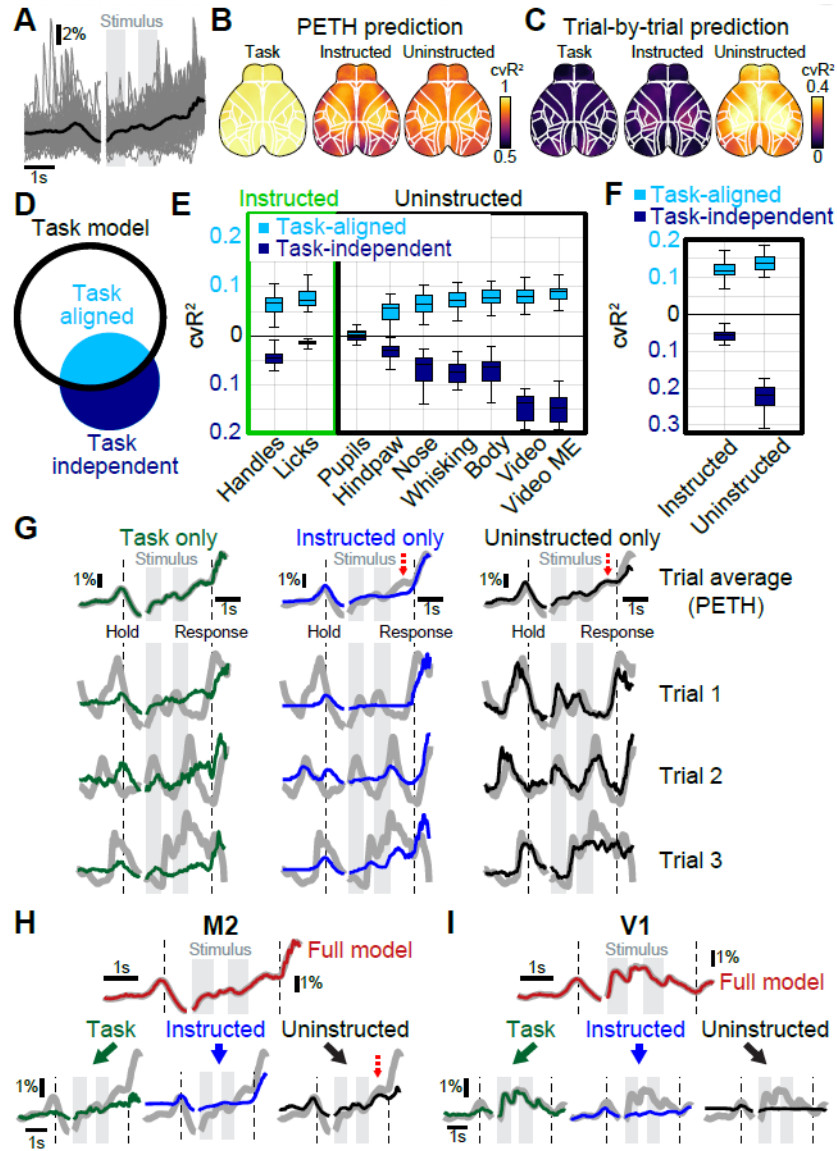
**Figure 3. Linear model predicts cortical activity with highly specific model variables.**

(A) Maps of cross-validated explained variance for two task epochs and the whole trial. (B) Example weight maps of the event kernel for right visual stimuli, 0.2 and 1 seconds after stimulus onset. Trace on the right shows the average weights learned by the model for left V1. Gray lines indicate times of the maps on the left. (C) Same as in (B) but for the event kernel corresponding to right handle grabs, 0 and 0.33 seconds after event onset. (D) Same as in (C) but for the event kernel corresponding to nose movements in the olfactory bulb (OB). (E) Weight map for the analog pupil variable. (F) Cumulative remaining variance for PCs of model's weight matrix for all video variables. Black line shows session average, gray line individual sessions. On average, >90% of all variance was explained by 8 dimensions (dashed line).  $n=22$  sessions. (G) Widefield maps corresponding to the sparsened top 6 video-weight dimensions for an example session. (H) Influence of each behavioral video pixel on widefield data. The opacity and color of the overlay were scaled between the 0th and 99th percentile over all beta values.



**Figure 4. Uninstructed movements dominate cortical activity.**

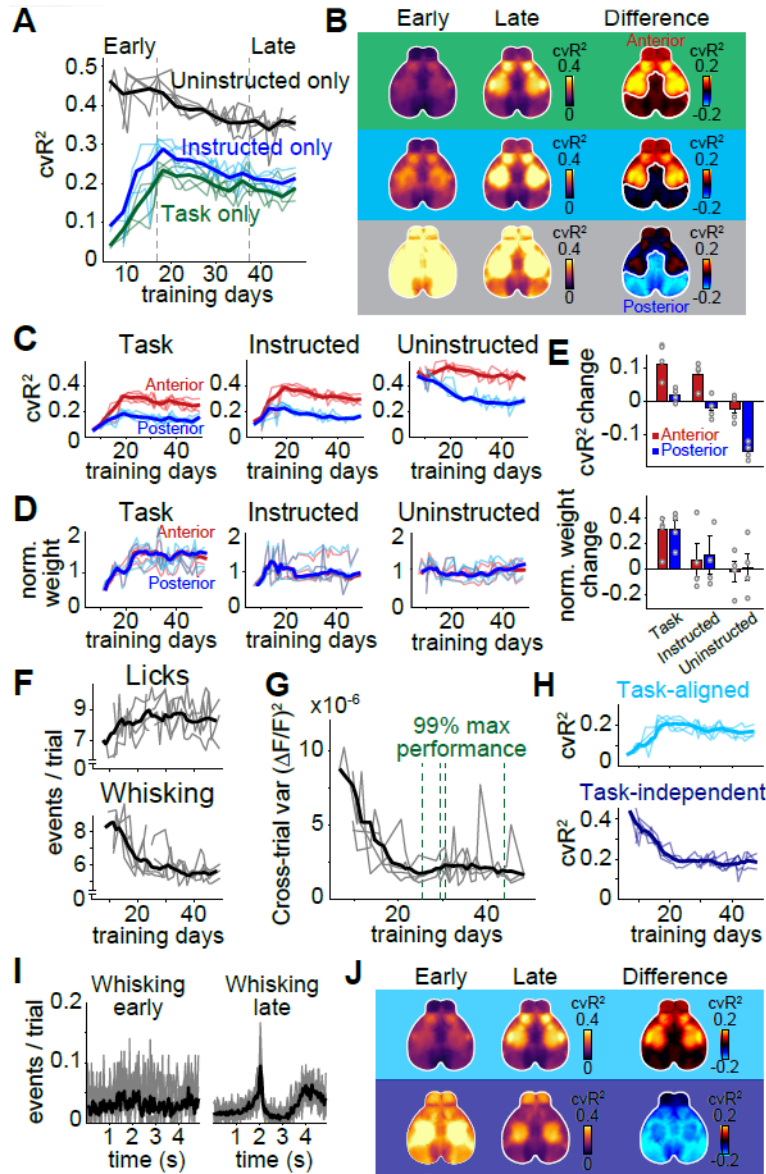
(A) Black circle denotes information of a reduced model, lacking one variable. The single variable (green circle) has information that overlaps (light green) with the reduced model and a unique contribution (dark green) that increases the full model's information. (B) Top row: Cross-validated explained variance ( $cvR^2$ ) maps for different single-variable models. Bottom row: Unique model contribution ( $\Delta R^2$ ) maps for the same variables. (C) Explained variance for single model variables, averaged across cortical maps. Shown is either  $cvR^2$  (light green) or  $\Delta R^2$  (dark green). The box shows the first and third quartiles, inner line is the median over 22 sessions, whiskers represent minimum and maximum values. Prev.: previous. (D) Explained variance for variable groups. Conventions as in (C). (E)  $\Delta R^2$  map for each variable group. (F)  $\Delta R^2$  for variable groups at each time-point, averaged across cortex. (B-F) ( $n=22$  sessions).



**Figure 5. Uninstructed movements make considerable task-aligned and task-independent contributions**

(A) Example M2 data from a single animal. Black trace shows average over all trials, thin gray traces activity from 100 randomly selected trials. (B) Predictions of trial-average by different models that were based on a single variable group. Maps show  $cvR^2$  after averaging over all trials. (C) Predictions of trial-by-trial variability. Maps show  $cvR^2$  after averaging over all time-points within each trial. (D) Black circle denotes task-model information. A movement variable (blue circle) has information that overlaps with the task (light blue) and information that is task-independent (dark blue). (E) Explained variance for all movement variables. Shown is task-independent (dark blue) and task-aligned explained variance (light blue). Values averaged across cortex. The box shows the first and third quartiles, inner line is the median over 22 sessions, whiskers represent minimum and maximum values. (F) Task-aligned and task-independent explained variance for movement groups. Conventions as in (E). (G) M2 data (gray traces), predicted by different variable group models. PETHs (top

row) are mostly well-explained by all three models. Single trials were only well predicted by uninstructed movements (right column, black). **(H)** The M2 PETH is accurately predicted by the full model (top trace, red). Reconstructing the data based on model weights of each variable group allows partitioning the PETH and revealing the groups' respective contributions to the full model (bottom row). Red dashed arrow indicates an example PETH feature that is best-explained by uninstructed movements. **(I)** Same as in **(H)** but for V1 data. **(A, G-H)** (n=412 trials).



**Figure 6. Cognitive and movement responses during learning**

(A)  $cvR^2$  for uninstructed movement (black), task (green) and instructed movement (blue) models, from early to late in training (dashed lines). (B)  $cvR^2$  maps for each model, either early (left column) or late (middle column) in training. Right column shows the difference between early and late  $cvR^2$ . White outlines show separation between ‘anterior’ and ‘posterior’ cortex, which behaved differently with regard to  $cvR^2$  change. (C) Changes in  $cvR^2$  for anterior (red) and posterior (blue) cortex for each model and training day. (D) Same as in (C) but for the sum of absolute weights in each model. Each model weight was normalized to the early training period. (E) Difference between early and late training data for  $cvR^2$  (top) and model weights (bottom). Error bars show the SEM. (F) Mean number of behavioral events per trial for each training day. Shown are event rates for licking (top) and whisking (bottom). (G) Trial-by-trial variance in widefield fluorescence across cortex for each training day. Green lines show the time when individual animals reached 99% of their

maximum task performance. **(H)** Task-aligned (top) and task-independent  $cvR^2$  (bottom) for uninstructed movements. **(I)** PETH for whisking early versus late during training.  $n=10$  sessions per condition from one mouse. Thin lines: sessions; thick lines: average. **(J)** Averaged cortical maps for task-aligned (top row) and task-independent  $cvR^2$  (bottom row). Left and middle columns show  $cvR^2$  early and late in training, right column shows their difference. Colors indicate the task-aligned contribution (top) or task independent contribution (bottom). **(A-H, J)** ( $n=4$  mice, thin lines/circles: animals, thick lines/bar plots: mean).

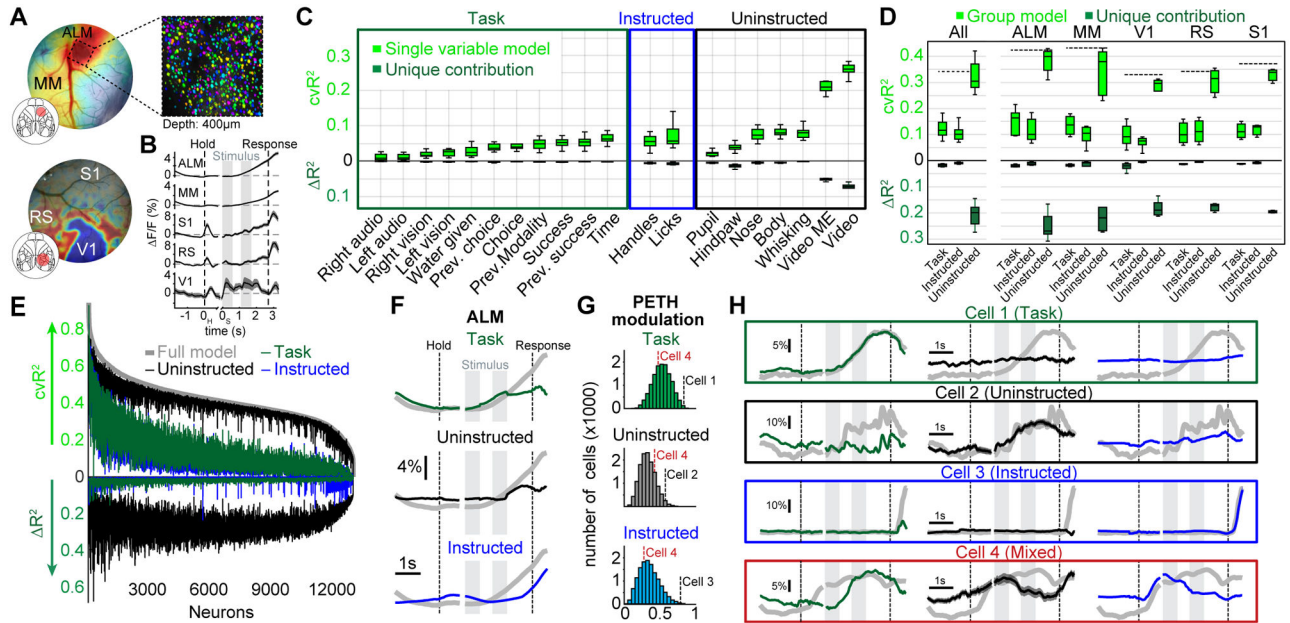
Author Manuscript

Author Manuscript

Author Manuscript

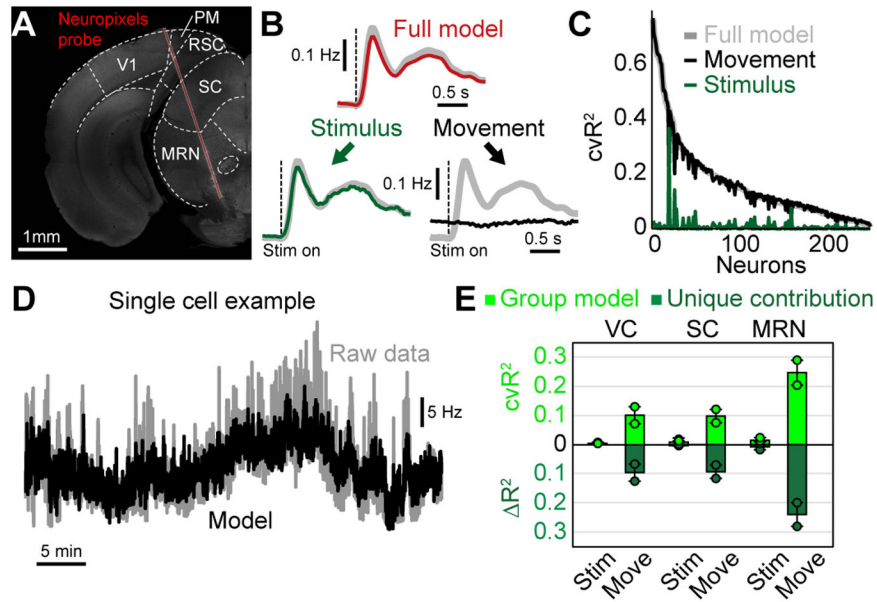
Author Manuscript





**Figure 7. Movements are important for interpreting single-neuron data.**

(A) Example windows in M2 (top) and posterior cortex (bottom). Colors in top window indicate neural response strength during licking. Colors in bottom window indicate location of visual areas, based on retinotopic mapping. Top right picture shows example field-of-view in ALM with neurons colored randomly. (B) Population PETHs from different cortical areas averaged over all trials and recorded neurons. Shading indicates SEM.  $n_{\text{ALM}}=4571$  neurons,  $n_{\text{MM}}=6364$  neurons,  $n_{\text{V1}}=594$  neurons,  $n_{\text{RS}}=206$  neurons,  $n_{\text{S1}}=252$  neurons. (C) Explained variance for single model variables. Shown is either all explained variance (light green) or unique model contribution (dark green). The box shows the first and third quartiles, inner line is the median over 10 mice, whiskers represent minimum and maximum values. Prev.: previous. (D) Explained variance for variable groups. Conventions as in (C). Left panel shows results for all areas, right panels show results for individual cortical areas. (E) Explained variance of variable groups for individual neurons, sorted by full-model performance (light gray trace). Traces above the horizontal axis reflect all explained variance and traces below show unique model contributions (similar to light and dark green bars in D). Colors indicate variable groups. (F) Partitioning of the ALM population PETH. Colors show contributions from different variable groups. Summation of all group contributions results in the original PETH (gray traces). (G) PETH modulation indices (MIs). Histograms show MIs for each variable group. Dashed lines show MI values for example cells in H. (H) PETH partitioning of single neurons. Boxes show example cells, most strongly modulated by the task (green), uninstructed movements (black), instructed movements (blue) or a combination of all three (red). Original PETHs in gray.



**Figure 8. Uninstructed movements predict single-neuron activity in cortical and subcortical areas**

(A) Coronal slice,  $-3.8$  mm from bregma, showing the location of the Neuropixels probe in an example recording. We recorded activity in visual cortex (VC), superior colliculus (SC) and the midbrain reticular nucleus (MRN). (B) Population PETH averaged over all neurons and visual stimuli,  $n = 232$  neurons. Gray traces show recorded data. Top red trace shows model reconstruction. Bottom: PETH partitioning of modeled data into stimulus (left, green trace) and movement (right, black trace) components. (C) Explained variance for all recorded neurons using either the full model (gray) or a movement- (black) or stimulus-only model (green). (D) Activity of a single example neuron over the entire session ( $\sim 40$  minutes). Gray trace is recorded activity, black is the cross-validated model reconstruction. (E) Explained variance for stimulus or movement variables in different brain areas. Shown is either all explained variance (light green) or unique model contribution (dark green). Bars represent the mean from 2 mice for VC, SC and MRN, respectively. Dots indicate the mean for each animal (on average 38.7 neurons per mouse and area). Error bars show the SEM.