nature portfolio

Peer Review File

Connectivity underlying motor cortex activity during goal-directed behavior

Corresponding Author: Dr Arseny Finkelstein

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
Finkelstein and colleagues explored the dynamics and functional connectivity of layer 2/3 of motor cortex in mice performing a novel sensory-guided task in which a water reward was retrieved at a single location in a lick spout array. They used two-photon calcium imaging to demonstrate that neurons are often tuned to location and timing. In an apparent difference between this study and other work, the timing of neuronal responses was often delayed, sometimes occurring after the movement. The authors then demonstrated that a subpopulation of neurons modulated their firing based on expected reward, indicating a role of the network on performance evaluation. To interrogate functional connectivity, the authors used an 'all optical' approach in which neurons expressing an opsin are activated, leading to changes in fluorescent responses in neighboring cells. Neurons that were functionally connected were shown to have more similar behavioral tuning and exhibit offline 'noise' correlations. Additionally, some neurons had larger than expected 'out-degree' functional connections, indicative of hub neurons. Overall, the experiments are impressive, and the analyses were carefully performed with appropriate controls. The scope of the study is also stunning (e.g., 22 million possible connections assayed, a Herculean effort!).

My primary concern with this study is that the whole does not rise above the sum of its parts. Each experimental result appears sound, but the story is somewhat fractured into a series of vignettes that are not particularly surprising. For instance, the result of nearby excitatory connectivity and more distant inhibitory responses has been observed across many different networks, and functional clustering of motor cortex has been described in papers that are not cited here (e.g., Georgopoulos et al., 2007; Dombeck et al., 2009). Reward responses in motor cortex have also been well-documented in both human and nonhuman primates as well as rodents. Finally, as the authors point out, others have seen hub neurons in cortical circuits as well as in models of connectivity. Therefore, the accumulated insight related to brain function does not – in my view – represent a significant singular advance over previous work. The study certainly has merit, but I think a specialized journal such as Nature Neuroscience would be a more appropriate platform.

Although my major concern is advancement over prior work, other issues are listed below:

That a head-fixed animal retrieving a reward from a lick spout array is 'naturalistic behavior' (e.g., Lns. 4 and 62) is certainly debatable. A more defensible term may be 'not overtrained'. However, I am not certain of that designation either. The authors point out that the motor cortex is in a state prior to learning and that experience will shape dynamics and connectivity. Because of the simplicity of this task, I wonder if mice have already become 'experts' prior to engaging with the experiment. Sensory-guided licking is one of the few behaviors that these animals experience in their home cages every day.

Mice are given typically 7 consecutive trials of repeated rewards at a single location ('Methods', Ln. 321-322); is there a change in responsiveness across these trials? Did responses or functional connections change across sessions?

The authors often point to instances of 'conjunctive tuning' (e.g., Ln. 100). I don't find these particularly compelling. For instance, 29.5% of the neurons changed activity due to reward, and 47.7% were tuned to location. A simple product of these numbers is 14.1%. Therefore, the fact that the authors find 13.8% are co-tuned is not surprising.

Throughout the document, the authors refer to the strength of the connections [e.g., 'weakly connected', 'strongly connected' (Ln 12), 'strongly influenced' (Ln 14), 'inhibitory interactions were… two orders of magnitude weaker than excitatory interactions' (Lns. 167-168)]. The all-optical approach provides a measure of the incidence of functional connections, rather than their strength, and such wording should be modified accordingly.

A major result is that the activity of layer 2/3 neurons in motor cortex appears to often outlast the stimulus. To illustrate this more directly, a variant of Fig. 1g should be plotted aligned to the end of the task (e.g., the removal of the lick spouts or the cessation of licking).

The number of trials of each type in the examples shown in Extended Data Fig. 5a-c should be listed in that figure. How do the authors interpret the differences between strong and weak target neuron responses? My concern is that strong responses may suggest spontaneous correlated activity that may lead to spurious activation of neurons and correlated fluorescence even in the absence of a functional connection.

The metric to determine stability of responses (even VS odd, with a 0.25 cutoff) seems potentially inappropriate given observed response distributions (Extended Data Fig 2f-h).

The 'reward outcome tuning was not explained by differences in licking' (Lns. 98-99) statement is seemingly violated by an example given in Extended Data Fig. 3b (Cell 6). Authors should quantify the number of lick responsive neurons recorded and perform a test (e.g., log-likelihood) to demonstrate that reward is more related to the responses than licking.

Lns. 119-120: 'We developed a method for two-photon photostimulation while evoked responses were measured in other excitatory neurons'. It is unclear how this statement can be made, given previously reported efforts (e.g., Packer, Tank, Histed, Harvey, Adesnik, others). This manuscript may have features not present in these existing studies, but it is certainly an overstatement to say that this approach was newly developed here.

What was the density of ChrimsonR+ neurons? Because the opsin was introduced virally, it would be helpful to know the percentage of expressing neurons and their relative spacing so that others can attempt this method, especially given the possibility of off-target stimulation.

Did the pairs with functional (polysynaptic) inhibition show any significant differences in tuning?

How were cell distances measured? Center-to-center or perimeter-to-perimeter in Euclidean space?

Descriptive statistics (e.g., number of mice, trials, etc.) should be moved to the results. How many male/female mice were used? Were results different across these populations?

Typos are scattered throughout the text: 'doral' (Ln 83), 'examples' (Ln. 110), 'experiment' (Ln. 292), 'lickpot' (Ln. 298), 'epoch' (Ln. 358), 'allowed to' (Ln. 408), 'shuffle' (Ln. 426), 'target' (repeated) (Ln. 489), 'individuals' (Ln. 530), '3.h' (Ln. 572), '1 <' (Ln. 648), '0' (Ln. 766). Also, an entire sentence was repeated (Lns. 31-33 and Lns. 34-37).

References are missing for the statement made in Ln 26: 'target location, muscle activation, effector velocity, and others'.

Referee #2

(Remarks to the Author)
Finkelstein et al combine behavior, large scale neural recording with all-optical connectivity mapping to test how task-locked neuronal representations associate with their local connectivities. The behavioral task requires naive mice to contact spouts at one of ~25 locations in a 5x5 grid of spout positions; spout contact results in water reward. Mice execute this task and the authors identify representations of target location, time-relative to first contact, reward, and reward-modulated gain of contact site representations. In the same session, the authors conduct connectivity mapping to examine how connectivities relate to the behavioral representations. With careful controls and forthright language about off-target activations and potential complications of polysynaptic activations, they quantify numerically the 3-D spatial profile of the photostimulation-associated follower network. First, follower photoactivation extended farther axially than laterally, consistent with the columnar nature of cortex. Second, tuning during behavior could subtly but significantly predict connectivity; e.g. neurons with similar location tuning tended to be nearby (<100 um), and, third, 'hub' neurons with outsized distant connectivity were identified and, interestingly, these neurons tended to have weaker tuning to behavioral variables such as reward and spout location. While I am mostly excited by the execution of the paper, I think there is room for improvement for clarifying/motivating what was actually discovered (see Main issue #1 below).

Strengths.

1. A true technical tour de force

To say this is a technical tour de force may be an understatement. Consider this section:

"we photostimulated neurons in the most superficial imaging plane while randomly switching photostimulation targets every 192 ms. This randomization of photostimulation targets, combined with volumetric imaging, allowed rapid (30 mins) mapping

of causal connections between ~300,000 pairs of neurons within a 700×700×120 127 µm ALM volume (~150 targeted × ~2000 imaged neurons) per session. In each session, we imaged the activity of 128 the same neurons during 1) baseline at rest, followed by 2) connectivity mapping at rest, and then 3) during multidirectional tongue-reaching behavior (Fig. 3c)."

This is an amazing set of experiments that few labs could pull off. Though past work (e.g. Chettih and Harvey, 2019; Rickgauer et al, 2014; Cossell et al, 2015; Daie et al, 2021) conducted mapping and recording in identical populations of neurons, the sheer scale of the experiments here is at a new level. The volumes of optically accessed tissue here hints at a new era in neuroscience which makes this paper very exciting.

2. Characterization of interesting neural representations in a new behavioral paradigm
*The mostly single peaks in location tuning echoed classic MC studies (e.g. Georgopoulos et al, 1981) and were satisfying to see in this task, even if not entirely conceptually novel in their own right (but see issue section 1 below for recommendations on improving this section).
*The reward outcome-modulation of the target responses was also interesting, if not entirely novel (Levy et al, 2020). Still, the dense sampling affords an important quantification of this phenomenon only hinted at by previous work.
Overall, the paradigm of naive mice licking to a grid of many targets elicited robust reward, timing and target selectivity in single ALM neurons, which was prerequisite to testing if representations predicted connectivity.

Main issue

1. Far and away the biggest area of improvement for this paper is that after all the hard-won work of characterizing interesting (though not entirely novel) neuronal responses and then correlating these responses to connectivity - it's still not clear what the big finding or discovery is. After enormous effort in experiment and analysis, the authors essentially report

(1) Numerical quantifications in Fig. 3 and in the paragraph starting at Line 164 report that laterally, excitation decayed over ~100um, beyond which inhibition dominated. Axially, excitation peaked at ~60um and extended further than laterally, supporting a widely acknowledged columnar structure of cortex going back to Mountcastle (Mountcastle, 1957) and supported even in brain slice studies (e.g. Yuste et al, 1992; Sakmann, 2017) .

(2) Stronger connectivity was associated with similar location tuning, consistent with observations that V1 neurons with similar orientation selectivity exhibit stronger connections.

(3) Perhaps the most novel and important discovery was that "Hub neurons showed weaker location tuning (Fig. 4g) and weaker reward modulation (Fig. 4h) than less-connected neurons" But this finding is tucked away deep in the results section without framing, setup, or interpretation.

Honestly at this point in reading the paper I felt like I was reading Hitchhiker's Guide to the Galaxy - when Deep Thought reveals that the answer to the "Great Question of Life, the Universe and Everything" is "forty-two."

I want to support this paper - and I doubt that this is how the authors want their reader to feel as they read through the quantifications of the mapping experiments. I fully understand that formatting for a Nature paper requires precision - but I think the paper will be dramatically improved if clearer hypotheses and/or ideas were provided to motivate these hard-won quantifications of representation-defined cortical connectivity.

Some examples: are there models of network function or RNNs where connectivity extends optimally co-vary with tuning? Are there tradeoffs between connectivity and tuning? If hub neurons exhibit reduced tuning specificity - what might their function be? To constrain a task- or behavioral state- pecific manifold in the network (e.g. as in Richman et al, 2023 Nature from Liqun Luo's lab)? Of course these are open questions for future studies but for this paper to achieve more widespread appeal for a general audience beyond the technical sophistication the paper could be strengthened if it was communicated what is actually at stake in the findings.

Critically, some text about an idea or hypothesis of cortical microcurcuit function that was RULED OUT by the present findings would go a long way in helping a general audience achieve a conceptual takeaway beyond the technical impressiveness of the combined mapping and measurement-during-behavior.

2. MINOR ISSUES

2.1. More description of the behavioral paradigm and associated representations

*The authors claim that the tuning of neurons is related to target location. While the authors make clear the benefits of using a 'naturalistic' task that requires no training, one drawback to this approach is that naive mice unfamiliar with the behavioral task/setup may exhibit large trial-to-trial and session-to-session variability in their licking relative to even a single, stationary target. For example, mice may lick 'straight' to a right target and still make contact on trial 1, but over the trials within a block, mice may learn to re-aim their licks 'more right' on trial 10 for example, to make better contact with the target. This leads to the question of whether single neurons in ALM are tuned to target location or current/upcoming lick direction? As the authors perform tongue tracking in this task along with connectivity mapping, this gives the authors the opportunity to answer this question. Are responses to 'target' similar independent of the mouse's lick direction, or is it dependent on lick direction? The various possibilities for how target sites are represented (e.g. is it related to lick direction, contact site, or both) are outside

the scope of the paper - as the authors primarily want to characterize reliable representations for the purpose of the circuit mapping. Still, a brief caveat in supplemental text or methods that states explicitly that target site is not a simple sensory input but has to be computed on the basis of integrating tactile, proprioceptive, and possibly efference copy signals would help the reader appreciate the nature of the target site representations. .

*Although behavior was not the primary focus of this paper, more panels in Extended data fig 1 on the behavioral performance and kinematics in the task across mice would be useful for specialists who care about details of the behavioral task. For example, what percentage of licks resulted in successful contacts? What was the variability in lick direction (lick angle) for each of the targets for each mouse? Did performance increase for trials within a block? Was there session-to-session increases in performance? Were there differences in lick kinematics dependent on target (e.g., did 'corner' targets - those to the upper/lower left/right - require longer pathlengths than the 'centered' targets) - and might this explain any of the results in the paper (e.g., Fig 1K)? It was not specified in the methods how many sessions an animal performed.

*Can the moment of water dispense be indicated in the main text and not only methods? Was dispensation on first, second, third contact?

*Fig 1k appears to pool data over left and right hemispheres and therefore information about ipsi/contra lateral target location bias in the representations is not seen, or perhaps I missed it. Can the authors please report ipsi/contra bias in their target representations?

2.2. Was there reward modulation of temporal tuned neurons?
The section on reward tuning immediately focuses entirely on the target location modulation by reward outcome - what about the modulation of the temporal tuning?

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors satisfactorily addressed my concerns. The resulting paper is considerably stronger in its present form, and I think it is now appropriate for publication in Nature.

Michael Long

Referee #2

(Remarks to the Author)
In this revision the authors address all of my initial concerns. The major issue in my initial assessment was not about the data, the analyses, or the claims, but that an opportunity was lost to link the major findings of cortical functional connectivity and representations to past work - so that the reader could know exactly what past models are supported or ruled out. This revision does a much better job here. In particular, the model in Figure 4 is welcome and links the findings of local recurrent connectivity and weakly tuned hub neurons to a conceptual model of cortical function.

The revised manuscript also more explicitly states which plausible models of cortical connectivity and function are ruled out. The relevant sections of the Discussion significantly enhance the accessibility of the work and strengthen the overall impact.

Extended Data Figure 1 now provides additional detail on tongue kinematics beyond spout location, which is important for interpreting the selectivity data. The authors present evidence that both spout position and endpoint tongue position contribute to selectivity, with neurons appearing to be tuned somewhat more strongly to spout location. However, the use of terms such as "predominantly" encoding spout position seems too strong, given the $r = 0.54$ vs $r = 0.37$ correlation values in the scatter plot comparing contributions of spout vs tongue position to neural coding. Clearly both factors play a role. I suggest tempering the language here; this is a minor but worthwhile point.
The revisions to the Discussion that clarify the implications for models of cortical connectivity and function are especially welcome.

Minor comments
Typo (Line 186): Do the authors mean "inability"?

Line 280: The finding of mixed representations of reward and preceding action is both novel and interesting. I suggest unpacking this more clearly for an advanced undergraduate neuroscience reader who may not be familiar with the distinction between vector and scalar feedback. In particular, these representations are really key for the implementation of model based RL.

Overall this is really a beautiful and timely paper that should be published with minimal delay.

Jesse Goldberg

Response to Reviewers

Referee #1 (Remarks to the Author):

Finkelstein and colleagues explored the dynamics and functional connectivity of layer 2/3 of motor cortex in mice performing a novel sensory-guided task in which a water reward was retrieved at a single location in a lick spout array. They used two-photon calcium imaging to demonstrate that neurons are often tuned to location and timing. In an apparent difference between this study and other work, the timing of neuronal responses was often delayed, sometimes occurring after the movement. The authors then demonstrated that a subpopulation of neurons modulated their firing based on expected reward, indicating a role of the network on performance evaluation. To interrogate functional connectivity, the authors used an 'all optical' approach in which neurons expressing an opsin are activated, leading to changes in fluorescent responses in neighboring cells. Neurons that were functionally connected were shown to have more similar behavioral tuning and exhibit offline 'noise' correlations. Additionally, some neurons had larger than expected 'out-degree' functional connections, indicative of hub neurons.

Overall, the experiments are impressive, and the analyses were carefully performed with appropriate controls. The scope of the study is also stunning (e.g., 22 million possible connections assayed, a Herculean effort!).

We thank the Reviewer for the encouraging words!

My primary concern with this study is that the whole does not rise above the sum of its parts. Each experimental result appears sound, but the story is somewhat fractured into a series of vignettes that are not particularly surprising.

In addition to developing a novel behavioral paradigm, our study presents six major findings. These findings offer insights into the organization and coding principles of the motor cortex. Although some of these principles, such as like-to-like connectivity, may have parallels in other brain regions, their discovery in the motor cortex is new.

1. **New type of tuning revealed in the naïve motor cortex**. We developed a multi-locational licking paradigm in mice, in analogy to arm-reaching or saccade tasks in primates. This behavior revealed a new type of tuning: target location tuned neurons, with responses that often occur after movement.
2. **Organization of tuning in three dimensions**. We uncovered the anatomical organization of temporal and location tuning.
3. **Reward and location tuning**. We found that reward modulates location-tuned neurons, which may have important implications for learning in neural circuits.
4. **Organization of excitation and inhibition *in vivo***: We uncovered columnar organization of excitation and inhibition in the motor cortex *in vivo*.

5. **Like-to-like connectivity in the motor cortex *in vivo***: We discovered "like-to-like" connectivity in the motor cortex, a new finding in the motor cortex.
6. **Connectivity motifs and their relation to tuning**: Our identification of hub neurons and their connectivity motifs *in vivo* is entirely new in relation to tuning and offers a deeper understanding of how connectivity relates to coding in the motor cortex.

We have reframed the manuscript, including a new conceptual model, to make the connections between the parts clearer. We believe that the whole now rises above the sum of the parts.

Point 1,1

For instance, the result of nearby excitatory connectivity and more distant inhibitory responses has been observed across many different networks, and functional clustering of motor cortex has been described in papers that are not cited here (e.g., Georgopoulos et al., 2007; Dombeck et al., 2009).

Local excitation and more distant inhibition has been observed in other brain regions, particularly in sensory cortices (Rosenbaum et al., 2017; Chettih & Harvey, 2019, Oldenburg et al 2024). However, much less is known about organizational principles in the frontal cortex. Compared to sensory cortex, motor cortex is in a different dynamical regime (Amsalem et al., 2024, Nature Communications) and has different thalamocortical organization. Furthermore, our work was performed in awake animals, whereas previous studies that characterized excitation/inhibition profiles with cellular resolution were done mostly in brain slices or under anesthesia, which alter the inputs and the state of the network. Beyond showing a Mexican-hat like motif of excitation/inhibition in the motor cortex with respect to distance in the cortical sheet, we also characterize its three-dimensional profile, including axial organization. Finally, we relate this profile of excitation and inhibition to the tuning of cells in the motor cortex. We added citations to other studies characterizing the anatomical profile of excitation and inhibition (Rosenbaum et al., 2017; Chettih & Harvey, 2019, Oldenburg et al 2024).

Regarding functional clustering, we thank the Reviewer for pointing out these references. We now include these in the Discussion section. By building on these results, we show for the first time a three-dimensional anatomical organization of the tuning, in a new behavior. This new behavior allowed us to characterize anatomical organization along different tuning dimensions (temporal and location tuning).

Most importantly, in our combined approach we relate for the first time functional clustering to the underlying connectivity in the motor cortex. We believe this mechanistic aspect was not emphasized enough in our previous submission. To emphasize this link between the mechanisms of functional organization of tuning and the underlying connectivity, we now discuss how different network mechanisms can contribute to local clustering. For instance, local clustering can be fully inherited from external inputs or have contribution from local recurrent connectivity. We show that local connectivity contributes to the observed clustering. We also include a model

(discussed in details in Reviewer 1 Point 3) that shows the advantage of specific network motifs, such as hubs, for tuning selectivity for cells organized in functional clusters.

Point 1,2

Reward responses in motor cortex have also been well-documented in both human and nonhuman primates as well as rodents.

We acknowledged in the original manuscript that:
'Neural activity in the parietal and the frontal cortex is correlated with action values [Platt and Glimcher 1999; Sugrue et al. 2004; Stuphorn et al. 2000; Bari et al. 2019; Hattori et al. 2019; Hirokawa et al. 2019]. ' We now emphasize this point more by citing additional reference showing reward-related activity in the motor cortex (Levy et al., Neuron 2020).

We believe that our paper is one of the first to highlight that reward outcome modulation is location (or action) specific, and thus may code for vectorial prediction error (i.e., 'what reward did I get after reaching *that specific* target?'). This is an important finding, because such vectorial error signals may be required for gradient deep learning in the brain (Lillicrap et al, Nat Rev Neurosc 2020).

Point 1,3

Finally, as the authors point out, others have seen hub neurons in cortical circuits as well as in models of connectivity.

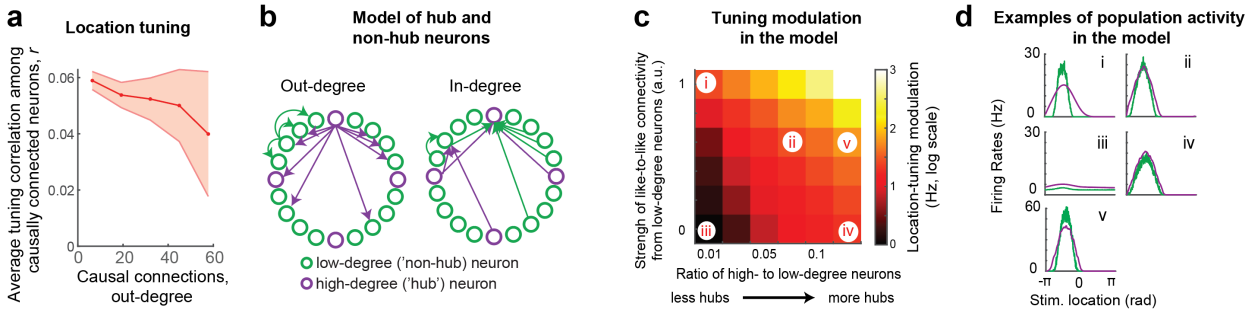
As we acknowledged in the original manuscript: 'Hub neurons were identified in the structural connectome of C. elegans [Towlson et al. 2013; Uzel et al. 2022], in rodent hippocampus and neocortex during development [Bonifazi et al. 2009; Bollmann et al. 2023], and predicted based on simulations of cortical columns [Gal et al. 2017]. ' Hub neurons have been previously reported in the hippocampus and the sensory cortex during development. Most of this work was done *in vitro* with no relation to neural coding. Our finding of hub in the adult motor cortex *in vivo* and the relation between hub neurons and tuning are new results.

We have added additional data analysis and modeling to make the link between the tuning properties and network connectivity clearer. In brief, we further analyzed the relation between tuning and degree distribution of local network connectivity (**Reviewer 1 Figure 1a**), and built a network model showing the possible functional roles of hubs versus non hubs in amplifying tuning selectivity (**Reviewer 1 Figure 1b-d**).

We examined the similarity in location tuning between a neuron and its connected neurons as a function of the neuron's out-degree. We found that neurons with fewer outgoing connections exhibited higher tuning similarity with their connected neurons compared to neurons with many outgoing connections (hub neuron; (**Reviewer 1 Figure 1a**). Importantly, despite this difference, hub neurons still showed considerable tuning similarity with the neurons they were connected to.

To get insights into potential computational role of hub neurons, we built a network model that captured the experimentally observed like-to-like connectivity and anatomical clustering of neurons with similar tuning. Specifically, we tested the hypothesis that weakly-tuned hub neurons can act as local amplifiers of tuning selectivity through recurrent connections to the rest of the neighborhood. To test this, we developed a simplified model of the motor cortex (**Reviewer 1 Figure 1b**), incorporating hub neurons based on our empirical findings: (1) most neurons exhibit unimodal location tuning, (2) neurons with stronger connectivity display more similar tuning, indicating like-to-like connectivity, (3) spatial clustering of similarly tuned neurons, and (4) a broad distribution of number of out-degree connections.

We found that hub neurons enhanced location tuning in the model (**Reviewer 1 Figure 1c-d**). Mechanistically, the hub and non-hub subnetworks act as two coupled ring-like networks (Ben-Yishai et al. 1995, Xie et al., 2002). The like-to-like connectivity within the non-hub subnetwork generated location-specific population activity, while hub neurons provided a broad but weak bias to nearby neurons, amplifying location-tuned activity and stabilizing bump-like dynamics. Thus, like-to-like connectivity in the non-hub subnetwork could be much weaker in the presence of connectivity from the hub subnetwork. This suggests that hub neurons may enhance network ability to encode spatial information by amplifying localized tuning (Douglas et al. 1995; Ben-Yishai et al. 1995).



Reviewer 1 Figure 1

a, The relation between number of causal outgoing connections (out-degree) of a neuron and its average location-tuning similarity with the tuning of neurons to which it is connected. Data is based on all significant causal connection pairs (n = 65,390 pairs; Methods); displayed as mean ± s.e.m. across neuronal pairs in bins.

b. A model of neurons tuned to location (one dimensional), with selective hub neurons. Neurons were divided into two subnetworks: low-degree ('non-hub', green) and high-degree ('hub', purple) neurons. Neurons were randomly connected to each other with a probability that decayed with their functional distance. Hub neurons were sparser but approximately ten times more connected than non-hub neurons. This resulted in hub neurons possessing higher in- and out-degree connectivity. Hub neurons summed their inputs from distant neurons, whereas non-hub neurons primarily connected locally.

c. Location-tuning modulation (max-min) of the non-hub neurons as a function of the like-to-like connectivity between non-hub neurons ($\bar{J}_{ll}$) and proportion of hubs in the network (α, Methods).

d. Examples of population activity in the hub and non-hub subnetworks under different conditions in (c); increasing the percentage of hub neurons amplifies the location-tuning in the network.

Together with the model and new data analysis we have a better intuition regarding possible computational advantages for the type of connectivity motifs and their relation to tuning observed in the motor cortex. These results are now part of **Figure 4** of the paper and accompanied by a new detailed Methods section.

Point 1,4

That a head-fixed animal retrieving a reward from a lick spout array is 'naturalistic behavior' (e.g., Lns. 4 and 62) is certainly debatable. A more defensible term may be 'not overtrained'. However, I am not certain of that designation either. The authors point out that the motor cortex is in a state prior to learning and that experience will shape dynamics and connectivity. Because of the simplicity of this task, I wonder if mice have already become 'experts' prior to engaging with the experiment. Sensory-guided licking is one of the few behaviors that these animals experience in their home cages every day.

Following Reviewer's comment we have removed the word 'naturalistic' from the manuscript title and Abstract. We now clarify in the Introduction that we define 'naturalistic' as a behavior that does not require training. We believe that this description accurately reflects the behavior under study – mice engage in goal-directed licking for water reward without the need for training, and perform hundreds of behavioral trials from the very first session. The lickport system used in our experiments is not present in the homecage, and all mice – regardless of the type of water dispenser available to them in their homecage – reliably perform the task.

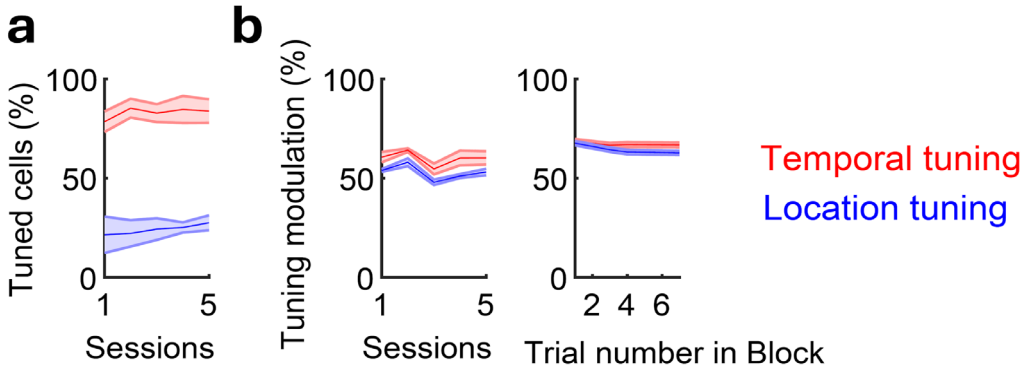
We acknowledge that the term 'naturalistic' is subject to interpretation, especially in laboratory settings where widely accepted naturalistic behaviors such as locomotion or foraging are studied in environments far removed from those encountered in nature. However, we believe that for practical purposes, the term 'naturalistic', as defined above, accurately conveys a key aspect of our behavioral task: the absence of a need for training, which differentiates it from highly structured, trained tasks typically used in neuroscience experiments.

Point 1,5

Mice are given typically 7 consecutive trials of repeated rewards at a single location ('Methods', Ln. 321-322); is there a change in responsiveness across these trials? Did responses or functional connections change across sessions?

We analyzed changes in temporal and location tuning across days and as a function of trial number within blocks of repeated locations. We found no change in the neural responses across sessions in terms of the percentage of significantly tuned cells (**Reviewer 1, Figure 2a**). We then assessed tuning modulation across sessions (**Reviewer 1, Fig. 2b, left**) and within blocks by comparing tuning in early, middle, and late trials (**Reviewer 1, Fig. 2b, right**), and again found no significant

changes. These results are now presented in **Extended Data Fig. 2e-f**.



Reviewer 1 Figure 2
a, Percentage of significantly tuned cells across sessions.
b, Tuning modulation across sessions (left) and trial number in block (right). Data is displayed as mean ± s.e.m. across mice (a, b left) or sessions (b, right).

Point 1,6

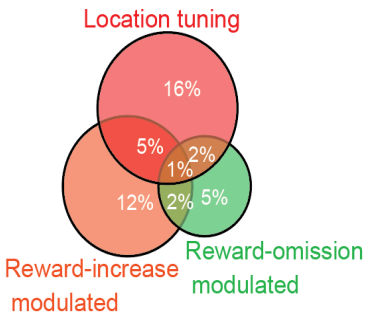
The authors often point to instances of 'conjunctive tuning' (e.g., Ln. 100). I don't find these particularly compelling. For instance, 29.5% of the neurons changed activity due to reward, and 47.7% were tuned to location. A simple product of these numbers is 14.1%. Therefore, the fact that the authors find 13.8% are co-tuned is not surprising.

We thank the Reviewer for this comment. We have implemented a shuffling-based approach throughout the manuscript to assess tuning significance (detailed in Methods and our response to Reviewer 1 Point 10 – where the Reviewer requested to use shuffling). This adjustment led to a lower percentage of neurons tuned to location and, consequently, a lower percentage of conjunctively tuned neurons to location and reward (see the updated Venn diagram in **Reviewer 1 Figure 3**). We then tested if the percentage of conjunctive cells can be expected from random overlap using chi-square test for independence.

For instance, after shuffling-based significance assessment, 22.8% of ALM neurons were tuned to location and 19.5% to reward-increase, with 5.7% showing conjunctive tuning – significantly higher than the expected random overlap ($p < 0.001$, chi-square test for independence). A similar pattern was observed for location tuning and reward omission: 9.9% of neurons were tuned to reward omission, and 2.7% were conjunctively tuned ($p < 0.001$, chi-square test for independence). Overall, 27.5% of all location-tuned neurons exhibited conjunctive tuning to reward outcome. We now report these statistical analyses in the main text.

Finally, as discussed in our response to Reviewer 1 Point 2 and in the Discussion, we consider conjunctive tuning to location and reward to be an important finding – regardless of whether it arises from random overlap or other specific mechanism. The presence of these neurons

indicates that reward outcome modulation is location (or action) specific, which suggests a potential role in encoding vectorial prediction error signals. Such signals have been proposed as necessary for gradient-based deep learning in the brain (Lillicrap et al., Nat Rev Neurosci, 2020).



Reviewer 1 Figure 3
Proportions of neurons with significant reward-outcome and location tuning in ALM.

Point 1,7

Throughout the document, the authors refer to the strength of the connections [e.g., 'weakly connected', 'strongly connected' (Ln 12), 'strongly influenced' (Ln 14), 'inhibitory interactions were… two orders of magnitude weaker than excitatory interactions' (Lns. 167-168)]. The all-optical approach provides a measure of the incidence of functional connections, rather than their strength, and such wording should be modified accordingly.

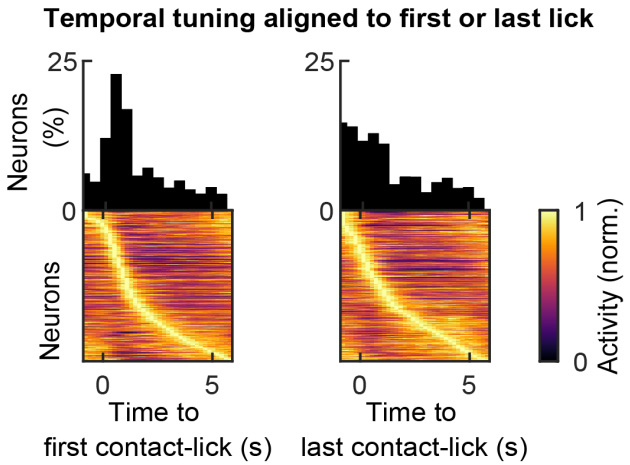
We now changed the wording throughout the paper to reflect that we are referring to the incidence of connections that corresponds to the number of connected neurons, rather than their strength. Specifically, the phrases previously describing neurons as 'weakly connected' or 'strongly connected' that the Reviewer mentioned have been replaced with wording such as *'abundant sparsely connected neurons'* and *'rare densely connected neurons'*, to more accurately reflect the prevalence of causal connectivity rather than its magnitude. Regarding the sentence referring to inhibitory interactions, we have added a clarification: *"… these weak inhibitory responses likely reflect our reduced ability to detect inhibitory connections rather than their actual strength."*

Point 1,8

A major result is that the activity of layer 2/3 neurons in motor cortex appears to often outlast the stimulus. To illustrate this more directly, a variant of Fig. 1g should be plotted aligned to the end of the task (e.g., the removal of the lick spouts or the cessation of licking).

We conducted the requested analyses, and computed the tuning aligned to 1) first contact lick (original analysis) and 2) last contact lick (**Reviewer 1 Figure 4**). In the original analysis, we found that 32.6% of neurons responded 2-6 s after first contact lick (typical licking duration was

<2 s). When aligning to last contact lick, we found that the peak response of 47.5% of neurons outlasted the last contact lick by at least one second, and for 32.5% of neurons by more than three seconds. This indeed suggests that for a large fraction of ALM neurons the neural activity outlasts the licking response. We now include these analyses in **Extended Data Fig. 2m** and mention them in the Results section.



Reviewer 1 Figure 4
Top, preferred response time of neurons. Bottom, neural activity of all neurons with significant temporal tuning; each row corresponds to the trial-average response of one neuron (of 82,148 neurons). Neural responses were aligned to the first (left) or last (right) contact lick.

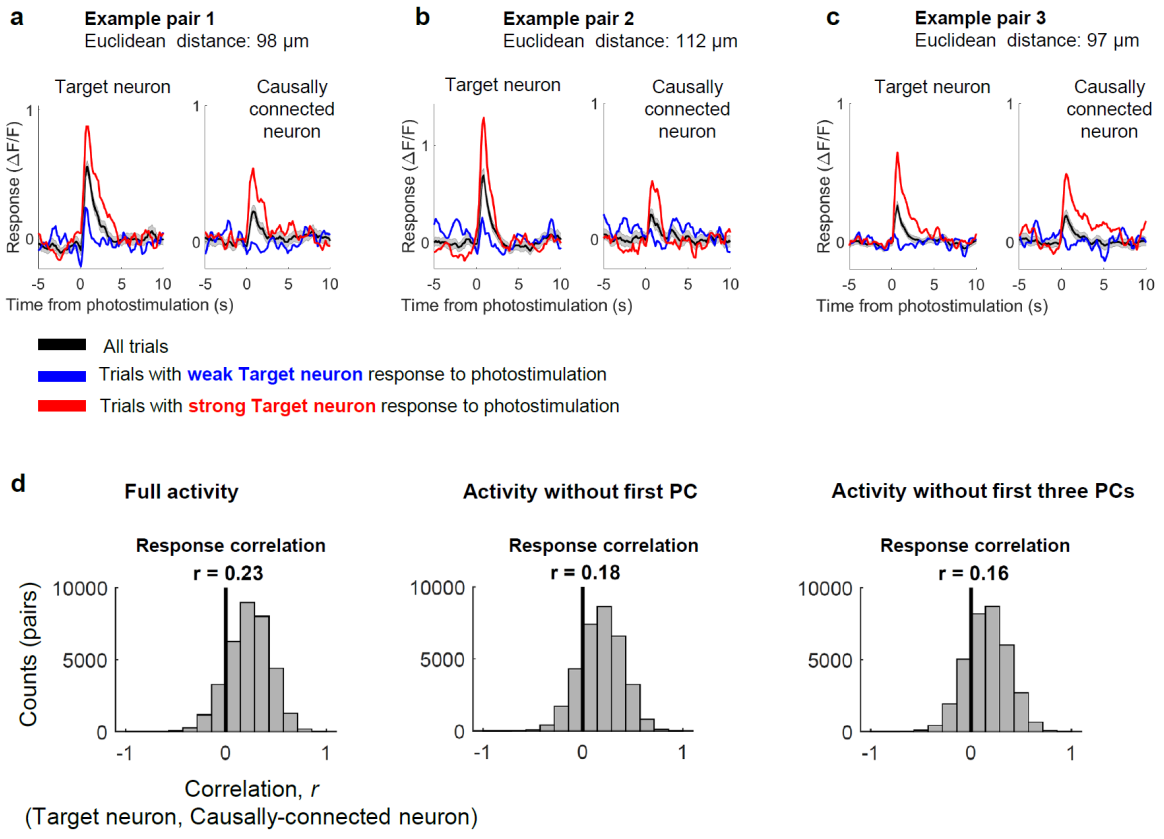
Point 1,9

The number of trials of each type in the examples shown in Extended Data Fig. 5a-c should be listed in that figure. How do the authors interpret the differences between strong and weak target neuron responses? My concern is that strong responses may suggest spontaneous correlated activity that may lead to spurious activation of neurons and correlated fluorescence even in the absence of a functional connection.

We thank the Reviewer for the comment. We have now added the number of trials of each type (strong and weak response trials) to the figure legend of **Extended Data Fig. 5a-c** (**Reviewer 1 Figure 5a-c**).

Variability in photostimulation-evoked responses can arise from multiple factors, including differences in photostimulation efficacy, intrinsic variability in neuronal excitability, and ongoing network activity. To address the potential contribution of spontaneous correlated activity to the observed evoked-response correlation, we performed an additional control analysis. Specifically, to remove shared global fluctuations that could reflect spontaneous network-wide activity we projected out the first principal component (PC; or the first three PCs) from the neural activity, and recomputed evoked response correlations. Even after removing these shared components, we

still observed positive correlation between the photostimulation evoked activity of target neurons and their causally connected neurons (**Reviewer 1 Figure 5d**). This suggests that the observed evoked response correlations are not merely a result of spontaneous network fluctuations but instead reflect causal connectivity. We added these results to the new **Extended Data Fig. 5d**.



Reviewer 1 Figure 5

a-c, Comparison of the amplitude of photostimulation evoked response of the target neuron and causally connected neurons, shown for 3 example neuronal pairs. Trials were categorized according to the strength of the evoked response (Δ Activity) of the target neurons, into trials with 'strong' or 'weak' response of the target (Methods). Responses of both target and causally connected neurons are shown by trial averaging the activity for strong (red), weak (blue), or all (black) trials. The number of trials in each category for example neuronal pairs: (a) 15 strong, 15 weak, 75 all; (b) 13 strong, 13 weak, 63 all; (c) 15 strong, 15 weak, 75 all.

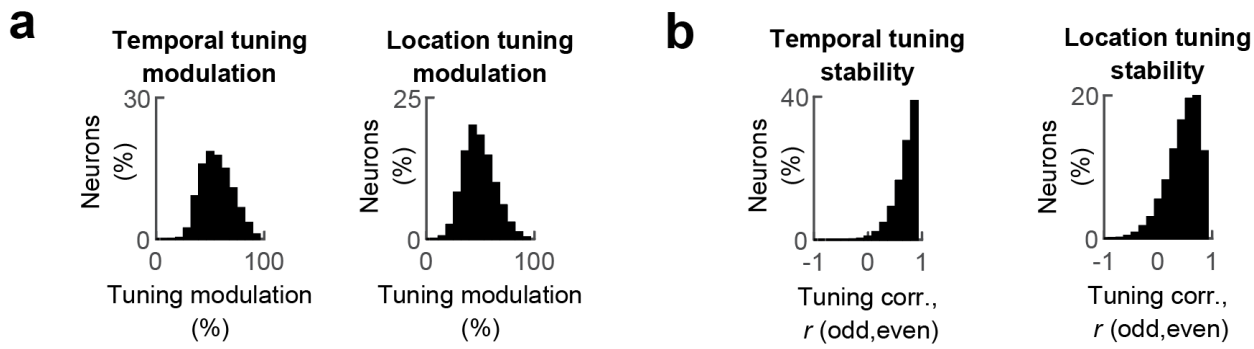
d, Distribution of Pearson correlation (*r*) between Δ Activity of target neurons and causally connected neurons on a trial-by-trial basis, shown for all neuronal pairs with significant causal connections (n = 33,830 connection pairs; Methods). To control for shared network activity that could lead to spurious evoked response correlations, we recomputed correlations after projecting out the first principal component of neural activity (middle) or the first three principal components (right). While the magnitude of correlations decreased, positive correlations persisted (average correlations $r = 0.24$ for full activity, $r = 0.18$ after removing the first PC, and $r = 0.16$ after removing the first three PCs), suggesting that the covariation in activity reflects functional interactions rather than spontaneous network fluctuations.

Point 1,10

The metric to determine stability of responses (even VS odd, with a 0.25 cutoff) seems potentially inappropriate given observed response distributions (Extended Data Fig 2f-h).

We thank the Reviewer for this comment. We now implemented a shuffling approach, and redefined the significance of tuning based on tuning modulation relative to the shuffled distribution (of each individual neuron). To independently verify this inclusion criterion, we also assessed the tuning stability metric (measured independently of tuning modulation). We updated the Results and Methods section in the manuscript accordingly.

Specifically, we defined the temporal tuning of a neuron to be significant if its temporal-tuning modulation was significantly different from shuffled distribution (at $p \leq 0.05$) obtained by circular permutations of the neural activity trace. Similarly, we defined the location tuning of a neuron to be significant if its location tuning modulation was significantly different from shuffled distribution (at $p \leq 0.05$) obtained by permutations of trial identity. This new criterion was more conservative and reduced the number of neurons with location-tuning (n = 23, 784; 22.8% of ALM neurons), without affecting much the number of neurons with temporal tuning (n = 82,148 neurons, 78.7% of ALM neurons). The resulting distribution of tuning modulation across all significantly tuned cells is shown in **Reviewer 1 Fig. 6a** (also in **Extended Data Fig. 2a-b**). This new shuffling-based criterion also increased the stability of the population of significantly tuned neurons compared to the previous criterion, with mean stability values of $r = 0.75$ for temporal tuning and $r = 0.53$ for location tuning (**Reviewer 1 Fig. 6b**, also shown as part of **Extended Data Fig. 2h-j**).



Reviewer 1 Figure 6

a, Distribution of temporal tuning modulation (left) and location tuning modulation (right) of all cells with significant tuning.

b, Distribution of stability of temporal tuning (left) and location tuning (right), computed as Pearson correlation (r) between tuning calculated based on odd versus even trials.

In both a-b we included cells with significant tuning based on tuning modulation shuffling (at $p \leq 0.05$).

Point 1,11

The 'reward outcome tuning was not explained by differences in licking' (Lns. 98-99) statement is seemingly violated by an example given in Extended Data Fig. 3b (Cell 6). Authors should quantify the number of lick responsive neurons recorded and perform a test (e.g., log-likelihood) to demonstrate that reward is more related to the responses than licking.

We thank the Reviewer for raising this point. First, we would like to note that the example highlighted by the reviewer (Cell 6) was shown deliberately as a *counterexample* – a neuron whose apparent reward tuning is indeed well explained by licking. We now further emphasize it in the figure legend.

Following Reviewer's suggestion to assess whether neural responses attributed to reward outcome could be explained by licking behavior, we replaced our original analysis by a more formal quantitative analysis using generalized linear models (GLMs) for each neuron, across the entire population. We compared two nested models:

- Model 1: time + licking
- Model 2: time + licking + reward

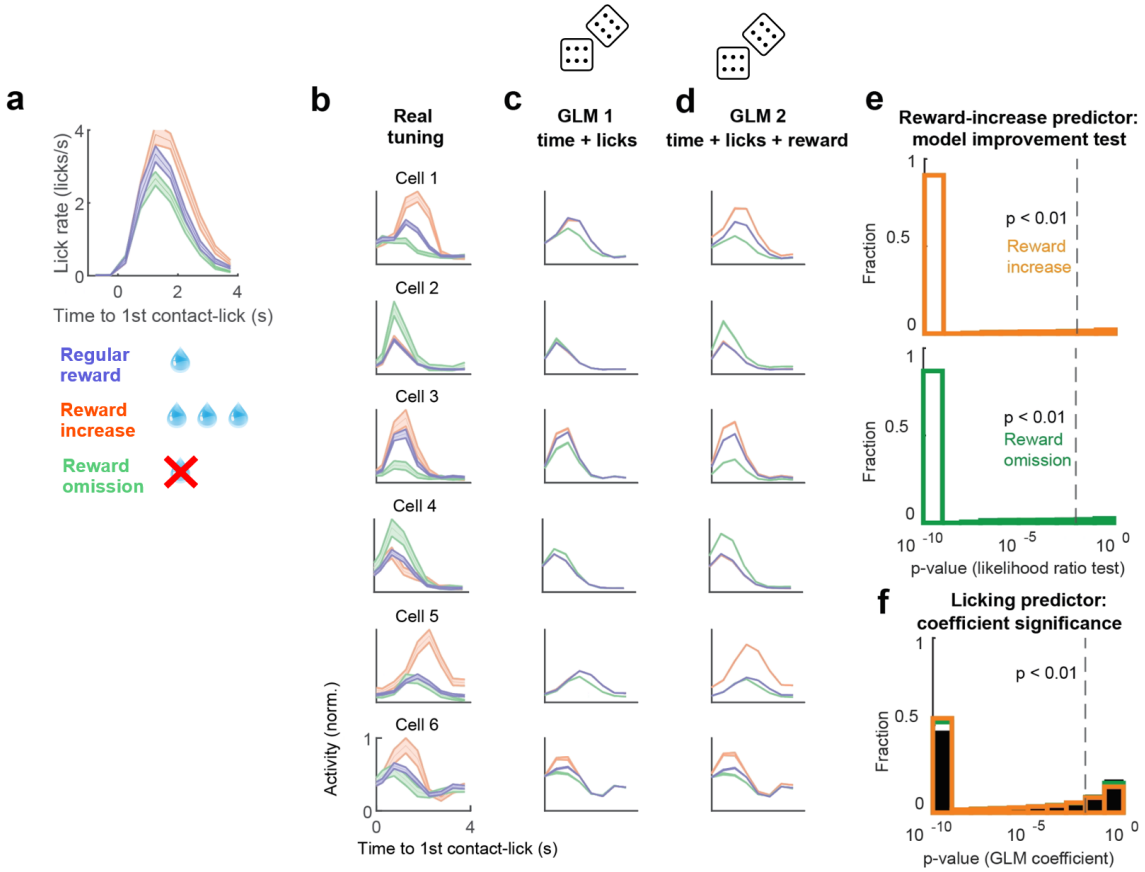
The first model included licking and trial time (represented using a cubic spline basis) as predictors, while the second model included the same predictors with the addition of the reward condition. Because Model 1 is a special case of Model 2 when the reward coefficient is zero, the models are nested and allow for statistical comparison via a Likelihood Ratio Test. Under the null hypothesis (that reward has no effect), the improvement in model fit when adding the reward predictor is expected to follow a chi-squared distribution with one degree of freedom (the degrees of freedom for this test are determined by the difference in the number of predictors between the models, which in this case is one).

Using this approach, we focused on neurons previously identified as reward-modulated based on their trial-averaged tuning curves (**Reviewer 1 Figure 7b, Fig. 2**), and asked whether the reward predictor significantly improved the GLM fit for these neurons (**Reviewer 1 Figure 7c-e**). We found that 95.5% and 96.3% of neurons modulated by reward omission and reward increase, respectively, also showed a significant effect of reward in the GLM ($p < 0.01$; **Reviewer 1 Figure 7e**). This demonstrates that, in the vast majority of these neurons (e.g., example cells 1-5), reward contributes significantly to explaining neural activity beyond licking and time. This analysis also confirms the simulated tuning curve approach presented originally, and provides a formal model-based quantification.

At the same time, licking was a significant predictor in 70.8% of all neurons ($p < 0.01$), including 74.1% and 76.5% of the reward-modulated neurons (omission and increase, respectively; **Reviewer 1 Figure 7f**), indicating widespread, but not exclusive, lick-related modulation. Among neurons that were defined as reward-modulated, significant lick-modulation was similar to the proportion of lick-modulation among the overall population (**Reviewer 1 Figure 7f**), again indicating that reward-modulation and lick-modulation are largely independent.

Together, these results support our original statement: reward tuning in the population is not explained by licking alone, but neurons are modulated by licking. We now included these

analyses in **Extended Data Figure 3**. We believe these analyses fully address the concern and strengthens the original conclusion that reward outcome tuning is not explained by licking differences alone.



Reviewer 1 Figure 7

a, Distribution of lick-rates at different times during the trial, for reward-increase, reward-omission, or regular trials.

b, Temporal tuning of six example cells with reward-outcome modulation, averaged across trials with different reward conditions.

c-d, Generalized linear model (GLM) predictions from Model 1 (c, time + licking) and Model 2 (d, time + licking + reward), for the same neurons shown in (b). Note that reward-related differences are captured in Model 2 but not in Model 1 for most neurons (Cells 1–5), while Cell 6 serves as a counterexample where licking does explains the apparent reward-outcome modulation

e, Histogram of p-values of the reward-increase (top) and reward-omission (bottom) predictors, assessed using a likelihood ratio test comparing GLM models with and without reward terms, for all reward-modulated neurons. In majority of neurons (e.g., Cells 1-5 in b), reward contributes significantly to explaining neural activity beyond licking and time.

f, Histogram of p-values for the licking predictor in Model 2, shown for all neurons (black), and for neurons modulated by reward-increase (orange) and reward-omission (green). Licking was a significant predictor in 70.8% of all neurons, including 74.1% and 76.5% of the reward-modulated neurons (omission and increase, respectively), indicating widespread, but not exclusive, lick-related modulation.

Data in a is shown as mean ± s.e.m. across sessions. Data in b-d are shown as mean ± s.e.m. across trials. A p-value threshold of 0.01 was used to determine significance in GLM-based tests.

Point 1,12

Lns. 119-120: 'We developed a method for two-photon photostimulation while evoked responses were measured in other excitatory neurons'. It is unclear how this statement can be made, given previously reported efforts (e.g., Packer, Tank, Histed, Harvey, Adesnik, others). This manuscript may have features not present in these existing studies, but it is certainly an overstatement to say that this approach was newly developed here.

We do not claim to have been the first to develop all-optical methods – in fact, we cited key prior work throughout the manuscript. Rather, we developed a specific protocol that modifies existing approaches to enable efficient and rapid probing of neural connectivity. In response to the Reviewer's comment, we revised the sentence in the Results section from "We developed a method" to "We adopted all-optical methods" to clarify that our approach builds on prior work. We also added the suggested citations throughout the paper.

Additionally, we modified the corresponding sentence in the Discussion to read: "We adopted all-optical methods based on two-photon optogenetics and calcium imaging [Rickgauer et al. 2014; Emiliani et al. 2015; Packer et al. 2015; Chettih and Harvey 2019; Daie et al. 2021; Randi et al. 2023; Oldenburg et al. 2024; LaFosse et al. 2024] to rapidly map causal connectivity between neurons in the intact brain."

Poin1,13

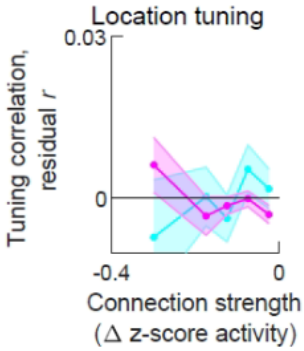
What was the density of ChrimsonR+ neurons? Because the opsin was introduced virally, it would be helpful to know the percentage of expressing neurons and their relative spacing so that others can attempt this method, especially given the possibility of off-target stimulation.

The percentage of ChrimsonR+ neurons was 74.1 ± 2.8 % (mean ± s.e.m. across sessions) out of all imaged neurons. The relative spacing between the photostimulated targets was 27.6 ± 0.6 µm (mean ± s.e.m. across sessions). In each experiment we photostimulated on average 28.4% targets (range 12.1- 43.1% across sessions) out of all imaged neurons in the top plane. We now added this info to the Methods section (Two-photon photostimulation protocol for causal connectivity mapping').

Minor points

Did the pairs with functional (polysynaptic) inhibition show any significant differences in tuning?

We checked the relationship between tuning and connection strength in pairs of neurons showing inhibitory connections. However, we did not observe a clear relationship between tuning and inhibitory connection strength. This may be due to the fact that measured inhibitory interactions between excitatory neurons were two orders of magnitude smaller than excitatory interactions.



Reviewer 1 Figure 8

Comparison of causal connectivity strength with residual correlations in location tuning, shown for target neurons (magenta) and control targets (cyan) as mean ± s.e.m. across sessions. Analysis was restricted to pairs of neurons showing inhibitory connections.

How were cell distances measured? Center-to-center or perimeter-to-perimeter in Euclidean space?

Distances were measured center-to-center. We now clarify this in the Methods: 'Here and throughout all the analyses, distances were measured between the centers of the cells.'

Descriptive statistics (e.g., number of mice, trials, etc.) should be moved to the results. How many male/female mice were used? Were results different across these populations?

We added descriptive statistics (number of male and female mice; number of sessions; number of trials per session) into the Results section. In this study we used 8 male and 7 female mice for imaging and behavioral experiments (61 sessions in total). We repeated all the analyses presented in the manuscript separately for male and female mice and found no apparent differences in behavioral performance, tuning, connectivity patterns, or other aspects. We now mention this in both Results and Methods section.

Typos are scattered throughout the text: 'doral' (Ln 83), 'examples' (Ln. 110), 'experiment' (Ln. 292), 'lickpot' (Ln. 298), 'epoch' (Ln. 358), 'allowed to' (Ln. 408), 'shuffle' (Ln. 426), 'target' (repeated) (Ln. 489), 'individuals' (Ln. 530), '3.h' (Ln. 572), '1 <' (Ln. 648), '0' (Ln. 766). Also, an entire sentence was repeated (Lns. 31-33 and Lns. 34-37).

We thank the Reviewer for pointing out these typos.

References are missing for the statement made in Ln 26: 'target location, muscle activation, effector velocity, and others'.

We decided to remove this sentence, because we do not focus on these aspects of the motor cortex in the paper.

Referee #2 (Remarks to the Author):

Finkelstein et al combine behavior, large scale neural recording with all-optical connectivity mapping to test how task-locked neuronal representations associate with their local connectivities. The behavioral task requires naive mice to contact spouts at one of ~25 locations in a 5x5 grid of spout positions; spout contact results in water reward. Mice execute this task and the authors identify representations of target location, time-relative to first contact, reward, and reward-modulated gain of contact site representations. In the same session, the authors conduct connectivity mapping to examine how connectivities relate to the behavioral representations. With careful controls and forthright language about off-target activations and potential complications of polysynaptic activations, they quantify numerically the 3-D spatial profile of the photostimulation-associated follower network. First, follower photoactivation extended farther axially than laterally, consistent with the columnar nature of cortex. Second, tuning during behavior could subtly but significantly predict connectivity; e.g. neurons with similar location tuning tended to be nearby (<100 um), and, third, 'hub' neurons with outsized distant connectivity were identified and, interestingly, these neurons tended to have weaker tuning to behavioral variables such as reward and spout location. While I am mostly excited by the execution of the paper, I think there is room for improvement for clarifying/motivating what was actually discovered (see Main issue #1 below).

Strengths.

Point 2,1

1. A true technical tour de force
To say this is a technical tour de force may be an understatement. Consider this section:
"we photostimulated neurons in the most superficial imaging plane while randomly switching photostimulation targets every 192 ms. This randomization of photostimulation targets, combined with volumetric imaging, allowed rapid (30 mins) mapping of causal connections between ~300,000 pairs of neurons within a 700×700×120 127 μm ALM volume (~150 targeted × ~2000 imaged neurons) per session. In each session, we imaged the activity of 128 the same neurons during 1) baseline at rest, followed by 2) connectivity mapping at rest, and then 3) during multidirectional tongue-reaching behavior (Fig. 3c)."
This is an amazing set of experiments that few labs could pull off. Though past work (e.g. Chettih and Harvey, 2019; Rickgauer et al, 2014; Cossell et al, 2015; Daie et al, 2021) conducted mapping and recording in identical populations of neurons, the sheer scale of the experiments here is at a new level. The volumes of optically accessed tissue here hints at a new era in neuroscience which makes this paper very exciting.

We thank the Reviewer for the encouraging words!

Point 2,2

2. Characterization of interesting neural representations in a new behavioral paradigm
*The mostly single peaks in location tuning echoed classic MC studies (e.g. Georgopoulos et al, 1981) and were satisfying to see in this task, even if not entirely conceptually novel in their own right (but see issue section 1 below for recommendations on improving this section).
*The reward outcome-modulation of the target responses was also interesting, if not entirely novel (Levy et al, 2020). Still, the dense sampling affords an important quantification of this phenomenon only hinted at by previous work.
Overall, the paradigm of naive mice licking to a grid of many targets elicited robust reward, timing and target selectivity in single ALM neurons, which was prerequisite to testing if representations predicted connectivity.

Thank you. Reward outcome modulation was indeed shown in the past, and we are now also citing Levy et al., 2020 in the Results and Discussion. We believe that out paper is one of the first to highlight that reward outcome modulation is location (or action) specific, and thus may code for vectorial prediction error (i.e., 'what reward did I get after reaching *that specific* target?'). This is an important finding, because such vectorial error signals are required for gradient deep learning in the brain (Lillicrap et al, Nat Rev Neurosc 2020). We now further emphasize this point in the Discussion.

Point 2,3

1. Far and away the biggest area of improvement for this paper is that after all the hard-won work of characterizing interesting (though not entirely novel) neuronal responses and then correlating these responses to connectivity - it's still not clear what the big finding or discovery is. After enormous effort in experiment and analysis, the authors essentially report
(1) Numerical quantifications in Fig. 3 and in the paragraph starting at Line 164 report that laterally, excitation decayed over ~100um, beyond which inhibition dominated. Axially, excitation peaked at ~60um and extended further than laterally, supporting a widely acknowledged columnar structure of cortex going back to Mountcastle (Mountcastle, 1957) and supported even in brain slice studies (e.g. Yuste et al, 1992; Sakmann, 2017).
(2) Stronger connectivity was associated with similar location tuning, consistent with observations that V1 neurons with similar orientation selectivity exhibit stronger connections.
(3) Perhaps the most novel and important discovery was that "Hub neurons showed weaker location tuning (Fig. 4g) and weaker reward modulation (Fig. 4h) than less-connected neurons" But this finding is tucked away deep in the results section without framing, setup, or interpretation. Honestly at this point in reading the paper I felt like I was reading Hitchhiker's Guide to the Galaxy - when Deep Thought reveals that the answer to the "Great Question of Life, the Universe and Everything" is "forty-two."

I want to support this paper - and I doubt that this is how the authors want their reader to feel as they read through the quantifications of the mapping experiments. I fully understand that formatting for a Nature paper requires precision - but I think the paper will be dramatically improved if clearer hypotheses and/or ideas were provided to motivate these hard-won quantifications of representation-defined cortical connectivity.

Some examples: are there models of network function or RNNs where connectivity extends optimally co-vary with tuning? Are there tradeoffs between connectivity and tuning? If hub neurons exhibit reduced tuning specificity - what might their function be? To constrain a task- or behavioral state- pecific manifold in the network (e.g. as in Richman et al, 2023 Nature from Liqun Luo's lab)? Of course these are open questions for future studies but for this paper to achieve more widespread appeal for a general audience beyond the technical sophistication the paper could be strengthened if it was communicated what is actually at stake in the findings.

We performed additional data analysis and modeling to link tuning properties and network connectivity, with an emphasis on hubs versus non-hub neurons. In brief, we further analyzed the relation between tuning and degree distribution of local network connectivity (**Reviewer 2 Figure 1a**), and built a network model showing the possible functional roles of hubs versus non hubs in amplifying tuning selectivity (**Reviewer 2 Figure 1b-d**).

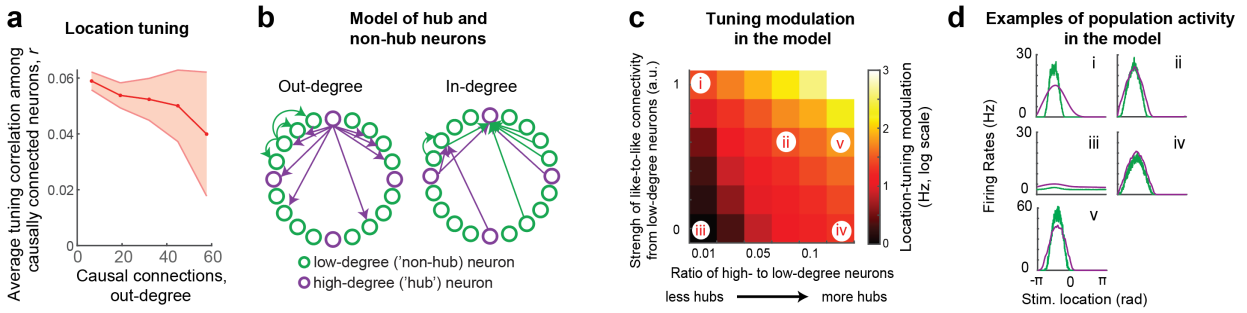
Specifically, we examined the similarity in location tuning between a neuron and its connected neurons as a function of the neuron's out-degree. We found that neurons with fewer outgoing connections exhibited higher tuning similarity with their connected neurons compared to neurons with many outgoing connections (hub neuron; (**Reviewer 2 Figure 1a**). Importantly, despite this difference, hub neurons still showed considerable tuning similarity with the neurons they were connected to.

To get insights into potential computational role of hub neurons, we built a network model that captured the experimentally observed like-to-like connectivity and anatomical clustering of neurons with similar tuning. Specifically, we tested the hypothesis that weakly-tuned hub neurons can act as local amplifiers of tuning selectivity through recurrent connections to the rest of the neighborhood. To test this, we developed a simplified model of the motor cortex (**Reviewer 2 Figure 1b**), incorporating hub neurons based on our empirical findings: (1) most neurons exhibit unimodal location tuning, (2) neurons with stronger connectivity display more similar tuning, indicating like-to-like connectivity, (3) spatial clustering of similarly tuned neurons, and (4) a broad distribution of number of out-degree connections.

We found that hub neurons enhanced location tuning in the model (**Reviewer 2 Figure 1c, d**). Mechanistically, the hub and non-hub subnetworks act as two coupled ring-like networks (Ben-Yishai et al. 1995, Xie et al., 2002). The like-to-like connectivity within the non-hub subnetwork generated location-specific population activity, while hub neurons provided a broad but weak bias to nearby neurons, amplifying location-tuned activity and stabilizing bump-like dynamics. Thus, like-to-like connectivity in the non-hub subnetwork could be much weaker in the presence of

connectivity from the hub subnetwork. This suggests that hub neurons may enhance network ability to encode spatial information by amplifying localized tuning (Douglas et al. 1995; Ben-Yishai et al. 1995).

With the model and new data analysis we now have a mechanistic view of computational advantages for the type of connectivity motifs and their relation to tuning as we observe in the motor cortex. These results are now part of **Figure 4** of the paper and accompanied by a new detailed Methods section.



Reviewer 2 Figure 1

a, The relation between number of causal outgoing connections (out-degree) of a neuron and its average location-tuning similarity with the tuning of neurons to which it is connected. Data is based on all significant causal connection pairs (n = 65,390 pairs; Methods); displayed as mean ± s.e.m. across neuronal pairs in bins.

b. A model of neurons tuned to location (one dimensional), with selective hub neurons. Neurons were divided into two subnetworks: low-degree ('non-hub', green) and high-degree ('hub', purple) neurons. Neurons were randomly connected to each other with a probability that decayed with their functional distance. Hub neurons were sparser but approximately ten times more connected than non-hub neurons. This resulted in hub neurons possessing higher in- and out-degree connectivity. Hub neurons summed their inputs from distant neurons, whereas non-hub neurons primarily connected locally.

c. Location-tuning modulation (max-min) of the non-hub neurons as a function of the like-to-like connectivity between non-hub neurons ($\bar{J}_{ll}$) and proportion of hubs in the network (α, Methods).

d. Examples of population activity in the hub and non-hub subnetworks under different conditions in (c); increasing the percentage of hub neurons amplifies the location-tuning in the network.

Critically, some text about an idea or hypothesis of cortical microcurcuit function that was RULED OUT by the present findings would go a long way in helping a general audience achieve a conceptual takeaway beyond the technical impressiveness of the combined mapping and measurement-during-behavior.

Combining naturalistic behavior with causal connectivity mapping allowed us to rule out three major hypotheses regarding the functional organization of motor cortical circuits:

Rejected Hypothesis #1: Neurons with similar tuning are randomly distributed ('salt-and-pepper' organization). Instead, we find that location-tuned neurons exhibit enhanced tuning similarity at short distances, particularly along the axial (depth) dimension (**Fig. 1n**), consistent with fine-scale columnar organization. This spatial pattern aligns with previous observations of functional clustering in the motor cortex [Georgopoulos et al., 2007; Dombeck et al., 2009; Komiyama et al., 2010] and contradicts the hypothesis that similarly tuned neurons are randomly intermixed, as seen in the rodent visual cortex [Ohki et al., 2005].
Thus, our data support a columnar organization of functional tuning in the motor cortex, and is inconsistent with salt-and-pepper anatomical distributions.

Rejected Hypothesis #2: Functional tuning and activity correlations arise solely from long-range inputs. We also rule out the hypothesis that tuning and coordinated activity patterns in the motor cortex arise exclusively from feedforward input without contributions from local recurrence [Douglas et al., 1995; Hubel & Wiesel 1962]. Instead, we find that causal connectivity strength predicts both tuning similarity (**Fig. 4a**) and activity correlations, indicating that local synaptic interactions contribute to functional properties. Moreover, the relationship between tuning, correlations, and connectivity strength cannot be explained by spatial proximity alone (**Extended Data Fig. 7**).
Thus, our results demonstrate that functionally specific local recurrent interactions contributes to tuning and coordinated activity in the motor cortex.

Rejected Hypothesis #3: Local connectivity follows random connection rules. Finally, we reject the hypothesis that local connectivity is random with respect to spatial and functional properties. Instead, we find:

I. Causal connectivity follows a Mexican-hat spatial profile – local excitation and longer-range inhibition (**Fig. 3g**). This spatial organization of connectivity mirrors the fine-scale columnar organization that we found for tuning (**Fig. 1n** – as discussed in relation to Hypothesis 1).
II. Like-to-like connectivity is confined within the 'column of excitation': neurons that are both tuned to similar target locations and located within short distances (<100 μm, where excitation dominates) are more likely to be causally connected (**Fig. 4a**; see also Hypothesis #2).

III. Outgoing connections distribution is inconsistent with an Erdős–Rényi network that is typically used in computational neuroscience. It consists of rare but highly connected hub neurons that are weakly tuned yet exert strong influence on local dynamics (**Fig. 4f-g**).

These findings indicate that local causal connectivity in the motor cortex is anatomically and functionally structured, rather than following random connection rules.

Taken together, our results reveal a cortical architecture in which functional tuning has a columnar structure with underlying non-random, spatially organized, and functionally specific local connectivity. The presence of weakly tuned but highly connected hub neurons further suggests that influence within the network is governed not only by selectivity but also by distinct roles in coordinating population dynamics.

These principles are now unified in a network model we added to the study (**Reviewer 2 Fig. 1,** now part of **Figure 4** of the manuscript), which synthesizes the findings that led to the rejection of Hypotheses 1-3. We also modified the text of the manuscript to highlight these hypotheses and the network model in the Results and Discussion sections.

2. MINOR ISSUES

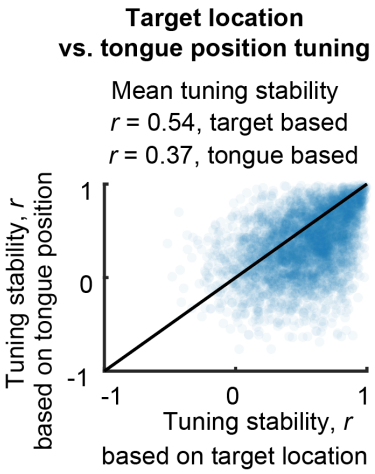
2.1. More description of the behavioral paradigm and associated representations

*The authors claim that the tuning of neurons is related to target location. While the authors make clear the benefits of using a 'naturalistic' task that requires no training, one drawback to this approach is that naive mice unfamiliar with the behavioral task/setup may exhibit large trial-to-trial and session-to-session variability in their licking relative to even a single, stationary target. For example, mice may lick 'straight' to a right target and still make contact on trial 1, but over the trials within a block, mice may learn to re-aim their licks 'more right' on trial 10 for example, to make better contact with the target. This leads to the question of whether single neurons in ALM are tuned to target location or current/upcoming lick direction? As the authors perform tongue tracking in this task along with connectivity mapping, this gives the authors the opportunity to answer this question. Are responses to 'target' similar independent of the mouse's lick direction, or is it dependent on lick direction?
The various possibilities for how target sites are represented (e.g. is it related to lick direction, contact site, or both) are outside the scope of the paper - as the authors primarily want to characterize reliable representations for the purpose of the circuit mapping. Still, a brief caveat in supplemental text or methods that states explicitly that target site is not a simple sensory input but has to be computed on the basis of integrating tactile, proprioceptive, and possibly efference copy signals would help the reader appreciate the nature of the target site representations.

We thank the Reviewer for this question. To address it, we analyzed tongue position using video tracking and compared the reliability of neural responses when computed based on tongue end-point positions (when reaching to different targets), versus target location regardless of tongue

position (as we did originally). To quantify tuning based on tongue position, we binned the tongue's position into the same number of bins as the target locations (e.g., 3×3) and measured the trial-averaged neural activity in each bin. We then compared the tuning stability across trials for both target location and tongue position. Tuning stability was assessed by dividing the trials into two subsets (odd and even) and comparing the correlation in the tuning maps for odd vs. even trials (defined as tuning stability). We found that tuning stability was significantly higher when computed based on target location compared to tongue position (see **Reviewer 2 Fig. 2**; most cells fall below the identity line). Specifically, the mean stability across cells was $r = 0.54$ based on target location, and $r = 0.37$ based on tongue position ($p \leq 0.001$, paired Student's t-test). This suggests that ALM neurons predominantly encode target location rather than tongue position. We have added these analyses to **Extended Data Fig. 2**, and the Results and Methods sections.

In addition, we conducted additional detailed behavioral analysis as described in response to Reviewer's next comment below.



Reviewer 2 Figure 2

Stability of location-tuning maps computed using two approaches: (1) based solely on target location, regardless of tongue position, and (2) based on tongue end-point positions when reaching for different targets. The scatter plot shows tuning stability (correlation between tuning maps from odd and even trials) computed based on tongue position or target location. Each dot represents a cell. Mean tuning stability across cells was significantly higher for target location compared to tongue position, indicating that neural activity is more reliably tuned to target location than to tongue position.

*Although behavior was not the primary focus of this paper, more panels in Extended data fig 1 on the behavioral performance and kinematics in the task across mice would be useful for specialists who care about details of the behavioral task.
For example, what percentage of licks resulted in successful contacts?
What was the variability in lick direction (lick angle) for each of the targets for each mouse?
Did performance increase for trials within a block?
Was there session-to-session increases in performance?

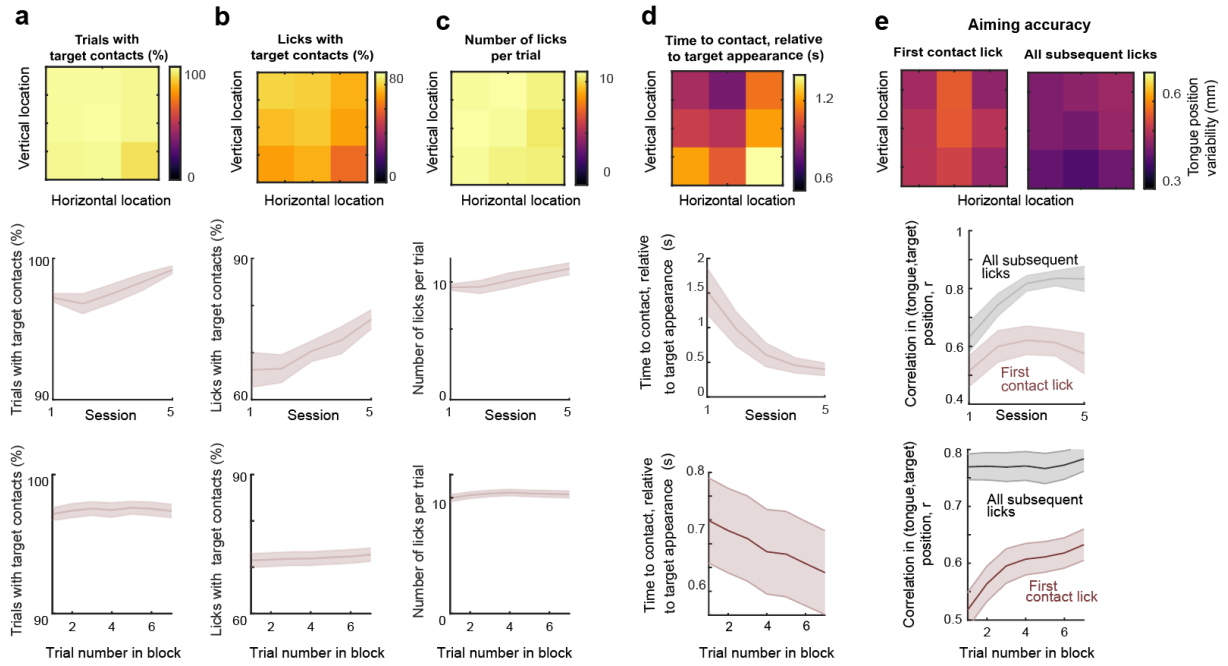
Were there differences in lick kinematics dependent on target (e.g., did 'corner' targets - those to the upper/lower left/right - require longer pathlengths than the 'centered' targets) - and might this explain any of the results in the paper (e.g., Fig 1K)?

Following the Reviewer's comment, we conducted an extensive analysis of behavioral performance and tongue kinematics during the task. These results are now summarized in **Reviewer 2 Figure 3** and included in **Extended Data Figure 1**.
To assess task performance, tongue kinematics and learning trends, we measured:

1. **Percentage of trials with target contacts** (**Reviewer 2 Figure 3a**). We found that nearly all trials resulted in successful target contacts, with performance close to 100% across all target locations already from the first session, confirming that the tasks did not require training.
2. **Percentage of licks with target contacts** (**Reviewer 2 Figure 3b**). While nearly all trials contained target contact, on average only ~70% of all licks resulted in contact with targets. Target at off-center locations, and in the bottom corners of the grid had higher incidence of licks that missed the target. We observed a session-to-session increase (but no improvement within a block) in the percentage of licks that resulted in contact with the target, suggesting a gradual learning effect.
3. **Number of licks per trial** (**Reviewer 2 Figure 3c**). There were on average ~10 licks per trial, and this number was rather constant across all target locations, blocks and sessions.
4. **Time to first contact (Reviewer 2 Figure 3d**) was longer for off-center targets and those in the bottom corners. Time to first contact decreased both within a block and across sessions (reaching a plateau after three sessions), suggesting both gradual learning and improved motor planning and execution on a trial-by-trial basis.
5. **Aiming accuracy and variability in lick direction** (**Reviewer 2 Figure 3e**). We examined aiming accuracy as correlations between tongue and target position for each lick, separately for the first contact lick and all subsequent licks. The first contact lick showed greater variability, particularly for central targets, whereas subsequent licks became more precise, with rather uniform accuracy across target locations. Interestingly, the accuracy of the first lick improved both within a block (i.e., across trials) and across sessions, reaching a plateau after approximately three sessions. In contrast, the accuracy of subsequent licks showed a steady improvement across sessions but did not significantly change within a block. Taken together, this suggests that initial tongue movements in each lick bout (of ~10 licks) within a trial are exploratory followed by corrective adjustments in subsequent licks that further improve with experience.

In summary, our results indicate that corner targets (upper/lower left/right), which are positioned farther from the center, were harder to reach. This did not demotivate the mouse from trying to reach them (no effect on overall number of licks; **Reviewer 2 Figure 3**c, top) but it was harder for the mouse to contact these targets accurately. Indeed, we observed smaller percentages of licks

resulting in contacts (**Reviewer 2 Figure 3b**, top) and longer times to first contact for off-center targets (**Reviewer 2 Figure 3d**, top), and less precise aiming (**Reviewer 2 Figure 3e**, top). These findings suggesting that off-center targets were harder to aim, may explain why the distribution of all location-tuning peaks across cells was not uniform, with more cells having peaks at the corners (**Fig. 1k**). Furthermore, as we show in **Reviewer Figure 2**, ALM neurons predominantly encode target location rather than tongue position (now shown in **Extended Data Fig. 2k**). Taken together, this suggests that a combination of factors – such as target location, motor planning, and sensory feedback – likely contributes to both the behavioral performance and the corresponding neural tuning. We added these results to the main text, **Extended Data Figure 1** and a detailed Methods section.



Reviewer 2 Figure 3

a-e, Lick statistics and tongue kinematics as a function of target horizontal × vertical location (top), session number (middle), or trial number in block of repeated locations (bottom). Data is displayed as mean ± s.e.m. across mice (middle) or sessions (top, bottom).

It was not specified in the methods how many sessions an animal performed.

We performed on average 4 behavioral sessions per animal, with a maximum of 7 sessions. We added this information to the Results and Methods.

*Can the moment of water dispense be indicated in the main text and not only methods? Was dispensation on first, second, third contact?

Water was dispensed after the first or second contact in 55.6% and 44.4% of trials, respectively. We now added this clarification to the first paragraph of the Results section.

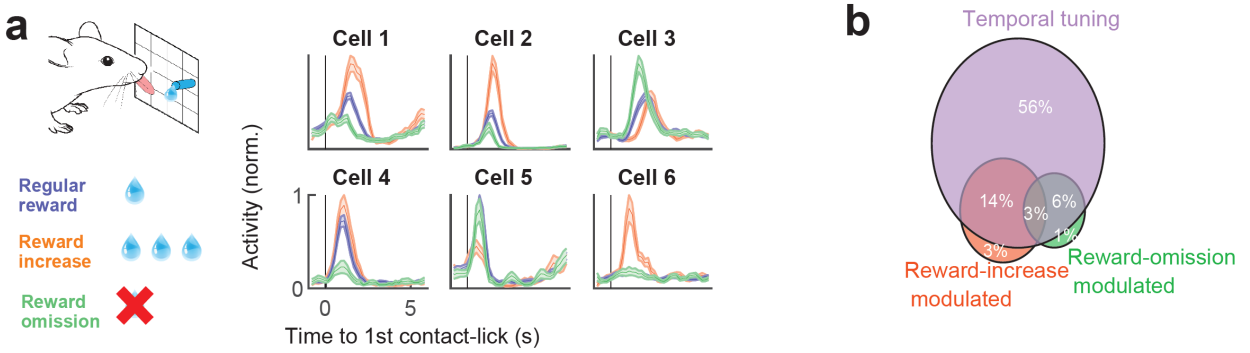
*Fig 1k appears to pool data over left and right hemispheres and therefore information about ipsi/contra lateral target location bias in the representations is not seen, or perhaps I missed it. Can the authors please report ipsi/contra bias in their target representations?

Following Reviewer's comment, we calculated the target location preference in neurons with significant location tuning, and found that 51.7% of cells were tuned to left targets, and 48.3 % of cells to right targets (we only imaged from the left hemisphere in this study). This lack of ipsi/contra bias is consistent with previous studies showing that ALM has equal proportion of neurons predicting ipsi- or contra lateral movements in directional licking task (Li et al., Nature 2015), which might reflect tongue's unique anatomy (i.e., in contrast to paws, eyes, whiskers etc., the tongue is not a paired organ). We now mention this result in the Results section.

2.2. Was there reward modulation of temporal tuned neurons?

The section on reward tuning immediately focuses entirely on the target location modulation by reward outcome - what about the modulation of the temporal tuning?

The majority of reward outcome neurons (85.9%) had a significant temporal tuning. We showed examples of such cells in **Fig. 2b** (shown in **Reviewer 2 Figure 4a**) and now also added a Venn diagram that quantifies the percentages of reward outcome cells with temporal tuning (new panel in **Fig. 2d**, right; shown in **Reviewer 2 Figure 4b**) and also mention this quantification in the Results section.



Reviewer 2 Figure 4

a, Reward was randomly increased or omitted on 20% of the trials. Temporal tuning of six example cells that are modulated by reward outcome, averaged across trials with regular reward (blue), reward increase (orange), or reward omission (green).

b, Proportions of neurons with reward-outcome modulation and temporal tuning in ALM.

Response to Editorial Comments

We updated the manuscript according to Editorial Request (our response is summarized in the Editorial requests table) and added all the missing information to the Reporting Summary and Methods.

We filled in the Third-Party Rights Table information for the mouse illustration in Fig. 1a and Fig. 2a and acknowledged the source in the figure legends. The remaining graphical illustrations in all figures submitted (e.g., dice, water droplets, and schematic illustrations) are now made by us.

Reporting summary comments:

Statistics

We added the required information to the figure legends and Methods.

Software and Code

We added the required information to the Reporting Summary and Methods and provided link to GitHub repository for custom code we used.

Data

We made the data publicly available in NWB format on DANDI archive, and added the required information to the Reporting Summary and Data Availability section in the Methods. Data is now available in NWB format on DANDI archive, and we now provided the link.

Life Science Study design

- We added the information about sample size to the Reporting Summary.
- We added this information about replication to the Reporting Summary and Statistics and Reproducibility section in the Methods.
- We added clarification about randomization to the Reporting Summary.
- We added clarification about blinding to the Reporting Summary.

Animals and Other Research Organisms

- We added information about dark/light cycle, ambient temperature, and humidity to the Methods.
- We added the information about mouse age to the reporting summary (it is also mentioned in the Methods).

Response To Remaining Comments by Reviewer 2:

Referee #2 (Remarks to the Author):

In this revision the authors address all of my initial concerns. The major issue in my initial assessment was not about the data, the analyses, or the claims, but that an opportunity was lost to link the major findings of cortical functional connectivity and representations to past work - so that the reader could know exactly what past models are supported or ruled out. This revision does a much better job here. In particular, the model in Figure 4 is welcome and links the findings of local recurrent connectivity and weakly tuned hub neurons to a conceptual model of cortical function.
The revised manuscript also more explicitly states which plausible models of cortical connectivity and function are ruled out. The relevant sections of the Discussion significantly enhance the accessibility of the work and strengthen the overall impact.

Thank you.

Extended Data Figure 1 now provides additional detail on tongue kinematics beyond spout location, which is important for interpreting the selectivity data. The authors present evidence that both spout position and endpoint tongue position contribute to selectivity, with neurons appearing to be tuned somewhat more strongly to spout location. However, the use of terms such as "predominantly" encoding spout position seems too strong, given the r = 0.54 vs r = 0.37 correlation values in the scatter plot comparing contributions of spout vs tongue position to neural coding. Clearly both factors play a role. I suggest tempering the language here; this is a minor but worthwhile point.

We tempered the language and write that: "Furthermore, ALM neurons showed more robust tuning to target location rather than to endpoint tongue position (**Extended Data Fig. 2k**), though both factors contributed. Taken together, this suggests that target location representation likely arises from an integration of target location, motor planning, and sensory feedback."

Minor comments
Typo (Line 186): Do the authors mean "inability"?

Thank you, we corrected it.

Line 280: The finding of mixed representations of reward and preceding action is both novel and interesting. I suggest unpacking this more clearly for an advanced undergraduate neuroscience

reader who may not be familiar with the distinction between vector and scalar feedback. In particular, these representations are really key for the implementation of model based RL.

We modified the related part in the Discussion to unpack this concept:

"In goal-directed behaviors, action selection is influenced by the expected outcome of specific actions. Neural activity in the parietal, retrosplenial, and the frontal cortex correlates with action values [46,25,47,48,26,27]. In ALM, we found neurons that conjunctively encode the target location and reward outcome (**Fig. 2**). This means they signal not only *how much* reward was obtained, but also *where in motor space* the action that led to that reward took place. Such 'vector feedback' contrasts with scalar feedback (e.g., more or less reward without specifying which action produced it). Vector feedback directly links outcomes and specific actions, allowing neural circuits to update action values at the appropriate locations in space [28,49]. Vector feedback is essential for efficient learning, since updating the behavioral policies by modifying the synaptic weights based on scalar feedback alone is notoriously slow [50]. Representations that mix reward outcome with preceding actions could provide a substrate for implementing model-based reinforcement learning, where predictions about future outcomes must be linked to specific actions to inform adaptive behavior."