

Form, Timing, and Behavioral Matching of the Population Response to ALM Photoinhibition

Zihan Yang

zihan.yang.zy366@yale.edu

Yale University

Research Article

Keywords:

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11447300/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Title:

Form, Timing, and Behavioral Matching of the Population Response to ALM Photoinhibition

Author:

Zihan Yang¹

Affiliation:

¹Department of Neuroscience, Yale University, 333 Cedar Street, New Haven, Connecticut 06510, USA

Correspondence:

zihan.yang.zy366@yale.edu

Abstract

Low-dimensional coding axes give a compact description of task-related population activity, but how they relate to the population change produced by a causal perturbation is less clear. We reanalyzed multi-probe Neuropixels recordings spanning anterior lateral motor cortex (ALM), striatum, thalamus, midbrain, and medulla in mice performing a memory-guided directional licking task, with bilateral ALM photoinhibition during the last 0.5 s of the delay. We compared the laser–control population difference with choice and action axes fit on control trials, and followed it from laser onset to the lick. In the 20–120 ms after laser onset, only 5–8% of the difference lay within the joint choice–action subspace, although the difference was far larger than that between size-matched splits of control trials. After the Go cue, single trials became increasingly predictive of lick direction, but the share of the mean laser–control difference on the choice axis did not rise with it. When laser and control trials were matched for lick direction and reaction time, laser–control distances still exceeded control–control distances in all five regions; in ALM, this residual was much smaller before the lick than during photoinhibition. Together, these results distinguish the population form of a perturbation from its temporal relationship to task information, and show that matching the measured behavioral outcome does not ensure matching of the population state.

Introduction

Neural population activity is often summarized by a small number of dimensions that track task variables such as sensory input, choice, and action.^{1–6} These descriptions capture how population activity covaries with behavior and have become a standard way to relate neural dynamics to task performance.^{7–11} Perturbation experiments approach the same circuits from the other direction, altering activity and measuring the consequences elsewhere.^{8,12–16} When the two approaches are combined, it is not obvious how the population change produced by an intervention should relate to the dimensions that describe task activity.

A perturbation has a form, meaning how the induced change is distributed across population dimensions, and a time course, meaning how that change evolves as later neural and behavioral processes unfold. Neither has to follow the task-coding dimensions. A perturbation can produce a large population change that is not directed mainly along a coding dimension,^{17,18} and an early displacement need not occupy the same dimensions as activity closer to an action.^{19,20} Likewise, growing information about an upcoming behavior does not imply that the mean population difference caused by the perturbation has moved onto the corresponding coding dimension. The geometry of a perturbation and the task information carried by the population are related, but one cannot stand in for the other.

A separate issue arises when trials are compared at the same behavioral outcome. The same action can be produced from different population states, particularly after an external perturbation.^{8,21} Matched behavior alone does not show that the population has returned to an equivalent state. Any remaining difference might simply record the history of the perturbation, or it might change how the system responds to later inputs. That second possibility cannot be examined until it is known whether a population difference remains once the measured behavior has been matched.

The distributed circuit underlying memory-guided licking is well suited to these questions.^{22–25} Chen et al.²⁶ used multiple simultaneously inserted Neuropixels probes²⁷ to record across ALM, striatum, thalamus, midbrain, and medulla. They found choice- and movement-related signals distributed across these regions, with subcortical choice coding depending on ALM. Other work has shown that activity outside explicit task-coding dimensions can still take part in population dynamics and shape later activity.^{17,28,29} What remains unclear is where the population change produced by the perturbation itself lies relative to the coding dimensions, how that relationship changes over time, and what difference remains when behavior is held equal.

We addressed these points by reanalyzing the publicly available recordings from late-delay bilateral ALM photoinhibition. We first measured how much of the early laser–control population difference was captured by choice and action dimensions fit only on control trials. We then asked whether the rise in trial-level information about lick direction was accompanied by a rise in the alignment of the laser–control difference with those dimensions. Finally, we matched laser and control trials for lick direction and reaction time and asked whether a population difference remained beyond matched control–control variability. The early response lay mostly outside the choice–action subspace, the growth of lick information was not matched by growth in choice-axis alignment, and population differences remained after behavioral matching in all five regions. Whether these remaining differences affect later neural activity is left open.

Results

We analyzed the population response to bilateral ALM photoinhibition during the last 0.5 s of the delay, using recordings from ALM, striatum, thalamus, midbrain, and medulla.²⁶ We examined three aspects of this perturbation: the direction of the population difference 20–120 ms after laser onset, the time at which single trials became predictive of lick direction, and the population difference that remained after lick direction and reaction time were matched. Analyses used late-delay bilateral sessions. Mouse-level estimates are means $\pm$ SEM (Table 1).

The early population difference lies mostly outside the choice and action axes

Most of the population difference 20–120 ms after bilateral laser onset lay outside the axes that distinguish lick direction. Those axes were fit on control trials only: lick right minus lick left, from 600 ms before Go until Go for the choice axis, and from Go to 300 ms after Go for the action axis. Jaw, nose, and tongue position were removed from the choice-axis fit. The early difference was laser minus control, 20–120 ms after laser onset, within instructed side (Fig. 1A).

Across regions, the joint fraction of that difference inside the two-axis subspace was 0.052–0.080 (Table 1; Fig. 1C,D). The choice axis alone held 0.030–0.059 and the action axis 0.021–0.032. Thalamus had the highest joint fraction (0.080 ± 0.016) and ALM the lowest (0.052 ± 0.011 ; 19 mice, 93 sessions). In a pre-specified ALM session near that mouse mean, laser and control means were separated mainly along the early difference orthogonal to both axes (Fig. 1B).

The axes are not orthogonal. Their cosine ranged from 0.29 ± 0.04 in thalamus to 0.52 ± 0.03 in ALM, which is why the joint fraction, rather than the angle, is the summary (Fig. 1-1A). The population difference itself was large. Relative to a size-matched split of control trials, the laser difference was larger in every mouse. The median ratio across mice was 152 in ALM and 3.3–8.7 in the other regions (Fig. 1E). After the control-split projection was removed, joint fractions remained low, with mouse means ranging from about 0.003 in medulla to 0.099 in thalamus (Fig. 1-1B). Joint fractions decreased with the number of neurons in a session (Fig. 1-1C). Without the video regression in the choice-axis fit, joint fractions remained low but were not unchanged; in thalamus the fraction rose from 0.080 to 0.109 (Fig. 1-1D).

A small fraction at laser onset could still grow as the perturbation unfolded. We therefore followed lick information and the coding-subspace fraction over time.

Lick decoding rises without a matching rise on the choice axis

Single-trial deviations from the control mean predicted lick direction much more accurately before the lick than in the early window. Balanced accuracy rose from 0.567–0.611 early to 0.788–0.839 in the 200 ms before the first lick (Table 1; Fig. 2C). On a Go-aligned grid the rise occurred after Go. Through the delay, balanced accuracy stayed near 0.55–0.58. In the 100 ms bin centered 250 ms after Go it was 0.73–0.77 (Fig. 2A). The pre-lick window is aligned to the lick. On the matched laser trials the median reaction time was 0.29 s (quartiles 0.22 and 0.39 s), so the pre-lick window fell mostly after Go. The two comparisons are reported separately.

The choice-axis share of the laser-minus-control vector did not rise with decoding accuracy (Fig. 2D). Early-window means were 0.035–0.046. Before the lick, the share fell or stayed flat in four regions and rose only in medulla, from 0.036 to 0.058. The pre-lick means were 0.025 ± 0.004 in ALM, 0.019 ± 0.005 in striatum, 0.012 ± 0.002 in thalamus, 0.034 ± 0.008 in midbrain, and 0.058 ± 0.014 in medulla. The action-axis share rose from 0.016–0.030 early to 0.040–0.122 before the lick, largest in medulla (0.122 ± 0.015 ; Fig. 2E; Fig. 2-1C). The joint fraction on the Go-aligned grid remained about 0.04–0.08 through the delay. In the bin centered 250 ms after Go it was 0.165 ± 0.028 in medulla and 0.102 ± 0.016 in midbrain, with smaller values in ALM (0.072 ± 0.010), striatum (0.075 ± 0.015), and thalamus (0.056 ± 0.012 ; Fig. 2B). The axes were the ones fit above. They were not refit in each bin.

Most of the population difference remained outside the coding subspace. In ALM, energy per neuron outside the subspace rose from 0.09 before laser onset to 26 ± 10 in the 100 ms bin that begins at onset, then fell. The other regions stayed near or below 1 on that scale (Fig. 2-1A,B).

Decoding accuracy increased, and the action axis took a larger share after Go, while most of the population difference stayed outside the subspace defined by lick direction.

The early difference and the later rise in decoding could still reflect differences in lick direction or reaction time between laser and control trials. We therefore compared the two after matching on both.

A population difference remains after matching lick direction and reaction time

Pairs were formed within session and instructed side, with the same lick direction and a reaction-time difference of at most 100 ms. Of 2,676 retained pairs, the median absolute difference was 1.3 ms and the largest was 99 ms (Fig. 3A). The residual was the laser–control distance minus the distance between two control trials drawn from the same pairs. The primary window was the 200 ms before the lick.

Every mouse had a positive pre-lick residual (Fig. 3D). The means were 1.52 ± 0.14 in ALM, 1.38 ± 0.51 in striatum, 2.99 ± 0.84 in thalamus, 0.76 ± 0.10 in midbrain, and 0.47 ± 0.13 in medulla (Table 1). Raw distance scales with the number of neurons, and the control–control distance is already larger in thalamus than in midbrain (Fig. 3-1A,B), so regions were compared on the residual divided by the control–control distance. That ratio was 0.111 in ALM, 0.093 in striatum, 0.142 in thalamus, 0.063 in midbrain, and 0.036 in medulla (Fig. 3E). Matching did not remove the pre-lick distance (Fig. 3-1D).

In ALM, most of the separation came earlier. The ALM residual was 12.6 ± 2.0 in the 20–120 ms window, compared with 1.52 before the lick (Fig. 3-1C). On the Go-aligned curve, the ALM residual peaked in the bin that begins at laser onset (21.3 ± 2.9) and then declined. Thalamus was negative in the two bins after laser onset (-1.2 ± 0.3 and -1.5 ± 0.7) and peaked at 11.2 ± 2.1 in the bin centered 25 ms before Go. Peaks in the other regions were smaller and occurred at different times: 2.1 ± 0.4 in midbrain shortly after laser onset, 2.8 ± 0.9 in striatum in the bin centered 25 ms after Go, and 1.3 ± 0.5 in medulla. From 0.23 to 0.38 s after Go, residuals were 0.8–1.4 in ALM, 0.9–2.1 in thalamus, 0.3–0.8 in striatum, and below 0.5 in midbrain and medulla (Fig. 3C). The pre-lick value is lick-aligned and is not a point on that curve. In the example session, matched laser and control trajectories separated along the early off-subspace direction and were closer again at Go (Fig. 3B). That plane held 25% of the session’s early matched difference. The first principal component of control-trial activity orthogonal to the choice axis held 4%.

After lick direction and reaction time were matched, laser–control pairs remained farther apart than control–control pairs, by 0.04–0.14 of the control–control distance before the lick. In ALM, this residual was much smaller than during the laser.

Materials and Methods

This study reanalyzed published recordings; no new animals were used. The summary unit, windows, and matching rule were fixed before the estimates were read as results.

Dataset

We reanalyzed the Mesoscale Activity Map dataset (DANDI 000363, version 0.231012.2129).²⁶ The original experiments recorded Neuropixels activity from mice performing a memory-guided lick task, with photoinhibition of ALM on a subset of trials. Surgical procedures, training, and the original inclusion of units are those of Chen et al.²⁶. We did not download the raw archive. Spike times, trial events, optogenetic locations, and video keypoints were read from the derived session files.

Analyses used the 168 sessions that contained photoinhibition. Bilateral trials were identified from the photostimulation control code whose optogenetic location list contained ALM on both sides. The main condition was bilateral inhibition in the late delay: a session entered only when the median bilateral laser onset was within 0.08 s of –0.50 s relative to the Go cue. Sessions whose bilateral laser began about 1.2 s before Go were not pooled with these. In the retained sessions the laser lasted 0.5 s and ended at Go, so there was no pre-Go rebound window to score. Control trials were trials with no photostimulation and no early lick.

Units and regions

Units whose classification label was “good” were kept. This is the classifier label used by Chen et al.²⁶, not the spike-sorting quality flag. Electrode locations were mapped to ALM, striatum, thalamus, midbrain, or medulla, and the two hemispheres were pooled. A region entered a session when that region had at least 20 good units. An instructed side entered when it had at least 10 laser trials and 10 control trials. Nineteen mice and 93 sessions contributed an ALM estimate. Counts for the other regions, and the smaller counts after lick matching, are in Table 1.

Rates and coding axes

Population activity in a window was the mean spike rate of each good unit, z-scored by the mean and standard deviation of control trials in that window. Units with a control standard deviation below 10^{-6} were dropped from that window.

The choice axis and the action axis were fit on control trials only, within session and region. The choice axis was the difference between lick-right and lick-left mean activity from 600 ms before Go until Go, after least-squares removal of mean jaw, nose, and tongue position in the same window. The action axis was the same lick contrast from Go to 300 ms after Go, without the video regression. Each axis was scaled to unit length. The early population difference was the mean laser activity minus the mean control activity, within instructed side, from 20 to 120 ms after laser onset. The two sides were combined in proportion to the number of laser trials.

The coding subspace was the plane spanned by the two axes after a QR factorization, so the reported fraction does not require the axes to be orthogonal. The joint fraction was the energy of the early difference inside that plane divided by its total energy. Fractions on each axis alone were reported the same way. The cosine between the axes was stored and was not used as the result.

As a noise reference, control trials on each instructed side were split into two groups of the same size as the laser set. The split was repeated 20 times. We report the energy of the laser difference divided by the energy of that control difference. The fraction after subtracting the control-split projection was used only to ask whether the low joint fraction was produced by sampling in a high-dimensional population.

Decoding

There is no paired trial without laser. The feature for each trial was its z-scored population vector minus the mean of control trials with the same instructed side. That mean was estimated inside the training fold. The label was lick direction. Decoding used L2-penalized logistic regression. For the comparison of the early window with the 200 ms before the first lick, the penalty was chosen from {0.1, 1, 10} inside the training fold. Performance was balanced accuracy under 5-fold cross-validation, stratified by instructed side and lick direction. A session needed at least 40 trials and 15 trials of each lick direction.

The pre-lick window is aligned to the lick. On these sessions the median reaction time is about 0.3 s, so that 200 ms lies mostly after Go. It was not treated as a late-delay bin, and it was not plotted on the Go-aligned curve.

Matched distance

Within each session and instructed side, each laser trial with a lick was paired to a control trial with the same lick direction and a reaction-time difference of at most 100 ms. Matching was without replacement, in random order. Ties were broken at random. Sessions with fewer than 10 pairs were dropped. For each retained control trial, a second control trial was drawn by the same rule from the controls not already used. The residual was the mean laser–control Euclidean distance minus the mean control–control Euclidean distance. The primary window was the 200 ms before the lick, divided into 50 ms bins. The same pairs were scored in the early window. Unmatched pairs, drawn without regard to lick direction or reaction time, were a reference only.

Distances were computed on z-scored population vectors. Because the Euclidean norm grows with the number of neurons, regions were compared on the residual divided by the control–control distance, not on the raw residual. Kinematic amplitude and duration were not rematched.

Time courses

Go-aligned curves ran from 600 ms before Go to 600 ms after Go. Distance used 50 ms bins. The joint fraction and balanced accuracy used 100 ms bins, formed by averaging adjacent 50 ms rates. All trials were aligned to Go. The choice and action axes were the axes already fit on their original windows; they were not refit in each bin. Decoding on the curve used a fixed penalty of 1. The early-versus-pre-lick comparison kept the penalty chosen inside each training fold. The two analyses were not overlaid.

For the distance curve, units had to pass the standard-deviation threshold in every bin. If fewer than 20 units did, units that passed in at least 80% of bins were kept and the remaining bins were set to zero. The z-score control for these bins was every non-early-lick control trial. Energy per neuron was the squared length of the laser-minus-control vector, inside or outside the coding subspace, divided by the number of units.

Example session

The session drawn in Figs. 1 and 3 was chosen before the panels were inspected for appearance. Candidates were ALM sessions that entered both the early-fraction analysis and the distance curve and had at least the median neuron count of that set. Among them we took the session whose joint fraction was closest to the ALM mouse mean, breaking ties by neuron count and then by session name. The displayed plane uses the choice axis and the early laser-minus-control direction orthogonal to the choice and action axes, held fixed across time. It is not a principal-component plane. The first principal component of control-trial activity orthogonal to the choice axis was computed only to report how much of the early matched difference it contained.

Statistics

The reported estimate is the unweighted mean across mice. Sessions were averaged within mouse first, with no weighting by session or by neuron count. Uncertainty is the standard error across mice. The same rule applies to the ratio of residual to control–control distance: the ratio was formed for each mouse and then averaged. Random splits, cross-validation folds, and match order used one seed, 17. No test was added after the means were known. The recovery analysis discussed for an earlier ALM dataset was not run.

Code and data

Analyses were run in Python 3.12 with NumPy and scikit-learn logistic regression (lbfgs). The dataset is DANDI 000363, version 0.231012.2129 (<https://dandiarchive.org/dandiset/000363/0.231012.2129>).

Code Accessibility

The analysis code described in the paper has been deposited in a private GitHub repository and will be made publicly available upon publication. Analyses were run on a Mac mini (Apple M4 Pro) running macOS 26.5.2.

Discussion

We asked how the population response to ALM photoinhibition was distributed relative to task-coding dimensions, how that relationship changed over the trial, and whether trials matched on behavior were also matched in population state. Shortly after photoinhibition onset, choice and

action dimensions defined on control trials captured only a small fraction of the laser–control population difference, even though the difference itself was large. After the Go cue, lick direction became much more decodable from single trials, but the choice-axis share of the mean laser–control difference did not rise with it. After laser and control trials were matched for lick direction and reaction time, population differences remained in all five regions; in ALM, the residual before the lick was much smaller than during photoinhibition. Together, these findings distinguish the population form of a perturbation, its temporal relationship to task information, and the degree to which matching the measured behavioral outcome is accompanied by matching of the population state.

The form of a perturbation extends beyond task-coding dimensions

The early coding-subspace fraction gives the sharpest separation between coding geometry and perturbation geometry. In all five regions, the joint choice–action subspace held only 5–8% of the early laser–control difference, even though that difference was well above the difference between size-matched splits of control trials, most strongly in ALM. The low fraction is not a failure to detect a perturbation response. Most of the response lay along directions other than the two used here to describe lick choice and action.

Coding dimensions and perturbation directions are defined differently. A coding dimension is defined by how activity varies with a task variable under a given set of conditions;^{2–4,8} a perturbation direction is defined by how activity changes after an intervention. Nothing requires the two to coincide.¹⁴ A dimension can separate behavioral conditions well and still account for little of an intervention-induced change. Equally, activity displaced outside a coding dimension is not necessarily functionally important, nor does it necessarily reflect an unidentified task variable. The result does not show that the choice and action dimensions describe task activity incompletely. It shows that the population response to a causal perturbation is not determined by the dimensions that best describe the task variables.

This complements earlier analyses of the same dataset. Chen et al.²⁶ mapped choice- and movement-related signals across the network and showed that subcortical choice coding depends on ALM, establishing both how task activity is organized and that ALM is causally important. Our analysis asks a narrower question: of the full population difference produced by the perturbation, how much lies along task dimensions defined from control activity?

Other studies have shown that activity outside explicit task-coding dimensions can have dynamical consequences.^{17,28,29} We make no such causal claim for the off-subspace activity

here: ALM was perturbed as a whole, and the components of the resulting displacement were not manipulated separately. Three questions are worth keeping apart: where task information is represented, where activity moves after an intervention, and which parts of that movement affect later activity or behavior. The present data address the second and relate it to the first. The third would require perturbations targeted to the population components identified here.

The expression of a perturbation changes over time

The relation between the perturbation response and the coding axes also depended on when it was measured. Lick direction became more predictable from single trials after the Go cue (balanced accuracy stayed near 0.55–0.58 through the delay), but the choice-axis share of the mean laser–control difference did not rise with it. That share stayed low, while the action-axis share increased after Go, most clearly in midbrain and medulla. Lick information and the alignment of the perturbation-associated difference followed different time courses.

Decoding and perturbation geometry measure different things. Decoding asks whether trial-to-trial variation carries information about a behavioral variable; the perturbation vector describes how the mean population state differs between laser and control conditions. Within each condition, activity can become more separable by upcoming lick direction while the difference between conditions stays largely off the choice axis. A perturbation could also shift activity along a coding dimension without changing trial-level decoding much. Neither measure can substitute for the other.

In a dynamical system, the effect of an intervention keeps changing as recurrent interactions, sensory input, task dynamics, and movement-related processes unfold.^{7,11,20,22,30–32} An effect measured shortly after the intervention and one measured near the behavior can both be real consequences of the same perturbation without having the same geometry. This limits causal descriptions based on one variable or one time window. Finding an effect along a behaviorally meaningful dimension shows that the intervention affects that dimension at that time. It does not show that the dimension captures the wider population response, or that the effect keeps the same form later. A weak early projection onto a task axis, likewise, does not mean the perturbation has no behavioral consequence.

Matched behavioral outcomes do not imply equivalent population states

The matching analysis addresses a third question. Laser and control trials were paired within session on lick direction and reaction time, so their measured outcomes were closely similar.

Even so, laser–control pairs were more separated than control–control pairs in all five regions before the lick. In ALM, this residual was much smaller than the separation during photoinhibition: ALM activity converged considerably over the trial but did not become indistinguishable at the level measured here.

Matching an observed output constrains one projection of the system's state, not every component of the population activity. Two trials can produce the same lick direction at almost the same time and still differ in ways that reflect what happened earlier. Here, one of those histories includes ALM photoinhibition.

The residual differences should be read narrowly. They do not show that the remaining activity contributes to the matched behavior, and they do not reveal a compensatory pathway. They could reflect unmeasured movements, internal state, parts of the perturbation response that have not yet decayed, or neural differences with little later consequence.^{33,34} What the analysis shows is state nonequivalence within the recorded population and measured variables, not functional nonequivalence. Two measurably different states could be functionally interchangeable if they respond alike to later inputs and evolve toward the same future states.^{21,35,36} A residual that is small relative to trial-to-trial variability could, on the other hand, still matter if it changes how the system responds to a later input. The size of the state difference alone cannot distinguish these cases.

The regional time courses are consistent with this point but do not reveal mechanism. ALM showed a large difference right after photoinhibition onset that then declined; thalamus was briefly negative after onset and peaked just before Go; the other regions had smaller peaks at different times. The response therefore evolved differently across regions. These profiles do not show direct propagation from ALM, recruitment of a specific pathway, or compensation. The ordering could arise from recurrent interactions, indirect circuit effects, changing task variables, or other network processes.^{20,22–24,37}

From state difference to subsequent consequence

The gap between state nonequivalence and functional nonequivalence suggests a direct experiment. Suppose a first perturbation leaves trials with matched behavioral output but distinguishable population states. A second input or perturbation could then be delivered conditional on those states. If later neural activity or behavior differs between them, the residual state difference changes the efficacy of the second intervention. If it does not, the difference may record the history of the first perturbation without altering what the system does next.

This would go beyond asking whether a single variable has recovered. Recovery is always defined with respect to some observable:^{8,38,39} a behavioral measure can return to the control range while population activity remains displaced, and one coding dimension can recover while other parts of the population do not. For functional equivalence, the relevant question is whether two states differ in how they evolve, not whether they are numerically identical. The present data do not test this. They show the precondition for testing it: matching the observed output does not guarantee matching of the measured population state.

Limitations

These conclusions have several limits. First, the choice and action axes examined here are specific, control-defined dimensions rather than an exhaustive description of task-related neural activity. The observation that most of the perturbation-associated difference lies outside their joint subspace should therefore not be interpreted as evidence for a discrete class of “noncoding” dimensions. Other task variables, alternative linear dimensions, or nonlinear representations could account for components of this activity.

Second, lick direction and reaction time provide only a partial description of behavior. Although matching these variables removes two prominent differences between trials, laser and control trials could still differ in lick kinematics, posture, preparatory movements, arousal, or other unmeasured variables.^{33,34} The residual population distance therefore cannot be attributed uniquely to perturbation history.

Third, Euclidean population distance establishes separation between recorded population states but does not specify the dynamical significance of that separation. A larger distance does not necessarily imply a larger difference in subsequent behavior or responsiveness to future input. For this reason, the behavior-matching result should not be interpreted as a measurement of perturbation efficacy.

Fourth, the off-subspace component was not independently manipulated. The experiment establishes a causal intervention at ALM and measures the resulting distributed population response, but it does not identify which components of that response are necessary or sufficient for later neural or behavioral effects.^{38,39}

Finally, the analyses focus on one task, one perturbation configuration, and a subset of the neural system.^{26,40} The bilateral laser in the analyzed sessions occupied the final 0.5 s of the delay and ended at Go, leaving no post-inhibition interval within the delay in which to test

whether recovery of a particular coding variable coincides with recovery of the broader population state. Whether the relationships among form, timing, and subsequent functional consequence generalize across tasks, circuit architectures, and perturbation regimes remains open.

Conclusion

What ALM photoinhibition did to population activity depended on which component was measured and when. The early laser–control difference lay mostly outside the choice and action dimensions. After the Go cue, lick information increased while the choice-axis share of the mean difference did not. With lick direction and reaction time matched, population differences remained in all five regions.

Matching behavior did not match the population state. Whether the remaining differences change how the circuit responds to later inputs is not answered here; answering it will require a second intervention that tests whether behaviorally matched but neurally distinct states respond differently to what comes next.

Acknowledgments

The author thanks Jessica Cardin for insightful discussions. The author also thanks the authors of Chen et al.²⁶ for generating and making publicly available the Mesoscale Activity Map recordings used in this study, and the DANDI Archive for hosting and distributing them (DANDI 000363, version 0.231012.2129; <https://doi.org/10.48324/dandi.000363/0.231012.2129>). During the preparation of this work, the author used ChatGPT (OpenAI) to assist with language editing and improve the readability of the manuscript. The author reviewed and edited the content as necessary and takes full responsibility for the final version of the manuscript.

Conflict of Interest

The author declares no competing financial interests.

Funding Sources

None.

References

1. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV (2012) Neural population dynamics during reaching. *Nature* 487:51–56.
2. Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503:78–84.
3. Cunningham JP, Yu BM (2014) Dimensionality reduction for large-scale neural recordings. *Nat Neurosci* 17:1500–1509.
4. Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi XL, Romo R, Uchida N, Machens CK (2016) Demixed principal component analysis of neural population data. *eLife* 5:e10989.
5. Gallego JA, Perich MG, Miller LE, Solla SA (2017) Neural manifolds for the control of movement. *Neuron* 94:978–984.
6. Saxena S, Cunningham JP (2019) Towards the neural population doctrine. *Curr Opin Neurobiol* 55:103–111.
7. Shenoy KV, Sahani M, Churchland MM (2013) Cortical control of arm movements: a dynamical systems perspective. *Annu Rev Neurosci* 36:337–359.
8. Li N, Daie K, Svoboda K, Druckmann S (2016) Robust neuronal dynamics in premotor cortex during motor planning. *Nature* 532:459–464.
9. Inagaki HK, Inagaki M, Romani S, Svoboda K (2018) Low-dimensional and monotonic preparatory activity in mouse anterior lateral motor cortex. *J Neurosci* 38:4163–4185.
10. Inagaki HK, Fontolan L, Romani S, Svoboda K (2019) Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* 566:212–217.
11. Vyas S, Golub MD, Sussillo D, Shenoy KV (2020) Computation through neural population dynamics. *Annu Rev Neurosci* 43:249–275.
12. Guo ZV, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, Feng G, Svoboda K (2014) Flow of cortical activity underlying a tactile decision in mice. *Neuron* 81:179–194.
13. Li N, Chen TW, Guo ZV, Gerfen CR, Svoboda K (2015) A motor cortex circuit for motor planning and movement. *Nature* 519:51–56.
14. Jazayeri M, Afraz A (2017) Navigating the neural space in search of the neural code. *Neuron* 93:1003–1014.
15. Chettih SN, Harvey CD (2019) Single-neuron perturbations reveal feature-specific competition in V1. *Nature* 567:334–340.
16. Pinto L, Rajan K, DePasquale B, Thiberge SY, Tank DW, Brody CD (2019) Task-dependent changes in the large-scale dynamics and necessity of cortical regions. *Neuron* 104:810–824.e9.
17. Kaufman MT, Churchland MM, Ryu SI, Shenoy KV (2014) Cortical activity in the null space: permitting preparation without movement. *Nat Neurosci* 17:440–448.
18. Kaufman MT, Seely JS, Sussillo D, Ryu SI, Shenoy KV, Churchland MM (2016) The largest response component in the motor cortex reflects movement timing but not movement type. *eNeuro* 3:ENEURO.0085-16.2016.
19. Elsayed GF, Lara AH, Kaufman MT, Churchland MM, Cunningham JP (2016) Reorganization between preparatory and movement population responses in motor cortex. *Nat Commun* 7:13239.
20. Inagaki HK, Chen S, Ridder MC, Sah P, Li N, Yang Z, Hasanbegovic H, Gao Z, Gerfen CR, Svoboda K (2022) A midbrain-thalamus-cortex circuit reorganizes cortical dynamics to initiate movement. *Cell* 185:1065–1081.e23.
21. Marder E, Goaillard JM (2006) Variability, compensation and homeostasis in neuron and network function. *Nat Rev Neurosci* 7:563–574.
22. Guo ZV, Inagaki HK, Daie K, Druckmann S, Gerfen CR, Svoboda K (2017) Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* 545:181–186.

23. Economo MN, Viswanathan S, Tasic B, Bas E, Winnubst J, Menon V, Graybiel LT, Nguyen TN, Smith KA, Yao Z, Wang L, Gerfen CR, Chandrashekar J, Zeng H, Looger LL, Svoboda K (2018) Distinct descending motor cortex pathways and their roles in movement. *Nature* 563:79–84.
24. Gao Z, Davis C, Thomas AM, Economo MN, Abrego AM, Svoboda K, De Zeeuw CI, Li N (2018) A cortico-cerebellar loop for motor planning. *Nature* 563:113–116.
25. Svoboda K, Li N (2018) Neural mechanisms of movement planning: motor cortex and beyond. *Curr Opin Neurobiol* 49:33–41.
26. Chen S, Liu Y, Wang ZA, Colonell J, Liu LD, Hou H, Tien NW, Wang T, Harris T, Druckmann S, Li N, Svoboda K (2024) Brain-wide neural activity underlying memory-guided movement. *Cell* 187:676–691.e16.
27. Jun JJ et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551:232–236.
28. Daie K, Svoboda K, Druckmann S (2021) Targeted photostimulation uncovers circuit motifs supporting short-term memory. *Nat Neurosci* 24:259–265.
29. Finkelstein A, Fontolan L, Economo MN, Li N, Romani S, Svoboda K (2021) Attractor dynamics gate cortical information flow during decision-making. *Nat Neurosci* 24:843–850.
30. Sauerbrei BA, Guo JZ, Cohen JD, Mischiati M, Guo W, Kabra M, Verma N, Mensh B, Branson K, Hantman AW (2020) Cortical pattern generation during dexterous movement is input-driven. *Nature* 577:386–391.
31. Remington ED, Narain D, Hosseini EA, Jazayeri M (2018) Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron* 98:1005–1019.e5.
32. Sussillo D, Barak O (2013) Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput* 25:626–649.
33. Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK (2019) Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci* 22:1677–1686.
34. Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364:eaav7893.
35. Prinz AA, Bucher D, Marder E (2004) Similar network activity from disparate circuit parameters. *Nat Neurosci* 7:1345–1352.
36. Gallego JA, Perich MG, Chowdhury RH, Solla SA, Miller LE (2020) Long-term stability of cortical population dynamics underlying consistent behavior. *Nat Neurosci* 23:260–270.
37. Semedo JD, Zandvakili A, Machens CK, Yu BM, Kohn A (2019) Cortical areas interact through a communication subspace. *Neuron* 102:249–259.e4.
38. Otchy TM, Wolff SBE, Rhee JY, Pehlevan C, Kawai R, Kempf A, Gobes SMH, Ölveczky BP (2015) Acute off-target effects of neural circuit manipulations. *Nature* 528:358–363.
39. Wolff SBE, Ölveczky BP (2018) The promise and perils of causal circuit manipulations. *Curr Opin Neurobiol* 49:84–94.
40. Steinmetz NA, Zatka-Haas P, Carandini M, Harris KD (2019) Distributed coding of choice, action and engagement across the mouse brain. *Nature* 576:266–273.

Figures

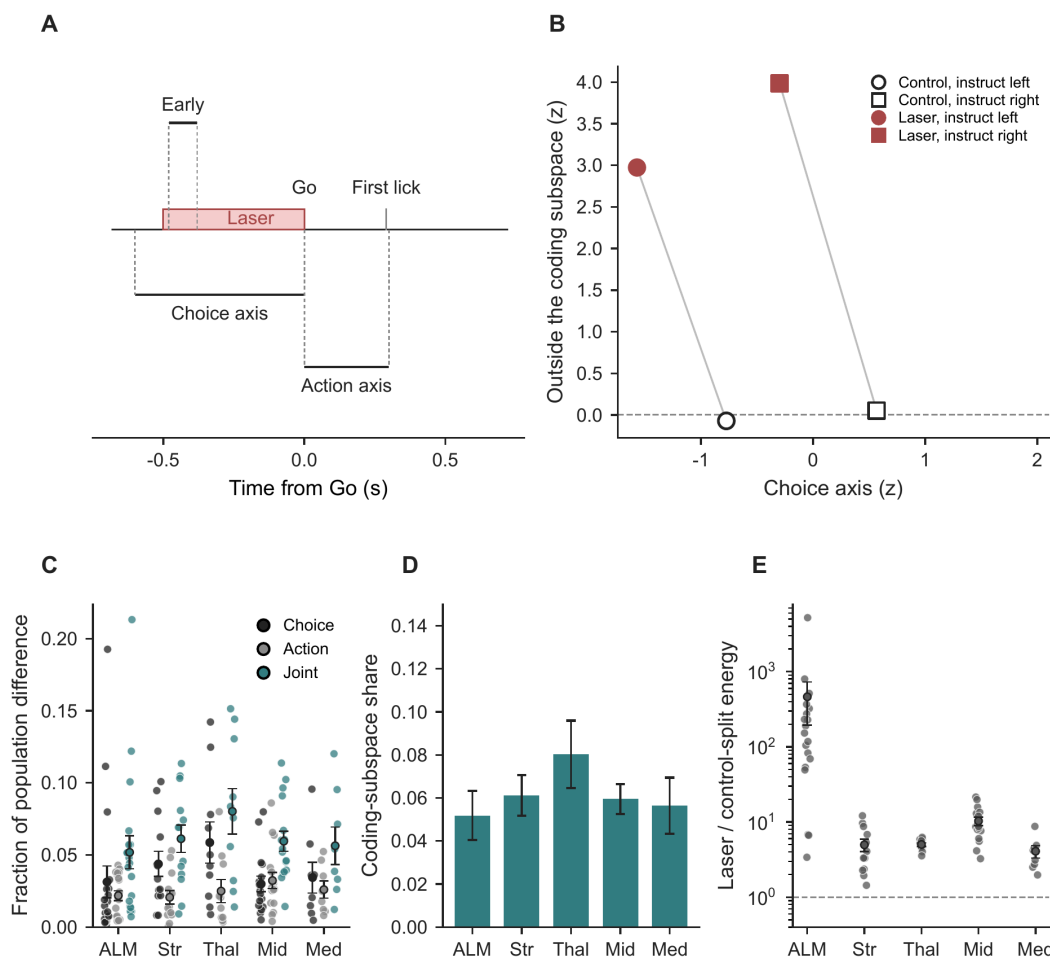


Figure 1. The early population difference during bilateral ALM inhibition lies mostly outside the choice and action axes. **A**, Task time. The bar is bilateral laser (−0.5 to 0 s from Go). Early is 20–120 ms after laser onset. The choice axis is fit on control trials from 600 ms before Go to Go, and the action axis on control trials from Go to 300 ms after Go. Dashed lines mark the ends of each interval. First lick is the median reaction time of matched laser trials. **B**, One pre-specified ALM session (ses-20200922T135318; 130 neurons; joint fraction 0.053) in the plane used to read that session. The horizontal axis is the choice axis. The vertical axis is the part of the early laser-minus-control difference orthogonal to both coding axes. Open symbols, control; filled symbols, laser; circles, instructed left; squares, instructed right. **C**, Fraction of the early population difference on the choice axis, the action axis, and their joint subspace. Each point is one mouse; bars are the mean $\pm$ SEM. **D**, The joint fraction alone, so the scale is the coding-subspace share rather than the remainder. **E**, Energy of the laser difference divided by the energy of a size-matched split of control trials (20 splits). The dashed line is 1. The fraction after the control-split correction, and with and without removal of jaw, nose, and tongue position from the choice axis, is in Fig. 1-1.

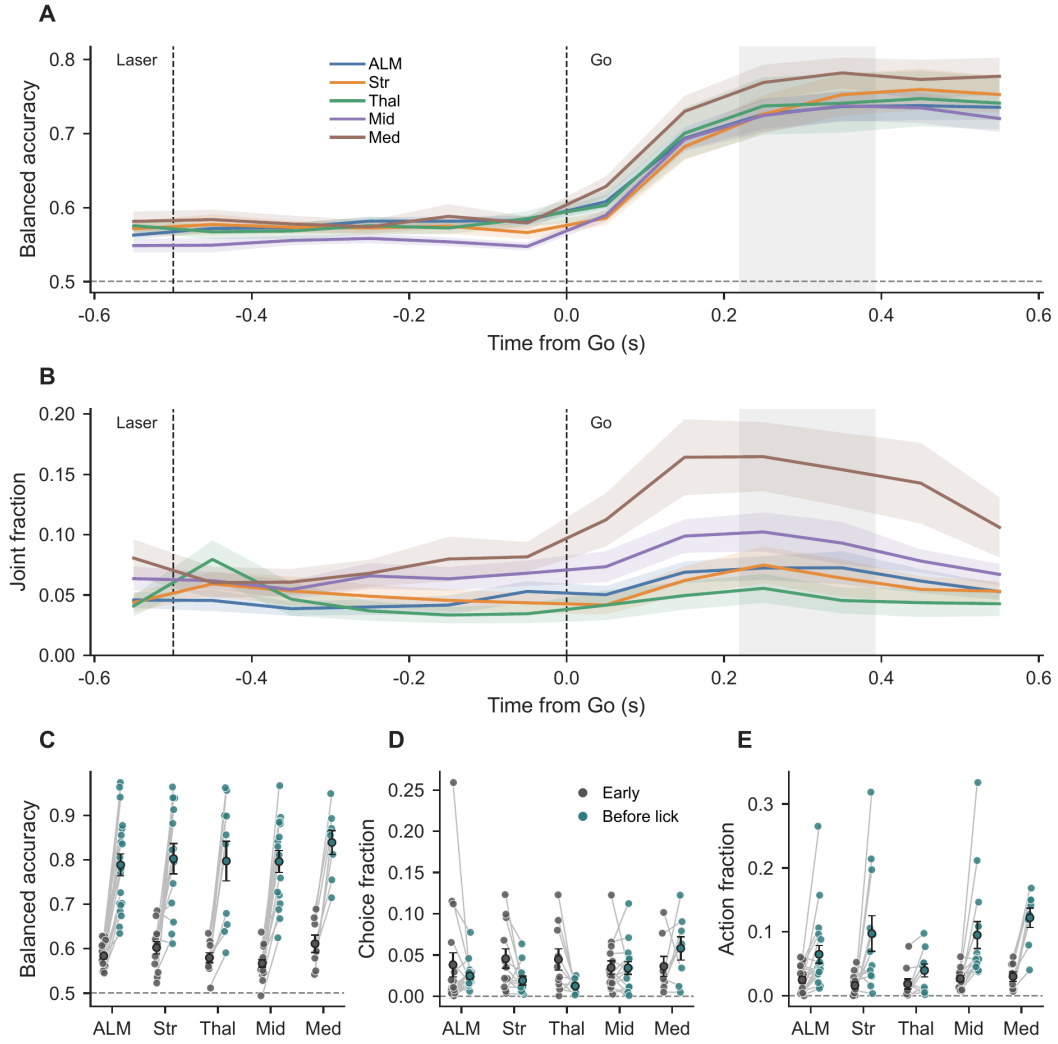


Figure 2. Balanced accuracy rises toward the lick without a matching rise on the choice axis. **A**, Balanced accuracy in successive 100 ms windows aligned to Go. The classifier uses a fixed penalty. The band is the SEM across mice. Vertical dashed lines mark laser onset and Go. The gray band is the 25th to 75th percentile of reaction time on matched laser trials. **B**, Fraction of the laser-minus-control vector in the joint subspace of the same choice and action axes, which were not refit in each window. **C**, Balanced accuracy in the early window and in the 200 ms before the first lick. This comparison chooses the penalty inside each training fold. It is not a point on **A**, because the pre-lick window is aligned to the lick. **D**, Fraction on the choice axis in those two windows. **E**, Fraction on the action axis. Lines in **C–E** are mice. The choice fraction does not rise with accuracy; the action fraction does (Fig. 2-1C).

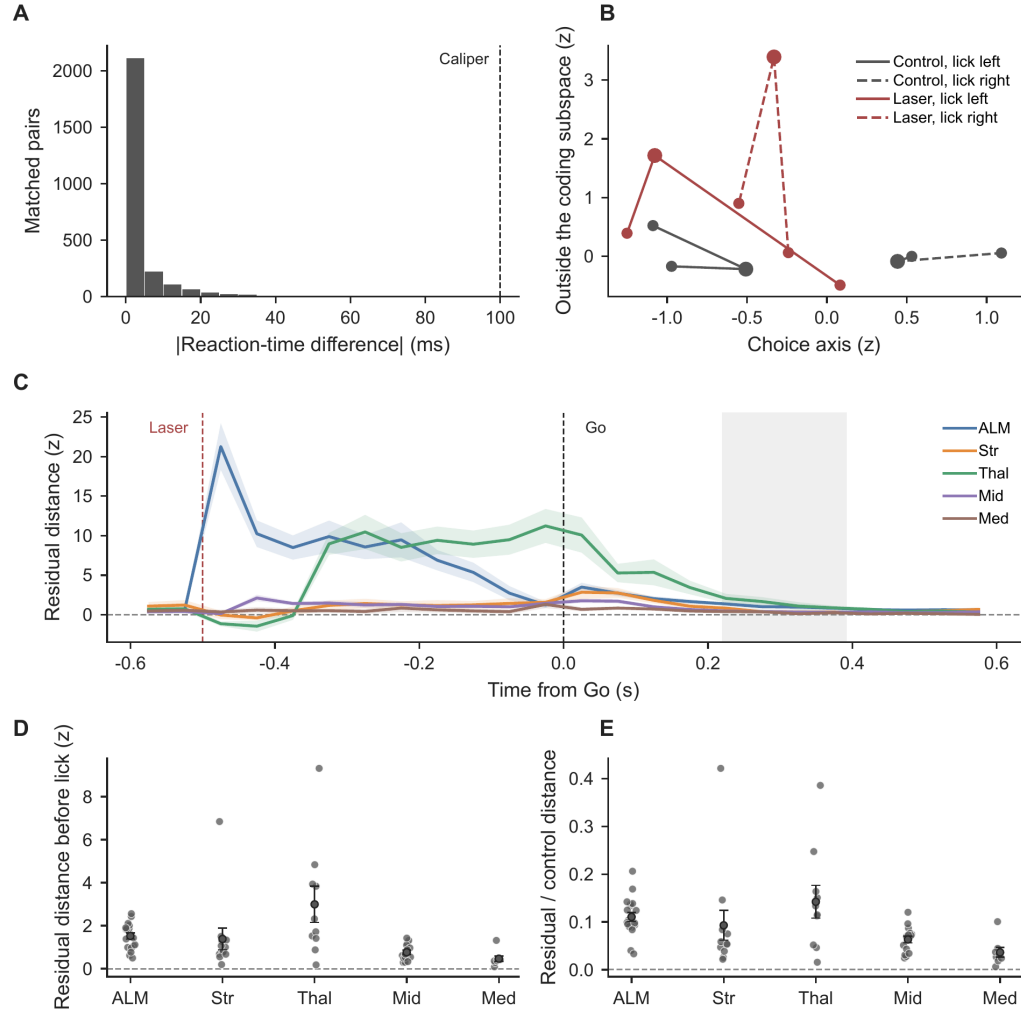


Figure 3. After lick direction and reaction time are matched, a residual distance remains, and it is small by the time of the lick. **A**, Absolute reaction-time difference of the retained pairs. The dashed line is the 100 ms caliper. **B**, Mean trajectory of the matched trials in the session in Fig. 1B. The horizontal axis is the choice axis. The vertical axis is the early laser-minus-control direction orthogonal to the choice and action axes, held fixed across time. Solid lines, lick left; dashed lines, lick right. On each path the three markers are before laser onset, the early window, and Go. In this session that plane contains 25% of the early matched difference. The first principal component of control-trial activity orthogonal to the choice axis contains 4%. **C**, Laser-control distance minus control-control distance on the same pairs, in 50 ms bins aligned to Go. The red dashed line is laser onset and the black dashed line is Go. **D**, That residual in the 200 ms before the lick. This window is aligned to the lick, so it is not a point on **C**. **E**, The pre-lick residual divided by the control-control distance. Points in **D** and **E** are mice (Table 1).

Tables

Table 1. Mouse-level summary of the three comparisons. Mouse means $\pm$ SEM. Each mouse is the unweighted mean of its sessions. Fractions and balanced accuracy use the late-delay sessions. The residual uses the sessions that retained matched pairs. The last column is each mouse's residual divided by its control–control distance, then averaged across mice.

Region	Mice / sessions	Joint fraction	Choice fraction	Action fraction	Early accuracy	Accuracy before lick	Matched mice / sessions	Residual before lick (z)	Residual / control distance
ALM	19 / 93	0.052 ± 0.011	0.031 ± 0.011	0.022 ± 0.003	0.584 ± 0.006	0.788 ± 0.025	19 / 87	1.52 ± 0.14	0.111 ± 0.009
Striatum	13 / 48	0.061 ± 0.009	0.044 ± 0.009	0.021 ± 0.005	0.602 ± 0.014	0.802 ± 0.034	12 / 42	1.38 ± 0.51	0.093 ± 0.031
Thalamus	10 / 31	0.080 ± 0.016	0.059 ± 0.014	0.025 ± 0.008	0.580 ± 0.011	0.797 ± 0.045	10 / 29	2.99 ± 0.84	0.142 ± 0.034
Midbrain	16 / 71	0.060 ± 0.007	0.030 ± 0.005	0.032 ± 0.006	0.567 ± 0.008	0.796 ± 0.025	15 / 64	0.76 ± 0.10	0.063 ± 0.007
Medulla	8 / 35	0.056 ± 0.013	0.034 ± 0.011	0.026 ± 0.006	0.611 ± 0.019	0.839 ± 0.027	8 / 32	0.47 ± 0.13	0.036 ± 0.010

Extended Data

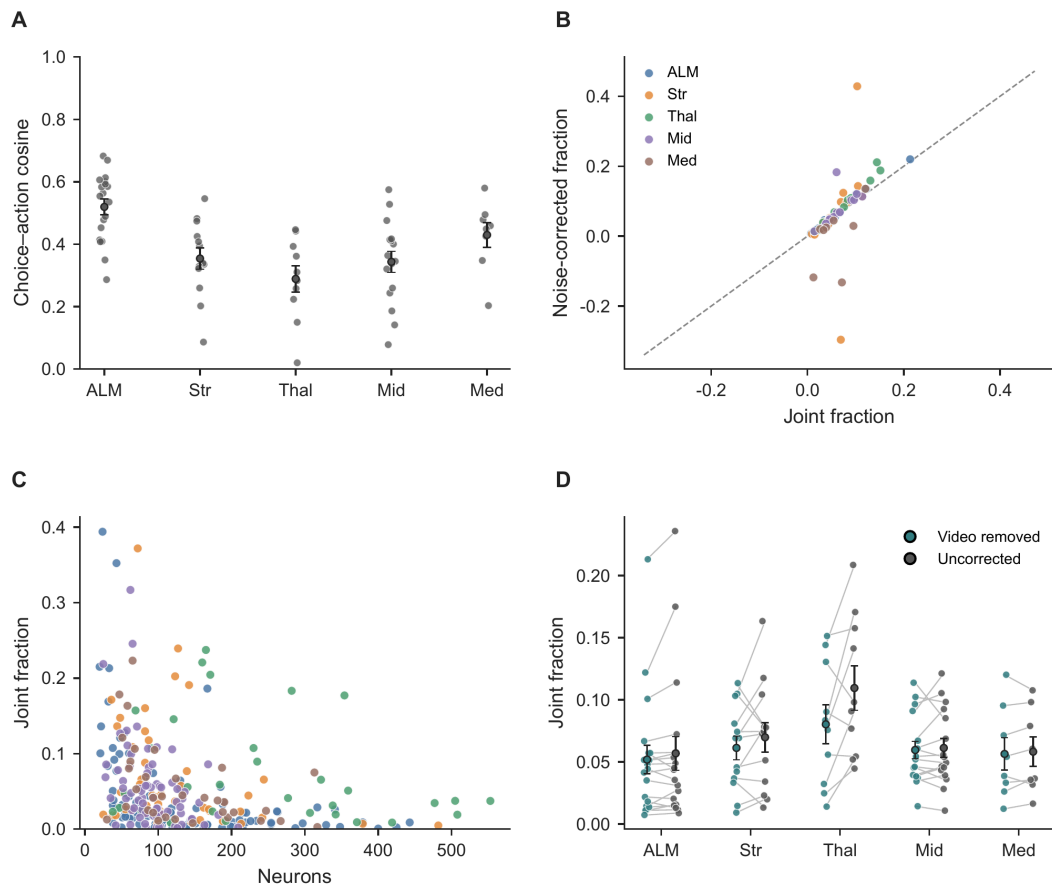


Figure 1-1. Robustness and recording properties of the early coding-subspace fraction. **A**, Cosine between the choice and action axes (mean $\pm$ SEM across mice). **B**, Joint fraction against the fraction after subtracting the control-split projection. The dashed line is equality. **C**, Joint fraction against neuron count, one session per point. **D**, Joint fraction with jaw, nose, and tongue position removed from the choice axis, and without that removal. Lines connect sessions.

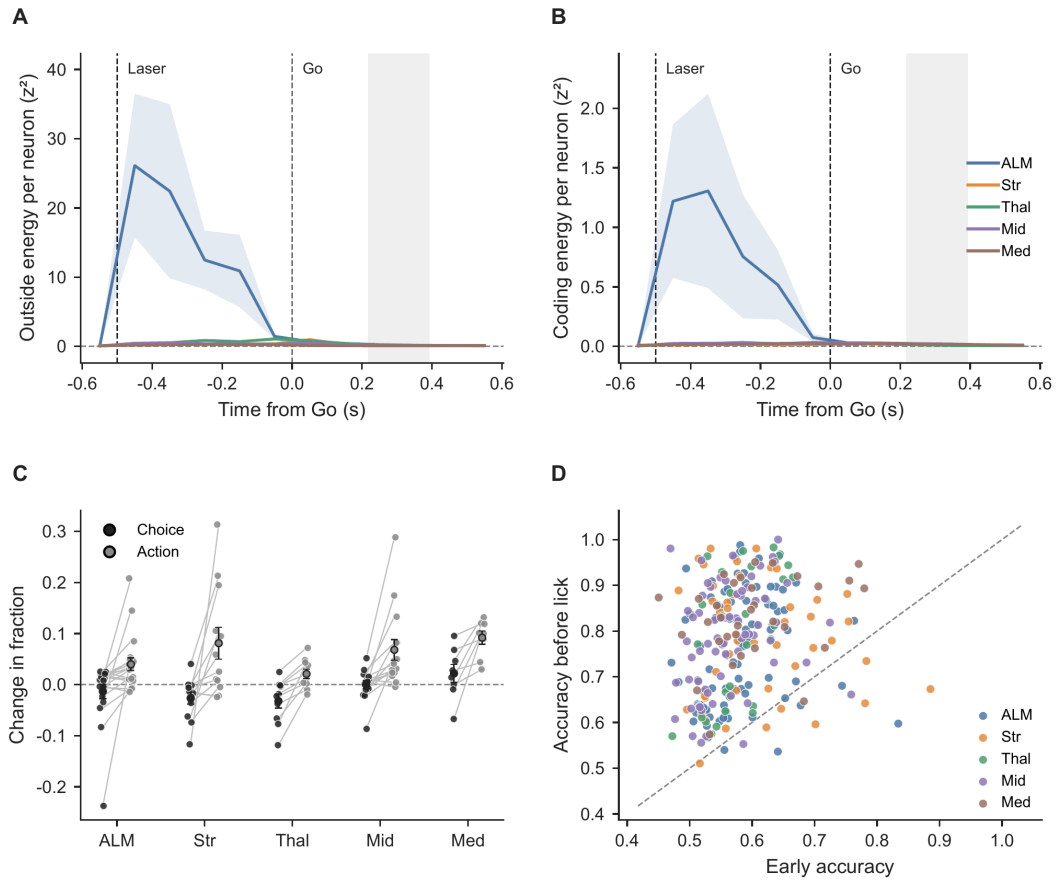


Figure 2-1. Most of the population difference remains outside the coding subspace. A, Energy per neuron outside the coding subspace. ALM sets the vertical scale. **B,** Energy per neuron inside the subspace, on the same sessions and the same vertical meaning. **C,** Change from the early window to the pre-lick window for the choice fraction and the action fraction. **D,** Session balanced accuracy before the lick against early accuracy. The dashed line is equality. Laser onset and Go are marked as in Fig. 2.

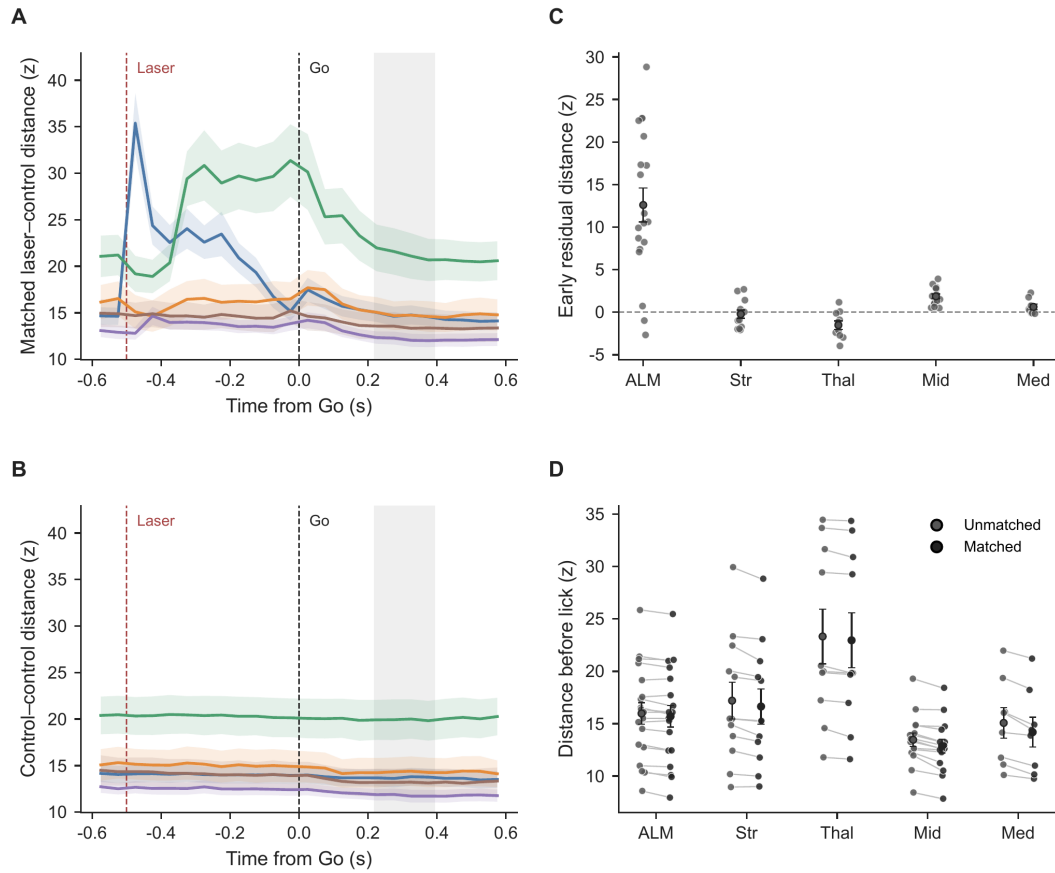


Figure 3-1. The residual is the difference of two distances, and matching does not remove either one. A, Matched laser-control distance. **B,** Control-control distance on the same pairs. **A** and **B** share one vertical scale. **C,** Residual in the early window. **D,** Distance before the lick without matching, and after matching lick direction and reaction time. Laser onset and Go are marked as in Fig. 3.