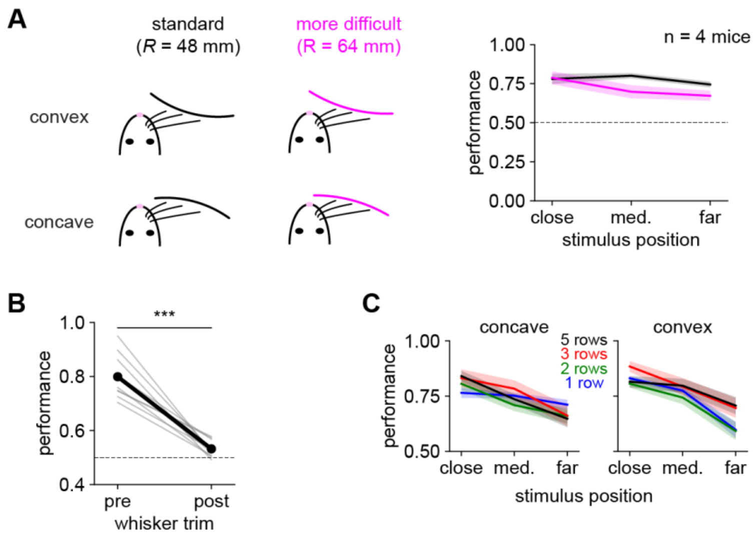
Supplemental Information

Supplemental Figure 1, related to Figure 1.

Mice can generalize to more difficult shapes and learn similar strategies with fewer whiskers



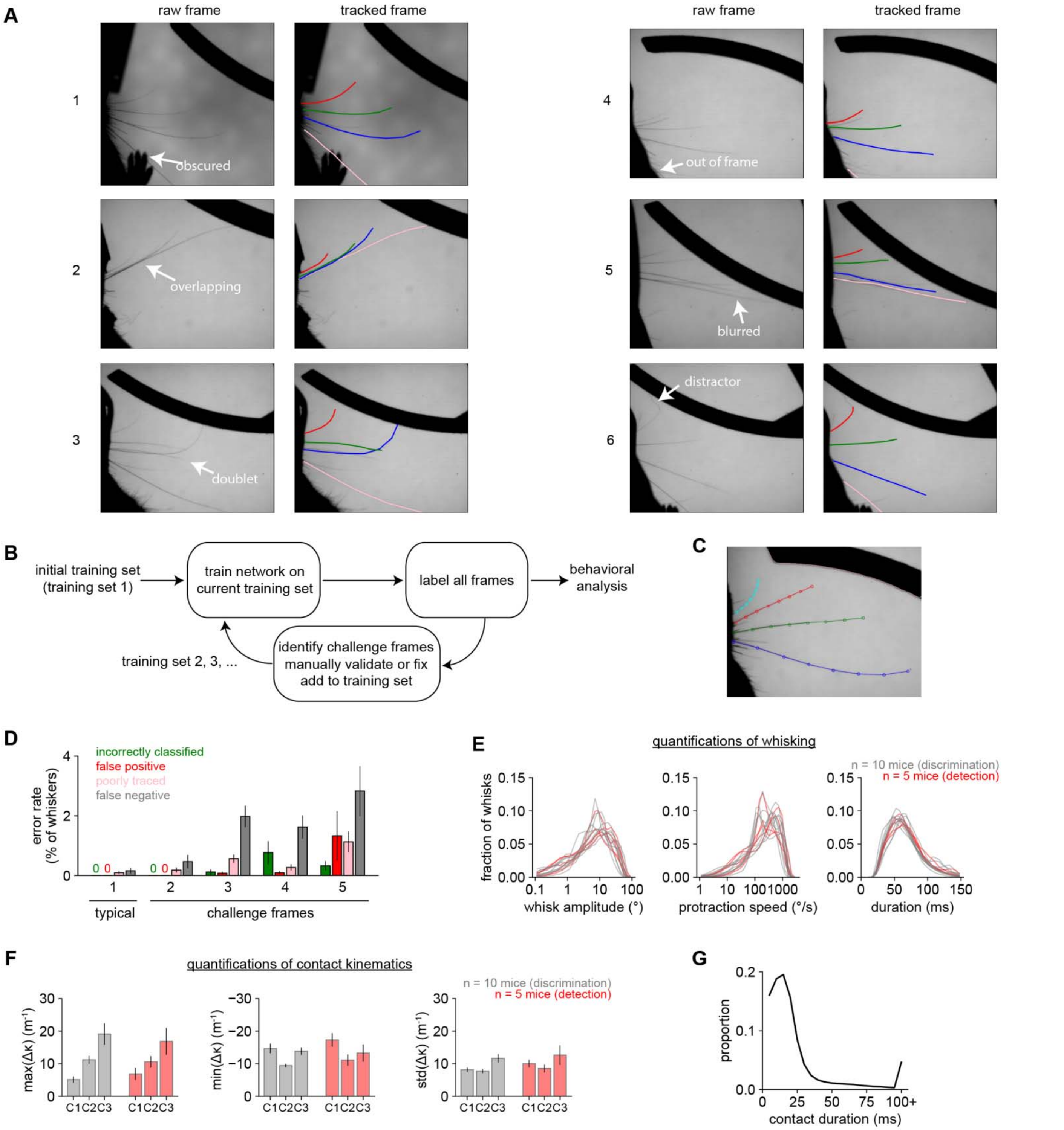
A) We trained a subset of mice on more difficult stimuli with a larger radius of curvature (pink). After some retraining, mice were able to perform well at these new stimuli, though at a slightly lower performance than for the original stimuli. Error bars: SEM over mice.

B) Effect of trimming all whiskers on the performance of n = 10 mice performing the shape discrimination or detection tasks. "Pre": average performance on the three sessions preceding whisker trim. "Post": performance on the first session after whisker trim. Thin lines: individual mice. Thick line: average. Performance significantly decreased (paired t-test, $p < 0.001$) from 79.9% to 53.2%, near chance (50%, dashed line).

C) Performance on shape discrimination throughout whisker trimming. During training on each task, we gradually trimmed the mice from 5 rows of whiskers to 3 to 2 and finally to 1, allowing them to recover to stable performance again after each trim. To address whether this trimming affects the strategy the mice use, we examined the performance on concave (left) and convex (right) shapes versus distance from the face for the shape discrimination task. Different colors indicate the number of whisker rows spared. Regardless of the number of whisker rows spared, their overall performance and pattern of errors was strikingly similar, suggesting that the strategy mice naturally use with all whiskers is not substantially or qualitatively affected by trimming. Error bars: SEM over n = 9 mice.

Supplemental Figure 2, related to Figure 2.

High-speed videography, whisker tracking procedure, and statistics of whisking and contacts.



A) Example frames demonstrating the quality of the whisker tracking. Within each pair of frames, the left frame is the raw frame (annotated with the region of interest) and the right frame shows the result of the whisker tracking algorithm. Performance was good (i.e., the correct whiskers were tracked throughout their extent) even when:

1) the whisker was obscured by a paw;
2) whiskers were nearly overlapping;
3) a "doublet" whisker emerged from the same follicle;
4) the whisker was nearly out of frame;
5) motion blurred the tips;
6) a similar-looking distractor hair was attached to the end of the whisker.

B) We used an iterative procedure to train the network, repeatedly identifying “challenge frames” on which it struggled and using those to generate the next training set.

C) Example frame showing the discrete “joints” that we tracked (circles).

D) Error rates for the initial training set of typical frames (training set 1), as well as each set of increasingly difficult challenge frames (training set 2-5).

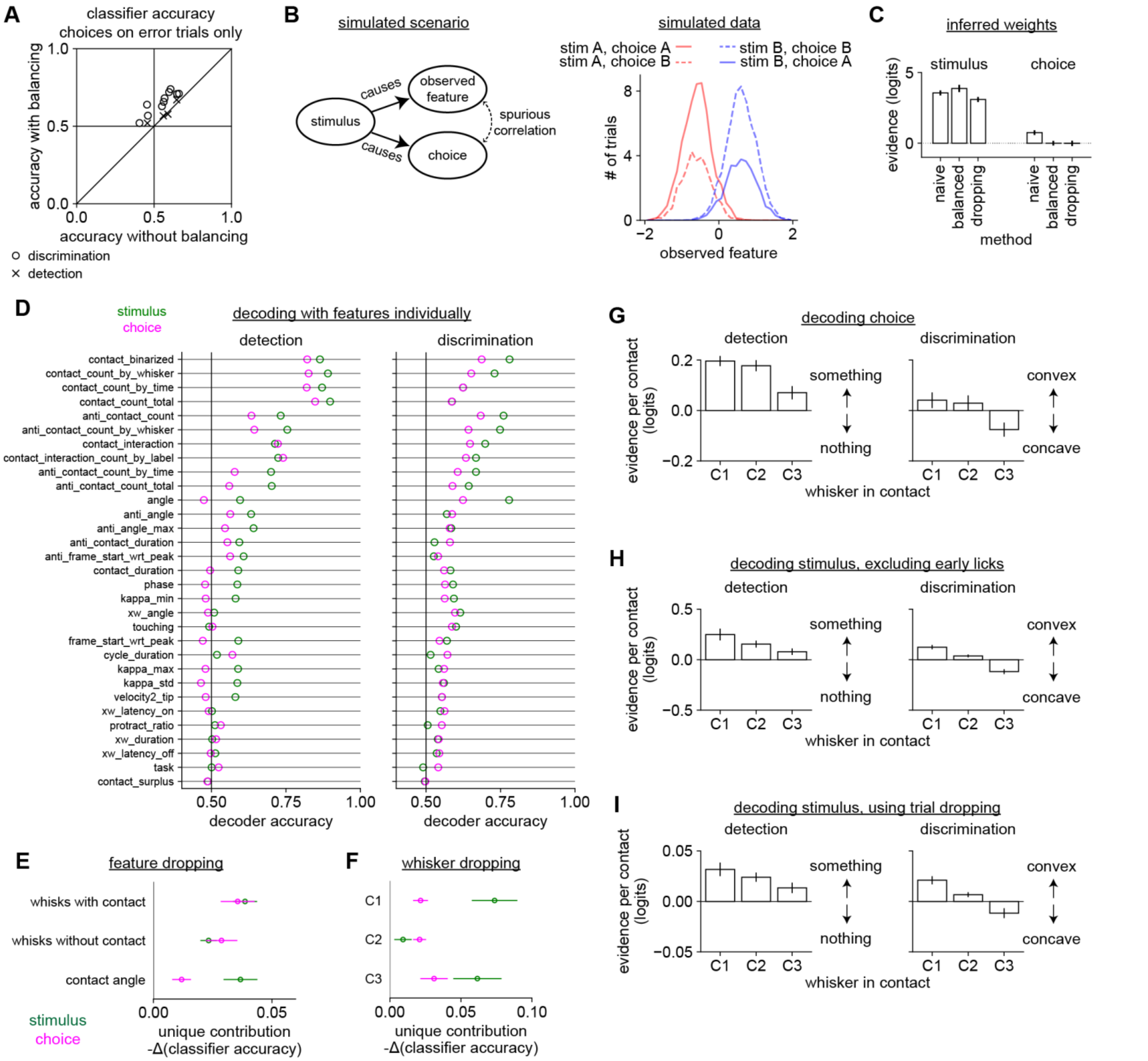
E) Whisking was broadly similar across mice regardless of task. Shown are the distributions of whisk cycle amplitude (left; trough to peak), whisk speed (middle; averaged over the time of protraction), and whisk cycle duration (right; trough to trough). The peak speed of the whiskers (data not shown) was much larger than the average speed. Each line shows an individual mouse performing discrimination (gray) or detection (red). Distributions were similar across mice, regardless of the task.

F) Contact kinematics were also similar across tasks (discrimination: gray; detection: red). Shown are the average values for each whisker of the maximum bend (left), minimum bend (middle), and standard deviation of the bend over the course of the contact (right). The minimal bend was negative and typically occurred during detachment. These are shown in the same m^{-1} units used in Fig 2F-G. Error bars: SEM over mice.

G) Distribution of contact durations, pooled across tasks and whiskers. Most contacts were short but there was a long tail of longer contacts. Contact durations are quantized at the frame rate of 5 ms.

Supplemental Figure 3, related to Figure 3.

Behavioral decoding validation and controls.



A) Accuracy of the behavioral decoders with and without trial balancing. The version described in the main manuscript uses balancing to equally weight correct and incorrect trials, and is plotted on the y-axis. We also trained a separate classifier that did not apply any balancing, but simply optimized its prediction of choice over all trials equally as in standard logistic regression; that accuracy is plotted on the x-axis. In both cases, the decoders had access to the full set of behavioral features. Here, we show the performance only on decoding choice on incorrect trials. Trial balancing improved the accuracy on this trial type in almost all cases and was always greater than 50%. On correct trials (data not shown), trial balancing slightly impaired performance, but was still quite high (*cf.* Fig 3D).

B) Simulated data. *Left*: In this simulation, we consider a hypothetical sensorimotor feature (*e.g.*, contact angle) that is purely driven by the stimulus identity and has no causal relationship with choice. However, because stimulus also drives choice, this sensorimotor feature will be correlated with choice as well. *Right*: Histograms of simulated data for this hypothetical feature. It is positive for stimulus B (blue) and negative for stimulus A (red), regardless of choice (dashed or solid). Note that a vertical line at 0 will separate the bulk of choice B on the right (dashed) from the bulk of choice A on the left (solid), even though this feature was constructed to have no causal relationship with choice.

C) Inferred evidence of stimulus (left) and choice (right) in the simulated data in panel B, using three different methods. All three methods reveal the presence of stimulus evidence (left three bars). The naive method spuriously reports the presence of choice evidence in the data, whereas trial balancing and trial dropping correctly reveal that the choice evidence is zero. Error bars: standard deviation over simulation repetitions.

D) Extended version of Fig 3B. The performance of classifiers trained on every individual feature. See Supplemental Table 1 on the following page for descriptions of each feature.

E) The unique contribution of each of the three features in the optimized behavioral decoder (Fig 3C), quantified by the decrease in accuracy of the full model after removing each feature individually.

F) The unique contribution of information from each of the three whiskers, quantified by the decrease in the accuracy of the optimized behavioral decoder after removing information from each whisker individually.

G) The weights for contacts made by each whisker, in order to predict choice. The results were qualitatively similar to the weights used to predict stimulus (Fig 3H). During discrimination, C1 and C3 weights have opposite sign.

H) The weights for contacts made by each whisker, excluding trials with early licks. Results are qualitatively similar to those calculated with all trials and presented in Fig 3H.

I) The weights for contacts made by each whisker inferred using trial dropping. Results are qualitatively similar to those calculated with trial balancing and presented in Fig 3H, though smaller due to increased noise of this method.

Error bars in E-I: SEM across n = 10 shape discrimination mice and n = 4 shape detection mice.

Supplemental Table 1, related to Figure 3.

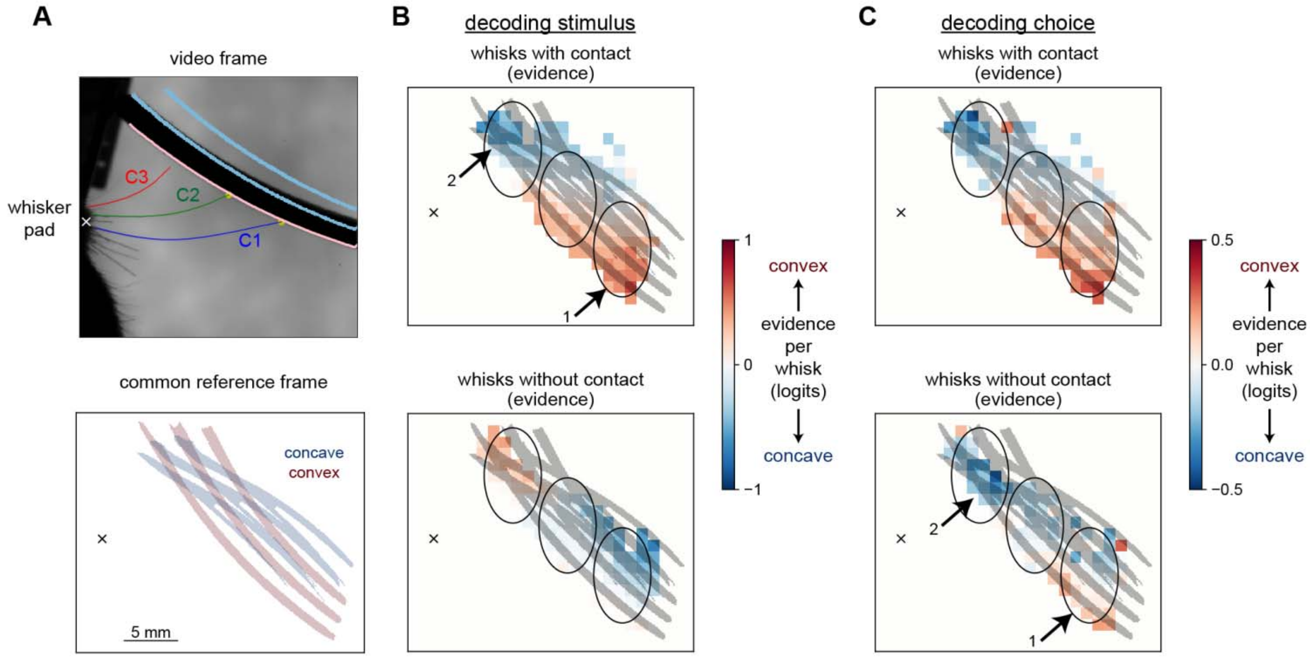
Complete set of features used in behavioral decoding.

This table defines the variables used in Supplemental Fig 3D, the entire list of features considered for inclusion in the behavioral decoders.

contact_binarized	Referred to as "whisks with contact" in the main text. A two-dimensional (whisker x time) binary array representing when each whisker was in contact over the course of the trial.
contact_count_by_whisker	The sum of the rows of "contact_binarized", so the number of contacts over the entire trial made by each whisker.
contact_count_by_time	The sum of the columns of "contact_binarized", so the number of contacts in each time bin
contact_count_total	The sum of the "contact_binarized", a single integer representing the total number of contacts made by all whiskers over all time points
anti_contact_count	Like "contact_binarized", but for whisks without contact.
anti_contact_count_by_whisker	Like "contact_count_by_whisker", but for whisks without contact.
contact_interaction	Like "contact_binarized", but for simultaneous contact made by two whiskers. The rows are all adjacent pairs of whiskers, and the columns are still time. We did not consider non-adjacent pairs as such contacts were quite infrequent.
contact_interaction_count_by_label	The sum of "contact_interaction" over time, so the total number of simultaneous contacts by each possible pair of whiskers over the entire trial.
anti_contact_count_by_time	Like "contact_count_by_time', but for whisks without contact.
angle	Referred to as "contact angle" in the main text. A two-dimensional (whisker x time) array representing the angle of contact of each whisker at each time in the trial. Contains null values where no contact occurred. When multiple contacts made by a whisker within a certain time bin, uses the average angle.
anti_contact_count_total	Like contact_count_total, but for whisks without contact
anti_angle_max	Like "angle", but for whisks without contact. Uses the peak angle of the whisk without contact.
anti_angle	Like "anti_angle_max", but instead of using the peak angle, uses the angle at which the whisk crossed the boundary where the shapes could have been present.
anti_contact_duration	Like "contact_duration", but for whisks without contact. The time during which the whisker was beyond the boundary where the shapes could have been present.
contact_duration	A two-dimensional (whisker x time) array representing the duration of contacts made by each whisker in that time bin. Contains null values where no contact occurred. When multiple contacts made by a whisker within a certain time bin, uses the average duration.
frame_start_wrt_peak	A two-dimensional (whisker x time) array representing the time of each contact relative to the peak of the whisk cycle on which it occurred. Contains null values where no contact occurred. When multiple contacts made by a whisker within a certain time bin, uses the average of each contact.
xw_angle	Cross-whisker angle. Like contact_interaction, but instead of a binary variable, it is the difference in angle between the two contacting whiskers at the time of contact. Contains null values where no such simultaneous contact occurred.
phase	Like "angle", but the phase of the contact.
kappa_min	Like "angle", but the minimum value of delta kappa achieved during the contact.
anti_frame_start_wrt_peak	Like "frame_start_wrt_peak", but for whisks without contact.
cycle_duration	A one-dimensional array versus timebins. The average duration of each whisk cycle.
kappa_std	Referred to as "contact-induced bending" in the manuscript. Like "kappa_min", but the standard deviation of delta kappa within the contact.
kappa_max	Like "kappa_min", but the maximum value of delta kappa within the contact.
touching	Like "contact_binarized", but nonzero whenever the whiskers is in contact, rather than just at the onset of the contact. Thus, it will be nonzero whenever "contact_binarized" is nonzero, but also on bins when the whisker maintained contact for more than one cycle.
velocity2_tip	Like "angle", but the angular velocity of the whisker averaged over the two frames preceding contact.
xw_latency_on	Like "xw_angle", but the time between contact onset of the two adjacent whiskers.
xw_duration	Like "xw_angle", but the mean duration of the contact onset of the two adjacent whiskers.
xw_latency_off	Like "xw_angle_on", but for the contact offset times.
protract_ratio	Like "cycle_duration", but the ratio between protraction time and the entire cycle duration.
task	Three variables: previous choice (left or right), previously rewarded side (left, right, or null if previous trial was not rewarded), and previously unrewarded side (left, right, or null if previous trial was rewarded).
contact_surplus	Like "contact_binarized", but the number of contacts on each whisk cycle minus one (instead of just a binary yes/no about contact).

Supplemental Figure 4, related to Figure 4.

The location sampled by the whiskers is informative about choice.

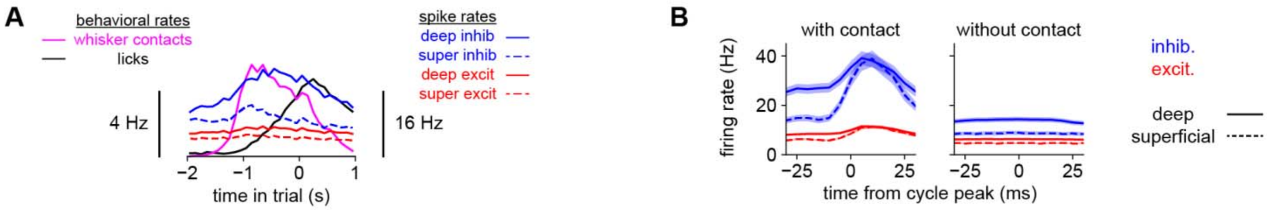


A,B) Replotted from Fig 4A,C for comparison.

C) The same data as in panel B (*top*: whisks with contact; *bottom*: whisks without contact) now colored by the evidence they contain about choice, *i.e.* whether the mouse would report the shape to be concave (blue) or convex (red). The top panel is similar to the analogous panel in B, indicating that mice correctly interpret whisks with contact in accordance with the ideal stimulus decoding. In the bottom panel, some whisks without contact are correctly interpreted by the mouse (that is, they look the same as in the analogous panel in B), but others are not. Specifically, in area 1, whisks without contact around the location of the closest convex shape indicate that the mouse would report convex. In area 2, whisks without contact around the location of the closest concave shape indicate that the mouse would report concave. This suggests that the mouse may be sampling these areas preferentially because it believes the corresponding shape is present.

Supplemental Figure 5, related to Figure 5.

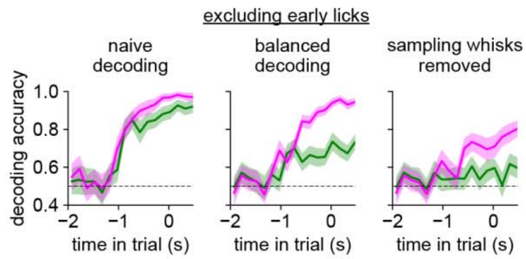
Barrel cortex neurons respond to whisker contacts, not licks.



A) The rate of contacts (pink) and licks (black), with scale bar 4 Hz, replotted from Fig 1I. We now also plot the spike rate of each cell type (blue: inhibitory; red: excitatory; solid: deep layers; dashed: superficial layers; scale bar: 16 Hz). On this timescale, the firing rates more closely track contacts than licks.

B) The same data from Fig 5F, but plotted in Hz instead of normalized to baseline. The results are qualitatively similar, but the inhibitory cells (blue) show a much higher baseline firing rate than the excitatory cells (red). Error bars: SEM over neurons.

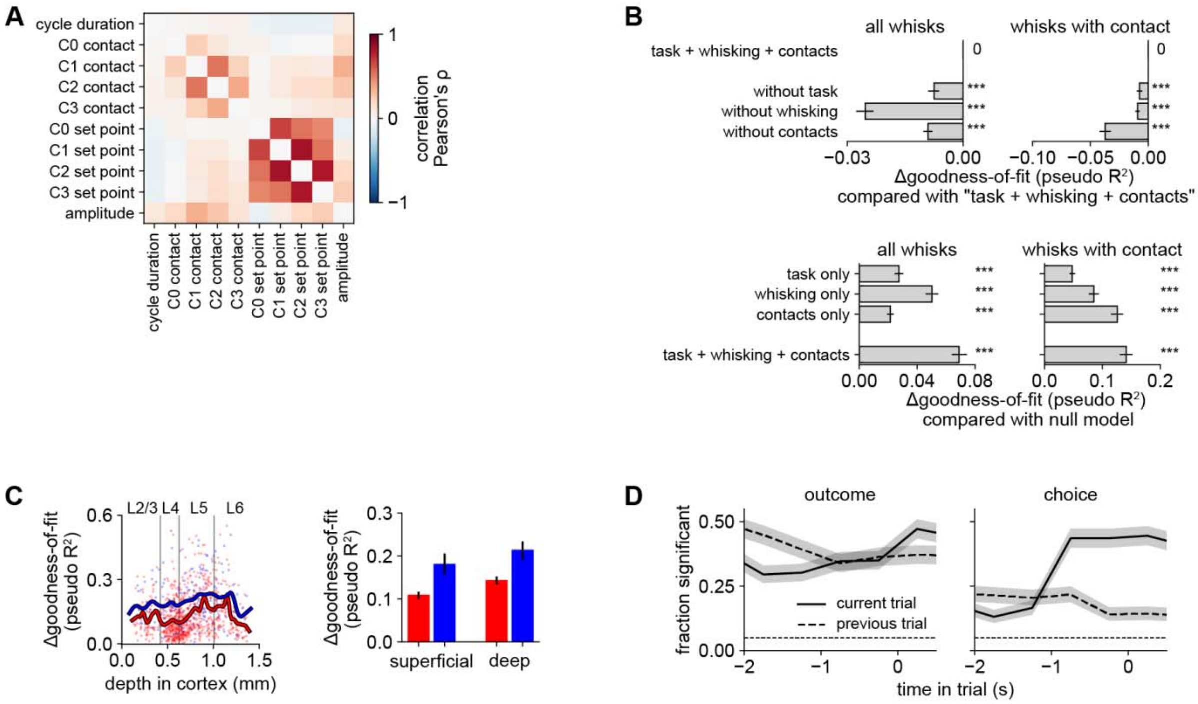
Supplemental Figure 6, related to Figure 6.
The persistent encoding of choice is not driven by licks.



Like Fig 6B, but excluding all trials in which even a single early lick was made. The results are almost exactly the same.

Supplemental Figure 7, related to Figure 7.

Additional characterization of the generalized linear model (GLM).



A) Feature correlation matrix. We concatenated the features over all sessions and calculated Pearson's ρ between every pair of features. For clarity, task-related temporal indicator variables are not included here because they are orthogonal by design, and auto-correlation values along the diagonal are masked.

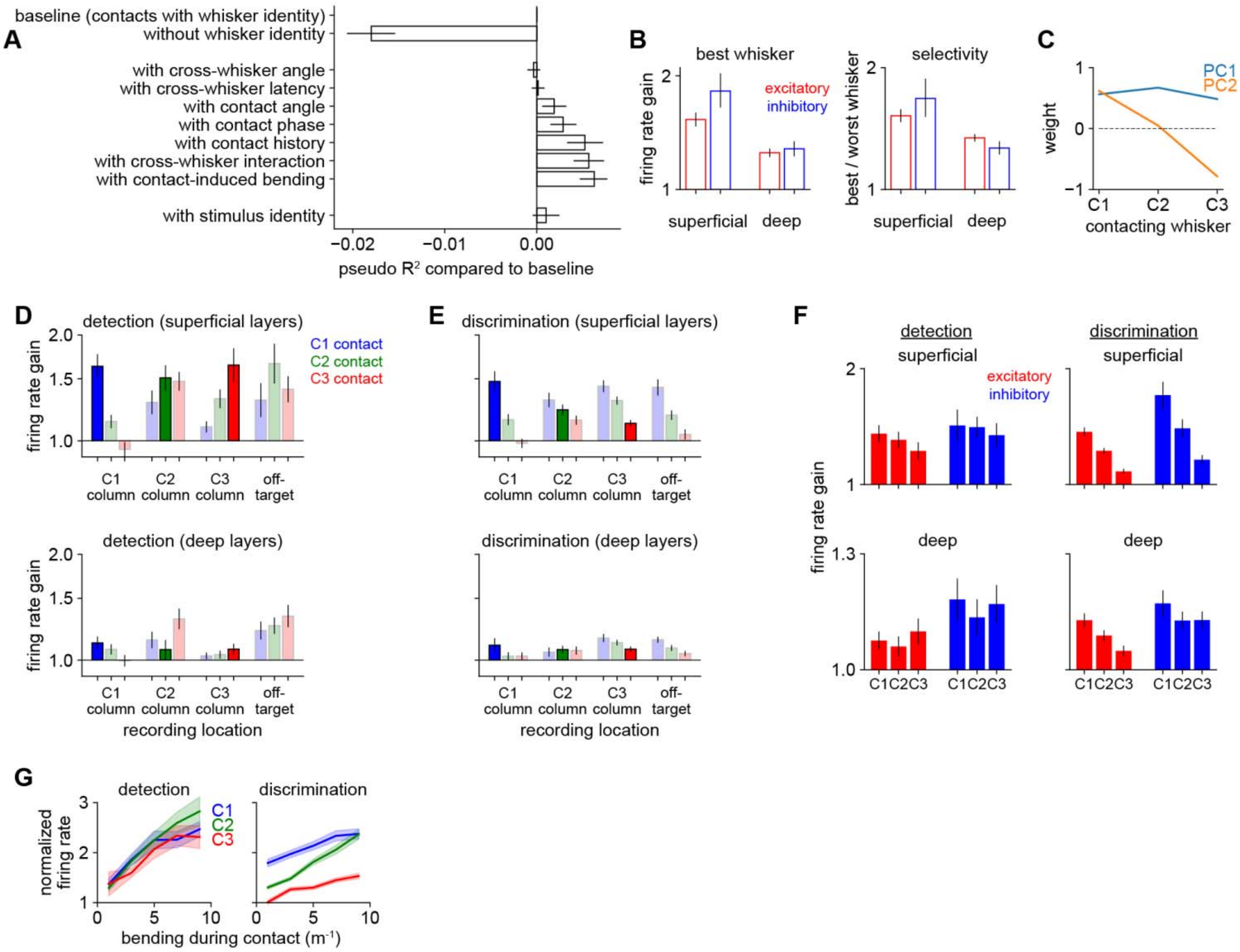
B) Like Fig 7B-C in the main manuscript, but plotted as the change in pseudo R^2 instead of log likelihood. The results are similar.

C) Like Fig 7D in the main manuscript, but plotted as the change in pseudo R^2 instead of log likelihood. The results are similar.

D) Like Fig 7F in the main manuscript (which already excluded any trials with optogenetic stimulation), but now also excluding sessions with any optogenetic stimulation at all. The results are similar.

Supplemental Figure 8, related to Figure 8.

Task-specific contact responses and somatotopy.



A) Like Fig 8G, but plotted as the change in pseudo R^2 instead of log likelihood.

B) *Left:* Similar to Fig 8E, but showing the response to the best whisker (*i.e.*, the whisker that evokes the strongest response in each neuron) instead of the average across C1-C3. The pattern is the same as in the main text: superficial neurons respond more strongly. *Right*: the selectivity of each neuron, parameterized as the response to the best whisker divided by the response to the worst whisker. Error bars: 95% bootstrapped confidence intervals.

C) Principle component analysis (PCA) applied to the response of each neuron to contacts made by each whisker. The first PC (blue) is nearly equal on all whiskers whereas the second PC tracks the anterior-posterior position of whiskers C1-C3. Thus, the first PC captures overall response strength and the second PC captures topographic preference (C1>C3 or vice versa). These two PCs capture 63% and 28% of the variance in contact responses.

D) The response to contacts by each whisker during shape detection, separately plotted by the cortical column in which they were recorded (C1, C2, C3, or "off-target", meaning all other columns). Superficial neurons (top) show somatotopy, typically preferring their topographically aligned whisker (dark bars) over all others (light bars). Deep neurons (bottom) show weaker responses and less somatotopy. Error bars: SEM over neurons.

E) Similar to panel D, but for the shape discrimination task. The preference for C1 contacts (blue bars) dominates somatotopic responses (dark bars). Error bars: SEM over neurons.

F) Similar to panels D and E, but pooling over recording locations and separately plotting by cell type (excitatory: red; inhibitory: blue). All populations prefer C1 contacts during discrimination (right). Error bars: SEM over neurons.

G) The neural response during each task to contacts made by each of the three whiskers, versus the amount of bending during that contact (a proxy for contact force). A whisker-specific effect is only observed in discrimination (*right*). The response is calculated as each neuron's evoked rate divided by its session mean firing rate, with 1.0 meaning no response. Error bars: SEM over neurons.

n = 235 neurons during detection and 675 neurons during discrimination. We excluded sessions in which too few contacts were made by any whisker (C1-C3) to estimate the response. Logarithmic y-axis in B (left) and D-F.