ALLEN INSTITUTE for BRAIN SCIENCE

Allen Brain Observatory: Visual Behavior Neuropixels

Technical Whitepaper

SECTION A: OVERVIEW

To further our understanding of the neural basis of behavior, the Visual Behavior Neuropixels project utilized the Allen Brain Observatory platform for *in vivo* Neuropixels recordings to collect a large-scale, highly standardized dataset consisting of recordings of neural activity in mice performing a visually guided task.

A key aspect of the experimental design involved the simultaneous recording of visual cortical, visual thalamic, hippocampal and midbrain structures across varying behavioral and sensory conditions, including stimulus novelty and fluctuations in motivation. Accordingly, this dataset can be used to investigate how sensory stimuli and task performance are represented by patterns of activity across the visual system, and how this activity is influenced by experience and engagement.

This dataset includes neural and behavioral measurements from 81 mice and 153 recording sessions, resulting in recordings from >200,000 units across visual cortical, thalamic, hippocampal and midbrain structures. These data are made openly accessible, with all recorded timeseries, behavioral events, and experimental metadata conveniently packaged in Neurodata Without Borders (NWB; Teeters et al, 2015) files that can be accessed and analyzed using our open-source Python software package, the AllenSDK.

This dataset builds upon 1) the recording pipeline established for the **Visual Coding Neuropixels** dataset and 2) the behavior training pipeline established for the **Visual Behavior Optical Physiology** dataset. We have focused this document on procedures that were not already described in the technical whitepapers for those previous datasets or required modification for this project. In the sections below, we have indicated where readers should consult the whitepapers for the previous projects for more experimental details. Here are links to the two relevant documents:

Visual Coding Neuropixels Technical Whitepaper: https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf

Visual Behavior Optical Physiology Technical Whitepaper: https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/4e/be/4ebe2911-bd38-4230-86c8-01a86cfd758e/visual_behavior_2p_technical_whitepaper.pdf

VISUAL BEHAVIOR ALLEN BRAIN OBSERVATORY PIPELINE

This dataset required the concerted effort of many teams to advance mice through a series of standardized procedures. Figure 1 summarizes these steps and indicates the appropriate reference where more experimental details can be found.

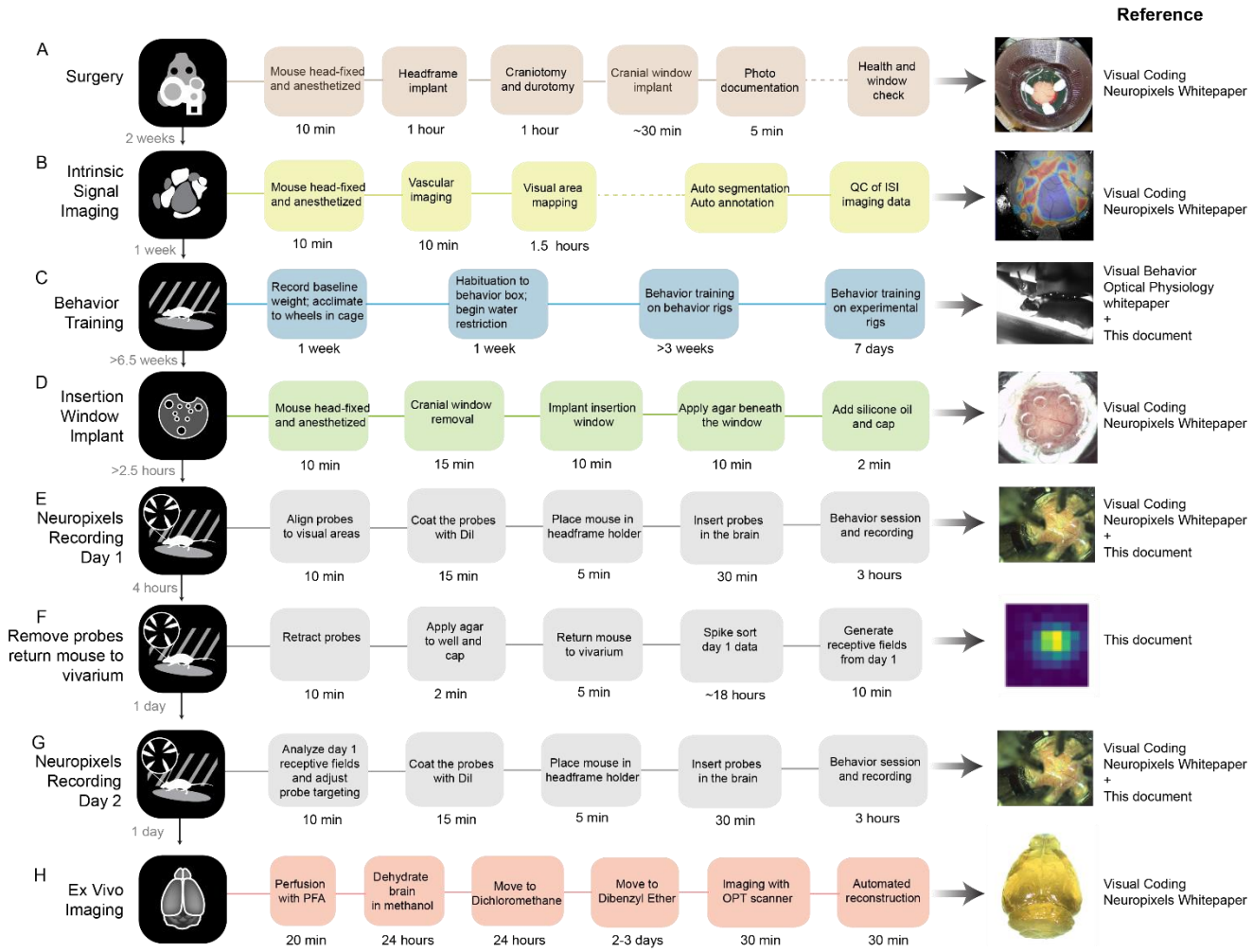


Figure 1. The Allen Brain Observatory Neuropixels workflow. The pipeline consists of 8 major steps: (A) a surgical procedure to implant a custom headframe and insert a glass window over visual cortex; (B) intrinsic signal imaging to identify cortical visual areas; (C) Behavior training of mice; (D) replacement of the glass window with a plastic window containing holes for inserting probes; (E) an in vivo extracellular electrophysiology experiment with behavior; (F) Removal of probes and processing of day 1 data; (G) a second in vivo extracellular electrophysiology experiment with behavior; (H) recovery of recording locations using optical projection tomography (OPT). The details and approximate duration of each step are described in each row. At the end of each row is a reference where readers can find more detailed information for each step.

CHANGE DETECTION TASK

The Visual Behavior dataset is built upon a change detection behavioral task. Briefly, in this go/no-go task, mice are presented with a continuous series of briefly presented stimuli and they earn water rewards by correctly reporting when the identity of the image changes.

Mice undergo standardized training in the behavior facility prior to imaging, in which they first learn to detect orientation changes using static grating stimuli, then transition to flashed gratings, and subsequently learn to detect changes in the identity of natural scene images. Each session included 8 images, for a total of 64 possible transitions. The images used in this study were a selected subset of the natural scenes stimulus set used in the Visual Coding Allen Brain Observatory.

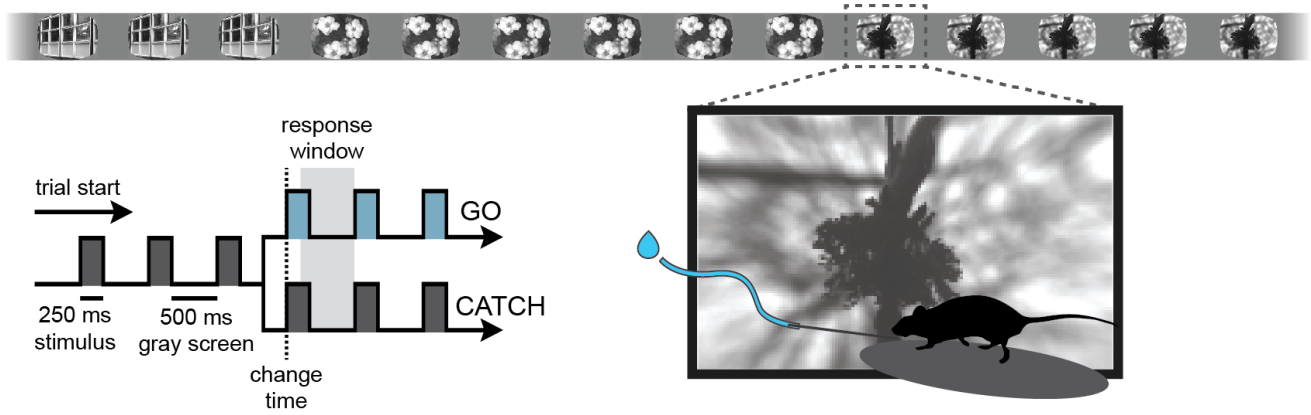


Figure 2. Change detection task. Top row shows the continuous stream of stimulus presentations (250ms per stimulus), with inter stimulus interval gray screen (500ms) that are displayed to the mouse during a behavior session. Lower left shows the two trial types in the task, "go" trials where the stimulus identity changes and the mouse must lick within a 750ms response window to earn a water reward, and "catch" trials where the image identity doesn't change and false alarm licking behavior is quantified. Bottom right shows the behavioral setup, with stimuli displayed on a monitor facing the right eye of the mouse, a lick spout for response detection via a capacitive sensor and water reward delivery, and a running wheel.

NEUROPIXELS RECORDINGS

This dataset includes recordings using Neuropixels 1.0 probes. We inserted up to 6 probes simultaneously in each mouse for up to two consecutive recording days. After the first recording day, probes were removed and the mouse was returned to its home cage. On the second recording day, probes were re-inserted. Probes were typically inserted 3-3.5 mm, resulting in 300-350 recording sites per probe.

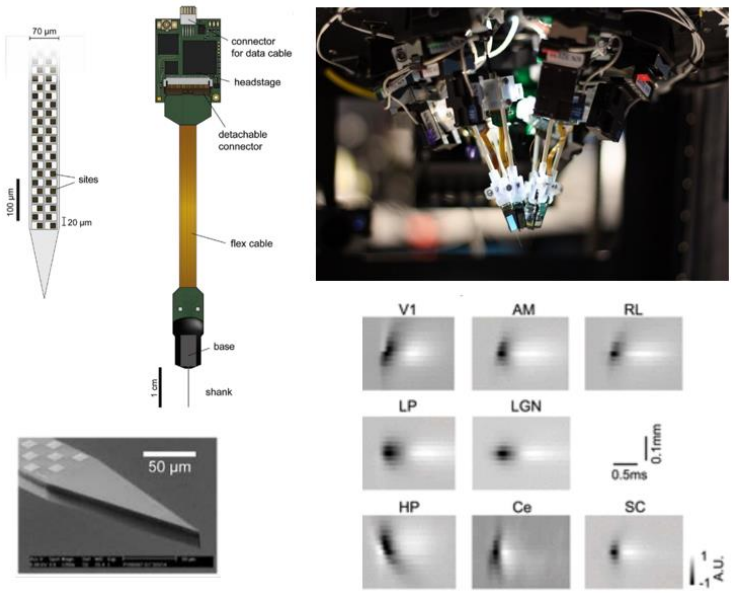


Figure 3. Neuropixels recordings. Left: Diagram of the probe at three different zoom levels, showing the recording site layout, the scale, and the overall package. Right top: Six probes mounted on the recording rig. Each probe is attached to an independent 3-axis manipulator and is angled to fit within a ~5 mm recording window. Right bottom: Example 2D waveforms for 8 example units recorded with Neuropixels probes across cortical and subcortical brain areas. For each of these plots, the x-axis indicates time and the y-axis indicates the recording site. The color indicates voltage in arbitrary units (plots have been normalized for comparison). Note that due to the density of recording sites for these probes, the waveform for a given unit extends across many channels.

EXPERIMENTAL DESIGN

Once mice are well-trained in the task, they transition to performing the task while brain activity is recorded with Neuropixels probes to enable simultaneous measurement of neural activity and behavior. Each mouse undergoes two experimental Neuropixels sessions. During the one recording session, mice perform the task with the set of eight natural scene images they viewed during training. During the other recording session, they perform the task with a novel set of images that they had not seen before. This allows evaluation of the impact of novelty on neural coding for stimulus and behavioral information. Mice also undergo a passive viewing block at the end of each recording sessions during which we present the task stimuli with the lick spout retracted so they are unable to earn water rewards (this allows analysis sensory activity during passive rather than task-engaged conditions). Lastly, during recording sessions only, stimuli were omitted with a 5% probability, disrupting the expected cadence of flashed image presentations. Stimulus changes, the stimulus immediately preceding the change, and the stimulus immediately following an omission were never omitted.

Recordings were made from a given mouse on two consecutive days. After the first experiment, the probes were removed and the well was covered with agar and sealed. The mouse was then returned to the vivarium. On the second recording day, any agar sitting on top of the implantation window was cleared and a fresh layer of silicone oil was applied. Probes were then reinserted. Since recordings were acute, units cannot be tracked across the two recording days.

Figure 4. Experimental design (below):

Image Sets used for behavior: We used two image sets, G and H, both comprising 8 natural images. Two images appeared in both image sets (im083_r and im111_r), allowing within-session analysis of stimulus novelty.

Training and Recording Workflow: Each mouse was trained in the behavior facility on one of the two image sets (G or H). Once mice reached the performance criterion (see Section D. Behavior) they were transferred to the Neuropixels team and trained for several days on the experimental rig to habituate them to the recording environment. Following habituation, neural activity was recorded while the mouse performed the behavior on two consecutive recording days. One recording day used the image set familiar to the mouse from training and the other used the novel image set. The numbers in each box indicate the number of sessions in the dataset for each training trajectory, as returned by the default SDK flags. We have given the total number of available sessions (without default filters) in parentheses.

Experimental Session Overview: Each recording session consisted of 4 main stimulus epochs: 1) active behavior during which the mouse performed the change detection task, 2) receptive field characterization using gabors and full-field flashes, 3) passive replay during which the stimuli encountered during the active behavior were replayed but without the lick spout present, and 4) optotagging.

Change Detection Task Structure. During the task, mice are presented with a continuous series of briefly presented stimuli and they earn water rewards by correctly reporting when the identity of the image changes. The regular cadence of stimulus presentations was occasionally disrupted by a stimulus omission, which occurred randomly on 5% of non-change stimuli.

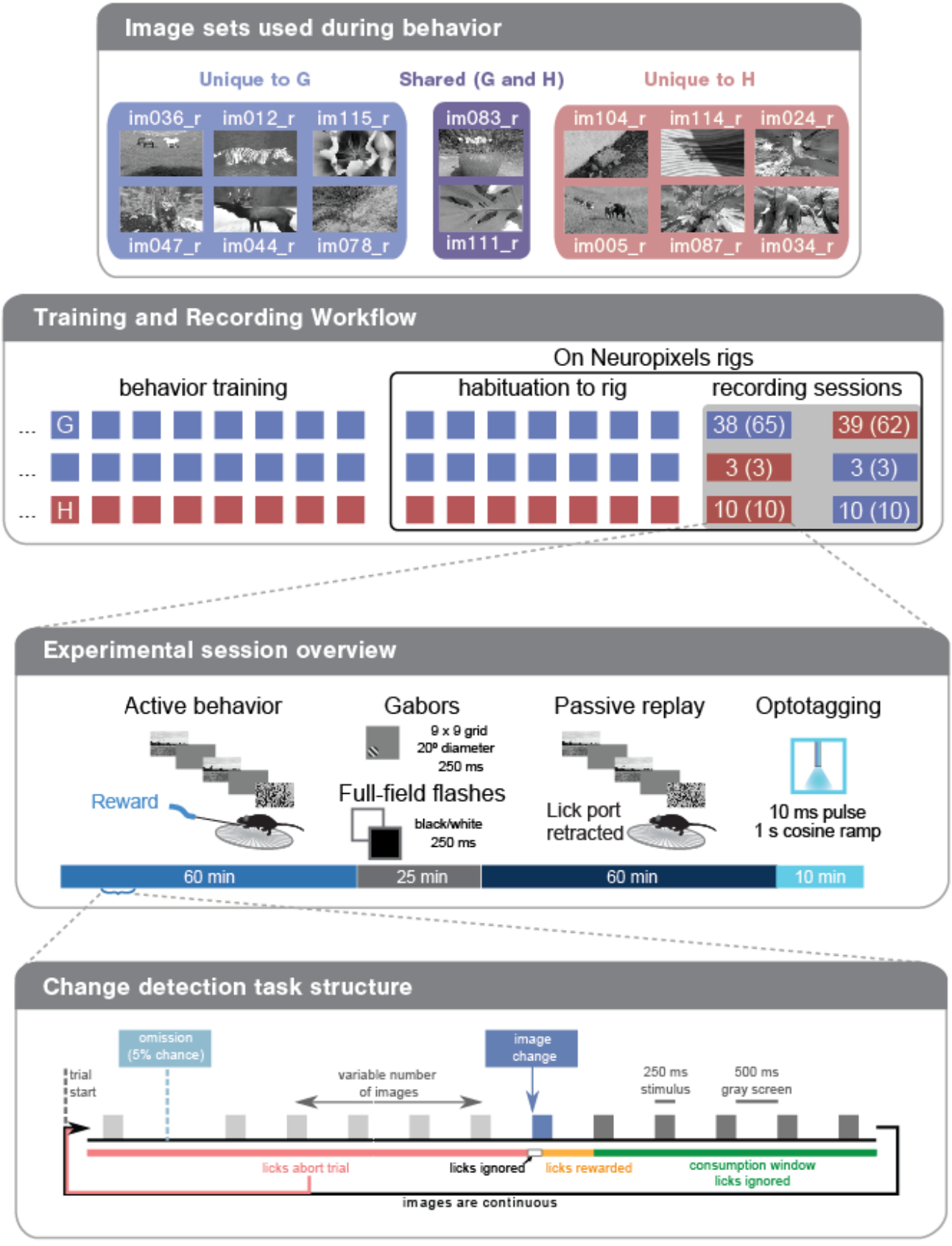


Image sets used during behavior
Unique to G
Shared (G and H)
Unique to H
im036_r
im012_r
im115_r
im047_r
im044_r
im078_r
im083_r
im111_r
im104_r
im114_r
im024_r
im005_r
im087_r
im034_r
Training and Recording Workflow
On Neuropixels rigs
behavior training
habituation to rig
recording sessions
G
H
38 (65)
39 (62)
3 (3)
3 (3)
10 (10)
10 (10)
Experimental session overview
Active behavior
Gabors
Passive replay
Optotagging
Reward
9 x 9 grid
20° diameter
250 ms
Full-field flashes
black/white
250 ms
Lick port
retracted
10 ms pulse
1 s cosine ramp
60 min
25 min
60 min
10 min
Change detection task structure
omission
(5% chance)
image
change
trial
start
variable number
of images
250 ms
stimulus
500 ms
gray screen
licks abort trial
licks ignored
licks rewarded
consumption window
licks ignored
images are continuous

DATA STRUCTURE

This dataset varies along the following dimensions:

Genotype: Mice were used from 3 genotypes: 1) Sst-IRES-Cre;Ai32 ChR2 reporter mice (**Sst**), 2) Vip-IRES-Cre;Ai32 (**Vip**), and 3) C57BL6J (**wt**). Recording in Sst and Vip mice allowed us to identify putative Sst+ and Vip+ interneurons which, due to the Ai32 cross, expressed ChR2 and thus could be activated with blue light (Zhang et al 2006).

Session number: Indicates whether a session was the first or second recording day for a given mouse.

Image set: Indicates whether image set G or H was shown during a particular session.

Experience level: Indicates whether the image set shown was familiar to the mouse from training or novel.

Mice took one of three paths through the possible behavior training and Neuropixels recording configurations (as pictured in the 'Training and Recording Workflow' bubble of Figure 4).

1) Train with image set G; see image set G on recording day 1 and image set H on recording day 2
2) Train with image set G; see image set H on recording day 1 and image set G on recording day 2
3) Train with image set H; see image set H on recording day 1 and image set G on recording day 2

Data Abnormalities

We noted two types of abnormalities in sessions which otherwise passed quality control assessment. These sessions are flagged in the metadata table accompanying this release and are not returned by the SDK by default. Nonetheless, we make these data available, as some analyses may benefit from their inclusion.

In 27 mice we noted abnormalities in the post-hoc OPT imaging, indicating potential tissue damage from probe insertion. Since we were unable to assign this damage to a particular session, we have flagged both sessions from these mice as having 'abnormal histology' in the sessions metadata table accompanying this release. In 12 sessions (11 of which came from mice with abnormal histology), we noted potential epileptiform activity during the recording session. We have flagged these sessions as having 'abnormal activity' in the sessions metadata table.

Dataset Summary

See Tables 1 and 2 below for a summary of the number of sessions available across the four dataset dimensions described above.

		genotype	SST	VIP	C57BL6J
session_number	image_set	experience_level			
1	G	Familiar	19.0	11.0	8.0
	H	Familiar	0.0	0.0	10.0
		Novel	3.0	0.0	0.0
2	G	Familiar	3.0	0.0	0.0
		Novel	0.0	0.0	10.0
	H	Novel	18.0	11.0	10.0

Table 1. Number of sessions for each condition as returned by default by the SDK (no abnormalities)

		genotype	SST	VIP	C57BL6J
session_number	image_set	experience_level			
1	G	Familiar	28.0	22.0	15.0
	H	Familiar	0.0	0.0	10.0
		Novel	3.0	0.0	0.0
2	G	Familiar	3.0	0.0	0.0
		Novel	0.0	0.0	10.0
	H	Novel	25.0	23.0	14.0

Table 2. Total number of sessions without filtering out abnormalities.

SECTION B: MICE & SURGERY

MICE

Wild-type C57BL/6J mice were bred in-house. For experiments involving optotagging of inhibitory cells, Vip-IRES-Cre, and Sst-IRES-Cre mice were bred in-house and crossed to an Ai32 reporter line (Madisen et al., 2012). Sst-IRES-Cre; Ai32 breeding sets (pairs and trios) consisted of heterozygous Sst-IRES-Cre mice mated to either heterozygous or homozygous Ai32(RCLChR2(H134R)_EYFP) mice. Vip-IRES-Cre; Ai32 breeding sets (pairs and trios) consisted of heterozygous Vip-IRES-Cre mice mated to either heterozygous or homozygous Ai32(RCL-ChR2(H134R)_EYFP) mice. Cre+ cells from Ai32 lines are highly photosensitive, due to expression of Channelrhodopsin-2 (Zhang et al., 2006). For more information on transgenic mice, see https://observatory.brain-map.org/visualcoding/transgenic. Prior to surgery mice were singly-housed and maintained on a reverse 12-hour light cycle (off at 9am, on at 9pm); all experiments were performed during the dark cycle.

SURGERY

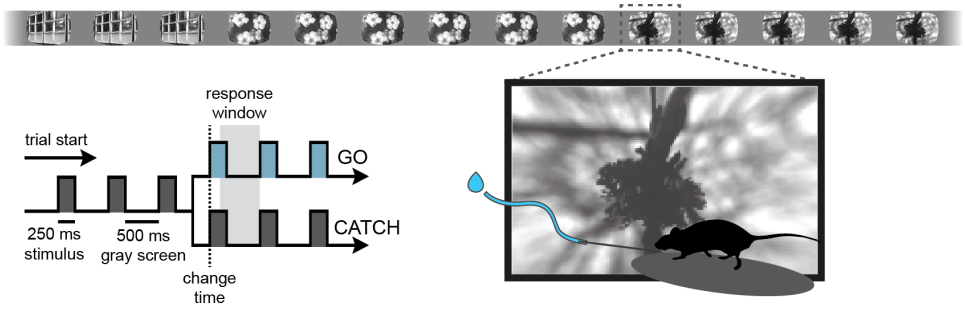
The headframe design and cranial window surgery for this dataset were identical to that described in the Visual Coding Neuropixels whitepaper: https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf

SECTION C: INTRINSIC SIGNAL IMAGING

Intrinsic signal imaging was performed to locate the visual cortical areas as described in the Visual Coding Neuropixels whitepaper: https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf

SECTION D: BEHAVIOR

The Change Detection task is a go/no-go task wherein mice are presented with a continuous series of briefly presented stimuli and they earn water rewards by correctly reporting when the identity of the flashed image changes.



(from Groblewski & Ollerenshaw, 2020)

Figure 1. Change detection task. Top row shows the continuous stream of stimulus presentations (250ms per stimulus), with inter stimulus interval gray screen (500ms) that are displayed to the mouse during a behavior session. Lower left shows the two trial types in the task, "go" trials where the stimulus identity changes and the mouse must lick within a 750ms response window to earn a water reward, and "catch" trials where the image identity doesn't change and false alarm licking behavior is quantified. Bottom right shows the behavioral setup, with stimuli displayed on a monitor facing the right eye of the mouse, a lick spout for response detection via a capacitive sensor and water reward delivery, and a running wheel.

WATER RESTRICTION AND HABITUATION

Throughout training mice were water-restricted to motivate learning and performance of the behavioral task. Mice had access to water only during behavioral training sessions or when provided by a technician on non-training days. During the first week of water restriction mice were habituated to daily handling and increasing durations of head fixation in the behavior enclosure over a five-day period. The first day of behavior training began after 10 days of water restriction. Mice were trained 5 days per week (Monday-Friday) and were allowed to earn unlimited water during the daily 1 hour sessions; supplements were provided in a home cage water dish if the earned volume fell below 1.0mL and/or body weight fell under 80-85% of initial baseline weight. On non-training days mice were weighed and received water provision to reach their target weight, but never less than 1.0 mL per day.

APPARATUS

Mice were trained in custom-designed, sound-attenuating behavior enclosures (see **Figure 2**) equipped with a 24" gamma-corrected LCD monitor (ASUS, #PA248Q) and behavior camera (Allied Vision). Mice were head-fixed on a removable behavior stage with 6.5" foam-covered running wheel tilted upwards by 10-15 degrees (see Figure

2. Running speed was recorded using an analog encoder (US Digital, MA-3, 0-5V) sampled via a data acquisition board (National Instruments 6001 DAQ). The center of the visual monitor was placed 15 cm from the eye (see Section E) and visual stimuli were spherically warped to account for the variable distance from the eye toward the periphery of the monitor. Water rewards were delivered using a solenoid (NI Research, #161K011) to deliver a calibrated volume of fluid (5-10μL) through a blunted, 82mm 18g hypodermic needle (Hamilton) mounted to an air cylinder with stroke of 67mm, and positioned approximately 2-3 mm away from the animal's mouth. The lick spout system is electrically connected to an Arduino for capacitive change lick detection. This system is mounted on a custom XYZ automated linear stage with 13mm travel in each axis enabling customizable and repeatable placement of the lickspout for each mouse during experimental sessions which span many days and across multiple scientific instruments. The lickspout retracts for safe load and unload of the mouse.

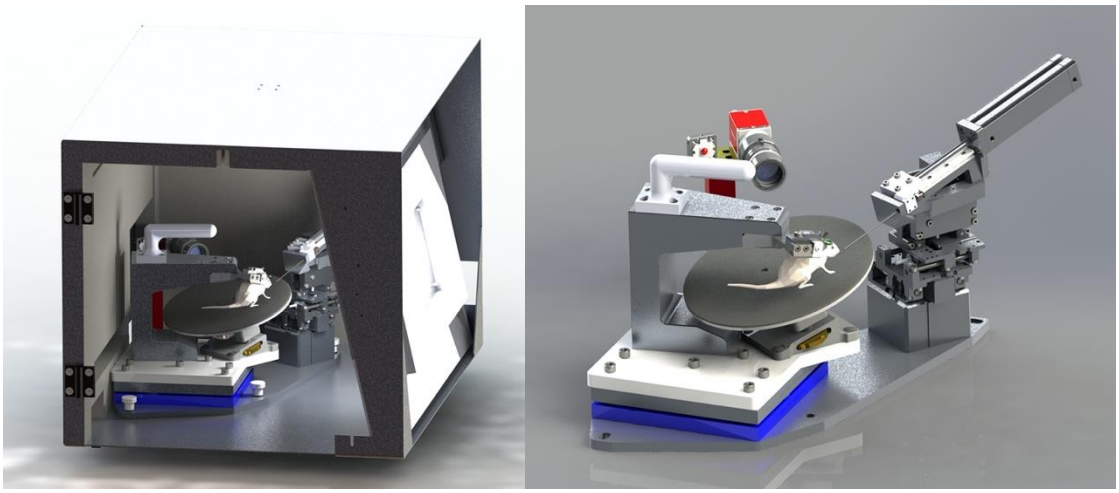
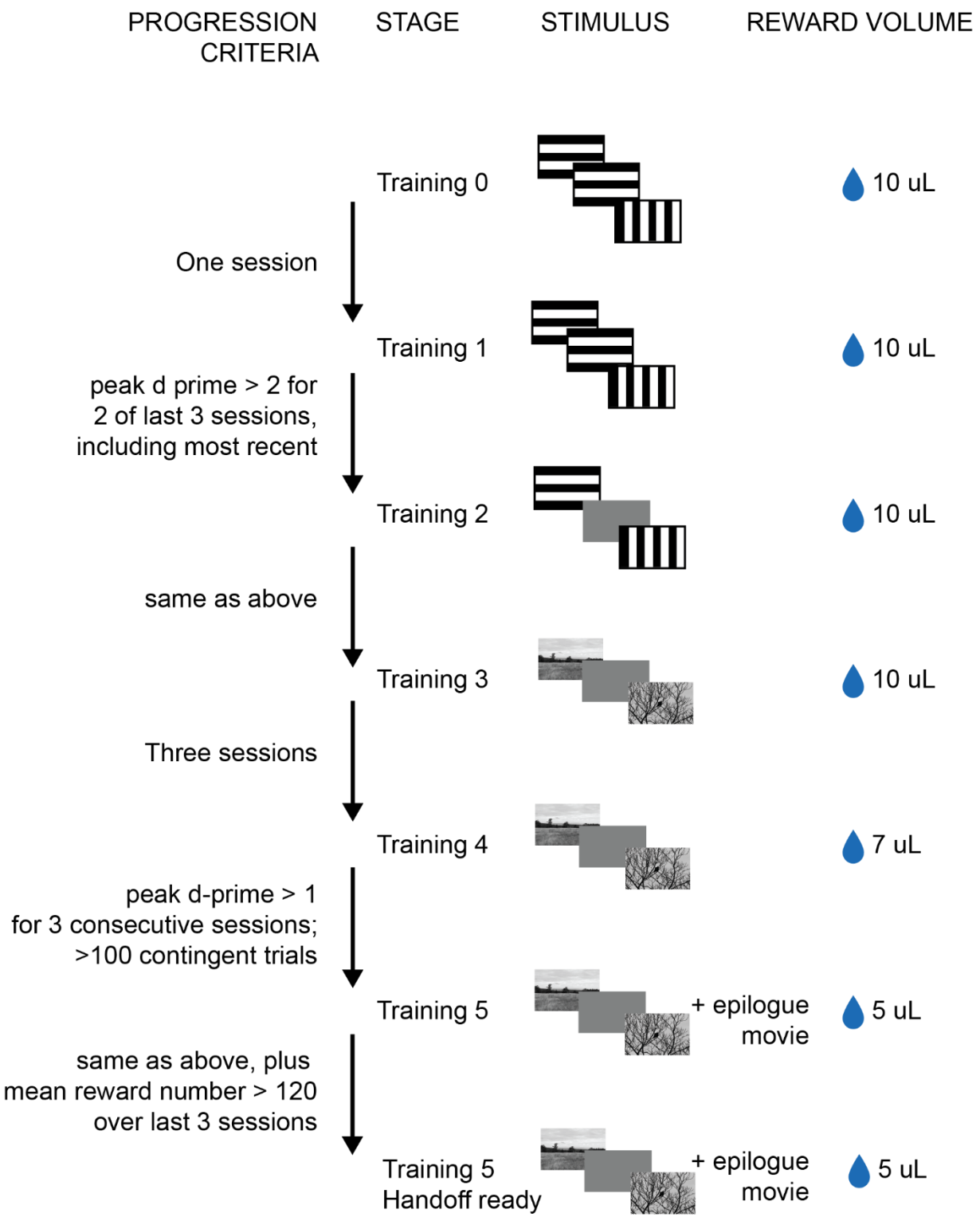


Figure 2. Behavior enclosure and stage. Custom behavior enclosure (LEFT) and removable behavior stage, shown with lick spout module and camera (RIGHT).

CHANGE DETECTION TASK

Overview: Mice were trained for 1 hour/day, 5 days/week using a behavioral program implementing a go/no-go change detection task schematized in **Figure 1**. Briefly, mice were trained to lick a reward spout when the identity of a flashed visual stimulus changed identity. If mice responded correctly within a short, post-change response window (150-750ms before compensating for monitor display lag) a water reward was delivered. The training stages, stimuli used, reward volumes given and progression criteria are diagrammed in **Figure 3**.

Figure 3 (below). Behavior training stages. Mice progressed through 6 training stages before qualifying to be handed off to the Neuropixels team.



On Day 1 of the automated training protocol (shown in Figure 3) mice received a short, 15-min “open loop” session during which non-contingent water rewards were delivered coincident with 90° changes in orientation of a full-field, static square-wave grating (Stage 0). This session was intended to 1) introduce the mouse to the fluid delivery system and, 2) provide the technician an opportunity to identify the optimal lick spout position for each mouse. Each session thereafter was run in “closed loop”, and progressed through 6 stages of the operant

task: **Training 1)** static, full-field square wave gratings (oriented at 0° and 90°, with the black/white transition always centered on the screen and the phase chosen randomly on every trial), **Training 2)** flashed, full-field square-wave gratings (0° and 90°, with phase as described in 1), and **Training 3 through 5 Handoff-ready)** flashed full-field natural scenes (8 natural images used in the *Allen Brain Observatory*, http://observatory.brain-map.org/visualcoding). As mice progressed through the training stages, the reward volume was gradually reduced to prevent mice from satiating during the session. Once in the final stage of the task, mice were introduced to the additional session components that would be included during recording sessions (5 min waiting period before the behavior session to acclimate mice to the time taken for probe insertion, as well as a 5 min epilogue movie exposing mice to gabors and full-field flash stimuli). During recording sessions, the reward volume for most mice was reduced further to 3 µL.

Progression through training stages: Mice progressed through each training stage by meeting stage-specific progression criteria, as shown in **Figure 3**. For these criteria, *peak d-prime* was defined as the maximum d-prime for a session calculated over a 100 trial rolling window; *contingent trials* were defined as any trials not aborted by premature licking.

Behavior session and trial structure: Each behavior session consisted of a continuous series of trials, schematized below.

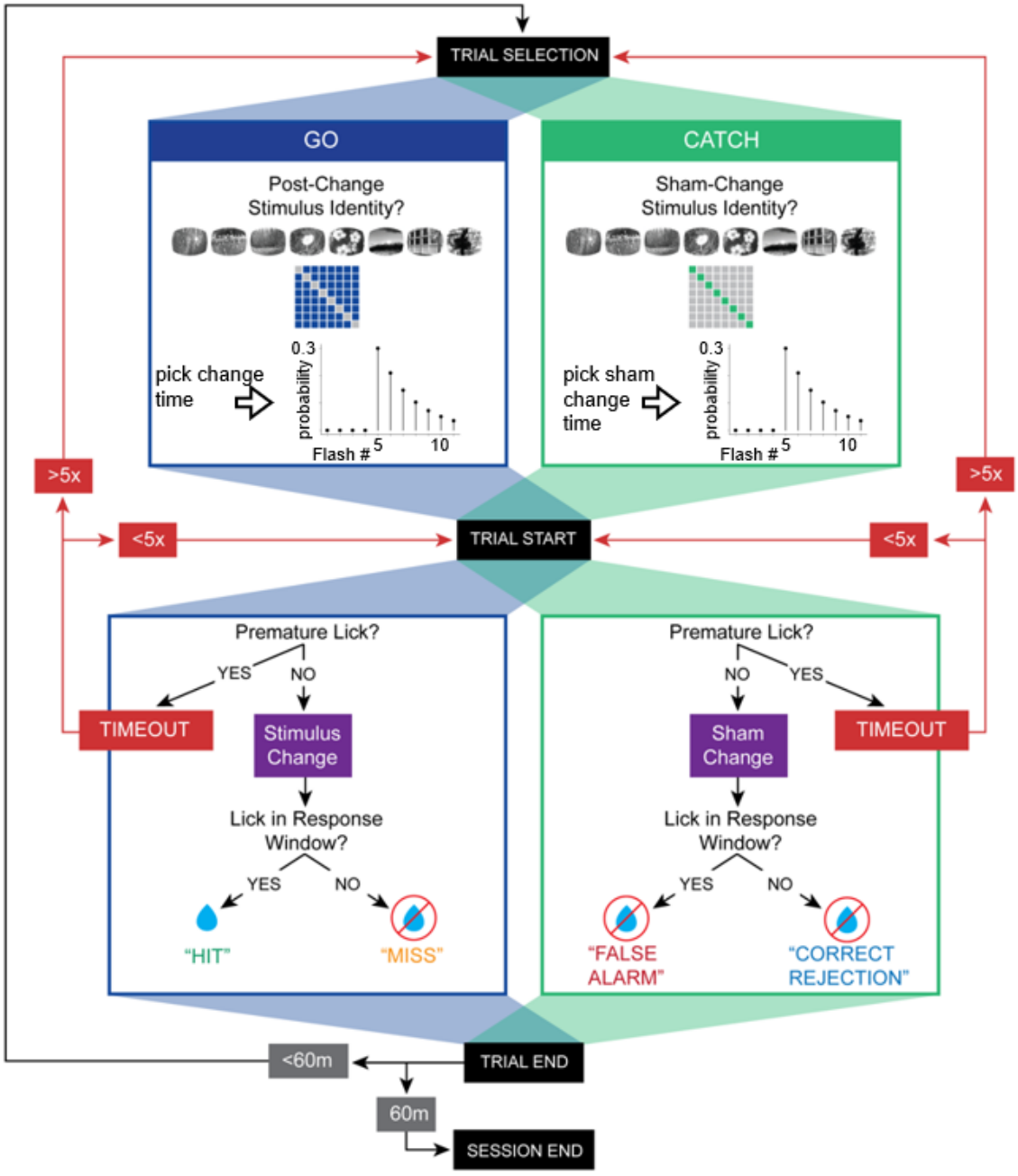


Figure 4 (above). Task flow diagram. The "Flashed Images" stage of the change detection task consists of 8 images, resulting in 64 possible image transitions, including both GO and CATCH trials. GO trials comprise 87.5% of all trials and are represented in the off-diagonal portions of the 8x8 change matrix. CATCH trials comprise 12.5% of all trials and are represented in the diagonal of the matrix. Each trial was first selected as either GO or CATCH and a post-change (or sham-change) image identity was chosen from the change matrix. The stimulus change (or sham-change) time was then selected from a truncated geometric distribution from 5 to 11 stimulus flashes after the trial start. Once a trial started, a premature lick (i.e., a lick that occurred prior to the predetermined change time) resulted in a 300 ms timeout and the trial was restarted at the time of the next scheduled stimulus presentation. If an animal caused a trial to timeout 5 times, new trial parameters were selected. If no premature licks were recorded, the trial progressed and the stimulus change (or sham-change) occurred at the predetermined change-time. On GO trials, a lick detected within a response window (150 ms to 750 ms following image change, before compensating for display lag) resulted in a "HIT" (and subsequent reward delivered) whereas a lack of response resulted in a "MISS". On CATCH trials, a lick within the window following the sham-change resulted in a "FALSE ALARM", whereas a lack of

response resulted in a “CORRECT REJECTION”. Following the stimulus change and response window the trial ended and a new trial was selected. The session ended after 60 minutes.

Prior to the start of each trial a trial-type and change-time were selected. Trial-type was chosen based on predetermined frequencies such that “GO” and “CATCH” trials occurred with specified probabilities. In stages 1 and 2, the catch probability was ~36%. In stages 3 and above, the catch probability was ~12.5%. Change-times were selected from a truncated geometric distribution from 5 to 11 stimulus flashes after the trial start. In trials when a mouse licked prior to the stimulus change the trial was reset, and a timeout period was imposed. The number of times a trial could be reset before re-drawing the timing parameter was limited to five. In all, this trial structure leads to a sampling of “GO” and “CATCH” trials, that when combined with mouse responding, yields “HIT”, “MISS”, “FALSE ALARM”, and “CORRECT REJECTION” trials.

In addition to the four trial types described above, behavior sessions contained a subset of “free reward” trials (“GO” trials followed immediately by delivery of a non-contingent reward). Behavior sessions across all phases began with 5 “free-reward” trials. Additionally, in order to promote continued task performance throughout the behavior session “free reward” trials were delivered after 10 consecutive “MISS” trials. These free rewards were not given during recording sessions. All non-contingent rewards were 5uL in volume.

Transition to Recording: Mice were deemed to be “handoff ready” when they had achieved the following performance criteria during the final phase of behavior:

1. A peak d-prime of >1 for three consecutive sessions
2. At least 100 contingent (non-aborted) trials on three consecutive sessions
3. Mean reward number >120 over last three sessions

If at any time during the final stage of behavior mice did not meet these criteria they would become “lapsed” until they met the criteria and returned to “handoff ready”. Additionally, due to operational constraints, mice were sometimes not immediately transitioned to the recording phase of the experiment despite having met the criteria.

BEHAVIOR METRICS

Hit and false alarm rates: The hit rate was calculated as the fraction of go-trials in which the mouse licked in a 0.150 to 0.750 second window following the non-display-lag-compensated image display time. On Catch trials, a response window was defined as a 0.150 to 0.750 second window following the sham change. False alarm rates were calculated as the fraction of catch-trials in which the animal licked in this response window.

Reaction times were recalculated post-hoc after accounting for the calculated display lag, which ranged from ~20-35 ms. Due to variability in the display lag calculations, the minimum and maximum reaction times that resulted in trial classification (hit/miss/false-alarm/correct-rejection) will vary from session to session.

Hit and false alarm rates were corrected to account for low trial counts by bounding their values using the following formula (Macmillan and Creelman, 2004):

$$\frac{1}{2N_H} \leq R_H \leq 1 - \frac{1}{2N_H}$$

$$\frac{1}{2N_F} \leq R_F \leq 1 - \frac{1}{2N_F}$$

Where R_H and R_f represent the hit and false alarm rates, and N_H and N_f represents the number of hit and false alarm trials.

D-prime (d'), which is a measure of the relative difference in response probabilities across the two trial types, is defined as:

$$d' = \mathrm{Z}(R_H) - \mathrm{Z}(R_F)$$

in which Z represents the inverse cumulative normal distribution function.

The AllenSDK returns hit rates, false alarm rates, and d-prime calculated over a rolling 100 trial window, excluding aborted trials (trials where the animal responded before the stimulus change). The reward rate is calculated over a 25 trial rolling window and provides a measure of the rewards earned per unit time (in units of rewards/minute). This provides a rough proxy for task engagement, with periods of high reward rates indicating engagement in the visual task. Low reward rates can indicate either disengagement (lack of responses), or extremely high rates of anticipatory licking (thereby continually delaying stimulus changes).

See detailed documentation of the SDK methods in the `get_rolling_performance_df` function in the behavior_session module, the `calculate_reward_rate` method in the trials_processing module, and in the readme.

RUNNING SPEED CALCULATION

Running speed was calculated as described in the Visual Behavior Optical Physiology documentation: https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/4e/be/4ebe2911-bd38-4230-86c8-01a86cfd758e/visual_behavior_2p_technical_whitepaper.pdf

SECTION E: NEUROPIXELS RECORDINGS

HARDWARE & INSTRUMENTATION

All hardware and instrumentation for the recordings in this dataset are identical to that described in the Visual Coding Neuropixels documentation (https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf), with the following exceptions:

Probes

Whereas the Visual Coding Neuropixels dataset included data collected with Neuropixels 3a as well as Neuropixels 1.0 probes, all data in this dataset were collected with Neuropixels 1.0 probes.

Experimental Rig

The original rig design was largely preserved (**Figure 1**), but several modifications were made to accommodate behavior data collection, as highlighted in **Figure 2**. These were:

1) The addition of a face camera in front of the mouse
2) Adjustment of the position of the body camera to capture a larger field of view
3) The addition of a lick spout to record licking and deliver rewards

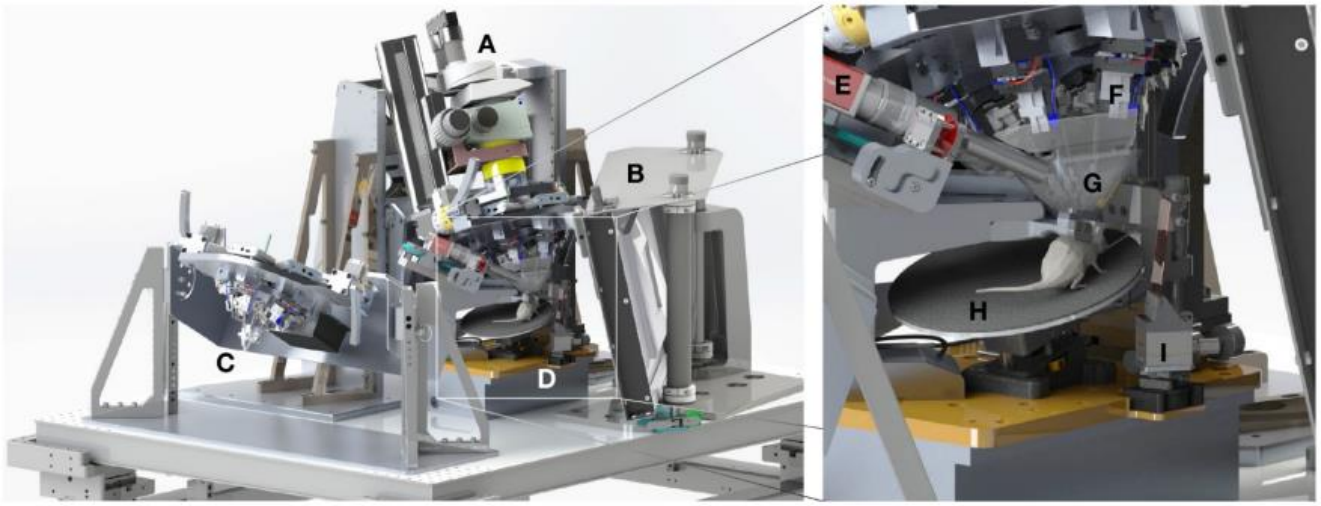


Figure 1. Visual Coding Experimental Rig. Following components were preserved as shown: (A), Stereomicroscope used to visualize the insertion site. The eyepieces have been replaced with cameras, to enable remote viewing while the rig is enclosed. (B), Visual stimulus monitor. (C), Maintenance stand used to repair and replace probes, headstages, manipulators, and ground wires. (D), Behavior stage. (E), Eyetracking camera. (F), Manipulators used to align probes with visual areas. (G), Protective cone, to prevent the mouse's tail from contacting the probes. (H), Running wheel. (I), IR dichroic, to reflect the mouse's eye without obscuring the view of the stimulus monitor.

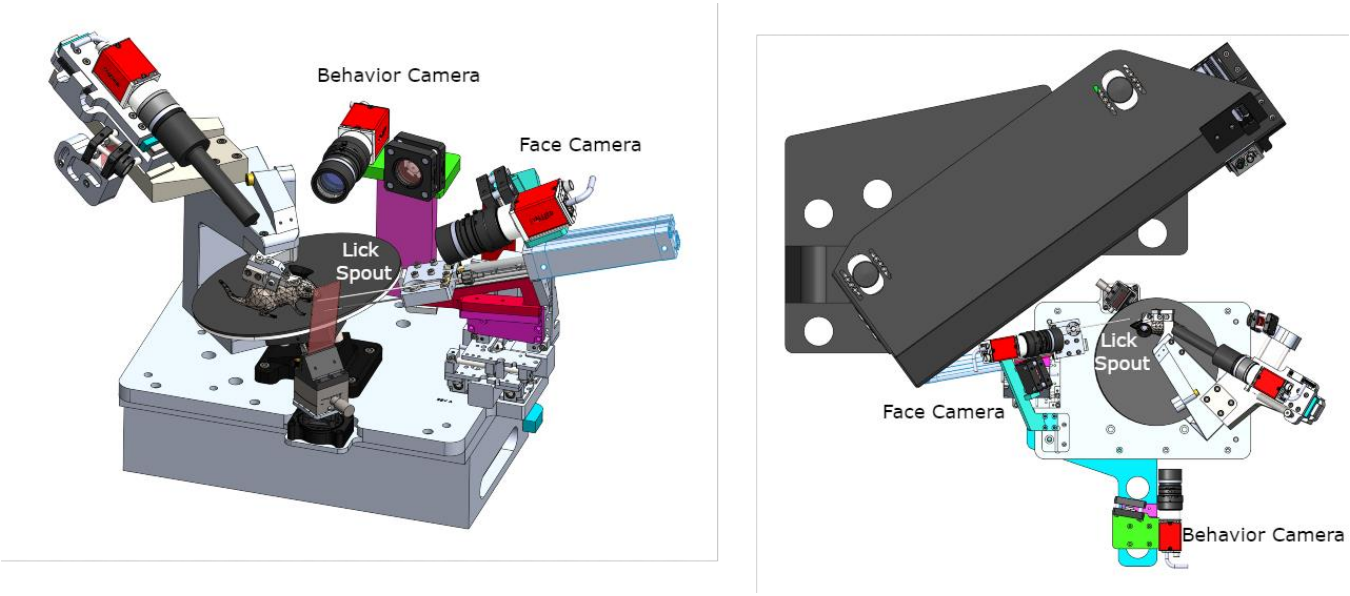


Figure 2. Rig upgrades for Visual Behavior Neuropixels: Left: Side view (from direction of stimulus monitor) showing position of the behavior (or body) camera, face camera and lick spout. Right: Top down view of same components.

Lick spout design

The capacitive lick sensor used during behavior training and the Visual Behavior Optical Physiology imaging sessions was found to be incompatible with Neuropixels recordings, as it caused large electrical artifacts when the tongue contacted the spout. To address this issue, a piezo lick sensor was designed for the Neuropixels rigs to detect licks by the motion imparted to the spout by the tongue. As diagrammed in **Figure 3,** the lick spout rests on top of a highly sensitive piezo microphone. The analog voltage from this microphone is then processed by a SAMD21 microcontroller to produce a TTL lick signal, which is used to determine lick times.

We found that even with this system, lick artifacts contaminated electrophysiological recordings when the spout was made from a conductive material. To eliminate these artifacts, we used a non-conductive PEEK needle for the lick spout. To add rigidity, this needle was pushed through a larger diameter, stainless steel sleeve.

Part Number	Vendor	Description
0195-110-01	Allen Institute	Stainless Steel Needle Reinforcement 90mm Long Make from McMaster 5560K36
8649-01-CUSTOM NEEDLE	Hamilton	PEEK 0.062 inch OD/ 0.030 inch ID, Kel-F Hub Needle, 2 in, point style 3 Make like 8649-01 w/ 110.5mm REFER TO HAMILTON QUOTE HKB101520-1
1007079-1	TE Connectivity	Contact Microphone
ADN-12-74-I-P-A	Festo	Compact Air Cylinder

Table 1. Parts list for piezo lick sensor.

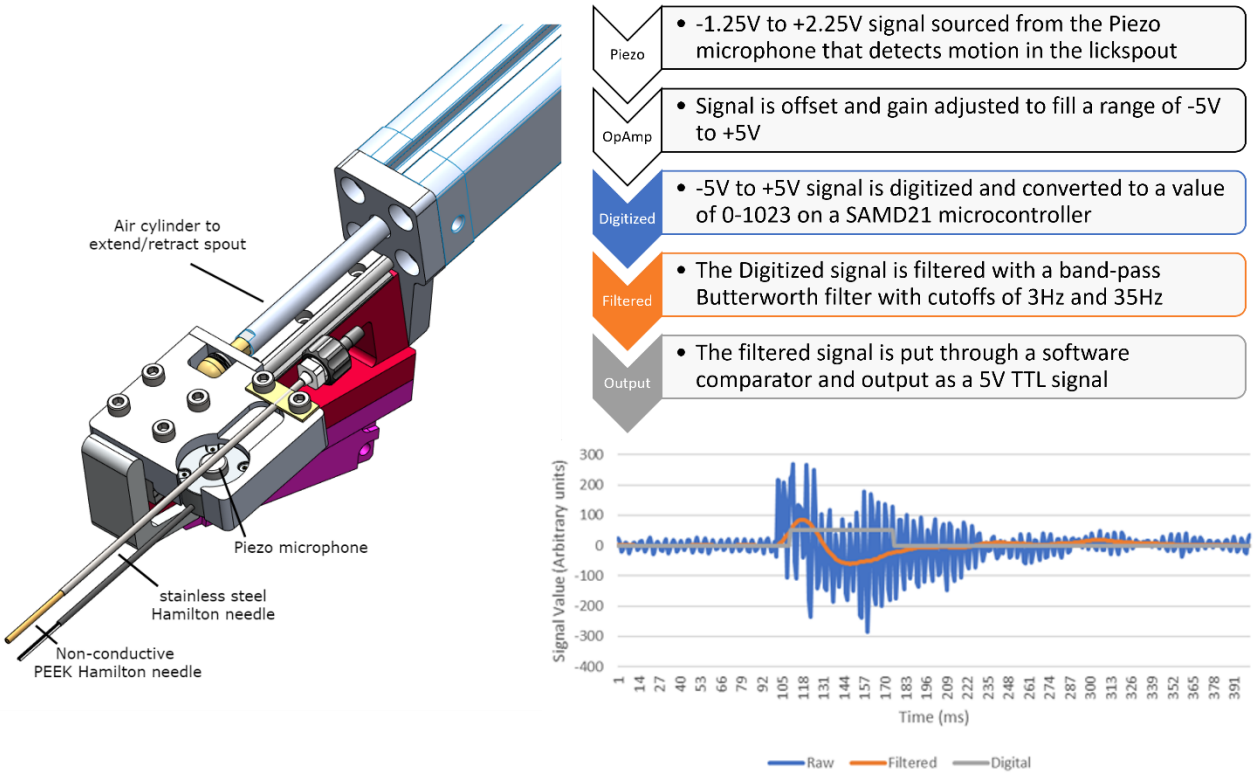


Figure 3. Piezo lick sensor design and signal processing: Left: Model of the piezo lick sensor assembly. Right: Signal processing steps to convert the raw piezo voltage into a digital lick signal. Bottom right plot shows the processing stages after manually tapping the spout: digitized raw signal (blue), the signal after applying a bandpass filter (3-35 Hz) (orange) and the final output of the comparator (gray), which was used to determine lick times.

Data acquisition and synchronization

Neuropixels data was acquired at 30 kHz (spike band) and 2.5 kHz (LFP band) using the Open Ephys GUI (Siegle et al., 2017). Gain settings of 500x and 250x were used for the spike band and LFP bands, respectively. Each probe was connected to a PXIe card inside a National Instruments chassis. Raw neural data was streamed to a compressed format for archiving which was extracted prior to analysis.

Videos of the eye, face and body were acquired at 60 Hz. The angular velocity of the running wheel was recorded at the time of each stimulus frame, at approximately 60 Hz. Synchronization signals for each frame were acquired by a dedicated computer with a National Instruments card acquiring digital inputs at 100 kHz, which was considered the master clock. A 32-bit digital "barcode" was sent with an Arduino Uno (SparkFun DEV-11021) every 30 s to synchronize all devices with the neural data. Each Neuropixels probe has an independent sample rate between 29,999.90 Hz and 30,000.31 Hz, making it necessary to align the samples offline to achieve precise synchronization. The synchronization procedure used the first matching barcode between each probe and the master clock to determine the clock offset, and the last matching barcode to determine the clock scaling factor. If probe data acquisition was interrupted at any point during the experiment, the experiment was failed. Because one LFP band sample was always acquired after every 12th spike band sample, these data streams could be synchronized automatically once the spike band clock rate has been determined.

To synchronize the visual stimulus to the master clock, a silicon photodiode (PDA36A, Thorlabs) was placed on the stimulus monitor above a "sync square" that flips from black to white every 60 frames. The analog photodiode signal is thresholded and recorded as a digital event by the sync computer. Individual frame times are reconstructed by interpolating between the photodiode on/off events.

Stimulus monitor

Visual stimuli were generated using custom scripts based on PsychoPy (Peirce, 2007) and were displayed using an ASUS PA248Q LCD monitor, with 1920 x 1200 pixels (21.93 in wide, 60 Hz refresh rate). Stimuli were presented monocularly, and the monitor was positioned 15 cm from the mouse's right eye and spanned 120° x 95° of visual space prior to stimulus warping. Each monitor was gamma corrected and had a mean luminance of 50 cd/m^2. To account for the close viewing angle of the mouse, a spherical warping was applied to all stimuli to ensure that the apparent size, speed, and spatial frequency were constant across the monitor as seen from the mouse's perspective (**Figure 4**).

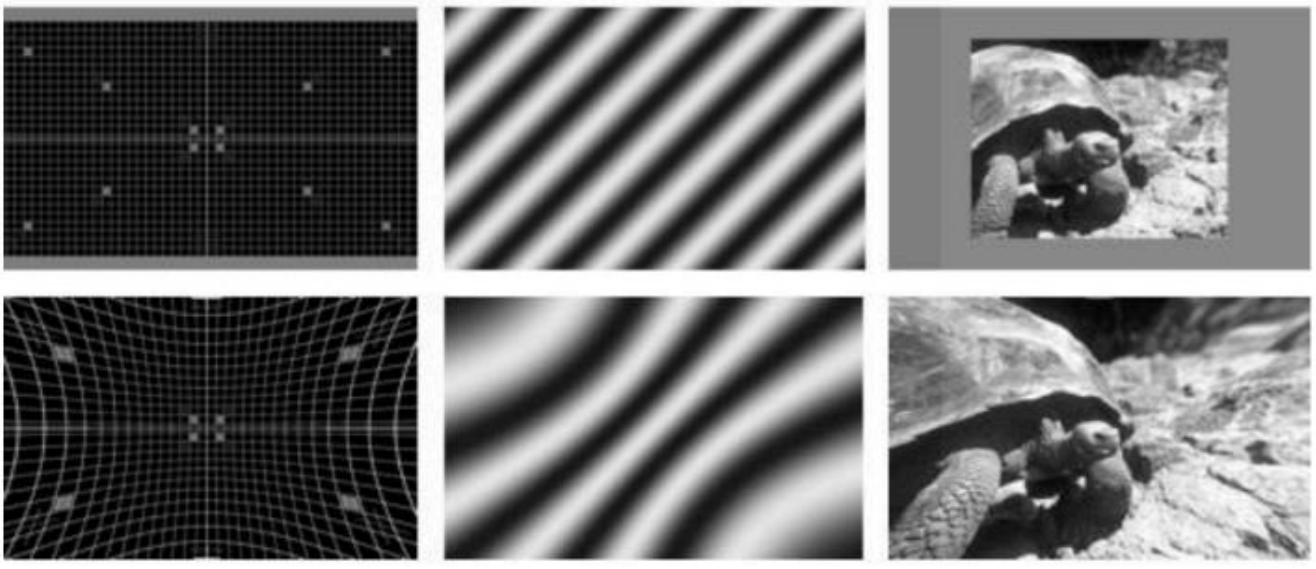


Figure 4. Demonstration of stimulus warping. Top, unwarped images, bottom, warped images. The center of the screen (origin of the grid in the left images), was located at 10 degrees elevation and 50 degrees azimuth.

Visual Stimuli

Every experiment comprised the same stimulus blocks in the same order, as diagrammed in **Figure 5**.

Active behavior block: the mouse performed the detection of change task described above. The images used during this block came from one of two image sets of 8 natural images (G or H, see **Section A Figure 4**). On one recording day, mice were exposed to the same image set they had seen during behavior training ('Familiar' sessions). On the other recording day, mice were shown a novel image set ('Novel' sessions). Note however, that two images were shared across these image sets and thus familiar on both sessions. Images were presented for 250 ms, with a 500 ms gray screen between images. During recording sessions, images were omitted (resulting in a gray screen) with a probability of 5%. However the change image, the image immediately preceding the change and the image immediately following an omission were never omitted, producing an actual omission probability of ~3%. At the end of this block, the lick spout was retracted for the remainder of the session.

Gabor stimuli: Gabor stimuli windowed to have a 20 degree diameter were presented in a 9x9 grid to map receptive fields. Grid positions were spaced by 10 degrees. Gabors had a temporal frequency of 4 Hz, a spatial frequency of 0.08 cycles/degree and were shown at 3 orientations (0, 45, 90 degrees). They were 250 ms in duration without gaps between stimuli. There were 15 trials for each condition (81 positions, 3 orientations).

Spontaneous: Following the gabor stimuli there was a brief gray screen (5 minutes)

Full-field flashes: Flashes were black or white at 80% contrast. They were 250 ms in duration with 2 seconds between flash starts. There were 75 trials for each condition (light and dark flashes).

Passive replay: The stimulus encountered by the mouse during the active behavior block was re-shown frame-for-frame. Note however that the lick spout was retracted for this block and the mouse was therefore unable to earn rewards.

Optotagging: The optotagging protocol was the final experiment block and is described in more detail below.

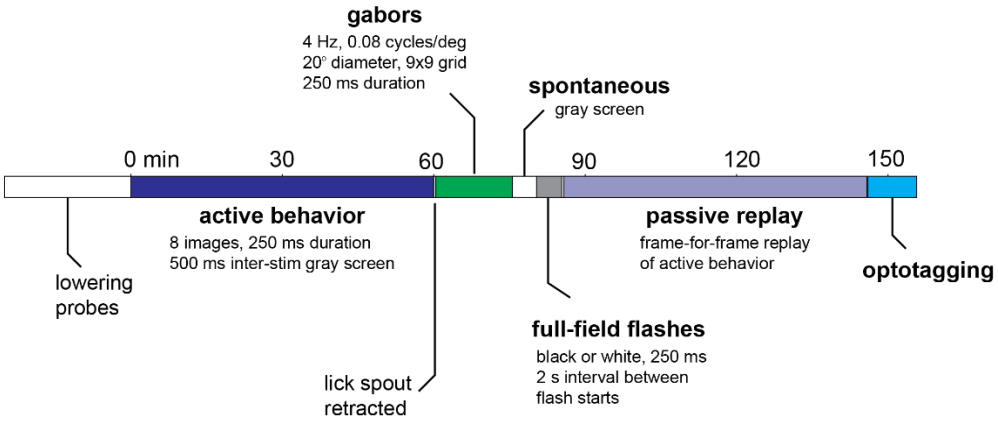


Figure 5. Stimulus set. Colors indicate the identity of each stimulus block. The session began with an active behavior block during which the mouse performed the detection of change task described above. After this block (~1 hour), the lick spout was retracted for the remainder of the session.

Optotagging

At the end of every experiment (regardless of genotype), an optotagging protocol was run during which the cortical surface was stimulated with blue light. In Sst and Vip mice, this protocol allowed us to identify putative Sst+ and Vip+ cortical interneurons by their dramatic increase in spiking activity timelocked to laser stimulation (and consequent ChR2 activation). We chose to focus on these two cell classes because they make up two of the three major non-overlapping populations of cortical interneurons (Rudy et al, 2011). The third major class (parvalbumin-expressing cells), can largely be distinguished by their narrow spike waveforms.

Blue light was delivered by a 473 nm laser (Laser Quantum, model ciel or Cobolt model 06-MLD). The light source was coupled to a 400 µm diameter fiber optic cable (Thorlabs) or bifurcated fiber bundle (Thorlabs, BFYL4LF01), with the tip(s) positioned such that blue light illuminated the entire cranial window as evenly as possible. Two types of stimuli at 3 different light levels were randomly interleaved: a 10 ms pulse, and a 1s raised cosine ramp. For the pulse stimulus, a 0.5 ms ramp was applied at the beginning and end of the pulse (**Figure 5**). Stimuli were presented at intervals of 1.5 s plus a uniformly distributed delay between 0 and 0.5 s. Mice had no exposure to the optotagging stimulus prior to the end of the first recording session.

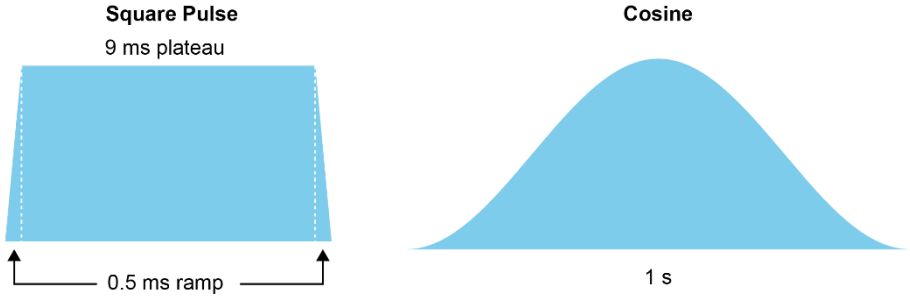


Figure 5. Optotagging stimuli. Laser command waveforms used for this dataset. Left: 10 ms pulse with ramped onset and offset. Right: 1 second raised cosine.

EXPERIMENTAL WORKFLOW

The experimental workflow for these recordings, including glass coverslip removal and plastic window insertion, head fixation, grounding, probe alignment, DiI application, and probe insertion, were identical to that described in the Visual Coding Neuropixels white paper: https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf. These procedures were extended for this dataset to allow for recordings on two consecutive days as described below.

Recording across two consecutive days

On the morning of the first recording day, the glass window covering the brain is removed and the plastic window implantation is performed as described in the Visual Coding whitepaper. After a minimum of two hours of recovery (typically 3-4 hours), the mouse is put on the rig and probes are inserted. After the recording, the probes are retracted and the well is sealed with agar and a cap. On the next day, the agar on top of the window is cleared, a fresh layer of silicone oil is applied and probes are re-inserted (**Figure 6**). To guide day 2 insertions, the receptive fields from day 1 were generated and an effort was made to maximize receptive field overlap with the center of the screen (**Figure 7**).

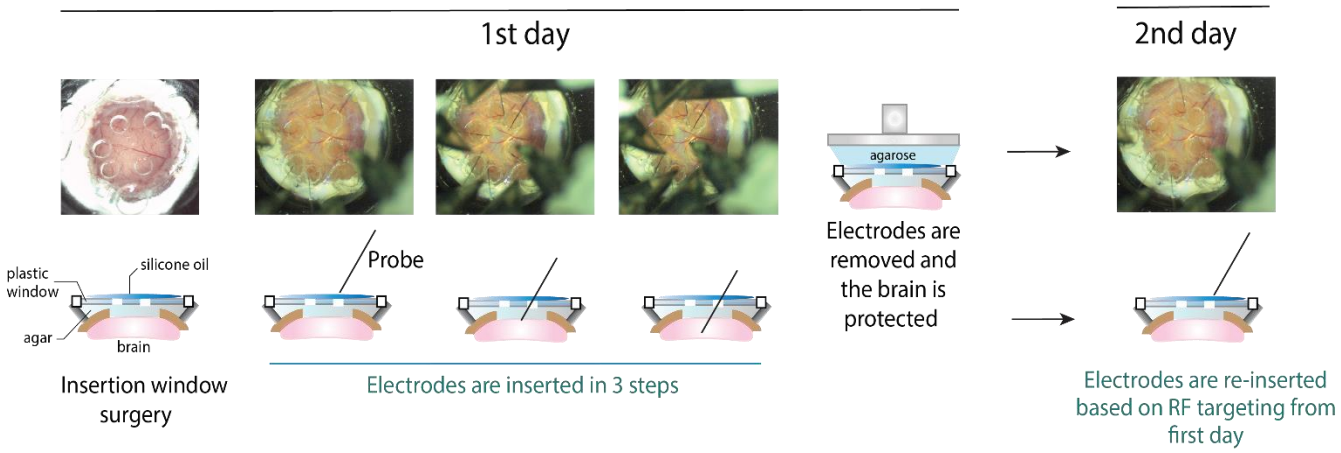


Figure 6. Recording on two consecutive days. Workflow for two day recordings. Time flows from left to right.

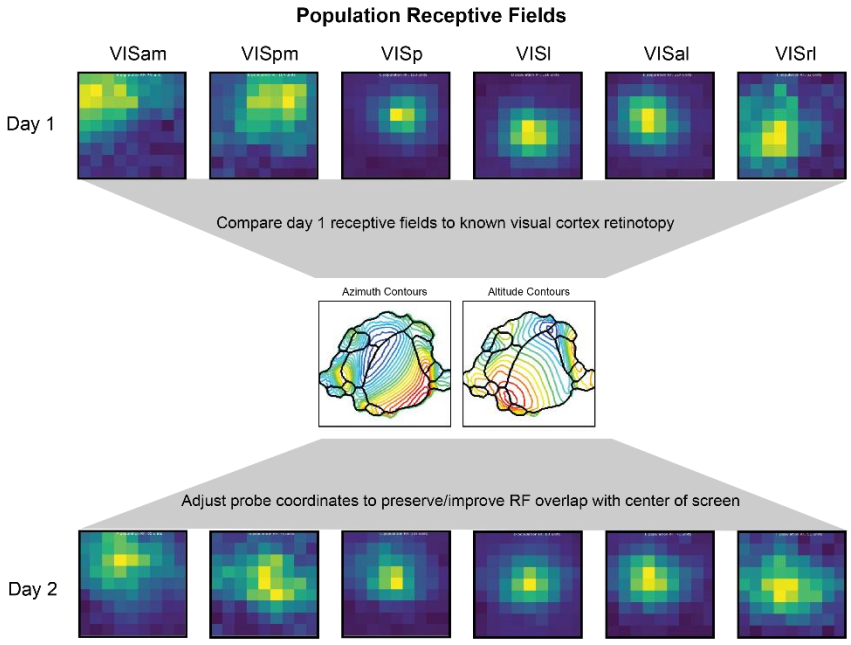
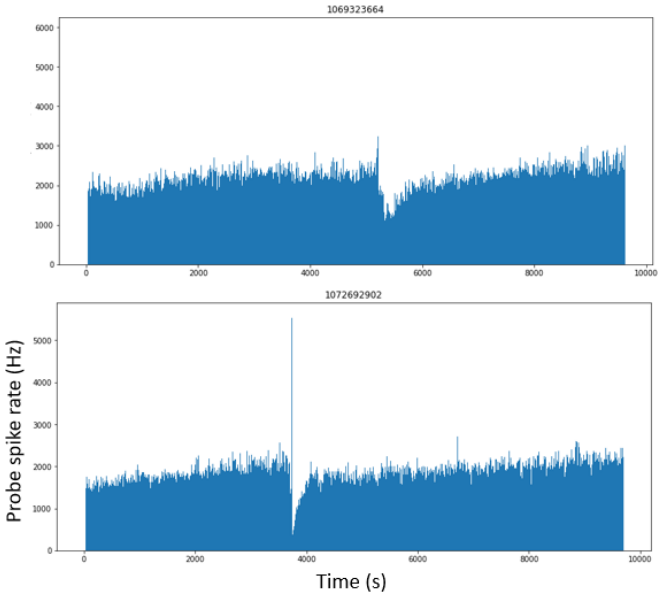


Figure 7. Probe targeting or day 2 recordings. Population receptive fields were generated for cortical units for day 1 before probe insertion on day 2. Receptive fields were compared to the known retinotopy in mouse visual cortex (Garrett et al 2014) and the probe targets were manually adjusted for day 2 to preserve or improve receptive field overlap with the center of the screen.

Probe Removal and Cleaning

Probes were retracted from the brain at a rate of 1 mm/s, after which the probe cartridge was lifted up to its full height. The 3D-printed protective cap was screwed into the headframe well, then mice were removed from head fixation and returned to their home cages overnight. Probes were immersed in a well of 1% Tergazyme overnight, which was sufficient to remove tissue and silicone oil prior to the next day's recording session.



Abnormal Activity

We flagged 12 sessions as having potential epileptiform activity. Flagged sessions exhibited a brief increase in spiking activity (lasting ~1-5 seconds) followed by a sustained suppression of activity (lasting for several minutes) for at least one probe during the session and were defined by the following procedure: 1) the spike rate across the entire probe was calculated in 1 second bins, 2) this vector was detrended by subtracting a median filtered copy (kernel of 1001 seconds), 3) candidate events were selected as bins exceeding 5 standard deviations above the mean, 4) these events were classified as abnormal activity if they were followed by a period of suppressed mean activity in a window 5 to 50 seconds after the event (2 standard deviations below the mean over the pre-event baseline). The events captured by this procedure were consistent with manual identification of seizure-like events.

Figure 8. Abnormal activity examples. Probe spike rate plots for two example sessions classified as having potential epileptiform activity. Note the brief increase in probe firing rate followed by sustained suppression.

Neuropixels Recording Quality Control

Out of a total of 180 sessions, 27 were excluded from the release due to the following QC failures: white foam buildup on the edge of the eye partially covering the pupil (1), software failures (2), photodiode failure (2), cortical bleeding or compromised brain health resulting in low unit activity and/or atypical visual responses (18), gap in data acquisition (2), discovery of purulent material over right hemisphere during ex-vivo imaging (2).

SECTION F: EX-VIVO IMAGING AND PROBE REGISTRATION TO CCF

Tissue clearing and **OPT imaging** were performed as described in the Visual Coding Neuropixels documentation (https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf).

Registering Probes to the Common Coordinate Framework

Reconstructed brains were downsampled to 10 µm per voxel and roughly aligned to the Allen Institute Common Coordinate Framework (CCFv3) template brain using an affine transform. The volume was then cropped to a size of 1023 x 1024 x 1024 and converted to Drishti format (http://sf.anu.edu.au/Vizlab/drishti). Next, 6-54 registration points were marked in up to 14 coronal slices of the individual brain by comparing to the CCFv3 template brain (Figure 20A). Fluorescent probe tracks were manually labeled in coronal slices of the individual brain, and the best-fit line was found using singular value decomposition (Figure 20B). The registration points were used to define a 3D nonlinear transform (VTK thinPlateSplineTransform), which was used to translate each point along the probe track into the CCFv3 coordinate space. Each CCFv3 coordinate corresponds to a unique brain region, identified by its structure acronym (e.g., CA3, LP, VISp, etc.). A list of CCFv3 structure acronyms along each track was compared to the physiological features measured by each probe (e.g., unit density, LFP theta power and visual responsiveness). The location of major structural boundaries were manually adjusted to align the CCFv3 labels with the physiology data, and each recording channel was assigned to a CCFv3 structure. In this way, each channel could be mapped to a unique brain region.

Distinguishing probe tracks from day 1 and day 2

Recording for two consecutive days posed an additional challenge in probe registration beyond the protocol established for the Visual Coding Neuropixels dataset. For each brain, up to 12 probe tracks were annotated (2 for each probe) and assigned to either the day 1 or day 2 recording. To make these assignments, we relied on two sources of information about the relative probe position on the two recording days:

1) Operator annotated where the probes were inserted relative to surface vasculature landmarks (red and blue cross hairs in **Figure 8)**.
2) A record was kept of the movements for each probe's manipulator, allowing us to calculate a predicted displacement vector from day 1 to day 2 (white arrows in **Figure 8**).

Operators manually assigned the probe tracks for each probe based on their relative position in the OPT volume and the expected displacement from (1) and (2). In the event that the probe tracks were too close to be disambiguated, they were annotated as one track.

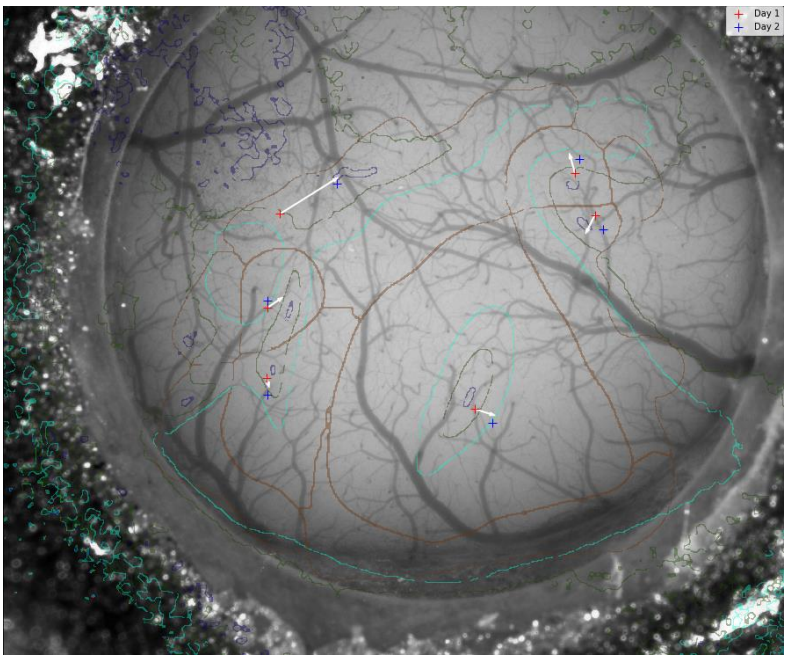
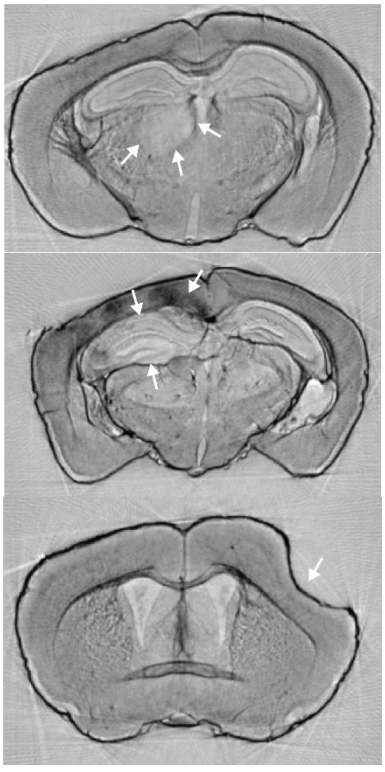


Figure 8. Calculating relative probe positions across day 1 and day 2. To assign probe tracks to day 1 or day 2 recordings, the expected displacement was calculated from a) operator annotations of the probe tip at insertion (red cross hair for day 1, blue cross hair for day 2) and b) the manipulator motor coordinates saved at insertion (relative displacement vectors from day 1 to day 2 are shown as white arrows, with the base positioned at the day 1 insertion annotation).

Identification of cortical visual area targets

To confirm the identity of the cortical visual areas, images of the probes taken during the experiment were compared to images of the brain surface vasculature taken during the ISI imaging session. Based on the vasculature patterns, the ISI map was overlapped with the insertion photo-documentation image. Key points were selected along the vasculature on both images and a perspective transform (OpenCV) was performed to warp the insertion image to the ISI target image. The probe entry points were manually annotated. Finally, an area was assigned to each probe. A small subset of penetrations were mapped to LI, MMA, or MMP. Penetration targets that could not be unambiguously associated with a particular visual area were classified as "VIS." If the cortical area label obtained via CCFv3 registration did not match the area identified via the ISI map overlay, the information in the ISI map took precedence.



Abnormal Histology

OPT imaging revealed potential tissue damage in 27 mice. This damage was characterized by one of the following abnormalities: 1) tissue discoloration indicating potential bleeding or hypoxic tissue (Figure 9 top), 2) swelling of the hippocampus (Figure 9 middle, left arrows) or 3) tissue bruising (Figure 9 middle, top arrow) or 4) skull deformity compressing the right cortex (Figure 9 bottom).

Figure 9. Abnormal histology. Three examples of mice flagged as having abnormal histology.

SECTION G: DATA PROCESSING

Electrophysiological and video data processing (including pre-processing, spike sorting and ellipse fits to the eye videos) were performed as described in the Visual Coding Neuropixels whitepaper (https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf). Please note however, that unlike the data for the Visual Coding Neuropixels pipeline, for which potential noise units were filtered from the released dataset, we have elected to return all units for the Visual Behavior Neuropixels dataset. We have flagged units that our algorithm identified as potential 'noise' units in the 'quality' column of the units attribute of the ecephys session object returned by the SDK. Please see our Quality Metrics Tutorial for more details: https://allensdk.readthedocs.io/en/latest/_static/examples/nb/visual_behavior_neuropixels_quality_metrics.html.

REFERENCES

Garrett ME, Nauhaus I, Marshel JH, Callaway EM. 2014. Topography and areal organization of mouse visual cortex. *J Neurosci*. doi:10.1523/JNEUROSCI.1124-14.2014

Groblewski PA, Ollerenshaw DR, Kiggins JT, Garrett ME, Mochizuki C, Casal L, Cross S, Mace K, Swapp J, Manavi S, Williams D, Mihalas S, Olsen SR. 2020. Characterization of Learning, Motivation, and Visual Perception in Five Transgenic Mouse Lines Expressing GCaMP in Distinct Cell Populations. *Front Behav Neurosci*. doi:10.3389/fnbeh.2020.00104

Macmillan NA, Creelman CD. 2004. Detection Theory: A User's Guide: 2nd edition, Detection Theory: A User's Guide: 2nd edition. doi:10.4324/9781410611147

Madisen L, Garner AR, Shimaoka D, Chuong AS, Klapoetke NC, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C, Gu H, Mills M, Cheng A, Tasic B, Nguyen TN, Sunkin SM, Benucci A, Nagy A, Miyawaki A, Helmchen F, Empson RM, Knöpfel T, Boyden ES, Reid RC, Carandini M, Zeng H. 2015. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron*. doi:10.1016/j.neuron.2015.02.022

Madisen, L., Mao, T., Koch, H., Zhuo, J., Berenyi, A., Fujisawa, S., Hsu, Y.-W.A., Garcia, A.J., Gu, X., Zanella, S., et al. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. Nature Neuroscience 15, 793–802.

Peirce, J.W. (2007). PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods 6.

Rudy B, Fishell G, Lee S, Hjerling-Leffler J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. Dev Neurobiol. 2011 Jan 1;71(1):45-61

Siegle, J.H., López, A.C., Patel, Y.A., Abramov, K., Ohayon, S., and Voigts, J. (2017). Open Ephys: an opensource, plugin-based platform for multichannel electrophysiology. Journal of Neural Engineering 14, 045003.

Teeters, J.L., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., et al. (2015). Neurodata Without Borders: creating a common data format for neurophysiology. Neuron 88, 629–634.

Zhang, F., Wang, L.-P., Boyden, E.S., and Deisseroth, K. (2006). Channelrhodopsin-2 and optical control of excitable cells. Nat Methods 3, 785–792.