dotCategoryLearn data documentation

This contains documentation specific for the **dotCategoryLearn** dataset – Evan Antzoulatos’s dot-category learning task. Subjects were Ani and Paloma. For more information on this dataset, see [Antzoulatos & Miller 2011 Neuron](about:blank) or [Antzoulatos & Miller 2014 Neuron](about:blank).

For general info about the Miller Lab data format, see MillerLabDataFormatSummary.docx

## Unique features

* Online stimulus-response and **category learning**
* **Simultaneous PFC & striatum** (head/body of caudate) recording during learning

# Experimental design

## Task design

Diagram

Description automatically generatedEach session, monkeys learned to categorize two categories of dot-pattern stimuli, each of which were defined by random distortions around a prototype. Each trial (Fig 1A), an exemplar from one category or the other was shown centrally for 600 ms, followed by a 1 s blank delay, then the response period. Each category was arbitrarily associated with a saccadic response to a target on the left or right, which was learned each session through trial and error. Correct trials were reinforced with drops of juice, while choice errors were corrected by redisplaying the sample exemplar over the location of the correct (unchosen) saccade target. The category prototypes were quasi-random “constellations” of 7 dots (Fig 1B).

The session was split into blocks of trials, with exponentially increasing numbers of category exemplars in successive blocks. The first block featured only a single exemplar of each category, and thus could be considered essentially a stimulus-response learning task. Each subsequent block doubled the number of exemplars (2,4,8,16,…), accumulatively keeping all exemplars from the immediately preceding block and adding a set of new, never-before-seen exemplars. Subjects moved from one block to the next when a learning criterion was achieved (16 correct trials over the previous 20 trials), up to a maximum of 8 blocks. This increase in novel exemplars gradually forced subjects to generalize to learn the underlying category structure (Fig 1C). In early blocks (“SR Learning”), performance increased through the block trials as the exemplars were learned. In later blocks (“Category Learning/Performance”), the category structure was learned and near-asymptotic performance was achieved even on the first trial with novel exemplars.

### Stimulus details

The category exemplars were defined within a 6 × 6 dva window around fixation, and consisted of seven 0.4 dva diameter dots. Dot locations for category prototypes and exemplar were chosen using a pseudo-random procedure described in the Supplements of Evan’s 2011 and 2014 papers. The fixation point was a red dot 0.4 dva in diameters. Saccade targets were green dots 0.6 dva in diameter, displayed at ±5 dva to the left and right of fixation. Stimuli were displayed on a CRT with a 100 Hz refresh.

## Chart, scatter chart Description automatically generatedRecorded areas

Recordings were in lateral PFC (dorsolateral and ventrolateral) and striatum (head/body of caudate nucleus) in the same hemisphere (left for Paloma, right for Ani; Fig 2). However, we currently only have ‘PFC’ vs ‘striatum’ area labels for this dataset; finer areal distinctions are not available.

## Recording setup

Most data was recorded with FHC single ‘G’ tungsten/parylene electrodes. Electrodes were acutely inserted and removed each day, so a different set of unit/LFP sites was recorded each session.

The data was recorded with the Plexon MAP system (through Plexon high-impedance headstages), and the Windows CORTEX behavioral system. LFPs were recorded at 1 kHz, spike thresholds were manually set online, and spikes manually sorted offline (Plexon Offline Sorter). Eye position was monitored at 500 Hz with an infrared-based tracker (Eyelink 1000, SR Research).

# Data organization

All data was cut into trials epochs from –1.5 to +3 s relative to the onset of the exemplar stimulus (‘sampleOn’) on each trial, and all data times are expressed relative to sampleOn on each trial. This epoch includes all trial events, from fixation through the saccadic response. For the 'withITI’ version of the dataset, the trial epoch was cut into trial epochs of –2 to +8 s, in order to include the post-trial inter-trial interval for the longest trials (this means that the cut data for shorter trials may include the start of the following trial).

## Preprocessing

In addition to the standard preprocessing (detailed in MillerLabDataFormatSummary.docx), two additional processing steps were run on this dataset:

1. An interpolation algorithm to try to fix clipping of LFP signals due to the limited dynamic range of the recording system (see *Caveats*). Sections of each raw LFP that were at the digital range of the system (± 2048) were identified and interpolated over using cubic spline interpolation (Matlab interp1() function with method=‘spline’, and extrapolation=‘extrap’).
2. In addition to removing line noise at 60 Hz + harmonics, we also removed line noise at 100.12 Hz \* [1:5] (found empirically in the data, presumably this was the refresh rate of the subject’s CRT).

## trialInfo fields

The trialInfo table-like struct contains per-trial task/behavioral data for each trial (in each table row), in each session. Some of the key columns/fields are:

|  |  |
| --- | --- |
| block | Trial block number (from 1–8) |
| learningStage | Stage of learning of each trial block, using criteria from papers: ‘SR Learning’ or ‘Category Learning’ or ‘Category Performance’ |
| category | Category of sample stimulus: 1 or 2 (associate with left or right response, respectively) |
| sample | Index of sample item (from 1–size of session sample pool). Internally consistent between trials, but otherwise arbitrary. Note: not all samples actually used, due to reaching block learning criterion quickly. |
| repetitionNum | Number of times each sample has appeared (1=first appearance of given sample [novel], 2=second appearance, …) |
| correctDir | Appropriate saccadic response direction (‘L’ or ‘R’) for trial category. 1:1 mapping with category. |
| responseDir | Actual saccadic response direction (‘L’ or ‘R’) |
| outcome | Trial outcome: ‘correct’ (choosing direction associated with sample category) or ‘choiceErr’ (choosing wrong direction) |
| reactionTime | Time (s) from target array onset (delayEnd) to first saccade away from fixation (responseTime) |
| numRewards | Number of juice drops delivered as post-trial reinforcement for correct response. In some sessions, this was increased for some trials later in the session (to maintain motivation?). |
| rewardOn | Onset time of each reinforcement juice drop |
| correctionOnset/Offset | Onset/offset of “correction stimulus” after choice error—exemplar redisplayed for 5 s over location of correct saccade target |
| feedbackOn | Onset of feedback to subject indicating whether choice was correct or incorrect (juice or correction stimulus, respectively) |
| badTimingTrials | Boolean labeling trials where event timing (sampleOff or delayEnd) is > 5 ms from expected value (this didn’t seem to happen at all) |
| badLFPTrials | Boolean labeling trials where LFP data had > 5% of datapoints clipped by the recording system’s dynamic range |
| badTrials | Trials with either bad timing OR LFPs (badTimingTrials | badLFPTrials) |

## *Caveats*

Relatively small dataset: 19 sessions, ~8-16 electrodes per session

Due to the limited dynamic range of the Plexon MAP system, LFP signals on some higher-amplitude channels were occasionally clipped at some timepoints. We’ve tried to recover the true unclipped signals using an interpolation method (see *Preprocessing*), but this is imperfect especially for electrodes with extensive periods of clipped data. Which datapoints were clipped across all (timepoints,channels,trials) is given in lfpSchema.mask. The proportion of all datapoints (across timepoints and trials) that were clipped for each electrode is given in electrodeInfo.LFPClipFracts and those with > 1% are flagged in electrodeInfo.badElectrodes. The proportion of all datapoints (across timepoints and electrodes) that were clipped on each trial are given in trialInfo.LFPClipFracts and trials with > 5% are flagged in trialInfo.badLFPTrials. *I recommend you not include the “bad” electrodes and trials in your LFP analyses.*

The fact that different exemplars are used across blocks and that some exemplars are used for only 1 trial (before learning criterion is reached) don’t match the assumptions of traditional analyses of stimulus-related neural information. This type of analysis would likely have to be run separately in each trial block and is likely not possible for later blocks.

In some sessions, the number of reward juice drops was increased on a (random?) subset of trials later in the session (during block 8), likely to maintain motivation and behavioral performance. But this might confound some analyses looking at reward or reinforcement learning.

For many sessions, category learning criterion was achieved in block 3, so learning stages (as defined in publications) jump from ‘SR Learning’ directly to ‘Category Performance’, without any intervening ‘Category Learning’ blocks.

## Data organization on server

|  |  |
| --- | --- |
| /data/common/datasets/dotCategoryLearn/mat/trialOnly | Data in lab standard format (trial events only) |
| /data/common/datasets/dotCategoryLearn/mat/withITI | **Data in lab standard format (includes post-trial ITI)** |
| /data/common/datasets/dotCategoryLearn/metadata | Additional behavioral & recording metadata |
| millerarchive.mit.edu:  /archive/data/common/datasets/dotCategoryLearn/plx/sorted | Spike-sorted Plexon datafiles |
| millerarchive.mit.edu:  /archive/data/common/datasets/dotCategoryLearn/plx/raw | Raw Plexon datafiles |
| millerarchive.mit.edu:  /archive/data/common/datasets/dotCategoryLearn/cortex | CORTEX task/behavioral datafiles |