MATLAB Data files and analysis scripts for:

Triple dissociation of attention and decision computations across prefrontal cortex

Laurence T. Hunt*, W.M. Nishantha Malalasekera*, Archy O. de Berker, Bruno Miranda, Simon F. Farmer, Timothy E.J. Behrens and Steven W. Kennerley, *Nature Neuroscience* 2018

Conditions for using this dataset:

These data required years of effort to collect. It is our lab policy that any preprint or peer-reviewed publication using these data lists as authors the people who collected the data (Laurence Hunt and Nishantha Malalasekera), and the group leader under whose responsibility the data has been collected (Steve Kennerley). In addition, if you publish any work using the data, please cite the publication given above (Hunt/Malalasekera et al., 2018) and the CRCNS DOI. Thank you.

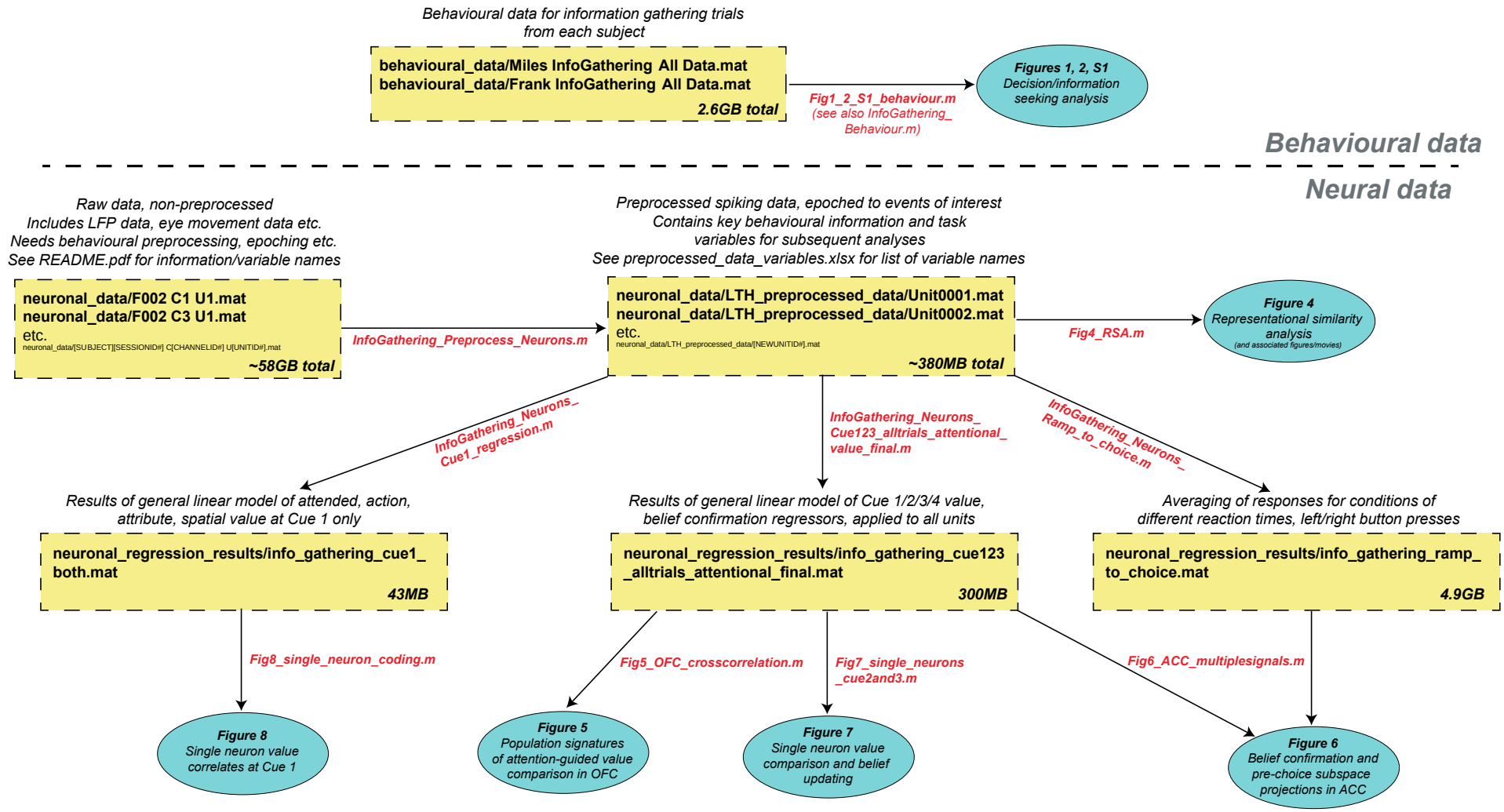
What would you like to do?

1. I want to reproduce the analyses from your paper.
   Before you begin, open the function 'get_basedir.m', and change the variable *bd* so that it points to top-level directory (i.e. the directory which contains 'get_basedir.m'). Once you've done that, you should be able to run all the different scripts that start with the word 'Fig' (i.e. 'Fig1_2_S1_behaviour', 'Fig4_RSA', etc.). This will reproduce the figures and movies from the paper.

2. I want to play with the raw data.
   Before you begin, open the function **get_basedir**, and change the variable *bd* so that it points to top-level directory (i.e. the directory which contains **get_basedir**). Once you've done that, have a look through the detailed documentation (starting on the next page) which tells you how the behavioural and neuronal data is structured.

   Probably the best place to start is to load in some of the 'preprocessed' data files, from 'neuronal_data/LTH_preprocessed_data'), which has already been epoched (e.g. 'Cue1_MatrixRaw' contains rasters timelocked to Cue1 onset) and useful variables precomputed. The way to understand what these variables mean is to look in **InfoGathering_Preprocess_Neurons**, which *generates* all the preprocesed variables (and is well commented), and look at the Excel document preprocessed_data_variables.xlsx that lists all of the variables that are loaded in (note there is occasionally redundancy in some of the variables). Also look at the documentation below.

   Note, however, that for some analyses (e.g. those based on LFP data), you'll need to load in the *non-preprocessed* data (which is stored in 'neuronal_data'). Again, have a look inside **InfoGathering_Preprocess_Neurons** for how to load in this data, and get started with it, and also look at the documentation below.


Behavioural data for information gathering trials from each subject
behavioural_data/Miles InfoGathering All Data.mat
behavioural_data/Frank InfoGathering All Data.mat
2.6GB total
Fig1_2_S1_behaviour.m
(see also InfoGathering_ Behaviour.m)
Figures 1, 2, S1
Decision/information seeking analysis
Behavioural data
Neural data
Raw data, non-preprocessed
Includes LFP data, eye movement data etc.
Needs behavioural preprocessing, epoching etc.
See README.pdf for information/variable names
neuronal_data/F002 C1 U1.mat
neuronal_data/F002 C3 U1.mat
etc.
neuronal_data/[SUBJECT][SESSIONID#] C[CHANNELID#] U[UNITID#].mat
~58GB total
InfoGathering_Preprocess_Neurons.m
Preprocessed spiking data, epoched to events of interest
Contains key behavioural information and task variables for subsequent analyses
See preprocessed_data_variables.xlsx for list of variable names
neuronal_data/LTH_preprocessed_data/Unit0001.mat
neuronal_data/LTH_preprocessed_data/Unit0002.mat
etc.
neuronal_data/LTH_preprocessed_data/[NEWUNITID#].mat
~380MB total
Fig4_RSA.m
Figure 4
Representational similarity analysis
(and associated figures/movies)
InfoGathering_Neurons_ Cue1_regression.m
InfoGathering_Neurons_ Cue123_alltrials_attentional_ value_final.m
InfoGathering_Neurons_ Ramp_to_choice.m
Results of general linear model of attended, action, attribute, spatial value at Cue 1 only
neuronal_regression_results/info_gathering_cue1_ both.mat
43MB
Results of general linear model of Cue 1/2/3/4 value, belief confirmation regressors, applied to all units
neuronal_regression_results/info_gathering_cue123 _alltrials_attentional_final.mat
300MB
Averaging of responses for conditions of different reaction times, left/right button presses
neuronal_regression_results/info_gathering_ramp_ to_choice.mat
4.9GB
Fig8_single_neuron_coding.m
Fig5_OFC_crosscorrelation.m
Fig7_single_neurons _cue2and3.m
Fig6_ACC_multiplesignals.m
Figure 8
Single neuron value correlates at Cue 1
Figure 5
Population signatures of attention-guided value comparison in OFC
Figure 7
Single neuron value comparison and belief updating
Figure 6
Belief confirmation and pre-choice subspace projections in ACC

Behavioural data

Behavioural data for both animals is loaded in from within the folder *behavioural_data*. For each animal, Frank and Miles, this directory contains a .mat file. This contains behavioural data from the Information Gathering trials, which are reported in Hunt/Malalasekera et al., Nat Neurosci 2018. The subjects also completed some 'Covert' trials in separate blocks of the same recording sessions (where all four cues are revealed to the subject simultaneously, rather than using overt attention to sample each cue individually).

Subjects completed blocks of 50 information gathering trials followed by 25 covert trials. This pattern repeated until the subject stopped performing the task, or an experimental time limit of 2 hours was reached.

Each file contains a cell array, *zInfoGatheringInfo*, which contains behavioural of all experimental sessions recorded for each subject (n=25 sessions for Frank, 32 sessions for Miles).

Structure of zInfoGatheringInfo

Each cell of *zInfoGatheringInfo* contains a structure (produced using MonkeyLogic) with the following fields:

- *CodeNumbers* – the identity of trigger code numbers on the current trial. This is the critical field for identifying epochs of interest for subsequent neural analysis, and also for obtaining other information about information sample etc. See Table of Trigger Code Numbers below for details.
- *CodeTimes* – the timing of trigger codes, in milliseconds since trial onset
- *TrialNumber* – indexes the current trial number (note that this is not simply 1:nTrials, as there are intervening 'Covert' trials, and also trials where subject aborted trial (e.g. moved joystick too early) or failed to respond have been removed)
- *ConditionNumber* – 3 if current trial is an 'Option' trial (i.e. second piece of information sampled is on *same option* as first piece of information); 4 if current trial is an 'Attribute' trial
- *TrialError* – 0 if subject chose option with higher expected value on current trial, 6 if subject chose option with lower expected value.
- *NumCodes* – the total number of trigger codes (see below) on the current trial
- *AnalogData* – structure for each trial containing:
    - *EyeSignal* – x and y locations for eye position, sampled at 1 kHz
    - *Joystick* – x and y locations for joystick, sampled at 1 kHz (note that joystick is constrained to only move along x-axis)
- *Uservars* – structure for each trial containing user-generated variables, which describe the values of pictures being presented, their locations, and the choice made by the subject. These include:
    - *TrialNumber* – same as *TrialNumber* field above

- *ProbabilityTracker* – the first two indices of this contain the probabilities associated with left/right options, respectively; the last two indices are always set to 0
- *PayoffTracker* – the first two indices of this contain the magnitudes associated with left/right options, respectively (expressed as a proportion of maximum juice available in training session, per trial); the last two indices are always set to 0
- *PayoffAmounts* – same as *PayoffTracker*, but now expressed as the duration of juice to be delivered by the reward pump (in ms)
- *EVTracker* – the first two indices of this contain the Pascalian expected value associated with left/right options, respectively (product of *ProbabilityTracker* and *PayoffTracker*); the last two indices are always set to 0
- *ProbTop* – indexes whether probability attribute was presented on the top picture of each option (1) or the bottom picture (2)
- *ChosenTarget* - indexes whether left option was chosen (1) or right option (2)
- *Rewarded* – indexes whether reward was delivered (1) or not (0)
- *Remove_index* – on information sampling trials, this indexes the location of the first presented cue (1 = top left, 2 = bottom left, 3 = top right, 4 = bottom right)
- *Sind* – on information sampling trials, this indexes the location of the second presented cue (1 = top left, 2 = bottom left, 3 = top right, 4 = bottom right)

Preprocessing data

The behavioural files require some preprocessing to obtain information that may be of interest.

For example, to obtain the number of cues sampled on each trial on information gathering trials (session n, trial t), we look in *zInfoGatheringInfo{n}.CodeNumbers{t}* for triggers 91, 92, 93 and 94 (see table below) – if three of these triggers appear, this indicates that three cues were revealed on that trial. We can also compare this to when the joystick was first moved (*zInfoGatheringInfo{n}.AnalogData{t}.Joystick*), and can also factor in how long the cue had been viewed for prior to moving the joystick (*zInfoGatheringInfo{n}.CodeTimes{t}*), to decide whether the cue has been viewed for a sufficient length of time to have any influence on the final joystick choice. Simiilar preprocessing is needed to obtain a range of other variables of interest.

Key pre-processing steps are performed by **InfoGathering_Behaviour.m**. This takes *zInfoGatheringInfo* as its input argument, and returns various statistics (collapsed across sessions) for plotting, as in Figure 1 and 2 of the paper.

Neuronal Data

Neuronal data is stored in two separate sets of .mat files. For convenience, all of the behavioural data is also stored alongside the neuronal data.

One set of files (the .mat files contained in *neuronal_data*, e.g. “F002 C17 U1.mat”) is spike sorted but *non-preprocessed* data. There is one file per sorted single unit. This includes raw spike times, and the LFP from the corresponding electrode sampled at 1 kHz (but not the raw 40kHz wideband data - this was not saved in the .mat file after spike sorting using Plexon Offline Sorter). It also includes eye data for that session. Each of these 724 .mat files is ~80-100MB in size (total of ~59GB). Loading all of these files into MATLAB takes time, but is essential if you want to do LFP analysis.

A second set of files (the .mat files contained in *neuronal_data/LTH_processed_data*, e.g. “Unit0123.mat”) contains spiking data that has been *preprocessed*. The spikes have been rasterised and aligned to various events of interest (Cue 1 onset, Cue 2 onset, Response time, etc.). A range of key regressors/events of interest have been pre-computed. Because these .mat files don’t contain analog data (LFP/eye tracking), the file size is much smaller (~0.5-1MB, total of 380MB); it is therefore much more manageable to load all of these files into RAM relatively quickly. Most of the analysis scripts load in this pre-prepocessed data, rather than the raw data.

You can find the script that converts the former, non-preprocessed set of files into the latter, pre-processed files in the base directory: it is titled **InfoGathering_Preprocess_Neurons.m**. This script is relatively well commented. Please use this script find out the meaning of variables that are pre-computed by the function, and then subsequently used by other scripts (such as InfoGathering_Neurons_Cue1_regression).

Structure of non-preprocessed data (in neuronal_data)

Each of the ‘non-preprocessed’ files contains a structure titled *BhvInfo*. This contains behaviour, spiking, LFP, eye-position and joystick data from all trials.

Behaviour: Several of the fields of *BhvInfo* are as in the behavioural data, as described in the previous section (e.g. *AnalogData*, *CodeNumbers, CodeTimes, ConditionNumber, NumCodes, TrialError, TrialNumber* and *Uservars)*. However, these data have not yet been separated into trials based upon whether they were information gathering/covert trials.

Spiking/LFP: BhvInfo also contains a **cell array of spike times** for each trial, in ms following trial onset (in *SpikeCodes*), and **LFP sampled at 1kHz** for this electrode (in *PlexonFieldPotentials)*. When aligning spiking/LFP data to trigger codes**, it is important to use timings obtained from *PlexonStrobes*** (registered directly by the Plexon recording system) rather than *CodeTimes* (registered by the behavioural control software MonkeyLogic). Whilst the trigger codes in *PlexonStrobes* are exactly the same as those in *CodeNumbers* (see previous section), the timings will be systematically

different from those in *CodeTimes*. Timings in *PlexonStrobes* reflect when strobes were actually sent to the Plexon recording software. Examples of how to align spiking data to cue events etc. can be seen in **InfoGathering_Preprocess_Neurons.m**.

Recording locations: To obtain information about where each recording was made, use “miles_area.mat” and “frank_area.mat”, both of which are stored in *other_data*. Each of these contains a N*12 matrix, where N is the total number of electrodes lowered across all recording days in that animal; columns are as follows:
Column 1 = recording session number
Column 2 = channel ID
Column 3 = area
Column 4 = hemisphere (1 = left hemisphere, 2 = right hemisphere)
Column 5 = AP position, in mm, stereotactic coordinates
Column 6 = X position (medial-lateral), in mm, chamber coordinates
Column 7 = number of rotations of microdive used to reach recording site (each rotation = 1/3 mm lowering of electrode)
Column 9 = distance electrode travelled to reach recording site (in mm, i.e. column 7 divided by 3)
Column 11 = length of guide tube (mm)
Column 12 = which session of a given stimulus set (0 = learning day when stimulus values were learnt, *not* information sampling session; 1 = first day with this stimulus set, 2 = second day; 3 = third day)

Structure of preprocessed data (in neuronal_data/LTH_preprocessed_data)

Each of the 724 units has been preprocessed using the script **InfoGathering_Preprocess_Neurons.m**, and the results stored as a .mat file in neuronal_data/LTH_preprocessed_data.

The best way to understand variables that are stored in these .mat files is to load an example file into MATLAB, and then examine the comments in **InfoGathering_Preprocess_Neurons.m** to understand how different variables have been computed/stored.

Information about each of the 724 units is also stored in “all_units_info.mat”, which is in the folder *other_data*. This contains a structure *UnitInfo* for each unit, containing information about recording site/day (e.g. *all_units{42}.UnitInfo* contains information that corresponds to *Unit0042.mat* in neuronal_data/LTH_preprocessed_data).

Table of Trigger Code Numbers (in *CodeNumbers* field)

Blue are triggers that only appear on information sampling trials; red are triggers that only appear on covert trials

1	Reward delivery epoch begins (still happens even if juice is not delivered)
2	Reward delivery epoch ends
3	End of trial
4	Joystick leaves central holding area (trigger not included on information sampling trials, use trigger 72 instead)
9	Repeated three times at beginning of each trial, to signal trial onset
10	Fixation point appears onscreen
11	First entering fixation point at beginning of trial
12	Fixation finishes
16	Trial start
18	Repeated three times at end of each trial, to signal trial end
62	Red 'hold joystick cue' disappears (subject permitted to move joystick at any point from now onwards)
72	Joystick enters target area, selected option highlighted, all four cues revealed
73	End of pre-feedback phase
80/81	End of block - switch to other trial type
89	First blue 'instruction cue' to tell subject where to direct first eye movement, information sampling trials
90	Stimulus removed on information sampling trials (300 ms after it has been first fixated), next instruction cue appears
91	Spatial location north-west (left option), time when first fixated on information sampling trials
92	Spatial location south-west (left option), time when first fixated on information sampling trials
93	Spatial location north-east (right option), time when first fixated on information sampling trials
94	Spatial location south-east (right option), time when first fixated on information sampling trials
162	No longer used (used to be 'green go cue appears onscreen', but go cue was removed)
200+	Can be ignored