Dataset Documentation

Dataset

Reinagel, Pamela, 2020, "Human and Rat Motion Discrimination Reaction Time Task Data", https://doi.org/10.7910/DVN/ATMUIF, Harvard Dataverse.

A replication data set for: *Nguyen, QN and Reinagel, P. (2020) A qualitative difference in decision-making of rats vs. humans explained by quantitative differences in behavioral variability. (bioRxiv)* *https://doi.org/10.1101/2020.01.05.895268*

These data come from, and were shared by the authors of: *Shevinsky, CA and Reinagel, P. (2019) The interaction between elapsed time and decision accuracy differs between humans and rats. Front. Neurosci. 13:1211. https://doi.org/10.3389/fnins.2019.01211*

File Names

allEpochsHUMAN200502.csv
allEpochsRAT200502.csv
BestUnbiasedPsychoEpochsHUMAN200502.csv
BestUnbiasedPsychoEpochsRAT200502.csv

Task

To briefly summarize the aspects of the task important to this analysis: subjects self-initiated a trial by licking a sensor centered relative to the display (rats) or by pressing the down arrow key on a numeric keypad (humans). The visual random dot motion stimulus started immediately, and persisted until a response was made by licking the L or R sensor (rats) or pressing the L or R arrow key (humans). There was no cue or deadline for responding but most subjects respond in < 2 s in most trials. Correct responses were indicated by an immediate beep (both species) and reinforced with an immediate reward: a water drop (for rats, which were water restricted) or incrementing the displayed "score" (for humans, whose monetary compensation was linked to their score). After correct responses, subjects could initiate another trial without delay (no enforced inter trial interval). Error responses were penalized by a time-out of a few seconds, during which initiation requests were ignored. The end of the time-out was indicated by a screen color change. As noted in Shevinski & Reinagel, for some groups of subjects the motion stimulus terminated after 3 or 5 s to encourage a rapid response strategy. These subjects were not distinguished in that study, nor in the modeling by Nguyen & Reinagel, but we acknowledge the possibility of an implicit deadline influencing decision time in those subjects.

Selection criteria of the data sets

1. allEpochs

These files include all the trials from all the available quasi-stationary epochs as defined in Shevinsky and Reinagel, 2019. Data inclusion criteria are described in detail there, but to highlight some relevant points: Although the epochs were identified as having stable experimental parameters (coherence

distribution, etc.) and relatively non-trending reaction time and accuracy, they may have weak linear trends, and do have significant temporal fluctuations in reaction time on intermediate timescales (tens of trials). Moreover some of these epochs have considerable lapse or bias, some have only a single coherence, and in some epochs the subject might be at chance or at ceiling due to the coherence range employed. Very long RT trials (>3s) are rare in both species, and we suspect these represent anomalous trials (e.g. subject was momentarily distracted). Generally analysis is restricted to RT<=3s. Nevertheless all trials up to RT=5 were included in this data set. An exception to this was made if the subject was trained with stimuli limited to 3s duration, in which case only trials with RT<3s were included. To avoid rare artifacts such as simultaneous keypresses, very short RTs (< 1st percentile of subject's lifetime RT distribution) were also excluded from the dataset. Finally we note that although the trials are listed in the order they occurred, they are not necessarily adjacent in time. Intervening trials may have been excluded, and for rats sequential trials are sometimes separated by gaps of minutes, hours or days.

2. bestUnbiasedPsychometricEpochs

These files contain a subset of the epochs in allEpochs, and should be considered the cleaner, more curated data set. These contain only one epoch per human or rat subject, which was chosen by the following criteria:

i. stationary (no linear trend)
ii. psychometric (>=4 coherences)
iii. low side bias at highest coherence
iv. if any remaining candidate epochs for a subject had no detectable temporal fluctuation in RT or Accuracy, the largest N trials among these was chosen
v. otherwise, the candidate with the least temporal fluctuation was chosen

Column explanations

The csv files have the following columns:

subj_idx integer: indicates the epoch ID
Note: this is a required column name for HDDM. Putting the epoch ID (rather than the subject ID) in this column causes HDDM to treat each epoch as a distinct "subject" with its own individual level parameters.

species string ('human' or 'rat')
Note: although this field is identical for every entry within a file, the column is required for HHDDM so that it will fit a hierarchical model with a species-level group posterior. One could also combine the human and rat trials in a single training file, in which case this field would identify group membership. We separated the species mostly for computational parallelization.

coh floating point number : indicates the motion coherence of the trial
Note: coherences is encoded as an unsigned value between 0 and 1. Motion direction is not encoded. Within an epoch, the distribution of coherence values should be i.i.d. Other stimulus parameters that were not varied within the epoch (such as contrast, dot size, dot density, dot speed, reward size, or penalty duration) are not indicated, though they also may affect performance.

rt floating point number : indicates response time, defined as time elapsed from trial initiation (always simultaneous with stimulus onset) until response commitment (selection of either R or L response, always simultaneous with stimulus offset and reinforcement). Any input at the request key/port after stimulus onset was ignored.

response binomial: indicates whether the response was correct (1) in the sense of choosing the rewarded option, or error (0) in the sense of choosing the penalized option.

subjectID integer: identifies the unique subject (rat or person), which can be the same for multiple epochs in the allEpochs data set. This was originally included with the idea of training nets with a nested hierarchy (epochs belong to subjects which belong to species). But to our knowledge HDDM does not support that functionality. Therefore in Nguyen & Reinagel 2020, *this field was never used by the HDDM fitting algorithm*, and is included only for documentation purposes.

Provenance

Original source

The experimental methods, ethics statements for humans and for rats, data processing and inclusion criteria and primary experimental findings are described in Shevinsky & Reinagel 2019. In that study the data were originally stored by the custom Matlab software that was used to run the behavioral trials (Ratrix v1.0.2). Raw data were saved in a custom trialRecords.mat data file format, one file per behavioral session. The analysis of Shevinsky & Reinagel worked directly upon the trialRecords data files. Data curation described in that manuscript includes criteria for exclusion of subjects, sessions, and trials, and the definition of "epochs" – sets of trials from a contiguous time range in which experimental parameters did not vary and subject performance was approximately stable.

Conversion

For purposed of the modeling study described in Nguyen & Reinagel 2020, new Matlab scripts were written (PR) to collate the relevant trial data directly from the original trialRecords files, based on the trial numbers defined as "epochs" in Shevinsky & Reinagel; and to save the necessary data fields into a CSV file formatted specifically for use as training data for the HDDM package (http://ski.clps.brown.edu/hddm_docs/). A few manual repairs were made to the CSV files to remove a small number of trials that were erroneously assigned to an epoch in the original study (rogue coherence values). The files provided here are identical to those used to fit models in Nguyen & Reinagel except that additional data columns that were not used by the models have been removed.

For the purpose of cross-validation for performance measures, Nguyen & Reinagel split each of these data files into two files (the "odd" and "even" trials) to fit parameters to each half of the data separately. Since this is a trivial operation, the split data files are not provided here. Parameters estimated from half the trials were slightly but reproducibly different between odd and even trials, indicating that the number of trials was limiting for sampling the reaction time distributions. Therefore Nguyen & Reinagel report the parameters estimated by fitting all the trials together.