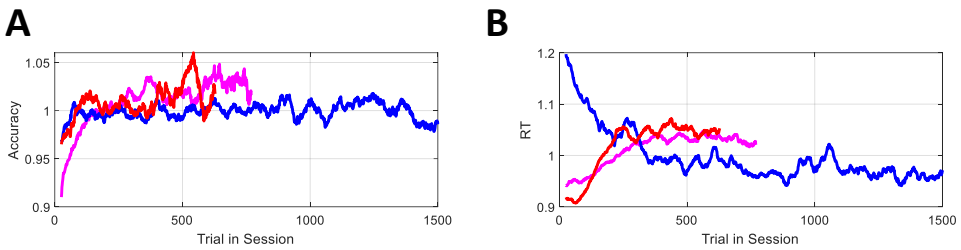
"Data not shown" from Shevinsky and Reinagel 2019, see published paper for context: https://doi.org/10.3389/fnins.2019.01211

Non-stationarities related to time within the session

For both rats and humans, average accuracy improved over the first few hundred trials of each session, even in 2nd and 3rd sessions (humans) or after dozens of sessions (rats). Average reaction time of rats increased over the beginning part of each session, while humans showed a decrease in reaction time. These trends are hard to resolve in individual sessions but are apparent after averaging over sessions and subjects.



Consistent trends over the course of a testing session. Red and magenta curves are for rats in mixed-coherence or single-coherence experiments respectively; blue curves are for humans in mixed-coherence experiments. **A.** Accuracy as a function of trial number in the session, normalized to the mean accuracy in the session, then averaged over sessions and subjects. **B.** Reaction time as a function of trial number in the session, normalized to the mean reaction time in the session, then averaged over sessions and subjects.

Rats are thirsty and working for water rewards, and therefore their motivation is highest at the session onset, moderating as the rat consumes water. This may explain why rats' reaction times increase with trial number. For humans there would be no such satiety effect. Humans, on the other hand, may need a warm-up or practice period at the beginning of each session. Rats typically preform many more sessions than humans, and thus are already highly practiced at the beginning of the session. For humans, accuracy and reaction time remain stable out to at least 1400 trials. Rats rarely performed more than 800 trials in a restricted-duration session.

For many types of analysis, these details may not be important. But for the purpose of determining if the response time distributions of errors and correct trials are the same (as predicted by the drift diffusion model or DDM), pooling data from early and late parts of sessions to could introduce spurious correlations. Therefore, such analyses should either exclude the first ~400 trials of each session, or use a temporally local comparison method that is robust to trending nonstationarity.

Analysis details

To obtain the average accuracy and reaction time as a function of trial number, vectors of binomial outcomes (correct/incorrect) and reaction times (seconds) were recorded for all qualifying trials in a session, and these vectors were normalized to the mean within session. These normalized session vectors were then averaged over sessions within subject. For the graphs shown we averaged over subjects up to the highest trial number for which we had data from at least 10 subjects; this average over subjects was then smoothed (moving average span 50) for the sake of graphical clarity. No average curve is plotted for humans in single coherence condition because we had data from <10 subjects; but the data from the 7 available subjects were consistent with the human results from multiple coherence experiments. In a separate analysis, we analyzed only the first session of every subject, or excluded every subject's first session. Results were qualitatively similar in first sessions and later sessions (not shown).