Supplemental information supporting "data not shown" in Shevinski and Reinagel 2019
See published paper for context: https://doi.org/10.3389/fnins.2019.01211

"You can't step in the same river twice" – Heraclitus, 535-475 B.C.

Slow fluctuations in behavior within sessions

Comparison of the response time distributions of error trials vs. correct trials requires pooling data over hundreds of trials. Despite best efforts to exclude nonstationary data sets, one's ability to detect non-stationarity is statistically limited, and to our knowledge a fully satisfactory method does not exist. Although we excluded any time series with significant overall linear trends in either accuracy or reaction time, as well as excluding the early trials of each session, it remains possible that subjects' state fluctuated over time on shorter timescales. We computed the autocorrelations of accuracy and of reaction time within the putatively stationary epochs. Within each epoch the experimental parameters (coherence and motion direction) were randomly and independently selected each trial. Therefore, if the subject's decision-making process was not changing, the probability of being correct in a given trial should not depend on or correlate with the correctness of adjacent trials. Similarly, the reaction time, however variable, should be uncorrelated from trial to trial.

We find only weak and insignificant short-range correlations (±32 trials) in our data from rats (e.g., Fig. 1A,B) and humans (e.g., Fig. 1C,D). Nevertheless, in both species some epochs had a weak positive autocorrelation in either accuracy or reaction time over the entire window (e.g., Fig. 1D) indicating correlations on slower timescales. To capture this, we computed the average accuracy and average reaction time in non-overlapping blocks of 16 trials. Example autocorrelations of these vectors are shown in Fig. 1E-H.

The variance of the average accuracy or reaction time across all non-overlapping 16-trial blocks can be used as a measure of the extent of fluctuation in these behavioral outcome measures. We chose 16-trial blocks as a compromise between averaging over enough trials to obtain approximately normal distributions of values, while still having enough blocks in an experimental epoch to estimate the variance over them (>30 independent estimates). Because stimulus values were randomly interleaved, trial outcome is binomial, and reaction time distributions are very broad, these 16-trial averages are expected to fluctuate considerably even if the subject's behavior were perfectly stationary. To quantify the amount of variation expected on chance from sampling alone, we repeated the analysis for 1000 shuffle controls of each epoch, computing the averages in randomly selected groups of 16 trials instead of sequential trials. Finally, we compared the mean shuffle-control variance to the observed variance on an individual epoch basis (Fig 1I-L).

This analysis reveals that the fluctuations in behavioral measures can be greater than expected of a truly stationary process, even in data sets that have no significant long-term trends and relatively flat autocorrelograms. In particular, significant fluctuations in reaction time were found in nearly all individual human and rat experiments in our study (Fig 1J,L). This is important because the effects we are studying – correlations between accuracy and reaction time – can arise just from mixing together trials collected under different behavioral states.

As experimentalists, we tend to test for trends on long timescales, or effects on very short timescales (dependence on previous trial). But fluctuations on any timescale would have the same consequences for the analysis of pooled trials; and fluctuations on short timescales are very hard to detect statistically or avoid experimentally. We conclude that it is unwise for models of decision-making (or any behavior) to rest on stationarity assumptions. Computational models that presume and explicitly model nonstationarity would be better.

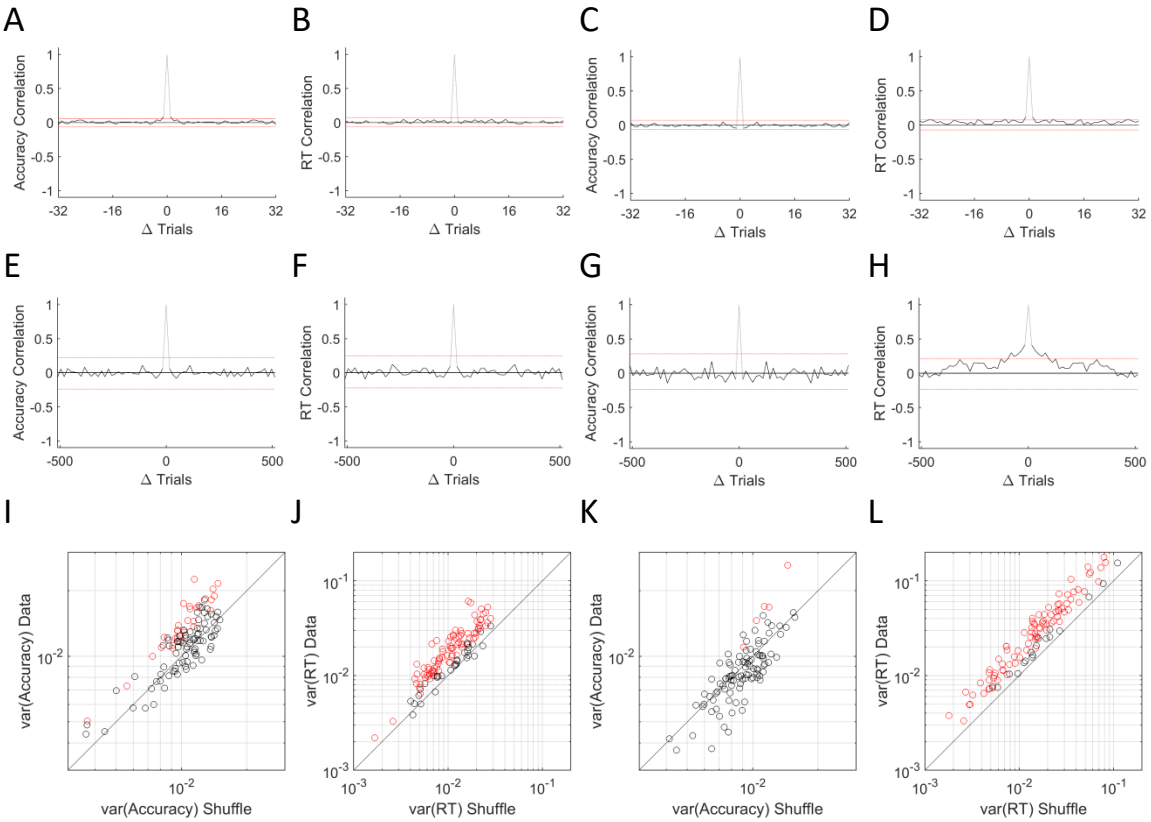

Figure 1. Evidence of fluctuations in behavioral state within "stationary" epochs. All left panels contain rat data, all right panels human data. **A.** Autocorrelation of accuracy as a function of distance between trials (±32 trials), from an example rat epoch. Correlation at 0 is 1 by definition. Red lines indicate the 99% confidence interval, computed from the autocorrelations of 1000 temporally shuffled controls of the same time series, with Bonferroni correction for multiple comparisons. **B.** Autocorrelation of reaction time, for the same rat epoch shown in A. **C.** Autocorrelation of accuracy for an example epoch from a human subject. **D.** Autocorrelation of reaction time for human epoch shown in C. **E-H,** like A-D, but autocorrelation is of ±32 non-overlapping 16-trial time bins. **I.** Each symbol represents the variance of accuracy over all 16-trial blocks in an epoch (vertical axis), compared to the average variance in 1000 shuffle controls, horizontal axis), for N=115 epochs from rats. The symbols in red were individually significantly more variable than chance (>99.5% of shuffle controls). At the population level, rats' accuracy variability was significantly higher than chance ($P=9.4\times10^{-9}$, one-tailed sign rank test). **J.** Variance of mean reaction time in 16-trial blocks, compared to shuffle control, for same rat epochs analyzed in I. At the population level rats' reaction time variability was significantly higher than chance ($P=4.5\times10^{-20}$). **K-L:** like I-J but for N=104 human epochs. At the population level humans' accuracy is no more variable than expected on chance (P=0.97), but reaction time variability was significantly higher than chance ($P=4.3\times10^{-19}$). *Computed 190325*