

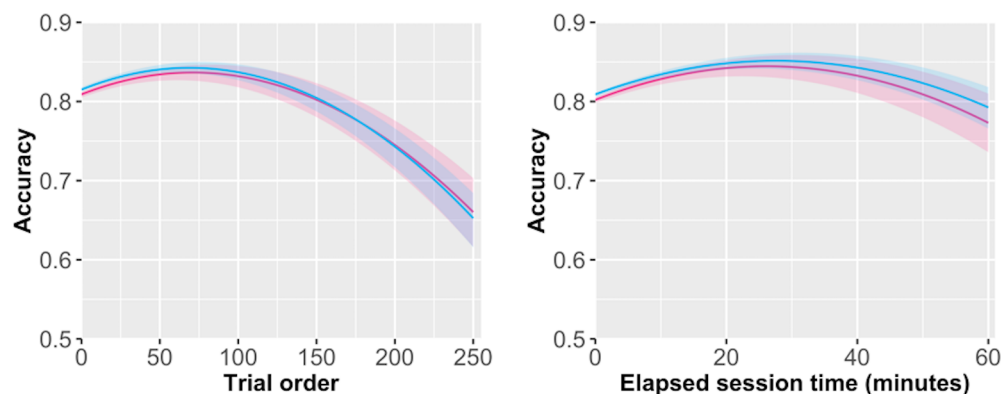


Fig 3. Quadratic regression of accuracy on trial length. Notes: Red = sample 1, Blue = sample 2. Average regression effect of (A) trial number and (B) elapsed time on accuracy. Shaded regions represent the 95% CI of the fixed effect.

<https://doi.org/10.1371/journal.pone.0182980.g003>

Session accuracy throughout the day

Our first analysis identified that users experience a decrement in ability as they persist in the task. This section considers whether there is an effect of time-of-day on ability, indicating more domain-general depletion of ability that wasn't directly caused by persistence on the task at hand. To assess this, we looked at the average score per session as a function of time-of-day. This involved finding the mean score for each session, and averaging all sessions within half-hour intervals across the day. Weighted linear regression was used to assess the relationship. In our analyses with sample 1, we found that many users provided either too few data points, or too little variance in terms of time-of-day, to run a MLM clustered within user.

We anticipated that if session accuracy had a measurable daily pattern, it would likely be sinusoidal, reflecting a daily pattern of increasing or decreasing ability, repeating itself each day. We tested sinusoidal regression lines from the first to fifth harmonic, as well as a linear, quadratic and intercept only (null effect) models. Harmonic functions are similar to exponential functions in that they allow the function to alter its acceleration to more precisely fit to the data. As with exponential transformations, harmonic transformations always improve the model fit at least slightly, and as such should be corrected for possible over-fitting. To account for this varying complexity across models, we used Watanabe-Akaike information criteria (or Widely-Applicable-Information-Criteria, WAIC) as the measure of model fit [11]. Models were fit with sample 1 and compared using WAIC. The exact models were then used to test out-of-sample deviance with sample 2 data based on root mean squared error (RMSE). This allows us to use the model from sample 1 and test how well it can predict new data relative to the other models. Finally, the models were re-fit with sample 2 data after adjusting for users' time-zone.

Our sample 1 results indicated that a 1st harmonic sine function was the best and most parsimonious fit for the data, with a WAIC score difference of at least 5.2 relative to all other models. RMSE from sample 2 largely supported this, though the small difference between all models highlight the small magnitude of the effect (RMSE ranged $(2.03-2.08) \times 10^{-4}$ across all models, the best model improved prediction by less than 4% against the intercept model). WAIC scores from the re-fit sample 2 data indicate that a 1st or 2nd harmonic function, or a quadratic model was the best fit. We focused on the 1st harmonic function as it was the simplest model that was ideal for cyclical data (Fig 4, S3 Table For full 1st harmonic regression models). This model indicates that performance drops throughout the day, albeit very slightly

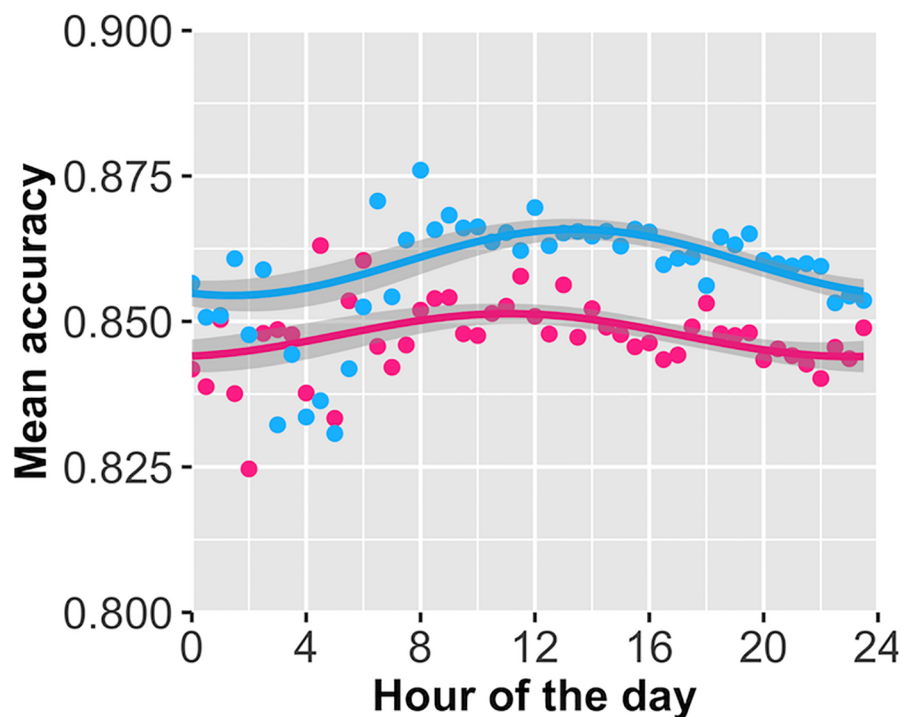


Fig 4. Daily session accuracy over 24 hours with 1st harmonic regression line. Notes: Both lines are 1st harmonic regression lines fit to data. Red = sample 1 data. Blue = sample 2 adjusting for users' time-zone. Shaded regions represent the 95% CI of the fixed effect. Dots represent the average score across sessions for each 30-minute window.

<https://doi.org/10.1371/journal.pone.0182980.g004>

(roughly 1% drop in performance from peak to trough). While the effect is small, it consistently emerged in both samples.

Fitting to our theoretical models (Fig 2), the data in sample 1 were best fit to the model assuming limited self-control. However, the sample 2 data were no more likely under either the limited self-control or evening preference model, which predicts the exact opposite (Table 1). The failure of our models to clearly distinguish between a pattern of early morning or late evening ability in sample 2 comes from our initial expectation that peak performance occurred early in the morning, whereas our 2nd sample (once adjusted for time zone data) suggests the peak is closer to 2pm. This is important, because our initial peak was drawn from an expectation that self-control begins to deplete across the sample shortly after waking. That the peak emerges at 2pm is somewhat at odds with this. Given this failure to replicate our sample 1 findings, the data do not speak strongly in favor of a domain-general resource model of

Table 1. Deviance of accuracy scores relative to theoretical models for daily patterns of accuracy.

Model	Sample 1	Sample 2 direct replication	Sample 2 adjusting time zone
Limited SC	1.1×10^3	8.84×10^6	8.84×10^6
Unlimited SC	2.8×10^3	8.97×10^6	8.97×10^6
Evening preference	1.5×10^3	8.84×10^6	8.84×10^6

Notes: SC = self-control. Values are the negative log likelihood of the data given the model. Smaller values imply less deviance and greater likelihood. See Fig 2 for representations of the theoretical models.

<https://doi.org/10.1371/journal.pone.0182980.t001>