

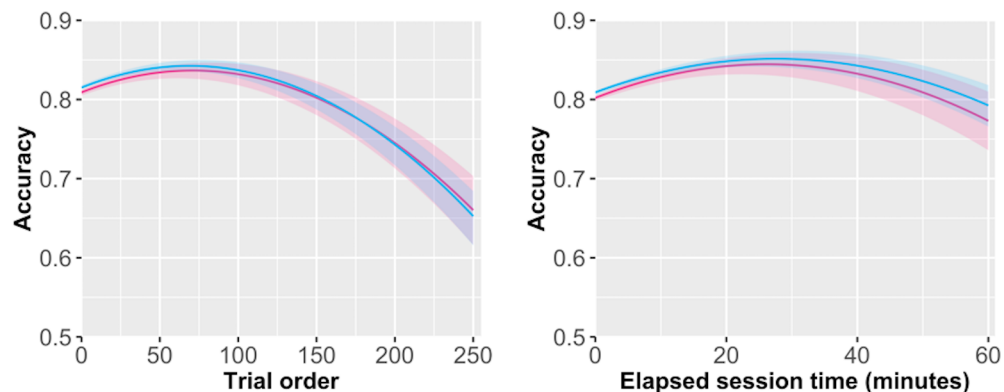


图3.准确率随试次长度变化的二次回归。注：红色=样本1，蓝色=样本2。(A)试次数和(B)经过时间对准确率平均回归效应。阴影区域表示固定效应的95%置信区间。

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

全天各时段的会话准确率

我们的首次分析发现，用户持续执行任务时，其能力会下降。本节考察一天中的时段是否会影响能力，这可能表明能力出现了更具领域普遍性的衰退，而这种衰退并非由持续执行当前任务直接导致。为评估这一点，我们考察了平均会话得分与一天中时段的关系。具体做法是计算每个会话的平均得分，并对全天每半小时区间内的所有会话取平均值。我们使用加权线性回归评估二者之间的关系。在对样本1的分析中，我们发现，许多用户提供的数据点过少，或一天中不同时段的数据变异太小，无法运行按用户聚类的多层线性模型。

我们预计，如果会话准确率存在可测量的日变化模式，这种模式很可能呈正弦曲线，反映能力每天逐渐增强或减弱的日变化，并且日复一日地重复。我们检验了从第一阶到第五阶谐波的正弦回归线，以及线性、二次和仅含截距（无效应）模型。谐波函数与指数函数类似，都允许函数改变其加速度，从而更精确地拟合数据。与指数变换一样，谐波变换总能在一定程度上改善模型拟合，因此需要校正可能出现的过度拟合。为考虑不同模型的复杂程度，我们使用渡边-赤池信息准则（亦称广适信息准则，WAIC）衡量模型拟合度[11]。我们使用样本1拟合模型，并根据WAIC进行比较。随后，使用完全相同的模型，根据均方根误差（RMSE）检验其对样本2数据的样本外偏差。这样，我们便可以使用基于样本1建立的模型，并检验与其他模型相比，它对新数据的预测效果。最后，在对用户所在时区进行调整后，我们使用样本2数据重新拟合模型。

样本1的结果表明，第一阶正弦谐波函数对数据的拟合效果最好，且最为简约；与其他所有模型相比，其WAIC至少低5.2。样本2的RMSE结果总体上支持这一结论，不过所有模型之间的差异很小，表明效应幅度也很小（所有模型的RMSE均在 $(2.03-2.08) \times 10^{-4}$ 范围内，最佳模型相比仅含截距的模型，预测效果提升不到4%）。使用样本2数据重新拟合后得到的WAIC分数表明，第一阶或第二阶谐波函数，或二次模型的拟合效果最好。我们重点关注第一阶谐波函数，因为它是适用于周期性数据的最简单理想模型（图4，补充表3列出了完整的第一阶谐波回归模型）。该模型表明，表现会在一天中逐渐下降，尽管下降幅度非常小

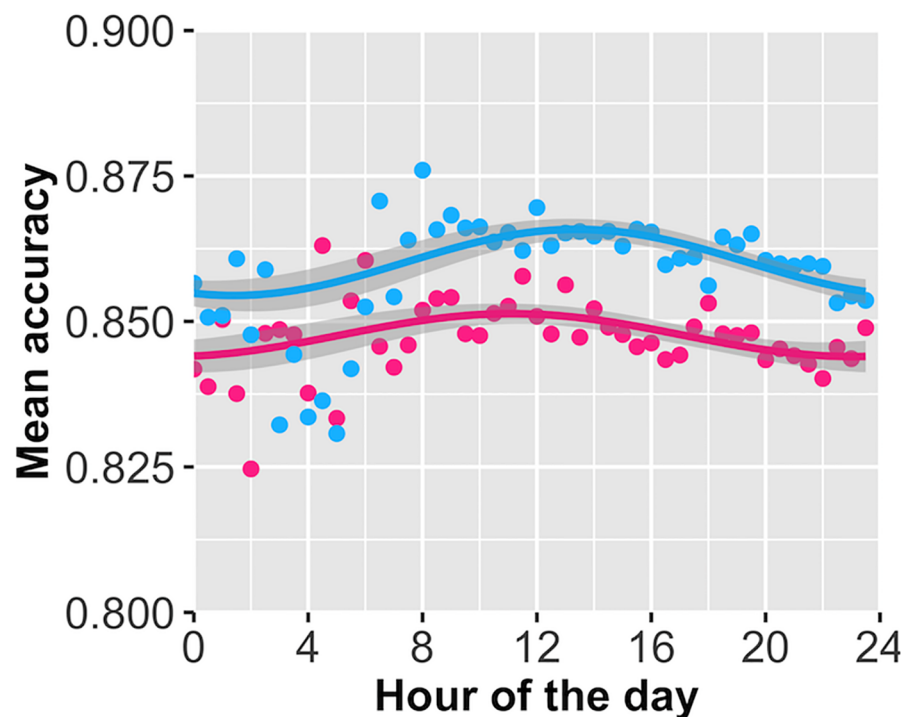


图4.使用第一阶谐波回归线呈现的24小时内每日会话准确率。注：两条线均为拟合数据的第一阶谐波回归线。红色=样本1数据。蓝色=按用户所在时区调整后的样本2数据。阴影区域表示固定效应的95%置信区间。圆点表示每个30分钟区间内各会话的平均得分。

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

（从峰值到谷值，表现大约下降1%）。虽然效应很小，但在两个样本中都稳定出现。

与我们的理论模型相符（图2），样本1的数据最符合有限自我控制模型。然而，样本2的数据在有限自我控制模型和晚间偏好模型下的可能性并无差异，而后者预测的结果恰恰相反（表1）。我们的模型未能明确区分样本2中清晨能力较强还是深夜能力较强的模式，这是因为我们最初预期表现峰值出现在清晨，而样本2（根据时区数据调整后）则显示峰值更接近下午2点。这一点很重要，因为我们最初对峰值的判断基于这样的预期：醒来后不久，自我控制能力便会在整个样本中逐渐消耗。峰值出现在下午2点与这一预期多少有些不符。鉴于样本1的发现未能得到重复验证，这些数据并不能有力支持领域通用的资源模型

表1.准确度得分相对于准确度每日变化模式理论模型的偏差。

模型	样本1	样本2直接重复验证	样本2调整时区
有限自我控制	1.1×10^3	8.84×10^6	8.84×10^6
无限自我控制	2.8×10^3	8.97×10^6	8.97×10^6
偏好晚间	1.5×10^3	8.84×10^6	8.84×10^6

注：SC = 自我控制。数值为给定模型时数据的负对数似然。数值越小，偏差越小，似然越高。有关理论模型的示意图，请参见图2。

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