APPENDIX

Multilevel models provide exceptional flexibility in how to model the relationship between the predictor and predicted variable. As in standard linear regression analyses, the predictor can be either a fixed or random effect. In addition to standard linear regression analysis, the predictor can be fixed or random at either or both the “day” and “participant” levels. A final degree of flexibility in this modelling framework is the capacity to have a covariance term to allow for a nonadditive relationship between the variances. Of course, different model structures can lead to different results. To allow our results to reflect the patterns in the data, rather than a particular choice of model structure, we compared the predictors across all possible model structures. There were altogether nine possible model structures for each predictor. We used Iterated Generalized Least Square (IGLS) to remove any bias introduced by autocorrelation in the predicted variable. This iterative method was able to find the maximum likelihood solution for 98% of the models, and these models were used in the remaining analysis. Instead of showing the maximum likelihood values, we presented the -log maximum likelihood values for better readability. The 2 SE variables were negatively skewed, so we monotonically transformed them to normally distributed variables, by replacing each score with its expected rank under a normal distribution. A total of 107 observations of the 830 total had 1 or more missing fields. To enable fair comparisons between models while using as much of the data as possible, we removed only those observations that had missing data for the predicted variable or any of the predicting variables associated with that predictor. On average, 36 observations of the 107 possible missing observations were excluded from each analysis.

The significance of individual predictors was assessed by between-model comparisons using maximum likelihood fits. As the constant model, \( M_c \), was nested within the constant + predictor model, \( M_{c,p} \), we could assess whether the addition of the predictor made the fit significantly better via a LRT, \( \text{LRT} = -2\log p(M_c) + 2\log p(M_{c,p}) \), where \( p(M_c) \) is the maximum likelihood of the constant model and \( p(M_{c,p}) \) is the maximum likelihood of the constant + predictor model. The probability of finding this LRT value was then determined by comparing it to a \( \chi^2 \) distribution with degrees of freedom equal to the difference in the number of parameters between the 2 models. The number of parameters here is not the number of predictors, but is instead the number of terms that were fit for each model. For example, the constant model, \( M_c \), had a total of 3 parameters; 1 parameter for the mean, 1 for the level 1 variance, and 1 for the level 2 variance.

Assessing the relative strength of the predictors requires a comparison of non-nested models. A standard method for doing so is the AIC, which rewards goodness of fit and penalizes a model for flexibility. The AIC value of a model is calculated from the maximum likelihood fit and the number of parameters, \( \text{AIC}_i = -2\log p(M_i) + 2V_i \) where \( p(M_i) \) is the likelihood of model \( i \) and \( V_i \) is the number of parameters in the model. AIC gives us a scale for comparing models: lower AIC values indicate a better model.

To assist in the interpretation of the raw AIC values, we transformed them to relative model probabilities where a relative probability is equal to the ratio of the exponentiated AIC value of a particular predictor against the sum of the exponentiated AIC values of all of the predictors:

\[
\text{Relative Probability} = \frac{\exp(-\frac{1}{2}\text{AIC}_i)}{\sum_{i=1}^{k} \exp(-\frac{1}{2}\text{AIC}_i)}
\]

These relative probabilities tell us how likely a model is given the data, in comparison with the other models that we are considering.

An alternative to taking the best model is to find structure-agnostic probabilities by averaging the probabilities for each predictor over the different model structures and renormalizing. The orderings of relative model probability, as we found, were the same between the 2 approaches.