Estimating Individual Differences in Working Memory through ACT-R Modeling and Resting State Connectivity

Patrick Rice (pjrice@uw.edu)
University of Washington Department of Psychology, Campus Box 351525 Seattle, WA 98195 USA

Andrea Stocco (stocco@uw.edu)
University of Washington Department of Psychology & Institute for Learning and Brain Sciences, Campus Box 351525 Seattle, WA 98195 USA

Abstract

A complete and holistic understanding of human cognition should be able to relate idiographic parameters representing cognitive functioning to interactions between connected brain networks identified by neuroimaging methods. Here, using the ACT-R cognitive architecture, we examine the possibility of producing idiographic parameterizations of cognitive functioning in a task environment and show that these parameterizations produce reasonable predictions of individual behavior. We then demonstrate a method of determining a subset of parameters that are adequate for prediction of behavior before confirming that the most critical of these task-based parameters is related to functional connectivity measures in individual resting-state fMRI data.

Keywords: Cognitive Modeling; Long-Term Memory; Resting-state fMRI; Functional Network

Introduction

One of the advantages of the utilization of computational models in the study of cognition is the possibility to estimate parameters that characterize behavior and/or cognitive performance on a per-individual, or *idiographic*, level. For example, reinforcement learning (RL) models can be fit to behavioral data, and the resulting parameter estimates can be used to make inferences about individual differences in dopamine function or to distinguish between healthy and pathological groups (Frank et al. 2004). Similar work has been performed with drift diffusion models (DDM), which model decision-making through a noisy information accumulation process that "drift" towards one of two decision boundaries. In addition to being more clearly interpretable than raw behavioral data, parameters inferred through DDM are often more reliable in detecting individual differences than behavioral metrics (White et al. 2016). In the past, ACT-R models have been used to make such inferences as well. For example, Daily et al. (2001) estimated goal spreading activation from behavioral data, used it as a proxy for working memory, and successfully predicted performance on a different task.

This individual-difference approach, however, has not been applied consistently - instead, the majority of modeling efforts have focused on fitting parameter values that are descriptive at the group level. Furthermore, ACT-R is a far more complex computational framework than RL or DDM, and it encompasses dozens of parameters. While this complexity makes it possible to capture complex tasks that lay outside the scope of RL or DDM models, it also poses some significant challenges: is it possible to identify idiographic parameter values that reliably characterize the behavior of a given individual? How many parameters are needed to characterize individual differences within a group? How can each parameter's contribution to predicting these differences between individuals be determined?

Here, we provide an empirical answer to this question. We created a model of the zero-back condition of the standard n-back working memory task, and then fit the model to behavioral data from ~150 participants. We show it is possible to use convex optimization techniques to identify points in multidimensional parameter space that accurately capture an individual's performance. We then provide a method to determine which estimated parameters contribute most meaningfully to the prediction of individual performance. Furthermore, we demonstrate that these idiographic parameterizations are predicted by the individual's resting-state functional connectivity, indicating that the parameterization captures fundamental aspects of individual cognitive function.

Materials and Methods

The study presented herein consists of an analysis of N = 178 individuals from the Human Connectome Project, the largest existing repository of young adult neuroimaging data. The analysis was restricted to the resting fMRI subset in conjunction with the zero-back condition of the "Working Memory" (WM) task component. The resting fMRI data collection consisted of two 30-min recording sessions, performed 24 hours apart; the task fMRI data collection consisted of two 30-min task sessions performed directly after each resting-state acquisition session. During each task session, participants performed six other tasks in addition to the WM component, per the HCP protocol. All subject recruitment procedures and informed consent forms were approved by the Washington University in St. Louis' Institutional Review Board. The present study met criteria for exemption at the University of Washington's Institutional Review Board.

Task Data

Each working memory task session consisted of four 0-back blocks, with each block containing 10 trials. Each block begins with a 2.5 s cue that informs the participant of the target stimulus for the proceeding block of trials. Each trial presents a single image centered on the screen, and participants are required to indicate if the trial's stimulus is identical to the cue stimulus by pressing one of two buttons. The stimuli belong to one of four possible categories: faces, places, tools, and body parts. These categories were presented in a block-wise fashion such that two of the eight blocks presented a given category. Each trial stimulus is presented for 2 s with a 500 ms ITI, for a total duration of 27.5 s per block. Additionally, 15 s fixation blocks were presented after the second and fourth task blocks within a session. This paradigm produces stimuli of three conditions: targets (match to the block cue); lures (non-targets that have been presented at least once before within the block); and non-target, non-lures (non-targets that are presented for the first time within a block).

fMRI Image Acquisition and Preprocessing

Functional neuroimages were acquired with a 32-channel head coil on a 3T Siemens Skyra with TR = 720 ms, TE = 33.1 ms, FA = 52°, FOV = 208 × 180 mm. Each image consisted of 72 2.0 mm oblique slices with 0-mm gap in-between. Each slice had an in-plane resolution of 2.0 × 2.0 mm. Images were acquired with a multi-band acceleration factor of 8X.

Images were acquired in the "minimally preprocessed" format (Van Essen et al., 2013), which includes unwarping to correct for magnetic field distortion, motion realignment, and normalization to the MNI template. The images were then smoothed with an isotropic 8.0 mm FWHM Gaussian kernel.

ACT-R Modeling of WM Task Data

An ACT-R task device and model were developed in order to characterize individual behavior in the zero-back task. The task device implements the zero-back task by updating the ACT-R visicon with a representation of the task elements in the form of three strings: one that identifies the category of the stimulus, a second representing the stimulus itself, and a third indicating the "kind" of the stimulus - either a block-cue or a trial-stimulus. The model automatically attends to this information before transferring the chunk representation of the display to the imaginal buffer. In the case of a cue, the model updates the goal buffer to represent that the target of future retrieval requests is the block-cue "kind", and then waits until a new visual display - automatic buffer harvesting ensures that the chunk representing the cue is entered into declarative memory. In the case of a stimulus, after the chunk representation is loaded into the imaginal buffer, the model attempts to retrieve a chunk to compare against the stimulus by making a retrieval request specifying the category of the stimulus and the block-cue "kind". If retrieval is successful, the model proceeds to determine if the stimulus identity represented by the chunks in the imaginal and retrieval buffers are matched. If so, it responds that the current stimulus is a target; otherwise, nontarget. In some cases, the retrieval process may not complete before the trial ends. If so, the model detects the presentation of the ITI and interrupts the ongoing retrieval attempt through a secondary retrieval request. A flowchart depicting the strategy of the model can be found in Figure 1.

The behavior of the model in Figure 1 ultimately depends on the parameters that influence memory retrieval. In ACT-R, retrieval is affected by a memory's *activation*, $A(m)$, which is the sum of a base-level term $B(m)$ and a contextual spreading activation $S(m)$. $B(m)$ is the log sum of the decaying traces of previous uses of m:

$$B(m) = \log \sum_i t_i^{-d}$$

where t_i is the time elapsed from the i-th time m was used and d is the decay rate. The spreading activation is defined as an additional boost coming from the information stored in a buffer:

$$S(m) = \sum_b \sum_j (W_b/N)\, s_{i,m}$$

where W_b is the amount of activation spreading from buffer b and $s_{i,m}$ is the association between slot i in buffer b and memory m. In our model, two such sources of activation exist, one for the goal buffer W_g and one for the imaginal buffer W_i. The strength of association $s_{i,m}$ is computed through a function which returns a scalar integer value equal to the number of source chunks j contained in chunk i; 1 if chunks j and i are identical; and 0 otherwise. Task accuracy depends on both the availability of a memory and the probability of unintentionally retrieving a wrong item; the latter is controlled by a partial matching similarity parameter c that determines the penalty between two slots. Thus, chunks that do not match the retrieval specification are penalized, but can still be retrieved.

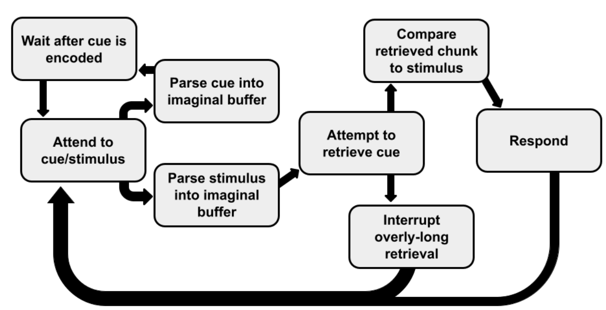

Figure 1. Flowchart of the ACT-R model strategy for performing the zero-back condition of the n-back task.

Finally, the relationship between the activation of a memory and the time *RT* it takes to retrieve is given by the equation:

$$RT(m) = Fe^{A(m)}$$

In summary, the dynamics of the model depend on five parameters: d, W_i, W_g, F, and c.

Individual-specific Estimation of ACT-R Parameters

Five model parameters were fit to individual participant behavior on an idiographic basis: d, W_i, W_g, F, and c. These parameters were chosen as they each have a strong effect on the model's response time or accuracy, the two participant measures that the model was fit against. The parameters d, W_i, W_g, and c all influence the likelihood that the correct cue-chunk is retrieved to be compared against (with additional minor influence on the RT due to changes in the activation level of the retrieved chunk), while the F parameter largely affects the retrieval time of the retrieved chunk (and therefore, the response time of the model). Parameters were only estimated for participants who demonstrated greater-than-chance performance on the target, lure, and non-target/non-lure conditions. To perform the fitting process, the *optimize.minimize()* method of the Python *scipy* package was utilized to minimize the RMSE between a set of participant measures and the commensurate model measures through the minimization function's *Powell* method. Bounds were placed on the five parameters (c: (-1,1); W_g: (0,2); W_i: (0,2); F: (1,3.5); d: (0.2,0.8)) to ensure that the minimization function remained within either reasonable or required ranges for these parameters. To compute a single RMSE across both RT and accuracy, these measures were placed on the same scale by dividing the trial-by-trial model and participant response times by 2 (as the maximum allowable RT by the task was 2 s). Missing RTs for both model and participant were replaced with the corresponding nan-meaned RT. Additionally, as the binary trial-by-trial accuracy outcomes had the potential to be exceedingly punishing to the model-fitting process, the aggregate block-wise and condition-wise (target/lure/non-target, non-lure) accuracies were used instead. Model and participant trial-by-trial scaled RTs and block-wise/condition-wise accuracies were then vectorized in order to compute the RMSE. Once the minimization algorithm converged to parameter estimates for each participant, model predictions were produced by running the model 100 times for each set of participant parameters, and then first taking the trial-by-trial average of the predicted RTs and accuracies over these runs before determining the average RT and accuracy for each participant.

Evaluation of Parameter Estimates

To evaluate the relative importance of each of the estimated parameters to the predictive efficacy of the model, a "decremental leave-one-out" (dLOO) procedure was applied. In this procedure, a set of models utilizing a subset of the estimated parameters are first produced from the full parameter set n by applying *n choose k*, where $k = n-1$. For each participant and each model in this set, the k chosen parameters are set to the participant's estimated values, while the "left out" parameter is set to the mean of that parameter's estimates (across participants). Model predictions are produced for each of the models in this set (as described above), and the R^2 between model predictions and participant measures are determined for both RTs and accuracies. The model with the largest mean R^2 (across RTs and accuracies) is determined to be the "best-fitting" model in this set, and the parameter that was "left out" of this model is "decremented" from the set of parameters. This procedure is then repeated for the remaining parameters, with both the "left-out" parameters and the "decremented" parameters set to the mean of that parameter's estimates, until only a single parameter remains.

Brain Parcellation

To calculate functional connectivity, each participant's brain was divided into discrete regions using a parcellation proposed by Power et al (2011). Although other parcellations have been proposed, this parcellation is notable for including both cortical and subcortical regions (see also Cole et al., 2016).

Statistical Learning Model

To identify the optimal combination of functional connectivity measures that reliably predicts individual parameters, resting-state functional connectivity was analyzed using a Lasso regression, a statistical learning method that combines feature selection and parameter fitting (Tibishirani, 1996). As a variant of linear regression, Lasso results remain interpretable in terms of beta weights that linearly scale a set of regressors. Unlike linear regression, Lasso reduces the complexity of the model by adding a penalty term that reduces to zero the weight of unnecessary variables, dramatically reducing the number of regressors provided. This feature is crucial for high-dimensional data such as the set of connectomes associated with a group of participants.

While in canonical linear regression the weights β are obtained by minimizing the quantity $||\mathbf{y} - \beta\mathbf{X}||_2$ (where the notation $||\mathbf{v}||_n$ represents the $L(n)$ norm of a vector $\mathbf{v}$), in Lasso the quantity to minimize includes a penalty term:

$$\beta = \text{argmin}(||\mathbf{y} - \beta\mathbf{X}||_2 + \lambda||\beta||_1)$$

The value of λ represents the tradeoff between model simplicity (captured by the first-order $||\beta||_1$ penalty) and accuracy (captured by ordinary least squares minimization term $||\mathbf{y} - \beta\mathbf{X}||_2$). When $\lambda = 0$, the model reduces to canonical linear regression. As λ grows, however, more and more regressors are eliminated to satisfy the constraints.

Results

Participant Task Performance

Participants who did not achieve greater-than-chance performance (binomial test) on the target, lure, or non-target/non-lure condition were not included in the analysis. This resulted in the exclusion of 36 out of 178 participants. Mean response times and accuracies were calculated for each participant for whom parameters were estimated. Mean RT was 0.78 ± 0.10 s, while mean accuracy was 0.94 ± 0.06. The distribution of RTs and accuracies across participants can be seen in Figure 2.

Idiographic Parameter Estimation and Prediction

For each participant, the set of parameter values that minimized the RMSE between trial-by-trial RTs and block-wise/condition-wise accuracies were estimated. The estimated cue-stimulus similarities c had a mean value of -0.43 ± 0.22, with a range of (-0.88, 0.19). While the majority of participants were found to have a negative c value, the c value for 2% of participants was estimated as slightly positive. The goal buffer spreading activation value W_g estimates had a mean of 0.93 ± 0.33 and a range of (0.18, 1.74), while the imaginal buffer spreading activation value W_i had a mean of 0.68 ± 0.34 and a range of (0.06, 1.71). As W_g and W_i provide a complementary but opposing influence on the retrieval process in this model (except in the case of target stimuli, for which they both promote the retrieval of the correct cue chunk), the difference between these two parameter estimates (W_g - W_i) was examined. The mean difference was 0.24 ± 0.51, with a range of (-0.85, 1.38). Over 70% of participants were estimated to have a W_g value greater than their W_i value, indicating that overall, information in the goal buffer drove the retrieval process. The mean of the latency factor F estimates was 2.49 ± 0.42, with a range of (1.31, 3.29).

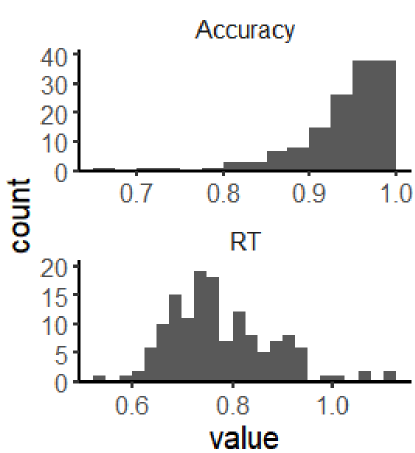

Figure 2. Histograms of accuracies and response times across participants in the zero-back condition of the HCP n-back task.

While F linearly affects retrieval times (and, by extension, response time) and the mean of the estimates was greater than the maximum allowable response time, the magnitude of this parameter compensates for retrieval time speeding caused by the influence of spreading activation and partial matching. Finally, for the decay-rate parameter d, the mean estimated value was 0.53 ± 0.10, and the range was (0.29, 0.71).

Once parameters for each participant were estimated, model predictions of participant performance were produced. Across predicted participants, the model's mean RT was 0.64 ± 0.09, and the model's mean accuracy was 0.89 ± 0.08. Individual participant RTs and predicted RTs were strongly correlated (r = 0.56, p < 0.001), while participant accuracies and predicted accuracies were moderately correlated (r = 0.23, p < 0.01). Scatterplots of participant measures versus predicted measures can be seen in Figure 3.

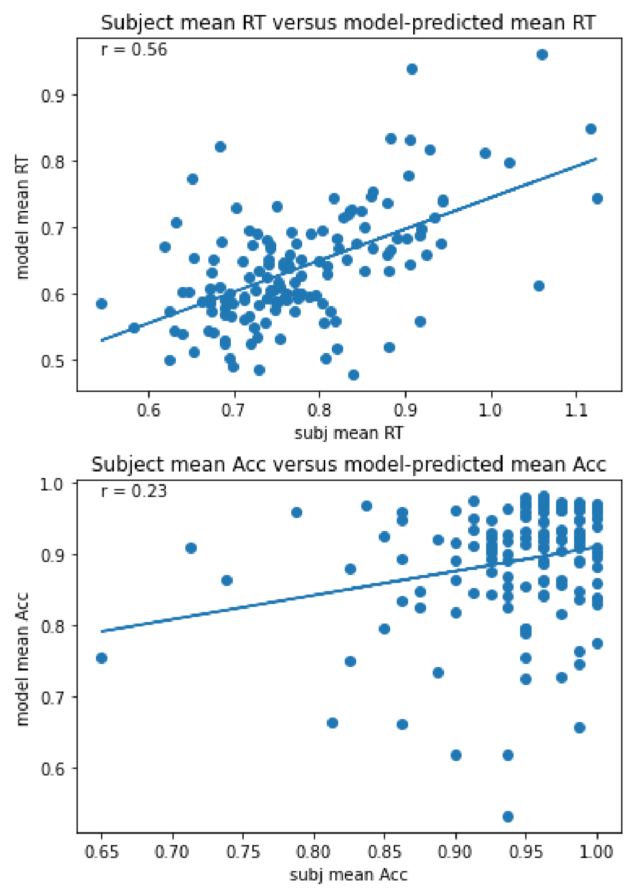

Figure 3. Scatterplots of participant mean RTs/accuracies versus model-predicted mean RTs/accuracies. Pearson's r between participant measures and model predictions shown in the upper left.

Decremental Leave-one-out Procedure

To determine which parameters contributed most strongly to the model's ability to predict individual participant RTs and accuracies, the parameter set was subjected to a "decremental leave-one-out" procedure. In the first round (five sets of four out of five parameters included, one parameter in each set assigned to the mean estimated value), it was found that the model containing the individual predictions of W_g, W_i, F, and d parameters had the largest mean R^2 (mean R^2 = 0.21; RT R^2 = 0.32; accuracy R^2 = 0.11); consequently, the c parameter was "decremented". In the second round, the model containing the W_g, W_i, and F parameters was the strongest predictor of participant behavior (mean R^2 = 0.21; RT R^2 = 0.33; accuracy R^2 = 0.09); the d parameter was dropped. In the third round, the model that included the W_g and W_i parameters was the most successful (mean R^2 = 0.18; RT R^2 = 0.27; accuracy R^2 = 0.09), and in the final round, the model including only the W_g parameter was the most predictive (mean R^2 = 0.18; RT R^2 = 0.24; accuracy R^2 = 0.11).

rs-fMRI Prediction of Individual-specific W_g

For each individual, we extracted a matrix of functional connectivity by calculating the Pearson correlation coefficient of each pair of the 264 x 264 regions in the Power (2011) parcellation. The group-level average of the individual correlation matrices, known as the *connectome*, was then visually inspected for comparison with similar functional connectivity studies. The connectome was found to be consistent with previous findings using the same parcellation scheme (compare, for example, to Cole et al. 2016). Because correlations between pairs of regions tend to be partially driven by common, unobserved factors (such as motion and physiological noise), the matrices were re-calculated using partial correlations (Cole et al., 2016), so that correlations between each region in the pair and the remaining 262 regions were partialled out. The resulting mean connectome is a much more sparse matrix (Figure 4B) and includes both negative and positive correlations (as expected from the spontaneous dynamics of brain activity: Fox et al., 2005).

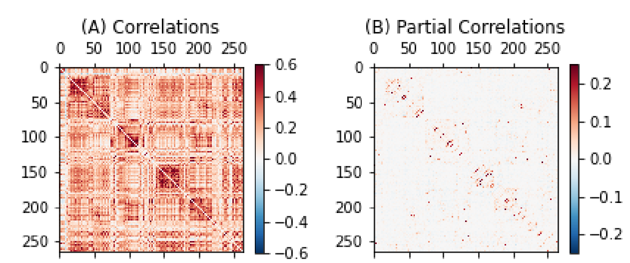

Figure 4: (A) Raw correlations between each of the 264 regions ; (B) Partial correlations between the same regions. In each matrix, rows and columns are ordered by network.

Each participant's sparse correlation matrix was then reshaped into a row vector of (264 x 263) / 2 - 264 = 34,452 elements. The number of possible regressors was further reduced by excluding connectivity measures related to three irrelevant networks (the Auditory, Cerebellar, and "Uncertain" networks in Power et al., 2011).

Lasso Fit and Cross-Validation

A cross-validation procedure was used to find the optimal value of λ. A sequence of possible λ values was generated, and, for each value, the performance of the Lasso algorithm in predicting the parameter W_g on a per-participant basis was measured using leave-one-out validation (LOOV). In LOOV, the algorithm is run 142 times, each time leaving out a different participant as the test set while the β values are fit to the remaining 141 participants as the training set. The mean error in predicting the parameter W_g for the left-out participant was then measured for all values of λ, and the value of λ that produced the smallest cross-validation error across all participants was chosen.

Resulting Connectivity

At the optimal level of λ, only 19 functional connections were left with a $\beta > 0$, involving a total of 36 brain regions from eight different functional networks. These connections and their regions are shown in Figure 5.

Notably, this list of regions includes the four ROIs in the Power parcellation that span the anterior cingulate cortex (ACC), corresponding to ACT-R's goal buffer (Anderson, et al. 2008). The list also includes five regions in the salience network, a set of regions involved, like the ACC, in the top-down control of attention. The functional connectivity values that best predict individual values of the W_g parameter include connections between the salience network and the default mode network, which is known to correlate with long-term memory function, and the sensorimotor network, including motor regions corresponding to the right hand.

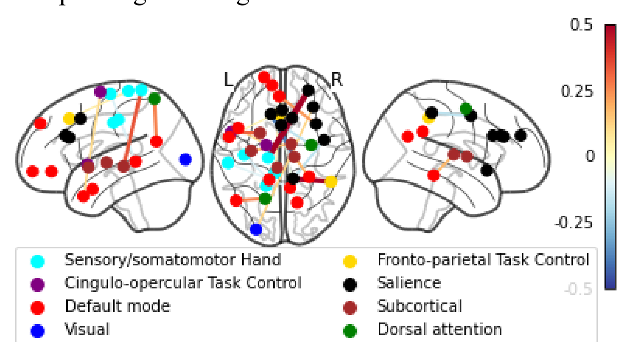

Figure 5: Functional brain connections predictive of individual rates of forgetting. Colored edges between nodes represent functional connectivity between the connected regions, while colors of nodes represent the network each region belongs to, using the Power et al (2011) scheme.

Together, connectivity between a group of networks including the salience network, the default mode network, and the sensorimotor network comprised a majority of the identified connections (16 out of 19, $\chi^2(1) = 15.474$ $p < 0.001$) and regions (24 out of 36, $\chi^2(1) = 10.667$, $p < 0.005$), significantly greater that what could be expected by chance.

As a final examination of this connectivity, we determined how well the parameter W_g can be recovered from functional connectivity alone. To do so, we multiplied each individual-specific set of functional connectivity values by the beta weights produced by Lasso, and compared them to the values inferred from the behavioral data by the ACT-R model. The predicted and observed values had a correlation of $r(142) = 0.775$, $p < 0.001$ (Figure 6).

Discussion

In the present study, idiographic parameterization of working memory function was investigated through the application of an ACT-R model. Values of five different parameters capturing various aspects of cognitive functioning were estimated for each participant through minimization of the RMSE between participant behavior and parameterized model predictions. A rank-ordering of the importance of these five parameters to the predictive efficacy of the model was determined through a "decremental leave-one-out" procedure, demonstrating that the goal-buffer spreading activation parameter W_g was critical to the model's predictive ability. Furthermore, it was shown that this essential parameter is predicted by an individual's resting-state functional connectivity.

This work makes it clear that ACT-R parameter estimates are capable of producing quality predictions regarding individual-level behavior. The correlation between participant RTs and model-predicted RTs was strong; while the predictions of accuracy were somewhat weaker, this can be partially attributed to the fact that the model was fit to the block-wise and condition-wise accuracies, instead of trial-by-trial accuracies (as RTs were). While this approach avoids the overly-punishing nature of RMSEs computed on binary outcomes, it reduces the amount of information the minimization algorithm has to fit the participant accuracy, relative to the participant RT. This effect was also apparent in the "decremental leave-one-out" procedure; the RT R^2 measures were overall larger than the accuracy R^2 measures across the iterations of the procedure. This procedure rank-ordered the "importance" of the five parameters to the model's predictive efficacy ($c < d < F < W_i < W_g$), and made it clear that W_g was by far the most valuable parameter for this model to predict individual differences in behavior, as both RT and accuracy R^2 measures changed negligibly as the other parameters were set to the mean values. As W_g was indicated to be the most crucial for individual prediction, it was chosen to be examined in relation to the participant's resting-state fMRI data.

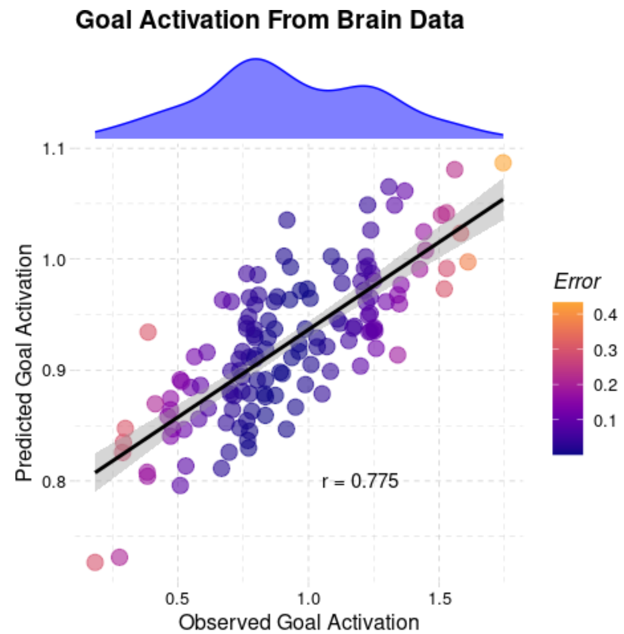

Figure 6: Correlation between observed values of the W_g parameter and the values of W_g predicted from functional connectivity (Figure 5).

A functional connectome for each participant was generated and subsequently used as the set of predictors for W_g in a Lasso regression. This resulted in the identification of a set of functional connections between regions inclusive of the salience, default mode, and sensorimotor networks as being maximally predictive of individual values of W_g. Moreover, this particular set of functional connections is entirely compatible with the putative role played by the goal buffer's spreading activation in the model, where it is used to assist in the retrieval of the correct cue from *long-term memory* (compatible with the salience-default mode connections). The result of this work would allow for the prediction of individual-specific W_g parameter values on individuals for whom resting-state measures exist, and through the ACT-R model, prediction of their behavior in a task environment.

In conclusion, this work exemplifies the potential of utilizing ACT-R modeling in conjunction with neuroimaging measures for the identification and prediction of signatures of cognitive functioning on an individual basis. Potential future efforts in this area of work include identification of a maximally predictive subset of parameters for each individual, as well as determination of resting-state nodes and functional connections that allow for the prediction of these parameters.

Acknowledgements

This research was supported by grant FA9550-19-1-0299 from the Air Force Office of Scientific Research to AS.

References

Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in Cognitive Sciences*, *12*(4), 136-143.

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718-1726.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*, *25*(3), 315-353.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673-9678.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, *306*(5703), 1940-1943.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665-678.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & Wu-Minn HCP Consortium. (2013). The WU-Minn human connectome project: an overview. *Neuroimage*, *80*, 62-79.

White, C. N., Curl, R. A., & Sloane, J. F. (2016). Using decision models to enhance investigations of individual differences in cognitive neuroscience. *Frontiers in Psychology*, *7*, 81.