Computational models with electrophysiological input predict behavior on a numerical cognition task

{ Draft Version 11.4.2016 }

Richard W Prather & Sara Heverly-Fitt

Abstract

The design of cognitive interventions is limited by the ability to predict participants' behavior and how it may change with experience. We present a novel computational model design that uses behavioral and electrophysiological input to predict participants' behavior on individual trials of a numerical comparison task. We model participants' behavior using independent model instantiations that are optimized for each participant using an evolutionary algorithm. The model is also then able to generalize to novel participant data without a significant reduction in prediction accuracy. We discuss both the potential and limitations of the current paradigm in developing training regimens for children with early math difficulty.

Introduction

The current interest in interventions and training paradigms includes areas of cognition such as language learning, memory, numerical cognition. The goal of such interventions typically is to improve learners' performance on a target task (Park, Bermudez, Roberts, & Brannon, 2016; Park & Brannon, 2014; Prather & Alibali, 2011), or in the case of older adults prevent decline (e.g., Anguera et al., 2013). There has been an increase in the automaticization (e.g., Iuculano et al., 2015) and individualization of interventions(e.g., Cohen Kadosh, Dowker, Heine, Kaufmann, & Kucian, 2013; Vanlehn, 2011). The design of effective interventions requires both accurate assessment of the learner's behavior and prediction of how that behavior may change with experience.

The current study is an evaluation of predictive computational model of learners' behavior on a numerical task. The long-term goal is to create individualized training procedures that are designed by the computational model. A significant limitation in intervention design is that there is large individual variation in participant behavior. A particular training sequence or task that is effective for participants of average skill may not be for participants with low-numeracy. As an initial step we developed a computational method for single-trial predictions of participants' behavior. We focus on evaluating the canonical non-symbolic numerical comparison task (Figure 1). Participant's skill in this task is used to index their *numeracy*; a construct that has been shown to correlate with important outcomes related to math performance (e.g., Libertus, Feigenson, & Halberda, 2013; Libertus, Odic, & Halberda, 2012; Mazzocco, Feigenson, & Halberda, 2011). Participants' skill in non-symbolic numerical comparison is correlated with their skill on symbolic arithmetic (e.g., Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014). The use of non-symbolic numerical comparison as a training task that may transfer to symbolic arithmetic has mixed results (Park et al., 2016; Park & Brannon, 2013, 2014).

The current study combines the behavioral and electrophysiological data as input for a predictive computational model. We focus on the non-symbolic numerical comparison task, which has well catalogued behavioral and neural correlates. Comparisons in which the two numbers are relatively closer in value tend to have more errors and longer reaction times. For example, a comparison between 20 and 22 objects produces more errors than a comparison between 20 and 50 (for alternate see Ratcliff, Thompson, & Mckoon, 2015). This phenomenon is termed the *distance effect* and depends on the ratio difference between the two values being compared. The neural correlates of the comparison task, as seen with electroencephalography, include a *neural distance effect* where the amplitude of event related potentials vary based on the distance of the comparison(e.g., Ben-Shalom, Berger, & Henik, 2013; Heine, Tamm, Wissmann, & Jacobs, 2011; Jiang et al., 2010). Our approach for the model predicts the specific responses for each individual participant on each individual trial. This allows the model to account for individual differences in performance.

Experiment 1

First we attempted to replicate the ERP neural distance effect using a non-symbolic numerical comparison task.

Method

Participants. Participants (n = 26) were adults recruited through a university participant pool. Protocols were approved by the University of Maryland Internal Review Board.

Procedure. We evaluated participants' performance on the non-symbolic numerical comparison task while measuring neural correlates via EEG. Participants were instructed that they would complete a series of computer-based tasks while simultaneous EEG recordings were taken. After a brief set up to initialize the EEG participants began the behavioral tasks. Instructions for task were given orally and again written on screen. Participants' responses were via key press. The experimental session was approximately 45 minutes.

Task and Stimuli. Stimuli were 90 visually presented pairs of shape arrays with a midline separator. Shape arrays ranged in number from 23 to 111 (see Appendix B) with a maximum ratio difference of 2.6. Participants were instructed to indicate which side contained more shapes via button press. Stimuli were displayed for 2 seconds after which the screen was blank. There was no response time limit; participants were instructed to respond as quickly as possible.

Electrophysiological Recordings Specifications. Recording was implemented using a 32-channel EEG cap. Twenty nine tin electrodes were held in place on the scalp by an elastic cap 1 (Electro - Cap International, Inc., Eaton, OH) in a 10 - 20 configuration (O1, FP2, O2, P7, 2 P3, Pz, P4, P8, TP7, Cp3, CPz, CP4, TP8, T7, C3, Cz, C4, T8, FT7, FC3, FCz, FC4, FT8, 3 F7, F3, Fz, F4, F8, FP1). Based on prior research we limited analysis to channels in the left parietal area (P3, P7, CP3, TP7). Bipolar electrodes were placed above and below the left eye and 4 at the outer canthus of the right and left eyes to monitor vertical and horizontal eye 5 movements. Additional electrodes were placed over the left and right mastoids. Scalp 6 electrodes were referenced online to the left mastoid and re - referenced offline to the 7 average of left and right mastoids. Impedances were maintained at less than 5 k Ω for all 8 scalp electrode sites, less than 2 k Ω for mastoid sites, and less than 10 k Ω for ocular 9 electrodes. The EEG signal was amplified by a NeuroScan SynAmps® Model 5083 10 (NeuroScan, Inc., Charlotte, NC) with a bandpass of 0.05 - 100 Hz and was continuously 11 sampled at 500 Hz by an analog - to - digital converter.

Results

Behavioral results. Participant performance on the numerical comparison task ranged from 83% to 97% correct (Figure 2). Median reaction times ranged from 352ms to 1604ms. Correlation between trial ratio difference and percent correct was non-significant, $r = 0.30$, $t = 1.54$, likely due to a ceiling effect (Figure 2). Correlation between reaction time and trial ratio difference was $r = -0.13$, $p < 0.01$

Electrophysiological results. We evaluated if there was an ERP distance effect in parietal areas. Mean ERP amplitude within the time window of 280-380ms was extracted. Trials were binned into equal sized groups of large, medium or small comparison distances. We compared the mean ERP amplitude for large and small comparison distances. Data were analyzed using repeated measures ANOVA, with trial comparison distance being the within-subjects variable. Our analysis shows that when evaluating data from two parietal channels there is a significant difference in ERP amplitude between large and small comparison differences. We found a significant effect of ratio difference on mean ERP amplitude for the P3 and CP3, $F(1, 25) = 6.33$, $p = .018$ and $F(1, 25) = 4.58$, $p = .042$ respectively. This analysis of the neural distance effect is consistent with prior results in terms of EEG channel location, ERP time window, and trial comparison.

The experiment 1 results demonstrate a replication of the neural distance effect (e.g., Ben-Shalom, Berger, & Henik, 2013; Heine, Tamm, Wissmann, & Jacobs, 2011; Jiang et al., 2010). We find that event related potentials for two parietal channels (P3 and CP3) significantly varied for close numerical comparisons versus far comparisons.

Experiment 2

For experiment 2 we used the behavioral and neural data from experiment 1 as input for a computational model of participants behavior. We used the first two-thirds of trials (60) as training data for the model and the last third of trials (30) as a prediction test. *The key comparison is if model predictions during the test phase are as accurate as model fit during the training phase.* The computational model is a multilayered dynamic field theory model (e.g., Erlhagen & Schöner, 2002; Sandamirskaya, 2014) that employs neural tuning curves related to numerical cognition (e.g., Prather & Heverly-Fitt, 2016; Prather, 2012, 2014) along with an evolutionary optimization algorithm. The evolutionary algorithm adjusts specifications of the model dynamics to fit each participant's training data. Thus, each participant's data was modeled by a separate and independent instantiation of the computational model. The resulting model specifications were then used to predict the remainder of each participant's data without further use of the evolutionary algorithm. We evaluate the model first by how well the model fits participant's data in the training phase, then by if that model instantiation can extrapolate to the test phase data without further adjustment.

Method

Model Specifications and Procedure. The model was implemented using MATLAB (Mathworks). The architecture is that of a multilayered dynamic systems model (e.g., Simmering & Perone, 2013; Spencer, Smith, & Thelen, 2001). Layers included two perceptual neural tuning curves and a decision layer. Layers modeled neural tuning curves associated with neural coding of stimuli (e.g., Prather, 2012, 2014; Tudusciuc & Nieder, 2007). On each trial the external inputs for the model were the two numerical values to be compared, taken from the behavioral task in experiment 1. The two values were represented by proportionally scaled Gaussian curves that reproduce the ratio dependent distance effect. Perceptual layers of the model reproduced the stimuli while activity was forwarded to the decision layer. The internal decision layer connections were specified to produce competition within the layer through lateral inhibition and self-excitation. Thus the two perceptual layers output created a competition within the decision layer. This dynamic corresponded to the "decision" which was the index of the first stable activation peak in the decision layer. Each trial was comprised of 500 time steps. The decision layer produced a reliable decision through a steady peak (activity with a peak value at the same layer index for 10 straight timesteps). The timestep of the decision was converted to the predicted reaction time of the decision. Thus on trials in which the model predicted a fast decision the steady peak was reached a relatively low timestep. On trials in which the model predicted a slower decision the steady peak was reach on a higher timestep.

Use of electrophysiological data. We used ERP data as reported in experiment 1 as model input. Data was limited to the P3 channel as data indicated a significant correlation between ERP amplitude and trial ration difference (i.e. the neural distance effect). For each trial we calculated the mean ERP amplitude across the time window used in experiment 1. Trial amplitudes were used to predict incorrect responses, which were mostly for trials with ratio differences in the lower half of the range. Thus ERP amplitude only affected model parameters on trials with a ratio difference of 1.7 or less (44 of 90 trials).

Model Fitness Calculation and Evolutionary Algorithm Procedure. Modeling procedure included two parts, training phase and test phase. During training the model was fit to the first 60 trials of the participants data. Fit was optimized using an evolutionary algorithm. After determining the specifications of the model that best fit we then used that model to predict the entirety of participants data without further adjustments via the learning algorithmic. The model must generalize what is learned during the training phase to novel trials during the test phase. This process is done independently for each participant.

Specifications that varied for each model instantiation included parameters for noise, tuning curve width, rate of activation change, inhibition in connections. The first generation of the algorithm included 10 instantiations of the model with randomly selected values of the specifications (within a predetermined range). The 10 models are then ranked by the resulting fitness. The bottom 5 performing models were discarded. The best performing model was duplicated for the next generation. Models ranked 2 – 5 were *mutated*, such that their specifications were altered slightly by a random process. The next generation also included 5 new models with randomly selected specifications. This process was repeated over 25 generations to produce a model instantiation that fit the participants' data. Since the model has a small amount of random noise performance does not replicate exactly. To best determine the fitness level of a given set of model specifications we then ran the model repeatedly and calculate the mean fitness level. For the testing phase we used the best-fit specifications to predict the final half trials of the participants data, which the model had not been training with.

Model fitness is calculated using an overall error terms that combines the number of correctly predicted decisions, the deviation of the predicted reaction times and correlation between predicted and true reaction times. The overall fit of the model was weighted towards minimizing response prediction error via the following equation:

Results

Model data addresses two questions. First, how well does the model fit to participant data using the evolution algorithm. Second, how well do the model instantiations generalize to predict novel participant data? Analysis of model performance was measured by deviation between empirical human data and model data for both training and test phases. We calculated the proportion of trials in which the model response matched the participant response. The key comparison is between the proportion of matches the training and test phase. That is, if the specifications of the model learned from the training can be used to predict future performance with no reduction in accuracy.

For the training phase deviations between the model response and participant response, as measured by proportion of mismatches ranged from 0.26 to 0.10 with a mean of 0.16. (see Figure 3). For the test phase, deviations between the model response and participant response ranged from 0.26 to 0.05 with a mean of 0.159. For the final third of trials specifically the deviation range was 0.26 to 0 with a mean of 0.19. (see Figure 4).

We then compared model performance on training and test phases. For response prediction of trained versus novel data we used a non-parametric comparison (Mann–Whitney U test) between initial training fit and test. The model error during training phase was not significantly different than test $W = 312$, $p = 0.626$. We then compared fit on the training data to only the novel portion of the test phase (i.e. the final 30 trials). The model performance on the novel trials was significantly better than the training data fit, $W = 427$, $p < 0.01$.

General Discussion

This study demonstrates that a multilayers dynamic field computational model using behavioral and neural data input can accurately predict behavior for individual participants. Specifically the model instantiations are able to generalize to future behavior. Experiment 1 results include a replication of the ERP numerical distance effect. Experiment 2 results demonstrate that the computational model is both able to accurately fit individual participants' behavioral data and extrapolate to predict novel data. The current model, using independent instantiations and an optimization algorithm was able to fit to a range of participant behavioral data. Fit to participant data was accurate within an error rate of 0.10 for responses. Extrapolation to novel data was of similar accuracy to the initial model fit.

Limitations and Future Directions. Evaluating the computational model results requires caution. There are several questions to consider. There was variation in the model fit to training data. Why might some participants have data that is better fit by the computational model? It could be the case that participants with low variability data are better fit by the model. For example a participant that correctly responded to 100% of trials would be trivial to predict. Participants performance on the task might correlate with model fit, where better performing participants had better fit models. A correlation between participant performance and model fit was significant $R = -0.54$, $t(22) = 3.009$, $p < 0.01$. We then looked more specifically of model predictions of incorrect answers. Model performance in predicting incorrect responses of participants did not correlate with participant performance, $R = -0.22$, $t(22) = 1.05$, $p = 0.30$. The model's performance somewhat depends on the characteristics of the participant data. The initial model fit was best for participants with a high degree of accuracy. However model prediction of which trials the participants answered incorrectly was not correlated with participant performance. Additionally there was no significant correlation between participant performance and how well the model generalized to the novel as measured by the difference in training and test error, $R = 0.03$, $t(22) = 0.141$, $p = 0.88$.

It is unclear if there is a limitation to the number of trials the model could predict based on the training. There may be an in principal limitation to the number of trials the model can predict. It may also be the case that a larger training set for the model results in more accurate extrapolation. Perhaps a multiple visit experiment with a much larger number of trials would result in greater accuracy in predictions. The current study only involved a brief experimental session. Longer and potentially multiple visit experimental sessions could address this question.

We use the canonical non-symbolic numerical task here. This task is central to the study of numerical cognition and has shown to correlate with other tasks, which are more relevant to educational applications. The question remains is such a modeling approach could be extended to non-symbolic arithmetic, symbolic comparison or even symbolic arithmetic. A similar style neural tuning curve computational model has been used for other tasks such as symbolic number line and symbolic arithmetic (Prather 2012; 2014; 2016)

Conclusions. In this study we present a novel computational method for predicting human behavior on a numerical comparison task. The computational method uses both behavioral and neural data to make predictions of participants' future behavior on individual trials. The method may potentially be expanded to include other numerical, arithmetic or general cognitive tasks. The method may also be adjusted to predict a particular stimulus set that leads to improved performance on the current task or other general cognitive tasks. Such expanded and adapted use of the computational method could prove to be a significant step in bridging research on cognitive intervention design and individualized training regimens.

References

Anguera, J. a, Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., … Gazzaley, a. (2013). Video game training enhances cognitive control in older adults. *Nature*, *501*(7465), 97–101. http://doi.org/10.1038/nature12486

Ben-Shalom, T., Berger, A., & Henik, A. (2013). My brain knows numbers! - an ERP study of preschoolers' numerical knowledge. *Frontiers in Psychology*, *4*(October), 716. http://doi.org/10.3389/fpsyg.2013.00716

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychologica*, *148*, 163–72. http://doi.org/10.1016/j.actpsy.2014.01.016

Cohen Kadosh, R., Dowker, A., Heine, A., Kaufmann, L., & Kucian, K. (2013). Interventions for improving numerical abilities: Present and future. *Trends in Neuroscience and Education*, *2*(2), 85–93. http://doi.org/10.1016/j.tine.2013.04.001

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, *109*(3), 545–572. http://doi.org/10.1037//0033-295X.109.3.545

Fazio, L. K., Bailey, D. H., Thompson, C. a, & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. http://doi.org/10.1016/j.jecp.2014.01.013

Heine, A., Tamm, S., Wissmann, J., & Jacobs, A. M. (2011). Electrophysiological correlates of non-symbolic numerical magnitude processing in children: joining the dots. *Neuropsychologia*, *49*(12), 3238–46. http://doi.org/10.1016/j.neuropsychologia.2011.07.028

Iuculano, T., Rosenberg-Lee, M., Richardson, J., Tenison, C., Fuchs, L., Supekar, K., & Menon, V. (2015). Cognitive tutoring induces widespread neuroplasticity and remediates brain function in children with mathematical learning disabilities. *Nature Communications*, *6*, 8453. http://doi.org/10.1038/ncomms9453

Jiang, T., Qiao, S., Li, J., Cao, Z., Gao, X., Song, Y., … Chen, C. (2010). Effects of symbol type and numerical distance on the human event-related potential. *Neuropsychologia*, *48*(1), 201–10. http://doi.org/10.1016/j.neuropsychologia.2009.09.005

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is Approximate Number Precision a Stable Predictor of Math Ability? *Learning and Individual Differences*, *25*, 126–133. http://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3), 373–379.

http://doi.org/10.1016/j.actpsy.2012.09.009

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability (Dyscalculia), *82*(4), 1224–1237. http://doi.org/10.1111/j.1467-8624.2011.01608.x

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Journal of Experimental Child Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, *152*, 278–293. http://doi.org/10.1016/j.jecp.2016.07.011

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*(10), 2013–9. http://doi.org/10.1177/0956797613482944

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, *133*(1), 188–200. http://doi.org/10.1016/j.cognition.2014.06.011

Prather, R. W. (2012). Connecting neural coding to number cognition: a computational account. *Developmental Science*, *15*(4), 589–600. http://doi.org/10.1111/j.1467-7687.2012.01156.x

Prather, R. W. (2014). Numerical discrimination is mediated by neural coding variation. *Cognition*, *133*(3), 601–610. http://doi.org/10.1016/j.cognition.2014.08.003

Prather, R. W., & Alibali, M. W. (2011). Childrens Acquisition of Arithmetic Principles: The Role of Experience. *Journal of Cognition and Development*, *12*(3), 332–354. http://doi.org/10.1080/15248372.2010.542214

Prather, R. W., & Heverly-Fitt, S. (2016). Neural coding partially accounts for the relationship between children's number-line estimation and number comparison performance. *Frontiers in Developmental Psychology*.

Ratcliff, R., Thompson, C. A., & Mckoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition*, *137*, 115–136. http://doi.org/10.1016/j.cognition.2014.12.004

Sandamirskaya, Y. (2014). Dynamic neural fields as a step toward cognitive neuromorphic architectures. *Frontiers in Neuroscience*, *7*(January), 1–13. http://doi.org/10.3389/fnins.2013.00276

Simmering, V. R., & Perone, S. (2012). Working memory capacity as a dynamic process. *Frontiers in Psychology*, *3*(January), 567. http://doi.org/10.3389/fpsyg.2012.00567

Spencer, J. P., Smith, L. B., & Thelen, E. (2001). Tests of a Dynamic Systems Account of the A-not-B Error : The Influence of Prior Experience on the Spatial Memory Abilities of Two-Year-Olds, *72*(5), 1327–1346.

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Neuron*, *104*(36).

Vanlehn, K. (2011). The Relative Effectiveness of Human Tutoring , Intelligent Tutoring Systems , and Other Tutoring Systems The Relative Effectiveness of Human Tutoring , Intelligent Tutoring Systems , and Other Tutoring Systems, (August 2014), 37–41. http://doi.org/10.1080/00461520.2011.611369

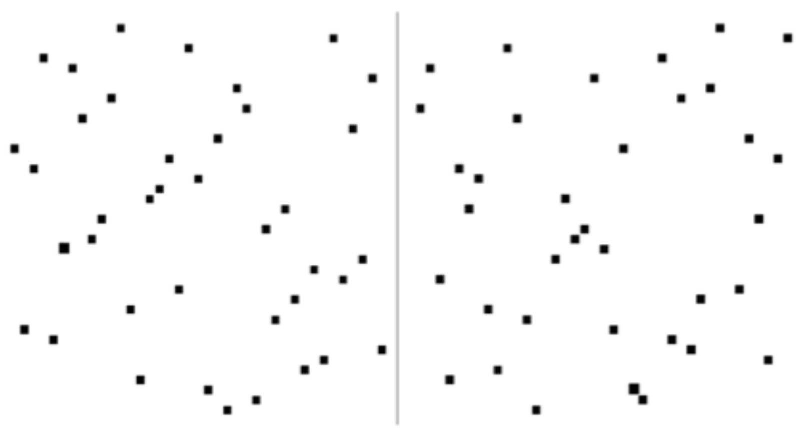

Figure 1. Example stimuli for a non-symbolic numerical comparison task. Participants are instructed to indicate which side of the mid-line has a larger number of objects. The stimulus is displayed for a limited amount of time to discourage counting.

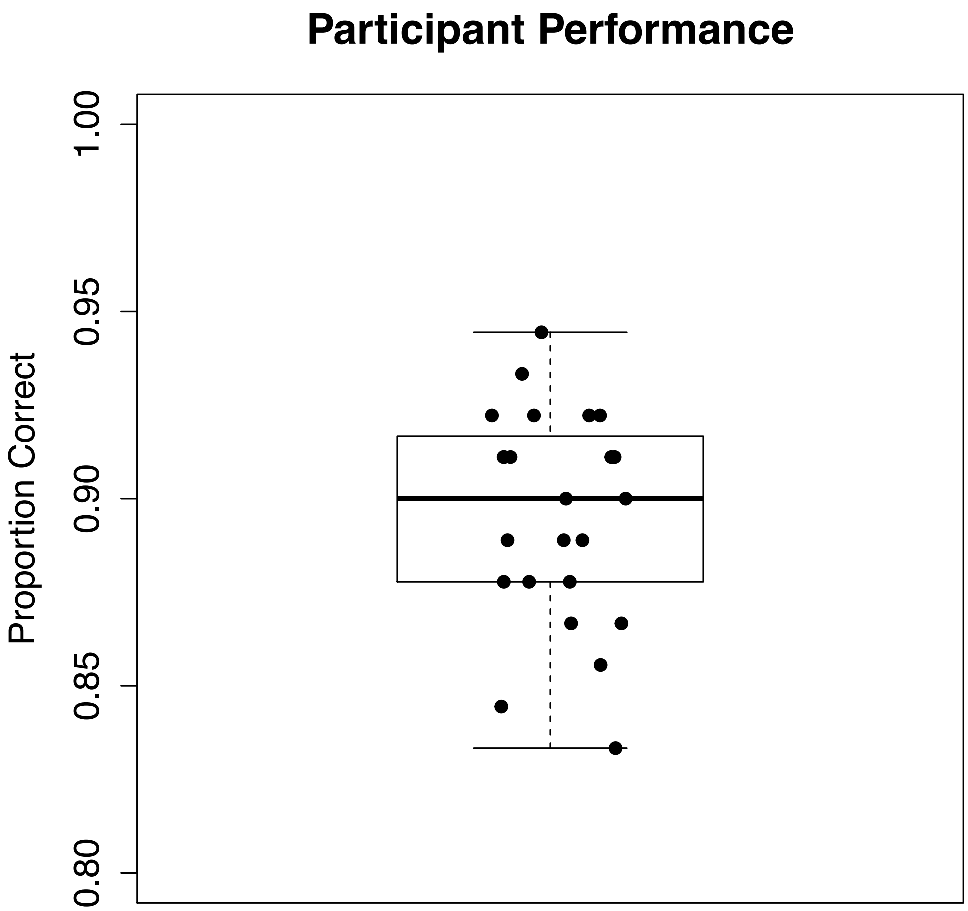

Figure 2. Participant performance on the numerical comparison task as proportion of trials correct.

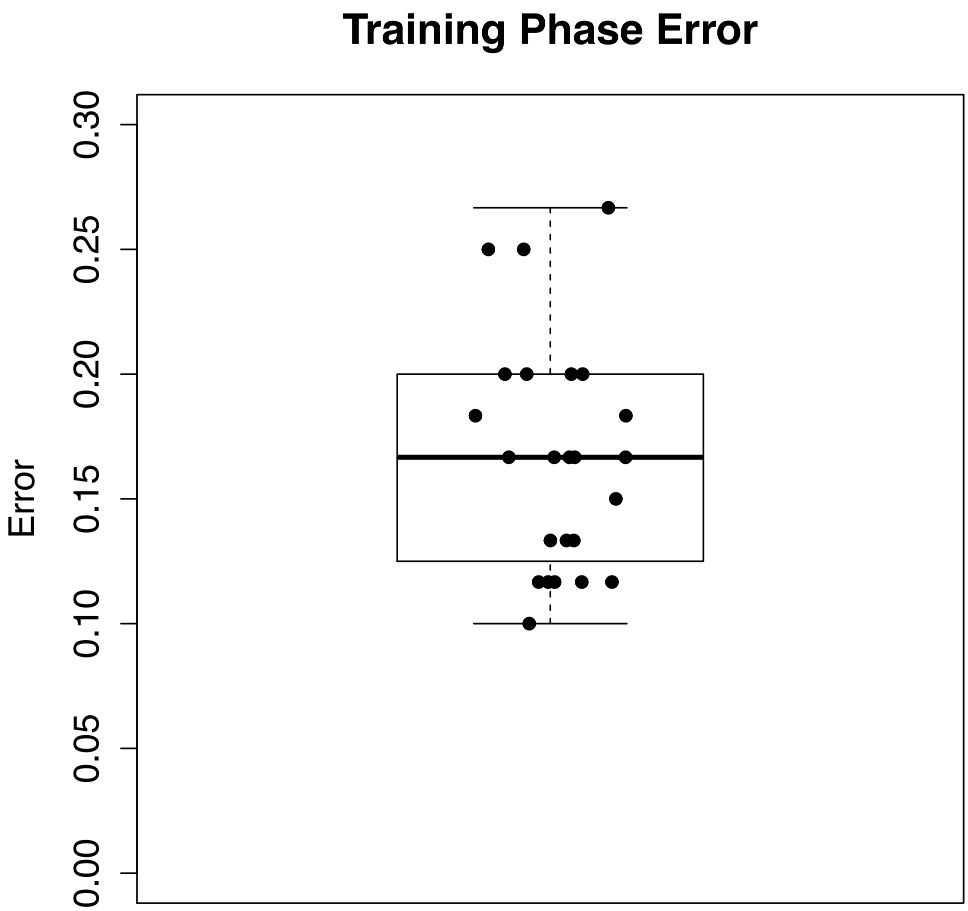

Figure 3. Model Error for the training phase in terms of proportion of mismatches.

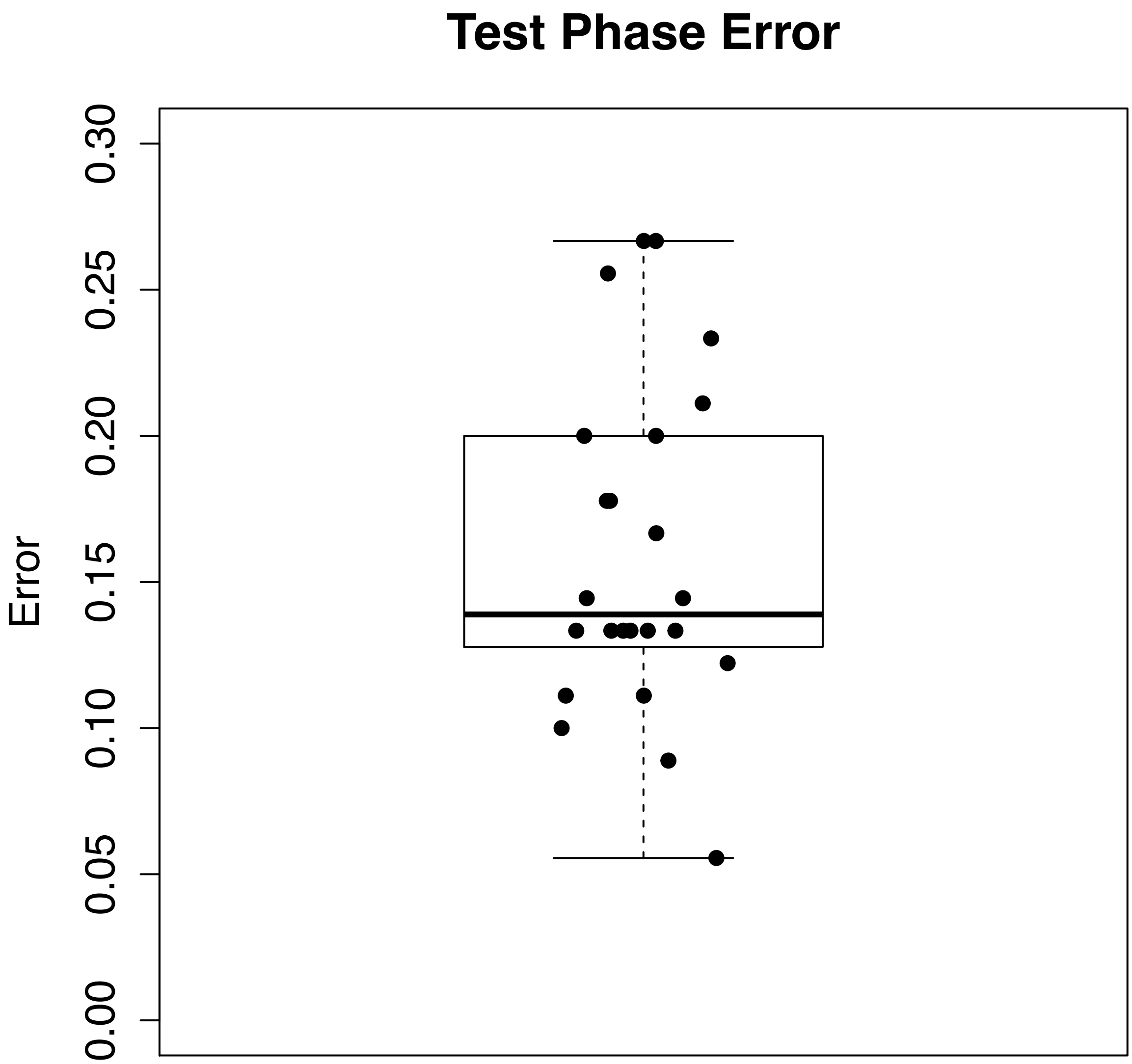

Figure 4. The mean correct response and reaction time deviation for the model for each participant. These are the model predictions from the testing phase.

Appendix A

Stimuli List

25	35
50	75
25	30
40	42
25	65
35	91
25	35
23	23
14	26
19	42
20	22
13	29
30	30
10	18
15	21
50	75
15	35
10	19
15	35
29	70
19	44
10	20
15	33
30	51
15	21
14	14
24	60
30	45
10	19
23	50
29	70
24	60
20	30
30	39
30	33
10	15
35	56
22	54
30	33
20	32
10	18
14	28
20	52
30	30
15	33
10	25
20	52
22	54
20	38
36	87
20	32
30	51
14	19
15	18
30	39
20	34
20	30
10	20
40	42
35	56
10	13
20	38
16	40
20	34
10	25
14	19
30	45
35	91
25	30
40	60
10	11
23	50
10	15
36	87
19	42
45	54
14	28
15	18
16	40
19	44
10	13
25	65
40	60
20	22
23	23
13	29
45	54
14	26
14	14
10	11