## Supplemental Figures

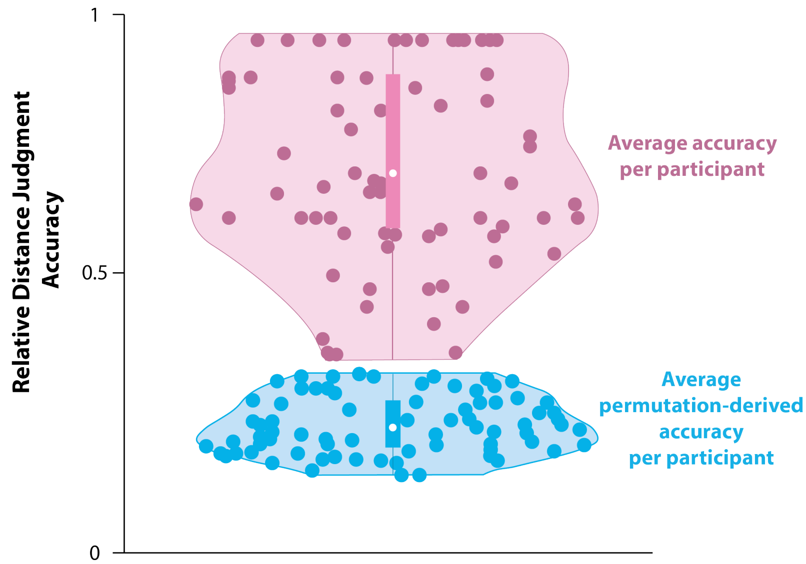


**Figure S1.** Distribution of (1) participant accuracy on graph reconstruction task ( the percentage of correctly identified edges) and (2) chance level (permutation-derived) accuracy for each subject. We estimated chance accuracy for each subject by randomly shuffling the graph 1000 times (such that connections between nodes differ from those in the ground truth graph), evaluating the accuracy of connected edges based on the adjacency matrix of the shuffled graph and averaging accuracy values in each subject. The edge connections based on the shuffled graph are a proxy of how accurate subjects would be if they were guessing at random while drawing the edges between nodes. Our results show that participants' edge connection accuracy evaluated based on the ground truth adjacency matrix is significantly greater than the chance accuracy based on the randomly shuffled graphs, confirming that on average participants were not randomly configuring the edges.

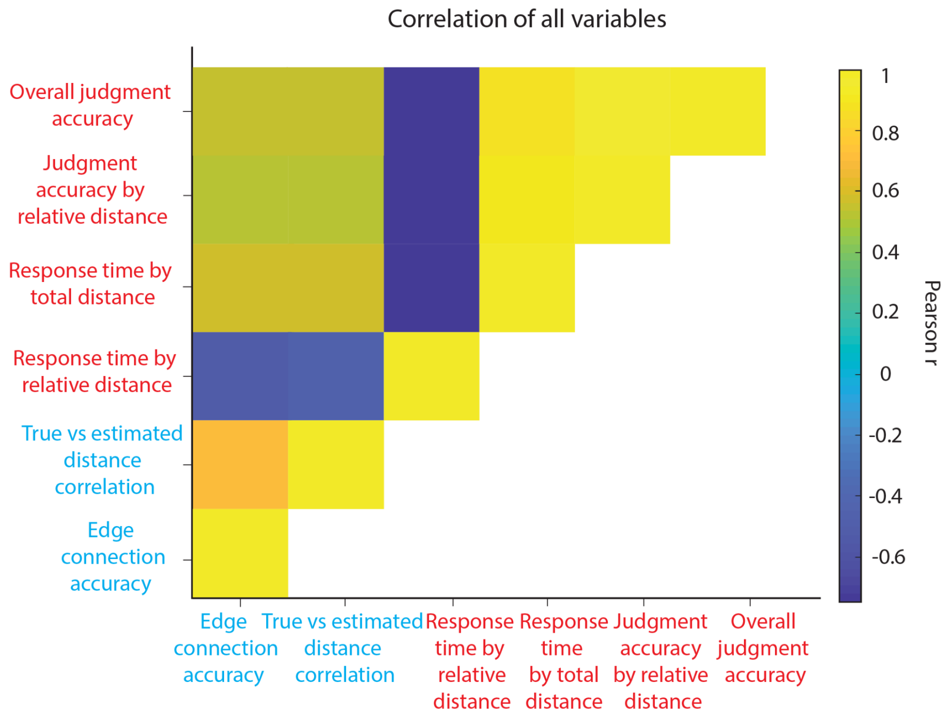


**Figure S2.** Pairwise correlations between all graph-task measures.

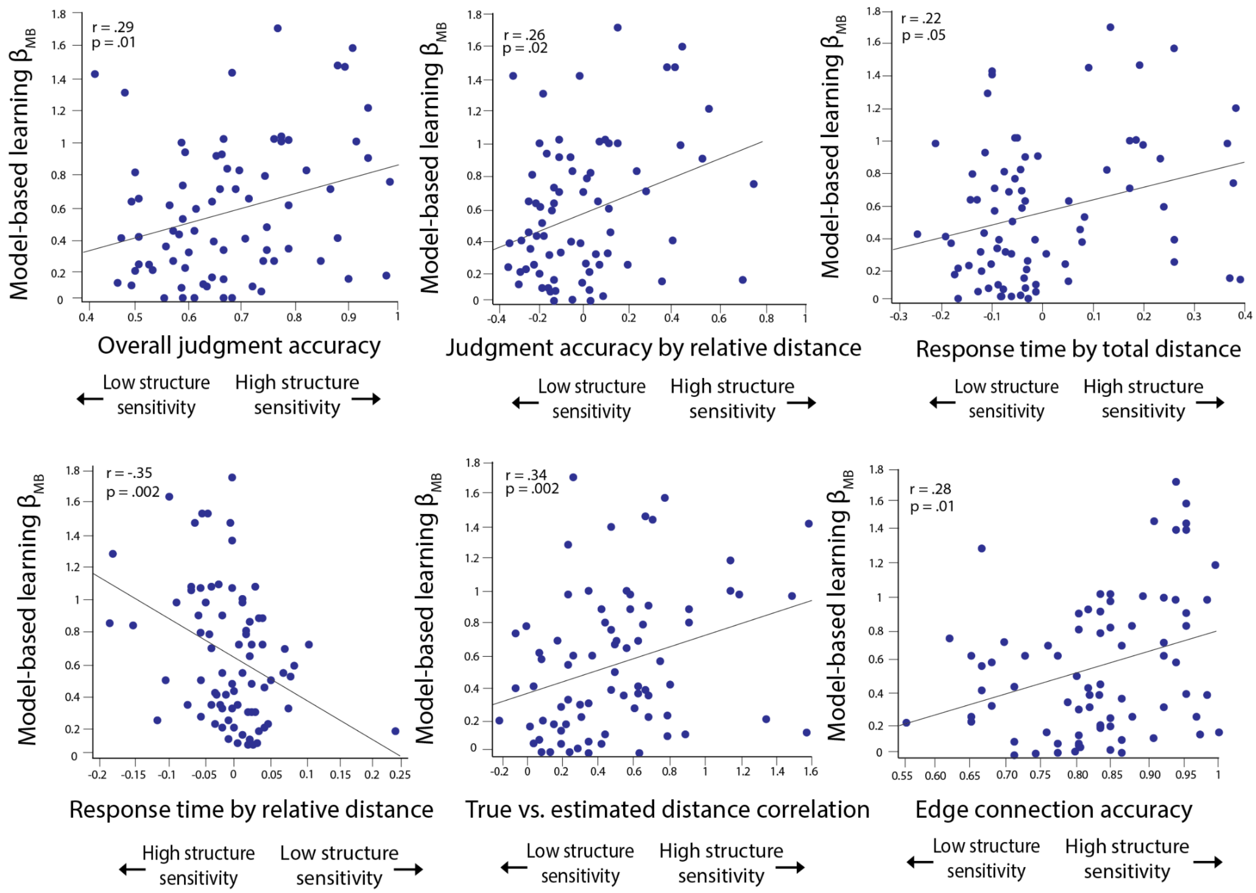


**Figure S3.** ***Model-based learning is associated with individual structure inference measures.*** The y-axis corresponds to the fit value of the model-based weighting parameter $\beta_{MB}$. The x-axis in the above figures plots the following: overall judgment accuracy, judgment accuracy by relative distance, response time by total distance, response time by relative distance, true vs. estimated distance correlation, edge connection accuracy.

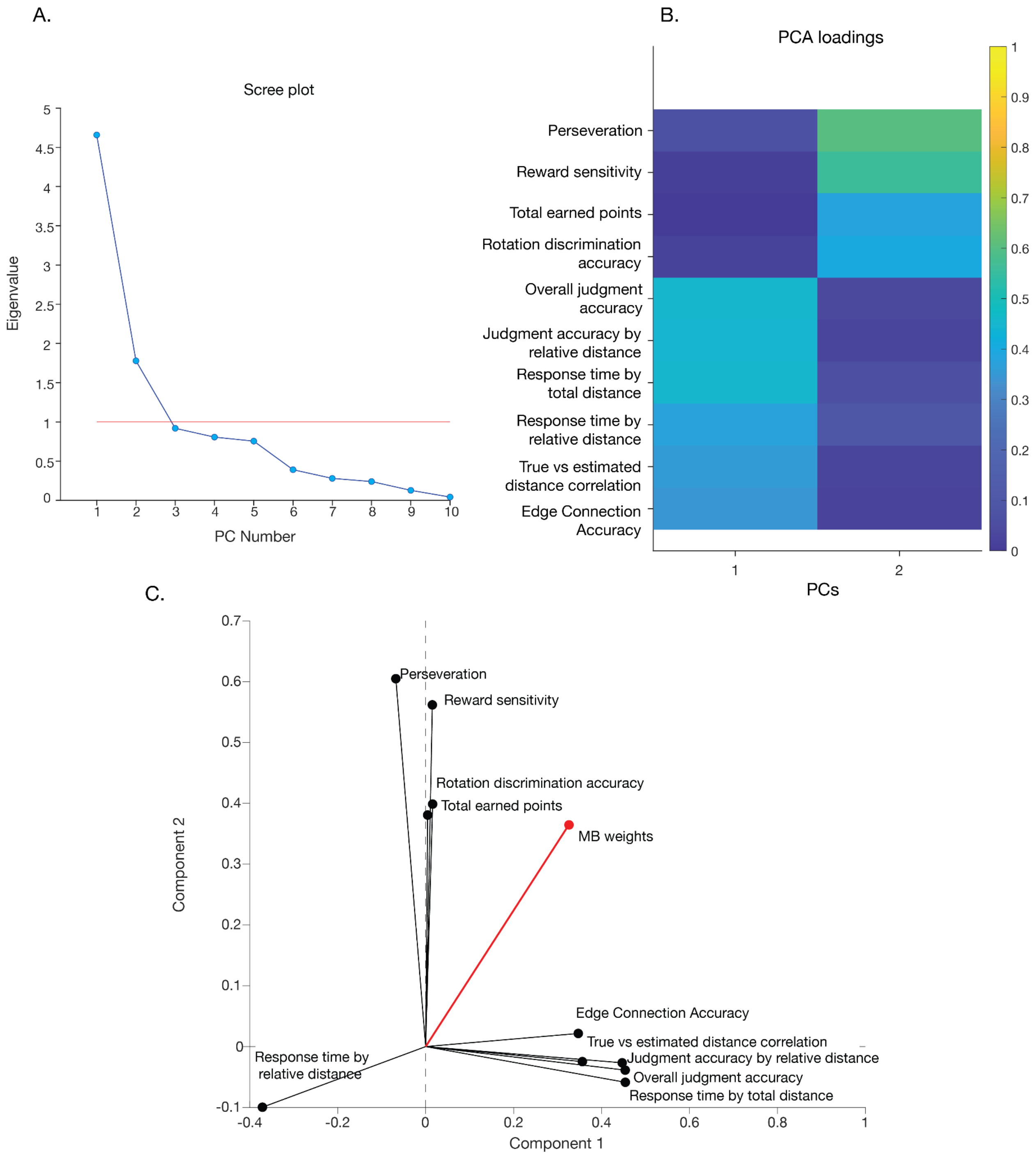


**Figure S4. Exploratory PCA on individual differences**. Our exploratory PCA (promax rotated) revealed that the data supported a two component model (A), with factors putatively related to structure learning (Component 1) and motivation (Component 2; B). (C) Our model-based index loaded on to both of these components.

**Supplemental Tables**

| | $\beta$(SE) | *T* | DF | *p* |
|---|---|---|---|---|
| **Transition type** | .10(.02) | 3.9 | 73 | .0001*** |
| **Reward** | .64(.04) | 14.38 | 73 | 2.9e-05*** |
| **Transition type * Reward** | .42(.04) | 9.17 | 73 | 1.5e-20*** |

**Table S1.** ***Modeling 1st-stage choices in the RL task as a function of model-free and model-based learning.*** Model statistics refer to the coefficients of the fixed main effect of reward, transition type and the reward-by-transition type interaction from the following model: *Stay ~ Reward × Transition type + (1 + Reward × Transition type* | *Participant)*. Here (SE) indicates the standard error of the mean (*p<.05; **p<.01; ***p<.001). Participants exhibit a mixture of model-free and model-based strategies, as shown by the significant main effect of reward and a reward-by-transition type interaction.

| | $\beta$(SE) | *T* | DF | *p* |
|---|---|---|---|---|
| **Estimated shortest path** | .42(.05) | 7.81 | 74 | 5.63e-11*** |
| **Euclidean/ distance** | .22(.02) | 8.36 | 74 | 1.22e-11*** |

**Table S2.** Predicting ground truth distance with Euclidean distance, controlling for reported shortest path. The Euclidean distance (estimated distance) predicts the ground truth distance over and above the reported shortest path. This suggests that the estimated distance model was not predictive of the ground truth simply as a function of edge connections participants drew during the reconstruction. DF here refers to Satterthwaite degrees of freedom approximation. (*p<.05; **p<.01; ***p<.001).

| | $\beta$(SE) | *T* | DF | *p* |
|---|---|---|---|---|
| **Model-based term random effects** | .49(.22) | 2.19 | 72 | .03** |
| **Perseveration percentage** | -.11(.14) | -.80 | 72 | .42 |

| | | | | |
|---|---|---|---|---|
| **Model-free term random effects** | -.04(.23) | -.17 | 72 | .86 |
| **Post-rare transition slowing random effects** | .06(.11) | .55 | 72 | .57 |

**Table S3.** Robust linear regression model predicting PC scores (latent component indexing structure inference) using different indices from the two-step task (percentage of perseverative response, model-free random effects, model-based random effects and post-rare transition slowing). Index of model-based planning is selectively significantly associated with the measure of structure inference. The beta coefficients here are estimated effect coefficients, SE is standard error of the mean. DF refers to error degrees of freedom. Positive terms indicate positive association with the structure inference measure (*p<.05; **p<.01; ***p<.001).

| | ***β(SE)*** | ***T*** | **DF** | ***p*** |
|---|---|---|---|---|
| Latent structure learning factor | .16 (.04) | 3.35 | 75 | .001** |

**Table S4.** ***Model-based weights from the computational model.*** The results from the robust linear regression model (*RL Model-based weights ~ 1 + structure inference score*)**.** The predictor in the model was z-scored. The results show that high PC scores (indexing high structure inference performance) predict increased model-based planning. (*p<.05; **p<.01; ***p<.001).

| **Structure inference measure** | ***β (SE)*** | ***T*** | **DF** | ***p*** |
|---|---|---|---|---|
| Overall distance judgment accuracy | .14 (.05) | 3.16 | 72 | .001** |
| Judgment accuracy by relative distance | .16 (.04) | 3.60 | 72 | .0003*** |
| Response time by total distance | .13(.04) | 2.95 | 72 | .003** |
| Response time by relative distance | -.16 (.04) | -3.69 | 72 | .0002*** |
| True vs estimated distance correlation | .18 (.04) | 3.98 | 72 | .00008*** |
| Edge connection accuracy | .12 (.04) | 2.67 | 72 | .007** |
| Latent structure learning factor | .17(.04) | 3.97 | 72 | .00008*** |

**Table S5*****. Modeling 1st-stage choices in the RL task as a function of structure inference ability and model-based planning.*** Each row reflects the results from an independent analysis where each covariate (z-transformed) was entered as Z in the following model: *Stay ~ Reward × Transition × Z + (1 + Reward × Transition* | *Participant)*. Model statistics refer to the coefficient of the fixed-effects interaction: *Reward × Transition × Z*. Positive values indicate an association with increased model-based planning Covariates with positive values are associated with increased model-based learning (except for the response time by relative distance effect). Here (SE) indicates the standard error of the mean. (*p<.05; **p<.01; ***p<.001).

| Performance measure names | Definition |
|---|---|
| Overall judgment accuracy | Proportion of trials on which participants correctly identified the closer node |
| Judgment accuracy by relative distance | The effect of distance discrepancy between the option nodes and the target node on accuracy. Greater distance discrepancy should lead to higher accuracy. |
| Response time by total distance | The effect of overall distance of two option nodes to the target on response times. The greater overall distance should lead to longer response times, since participants would need to perform a longer 'search' to make their response. |
| Response time by relative distance | The effect of distance discrepancy between the option nodes and the target node on response times. Greater distance discrepancy should lead to faster responses. |
| True vs estimated distance correlation | In the edge reconstruction task, we correlated the pairwise node distance given by 1) number of pixels between two node images in the reconstructed graph, and 2) the pairwise shortest path (number of edges) between all nodes in the underlying graph |
| Edge connection accuracy | Proportion of correctly identified edges, based on the comparison of node connections in the recovered graph to ground truth adjacency matrix |

**Table S6.** Graph task performance variables index.

| **Parameters** | $\beta^{s2}$ (Inverse temperature) | $\beta^{MB}$ | $\beta^{MF0}$ | $\beta^{MF1}$ | $\alpha$ (Learning rate) | $\gamma$ (Decay) |
|---|---|---|---|---|---|---|
| **Estimate** | 1.03 | 0.56 | 0.52 | 0.37 | 0.46 | 0.82 |
| **SD** | 0.56 | 0.43 | 0.32 | 0.46 | 0.27 | 0.20 |

**Table S7.** Reinforcement learning parameters point estimates and standard deviations.