Individualised dynamic internal representations from response times

Balázs Török[1,2,3], Dávid G. Nagy[1,4], Mariann M. Kiss[2,3], Karolina Janacsek[3,5,6,*], Dezső Németh[3,5,7,*], and Gergő Orbán[1,*]

[1]Department of Computational Sciences, Wigner Research Centre for Physics, Konkoly-Thege Miklós út 29-33, 1121, Budapest, Hungary
[2]Department of Cognitive Science, Budapest University of Technology and Economics, Egry József utca 1., Building T, 1111, Budapest, Hungary
[3]Brain, Memory and Language Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
[4]Institute of Physics, Eötvös Loránd University, Budapest, Hungary
[5]Institute of Psychology, ELTE Eötvös Loránd University, Izabella utca 46, 1064, Budapest, Hungary
[6]School of Human Sciences, Faculty of Education, Health and Human Sciences, University of Greenwich, Old Royal Naval College, 150 Dreadnought, Park Row, Greenwich, London, SE10 9LS, United Kingdom
[7]Lyon Neuroscience Research Center (CRNL), INSERM, CNRS, Université Claude Bernard Lyon 1, Centre Hospitalier Le Vinatier - Bâtiment 462 - Neurocampus 95 boulevard Pinel 69675 Bron, France
[*]Senior author

June 21, 2020

Abstract

Internal models are central to understand how human behaviour is adapted to the statistics of, potentially limited, environmental data. Such internal models contribute to rich and flexible inferences and thus adapt to varying task demands. However, the right internal model is not available for observers, instead approximate and transient internal models are recruited. To understand learning and momentary inferences, we need tools to characterise these approximate, yet rich and potentially dynamic models through behaviour. We used a combination of non-parametric Bayesian methods and probabilistic programming to infer individualised internal models from human response times in an implicit visuomotor learning task. Using this Cognitive Tomography approach we predict response times on a trial-by-trial basis and validate the internal model by showing its invariance across tasks and sensitivity to stimulus statistics. By tracking the performance of participants for multiple days, individual learning curves revealed transient subjective internal models and pronounced inductive biases.

Keywords: reaction time, learning, Bayesian inference

Introduction

Building internal models of the environment has been a key concept for understanding human inferences in complex settings. In particular, Bayesian models of cognition provide a normative framework for understanding how humans can perform complex computations flexibly. Bayesian hierarchical models learn efficiently from limited data, and their internal representations can be flexibly used when goals are changing[1–4]. However, Bayesian models of cognition operate under two strict conditions. First, domain-specific expert knowledge is exploited[5,6]. The experimenter has to select the relevant variables to be included in the ideal observer model's representation. This expert knowledge is regarded as the basis of useful and essential inductive biases reflecting scientific truths about our environment (e.g. physics), assumed to be readily available to humans partaking in these experiments. Earlier research has shown that in domains where the underlying mechanisms are scientifically known, an intuitive model approximating the true model can account for humans' judgements and their systematic deviations from the precise outcomes[7]. However, the constraint of relying on expert knowledge to determine the properties of the internal model is challenged in complex environments where the underlying exact mechanism certainly differs from the internal model maintained by human observers. This is particularly true when the observer is exposed to new environmental statistics and the model is gradually built up as information is acquired [8–10]. Second, individual performances are aggregated into a mean behaviour and the ideal observer models account for the aggregated behaviour. Such an account means that individual differences are eliminated and subjective inductive biases are left unidentified and neglected. When people approach a new, previously unseen problem, they rely on a wide variety of tools, regarded as priors to form hypotheses about the nature of the problem. The underlying true model is inherently unknown and the path people take to acquire the model can vary wildly[11,12]. It has been demonstrated that priors governing reasoning can have large deviations among individuals due to different experience[13]. To truly understand the nature of human learning, we ultimately need access to the temporarily diverging assumptions of individuals even when presented with identical observational data. Furthermore, testing whether fine grained individual variance in subjects' internal model is successfully captured requires moment by moment predictions of the individual's behaviour.

To capture individual differences with high-fidelity we need both a (i) highly expressive class of internal models and (ii) behavioural measurements that are highly informative about

the internal model. We proceed by choosing an experimental paradigm that satisfies (ii). In an experiment where trials are governed by a temporal dynamics and therefore individual trials are not independent, the joint set of behavioural measures in a number of trials have information content that far exceeds that of an independent and identically distributed (i.i.d.) experimental setting. In this paper our goal is to recover dynamical internal models of individuals in a dynamic task novel to participants. It has been extensively documented that participants do pick up temporal regularities in experiments with stochastic dynamics[14–16]. Furthermore, individuals show high variation in their initial assumptions[14,17]. In order to satisfy (i), we propose to use infinite Hidden Markov Models (iHMMs)[18] as a model class to characterize the internal models maintained by individuals. To infer the structure and dynamics of the iHMM we adopt and extend the Cognitive Tomography framework[13]. Cognitive Tomography aims to reconstruct high-dimensional latent structure using low-dimensional behavioural measurements. Critically, Cognitive Tomography distinguishes two components underlying behavioral responses: (1) the internal model which captures the statistical regularities of the environment, independent of the behavioral responses to be devised in a particular task, and (2) a response model that links the individual's inferences to their observable behaviour. In its original formulation the internal model structure was assumed to be unaffected by the observations during the experiment, the internal model was not dynamical, and behaviour was characterized solely by the key presses participants performed essentially providing roughly one bit of information in each trial. According to our model, participants filter the information gained from their observations over time, combine this information with their estimate of the dynamics to predict the next stimulus. We relate their subjective predictions to response times using the linear ascend to threshold with ergodic rate[19] (LATER) model. We invert this generative model of behaviour to infer individual's internal models.

We analyzed human performance in a dynamical response time paradigm. Human participants were exposed to a novel stimulus statistics and their performance was tracked for eight days. We used the response times of individuals to infer a dynamical probabilistic latent variable model underlying their behaviour. We constructed three benchmark models: one used in the literature for the task and two alternative ideal observer models which lack different aspects of flexibility of our model class. By using the response times of the correctly executed trials in one period of the experiment, we predict response times and errors in another period. We show that our method surpasses all compared models for all individuals

and explains a large share of variance in individual trial response times in a held-out dataset, reaching R^2 values of $R^2 = 0.248$ (s.e.m. 0.0183) on average and over 0.4 for some individuals. Our model also outperforms the benchmark models on predicting erroneous trials. Our model both performs above chance in predicting whether an error will occur in a held-out test dataset (even though incorrect trials are excluded from the training data), as well as predicts actual choices above chance. We also analyse the diversity of the inferred internal models of individuals. One key feature of our approach is that it automatically captures the structure of the individual's internal model without exploiting expert knowledge. When a participant only learns a simpler statistical structure from the data, our model naturally captures the simplified model structure. Our model also succeeds in accounting for counter-intuitive findings, that cannot be accounted for by a phenomenological model that is only sensitive to summary statistics, nor an ideal observer model equipped with a noisy realisation of the ground truth model. We construct individual learning curves and conclude that learning trajectories are highly non-linear in the sense that some individuals in initial stages obtain false assumptions about the underlying model but then reject them at later stages. Another key feature of our model is that it tracks the participant's uncertainty over the latent state. The true structure of the task exhibits systematic variation in this uncertainty which can be spotted in trained individuals' internal models but not in untrained ones. Our findings demonstrate that using expressive models to capture the complex internal models when exposed to complex stimuli enables us to assess both the contribution of momentary context to individual choices and to characterize the inductive biases of individuals as they learn about the environment.

Results

In order to see how behavioural data from individuals can be used to infer a dynamical probabilistic latent variable internal model, we used an experimental paradigm that could fulfil a number of key desiderata. First, the paradigm relies on across-trial dependencies; second, as in everyday tasks, the state of the environment cannot be unambiguously determined from the observation of momentary stimuli; third, the structure of the task is new to participants; fourth, the complexity of the task is relatively high, i.e. an a priori unknown number of latent states determine the observations; fifth, the task provides rich behavioral data, preferably continuous data with high temporal resolution. In the alternating serial response time task[20]

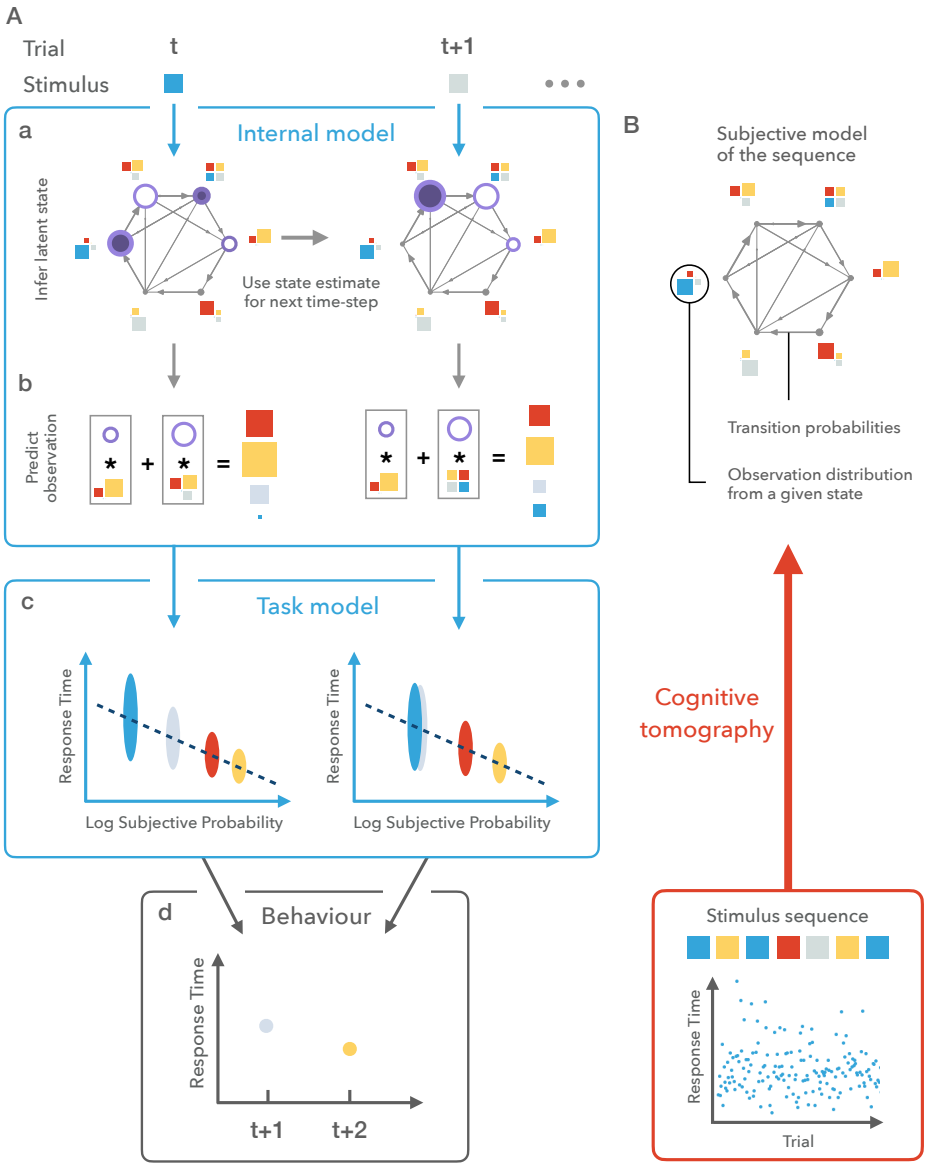


Figure 1: Ideal observer model and Cognitive Tomography. **Aa** The participant is assumed to maintain a model for the latent dynamics. The model posits that there is a latent sequence of states that govern observation probabilities. Grey dots represent latent states (dot sizes correspond to self-transition probabilities) and arrows stand for transition probabilities between states, with thickness proportional to the probability of the particular transition. Coloured squares represent the link between latent states and observations. Sizes of the squares are proportional to the probability of a particular observation in a particular state. **Aa-d** The generative model of behaviour consists of two parts, an internal model and a task model. Participants use the internal model components to update their beliefs over the current state of the latent variables (**Aa**, size of dark purple discs represent the posterior belief of the latent state based on the current observation). Then, participants play the model forward into the future (open purple circles). Finally, they generate predictions for the upcoming stimulus (**Ab**, rectangles) by marginalizing over (summing over latent beliefs of) the possible future states. **Ac**, According to the task model, the median and standard deviation of response times decreases linearly with log subjective probability [19]. Participants use previous state beliefs and the new stimulus to update latent state beliefs. Here only one of the possible states can generate the observation, hence there is only one dark purple disk at trial $t+1$. Again, they play the dynamics forward and predict the next stimulus. **Ad**, The observed response times are assumed to come from the distributions described by the LATER model in **Ac**. **B**, Cognitive Tomography is the method of inverting this generative model of behaviour. We inferred the internal representations of individuals from the stimulus sequences and the response times. Response times presented are 400 trials from our experiment from an example participant.

(ASRT) a stimulus can appear at four locations of a computer screen and the sequence of locations (untold to participants) follows a pre-specified structure. In odd trials, the stimulus follows a 4-element sequence, while in even trials the stimulus appears at random at any of the positions with equal probability independently of all other trials. Such stimuli precluded unambiguously determining the state of the task solely based on a single trial's observation. There are an additional 5 random trials at the beginning of each block. Participants are tasked to give fast and accurate manual responses through key presses corresponding to the locations of the stimuli. We collected response time measurements for sequences of stimuli organized into blocks of 85 trials. A session consisted of 25 blocks and the performance was tracked for 8 days with one session on each day, during which the same stimulus statistics governed the stimuli, followed by two additional sessions on subsequent days where the statistics of stimuli was altered (Extended Data Fig. 1).

We used the response times of individuals to infer a dynamical probabilistic latent variable model underlying their behaviour (Fig. 1). According to our model, participants filter the information gained from the observations over time to estimate possible the latent state of the system (Fig. 1Aa, *filled purple circles*). That is, they infer what history of events could best explain the sequence of their stochastic observations. Then, they use their dynamical model to play the latent state forward (Fig. 1Aa, *open purple circles*) and predict the next stimulus (Fig. 1Ab). By relying on the iHMM framework, a non-parametric Bayesian model, we can infer internal models of differing complexity[18]. Learning the internal model entails learning a transition matrix that defines the latent state dynamics and the emission matrix, which defines the probability of possible observations in any given state (Fig. 1B). Cognitive Tomography uses the response model to connect subjective probabilities of predicted stimuli to observed response times (Fig. 1Ad), for which we used the LATER model[19,21] (Fig. 1Ac). The resulting Cognitive Tomography (CT) model is implemented as a probabilistic program with components implemented in Stan[22].

To test that the proposed inference algorithm is capable of the retrieval of the probabilistic model underlying response time sequences, we validated our model on synthetic data (Extended Data Fig. 3). We used three different model structures for validation, which were HMMs (one sample from the iHMM inference in [18]) inferred from three different-length stimulus sequences. Similar to our human experiment data, we assessed CT by computing its performance on synthetic response times. Further, since synthetic participants provide access to true subjective probabilities we also calculated performance on the ground truth subjective

probabilities. We showed that the subjective probabilities can be accurately recovered from response times. As shown on Extended Data Fig. 3C, standard deviations of participants' response times are within the range of successful model recovery.

We constructed three benchmark models: a phenomenological trigram model, capturing summary statistics of the stimuli, and two ideal-observer models, the ground truth HMM model which is formulated as a noisy version of the true model generating the experimental stimuli, and another with a Markov model as internal model which can represent dynamics but not latent states (Table 1). The former ideal observer model could capture the full stimulus statistics but had a fixed structure therefore lacking the flexibility to capture variations across individuals and especially discover early-training internal models. The latter ideal observer model could track differences between individuals but was only sensitive to the dynamics defined over the observations, thus ignoring the possibility of learning dynamics over latent states.

Trial-by-trial prediction of response times for trained participants

To investigate how well we can predict response times on held-out data, we inferred the internal model as well as the parameters of the response time model on 10 blocks of trials measured at the second half of the session that was recorded on the eighth day of the experiment and predicted response times in 10 blocks of trials from the first half. Model performance was measured by evaluating the amount of variance of response times explained (R^2) on the held-out data (Fig. 2A), which provides a levelled comparison for the Bayesian and non-Bayesian accounts. Response times could be predicted efficiently even for individual trials as shown by the analysis of the response times from a single participant ($R^2(550) = 0.284,\ p < 0.001$, Fig. 2A). It is important to note that the predictive power was substantially increased by averaging over trials in the same positions of the sequence (0.54, Extended Data Fig. 4). Despite the significant advantage of trial-averaged predictions, we believe that single trial predictions provide a more rigorous and important characterization of human behaviour therefore we evaluate model performances on an individual trial basis in the rest of the paper. Predictive power of CT varied across participants but correlations between measured and predicted response times was significantly above zero for all participants ($M = 0.248$ ranging from 0.1 to 0.42, Fig. 2B).

We assessed three alternative models in an analogous way (Fig. 2A). Note that subject-by-subject analysis enables us to directly compare the relative predictive power of the three

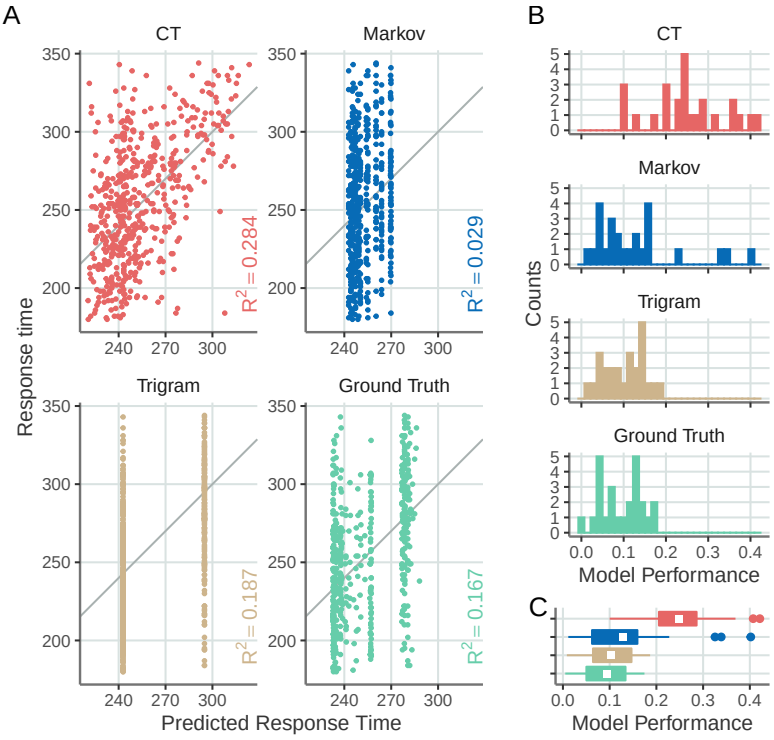


Figure 2: Response time predictions for individual trials on a held-out test dataset. Training our inference algorithm on a training dataset of 10 blocks, we predict response times of another 10 blocks on the same day by the same participant. **A** Individual trial predictions against measured response times (*dots*) for an example individual by the four considered models: the Ground Truth model (the HMM generating the data with nonspecific noise superimposed); the iHMM underlying Cognitive Tomography; the Markov model; and the Trigram model. Response times shown between 180msec and mean+3 s.d. for visual clarity. **B** Model performance histograms (R^2 as a measure of explained variance) of the different models for all participants on Day 8 of the experiment. **C** Summary of model performances (R^2 values) of the experiment for all participants on day 8. White rectangles show means, hinges of boxplot show 25th and 75th percentiles. Upper whiskers show largest value or 75th percentile $+ 1.5 \cdot IQR$. Similarly for lower whisker.

models. For the example participant analysed here, the trigram model had a slightly lower performance than the CT model but the performance of the Markov model was substantially lower, still above zero ($R^2(550) = 0.0295,\ p < 0.001$). Both of the models could predict response times on a trial-by-trial basis with significantly above zero performance (Fig. 2B), but on average controls had substantially lower performance than that of CT (CT vs Trigram two-sided $t(24) = 6.728,\ p < 0.001,\ CI = [0.125, 0.236],\ d = 1.99$ CT vs Markov two-sided $t(24) = 8.824,\ p < 0.001,\ CI = [0.118, 0.189],\ d = 1.34$ CT vs Ground Truth two-sided $t(24) = 7.331,\ p < 0.001,\ CI = [0.139, 0.248],\ d = 2.09$, Fig. 2C).

Evolution of the internal model with increasing training

Learning the model underlying observations entails that participants need to learn the number of states, the dynamics, and observation distributions, and is therefore challenging in the ASRT paradigm requiring substantial exposure to stimulus statistics. We tracked the evolution of the internal model by learning separate models for different days (Fig. 3A). Note, that this approach can identify changes in the internal model on the time scale of hundreds of trials (corresponding to different days in our paradigm) but assumes constancy of the internal model within this particular number of trials. Second-order statistical structure captured by the trigram model is only weakly present after a day of training (one-sided $t(24) = 0.664,\ p = 0.257,\ CI = [-0.00942, Inf],\ d = 0.133$, Fig. 3A) but the Markov model can capture a significant amount of variance from response times ($M = 0.0679$ ranging from 0.00046 to 0.13, one-sided $t(24) = 11.73,\ p < 0.001,\ CI = [0.205, Inf],\ d = 2.35$), and it is not different from CT (paired t-test on correlation values two-sided $t(24) = 1.426,\ p = 0.167,\ CI = [-0.00331, 0.0181],\ d = 0.0695$). Since CT contains the Markov model as a special case, this result suggests that after a day of training it is a simple Markov structure that dominates the responses of participants. Performance of all models increases in the first few days of the experiment but the advantage of CT gradually increases as time passes ($F(7, 168) = 32.21,\ p < 0.001,\ \eta^2_{partial} = 0.573$) and by day eight the dominance of CT model is very pronounced, reaching $R^2 = 0.248$ (s.e.m. 0.0183) on average.

By using a model trained on a specific day to predict responses on subsequent days (forward prediction) or previous days (backward prediction), we can obtain insights into the dynamics of learning the internal model. Participant-averaged backward prediction in CT indicates a gradual build-up of the model as prediction of earlier day responses strongly depended on the time difference between the training day of the model and the test day

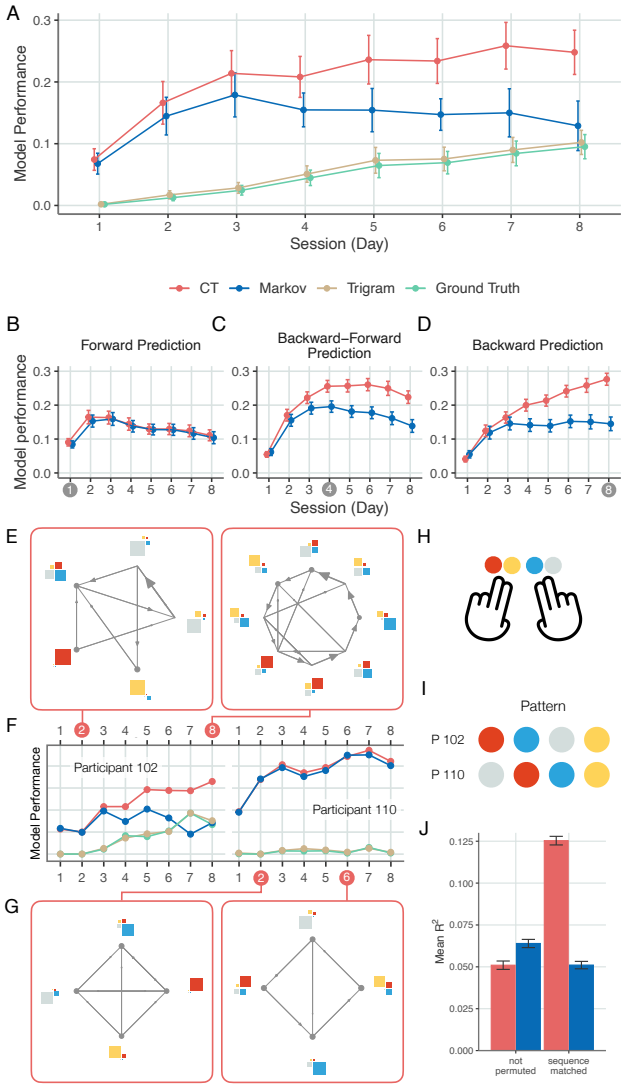


Figure 3: Evolution of the internal model with increasing training. **A** Mean explained variance (*dots*, averaged over participants) in held-out response times in sessions recorded on successive days for the CT (red), Markov (blue), and trigram (yellow) models. Error bars denote 2 standard error of the group mean. **B** Forward prediction performance of the CT and Markov models across different days of the experiment. Generalization capacity of models were tested by training models on the first day of experiment and testing them on successive days. We inferred internal models on the first day of the experiment and used this internal model to predict further days' response times of the same participant. Dots show means of model performances (R^2 values) over participants. Performance of the training day is measured on held-out trials. **C** Forward-backward prediction of CT and Markov models. Same methodology is used as on panel B but for models trained on Day 4 and tested both on preceding and succeeding days of the experiment. **D** Backward prediction of CT and Markov models. Model training was performed on Day 8 and tested on preceding days. **E-G** Learning curves of individual participants characterized by different models. **E** Participant 102 finds a partially accurate model by Day 2 and a model close to the true model by Day 8. **F** Model performances for two example participants. **G** Participant 110, sticks to a Markov model and their behaviour is ever better explained by this type of model. This is further supported by the trigram model's floored performance, indicating that no higher-order statistical structure was learned. **H** Color coding of response buttons used in this figure. **I** Color coding of sequence showed to participants. **J** Model performances across participants. Day 8 models are swapped across participants and performance of swapped models is measured (*left*) both for CT (*red*) and the Markov model (*blue*). We construct sequence-matched models by inverting the permutation (*right*). Error bars show 2 s.e.m.

(Fig. 3C,D). Forward prediction did not exhibit such strong dependence on time difference, indicating that on group level, the statistical structure acquired up to a particular day was incorporated in later models (Fig. 3B,C). Importantly, group averages can hide differences between individual strategies but CT provides an opportunity to investigate these through the patterns of forward an backward predictions (Extended Data Fig. 8). For instance, participant 119 initially entertains a Markov-like internal model up until day 2-3, only to dispose of it and acquire a different structure later in the training.

CT offers a tool to investigate the specific structure of the model governing behaviour at different points during training (Fig. 3E,G). We computed individual learning curves (Fig. 3F & Extended Data Fig. 5) by measuring the performance of the different models on each day of the experiment. Furthermore, we analyse the internal model structure associated with behaviour by taking posterior samples from the CT model. Early during training where the performance of the Markov model is close to that of CT, the inferred iHMM indeed has a structure close to that of the Markov, which is characterized by a strong correspondence between observations and states. Later in the experiment, however, the performance of CT deviates from that of the Markov model for most of the participants (Extended Data Fig. 5) and the model underlying the responses is close to the ground truth model (Fig. 3E). There are participants, however, where improved predictability of response times does not correspond to adapting a model structure that reflects the real stimulus statistics, but the model underlying response times is still featuring a fully observed structure (Fig. 3G).

We characterized differences between the internal models inferred from individuals on the eighth day of the experiment by swapping the models across subjects and predicting response times using swapped models on the same day. As a control, we performed across-subject swap not only for CT but the Markov model as well. The swapped CT models perform substantially worse than CTs matched to the individuals (two-sided $t(599) = 25.19,\ p < 0.001,\ CI = [0.0686, 0.0802],\ d = 1.2$) and these are outperformed even by the swapped Markov models (two-sided $t(599) = 5.51,\ p < 0.001,\ CI = [0.00834, 0.0176],\ d = 0.214$, Fig. 3J). This latter result can be a result of picking up task-independent factors from data, e.g. physiological constraints for finger movements, or biases, but being ignorant to across-individual differences (Fig. 3H,I) in the structure of pattern elements spares its performance. We tested whether across-individual differences in the pattern sequences can account for across-individual differences in the internal models by sequence matching, i.e. regularizing the internal models through inverting the effects of sequence permutations in

the iHMM model. The regularized swapped CT performs substantially better than any other swapped model (Fig. 3J) and in particular the regularized Markov model (two-sided $t(599) = 36.07,\ p < 0.001,\ CI = [0.0702, 0.0783],\ d = 1.26$), indicating that individual differences between the models inferred by CT indeed reflect variations in stimulus statistics.

Statistical structure learned by Cognitive Tomography from response times

Predictive performance of subject-averaged and individual models inferred by CT indicate that there is substantial improvement over the course of training. It remains elusive, however, what aspects of the stimulus statistics the internal model reflects. We investigated this question in three steps.

First, we directly contrasted the performances of CT, Markov and trigram models on a participant-by-participant basis. Gaining advantage by CT over the Markov model coincides with increasing performance of the trigram model (Fig. 4A). Therefore an important question is if the statistical structure captured by CT is merely a combination of the predictions of the Markov and trigram models. The Markov and trigram models capture orthogonal aspects of the response statistics: while the Markov model can only account for first order transitions across observed stimuli, the trigram model is insensitive to this aspect of the response statistics but can account for effects caused by second-order transitions (Table 1). As a consequence, the variances in response times explained by these two models are additive. We used this property to assess if the high-order statistical structure captured by CT goes beyond that captured by the trigram model. We compared the normalized CT performance (reduced by that of the Markov model) to the trigram model (Fig. 4A). Normalized CT performance of participants was aligned with trigram model performance but the normalized CT showed a small but significant advantage over the trigram model on day eight of the experiment (one-sided $t(24) = 2.395,\ p = 0.0124,\ CI = [0.00472, Inf],\ d = 0.479$, Fig. 4B).

Second, we sought to identify the fingerprints of the statistics acquired by CT but not by the trigram model. We analyzed response times to the third element of three-stimulus sequences which the trigram model is unable to distinguish but more complex models could. In one of the analysed conditions, the first and third elements were pattern elements and we compared these to a condition where the first and third elements are random elements but the actual observations were the same. Since only the latent state differed between the two conditions, these cannot be distinguished by the trigram model. Higher order learning,

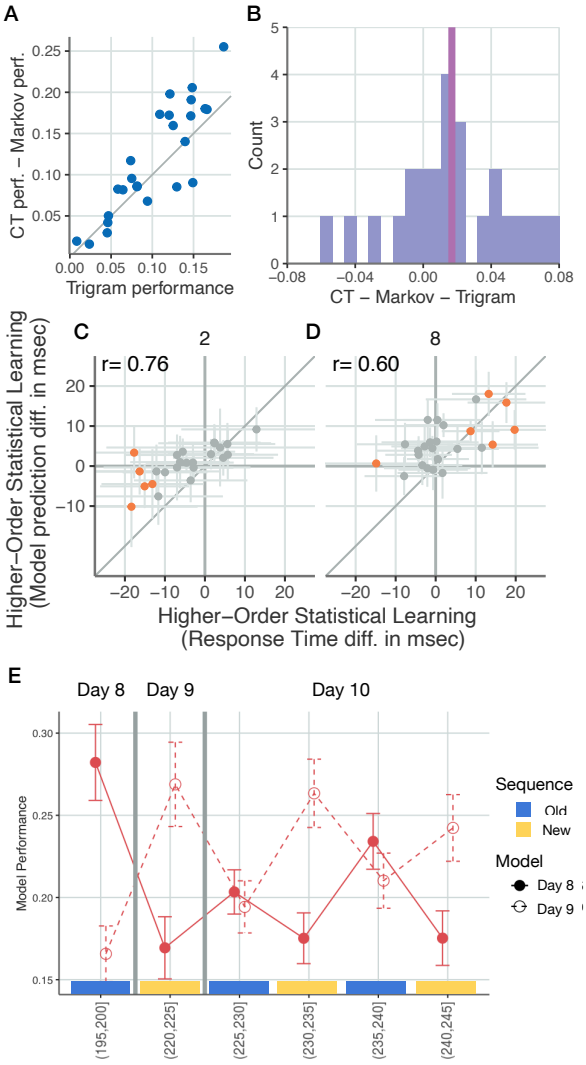


Figure 4: Analysis of the statistical structure learned by CT with that learned by trigram and Markov models. **A** Subject-by-subject comparison (*dots*) of triplet model performance and normalized CT performance (the margin by which CT outperforms the Markov model) on Day 8. **B** Histogram of the advantage of normalized CT performance over the trigram model. Red line marks the mean of the histogram. **C** Higher order statistical learning in CT on Day 2 of the experiment. Dots show individual participants. Orange dots represent participants with higher-order learning score significantly deviating from zero. **D** Same as C but for Day 8 of the experiment. CT can capture both negative deviations (Day 2, **C**) and positive deviations (Day 8, **D**) in this test, indicating that subtle and nontrivial statistics of the internal model is represented in CT. **E** Investigating new internal models emerging as new stimulus sequences are presented. Performance of predicting response times on Day 8-10 using CT-inferred models that were trained on Day 8 (*filled symbols*) and Day 9 (*open symbols*). On Day 9 a new stimulus sequence was introduced, therefore across-day prediction of response times corresponded to across sequence predictions. Training of the models was performed on 10 blocks of trials starting from the 11th block and prediction was performed on the last five blocks of trials. Stimulus sequence was switched across the 5-block segments between sequences used during Day 8 and Day 9 (*colours of bars* indicate the identity of stimulus sequence with colours matching the bars used in Day 8 and Day 9. Error bars show 2 s.e.m. Stars denote $p < 0.05$ difference.

characterised by response time difference between the two conditions, was highly correlated with higher-order statistical learning measured directly from response times both early in the training (Fig. 4C, $r(22) = 0.756,\ p < 0.001$) and on the last day of training (Fig. 4D, $r(23) = 0.603,\ p = 0.0014$). Interestingly, early in the training most of those participants whose higher-order statistical learning measure was significantly different from zero had negative score (Fig. 4C, orange dots), a counter-intuitive finding termed inverse learning[23,24].

Third, in order to test how participants use and update their internal model when stimuli are generated using a new stimulus statistics, participants were tested in two additional days. On Day 9 of the experiment, 25 blocks of trials were performed with a new pattern sequence. On Day 10, another 20 blocks of trials were performed switching between the pattern sequences of Day 8 and Day 9 every five blocks. Comparison of within-day and across-day predictions on Days 8 and 9 demonstrates that response time statistics change across days, as reflected by significant differences in within-day and across-day predictions (one-sided $t(24) = 4.958,\ p < 0.001,\ CI = [0.0746, Inf],\ d = 1.06$ and one-sided $t(24) = 4.9,\ p < 0.001,\ CI = [0.0616, Inf],\ d = 0.963$ for forward and backward predictions, respectively). Specificity of response time statistics to the stimulus statistics is tested by predicting Day 10 performance using Day 8 and Day 9 models. The Day 8 model more successfully predicts response times in blocks relying on Day 8 statistics than on blocks with Day 9 statistics (one-sided $t(24) = 3.734,\ p < 0.001,\ CI = [0.0236, Inf],\ d = 0.594$) and the opposite is true for the Day 9 model (one-sided $t(24) = 3.528,\ p < 0.001,\ CI = [0.0261, Inf],\ d = 0.575$). Oscillating pattern in the predictive power of Day 8 and Day 9 models on blocks governed by Day 8 and Day 9 statistics indicates that participants successfully recruit different previously learned models for different stimulus statistics (Fig. 4E and Extended Data Fig. 6).

Assessment of the internal model for response times on behavioral errors

The core principle of Cognitive Tomography is to distinguish an internal model that captures a task-independent understanding of the statistical structure of the environment and a response model which describes task-specific behavioural responses. Validity of this distinction hinges upon the validity of both models and the possibility to manipulate them independently. To validate this distinction, we use the inferred internal models and combine it with a different behavioural model for error prediction. In this setting, we predict on which trials is the participant likely to commit errors and if they do so what the erroneous response will be.

Note, that the internal models of CT were inferred only on correct trials using the response time model.

We investigated two relevant hypotheses. First, a participant will more likely commit an error when their subjective probability of the stimulus is low. Second, when committing an error, their response will be biased towards their expectations. We compared the rank of the subjective probability of the upcoming stimulus both for correct and incorrect trials (Fig. 5A). CT ranked highest the upcoming stimulus in correct trials above chance (0.461, $n = 18473$, $p < 0.001$) and significantly below chance for incorrect trials (0.175, $n = 2777$, $p < 0.001$). The Markov model's performance was close to chance in predicting the upcoming stimulus in correct trials (0.271, $n = 18473$, $p < 0.001$). In incorrect trials the Markov model, too, showed a tendency to not rank the upcoming stimulus the highest, therefore showed significantly below chance performance on ranking (0.137, $n = 2777$, $p < 0.001$). The trigram model behaved as expected for correct trials, with significantly above-chance performance (0.629, $n = 18473$, $p < 0.001$) but it also assigned the highest probability to the upcoming stimulus in incorrect trials (0.312, $n = 2777$, $p = 1$). Ranking of incorrect responses was above chance for all models (Fig. 5B).

We obtained a participant-by-participant assessment of the difference between model performances in predicting error trials by calculating ROC curves of the models based on the subjective probabilities assigned to upcoming stimuli (Fig. 5C, Extended Data Fig. 7). Note, that the trigram model does not have a continuous range of subjective probabilities therefore only two points are available from the ROC curve and area under the curve is calculated by linear extrapolation between the extreme values and the available points. Area between two ROC curves characterizes the performance difference between models and CT is shown to consistently outperform both the Markov and trigram models (paired t-test on AUC values one-sided $t(24) = 9.584$, $p < 0.001$, $CI = [0.0858, Inf]$, $d = 2.48$ and paired t-test on AUC values one-sided $t(24) = 7.976$, $p < 0.001$, $CI = [0.0462, Inf]$, $d = 1.49$ for Markov and triplet models, respectively, Fig. 5D).

Comparing CT to the ground truth model

In order to compare how the inferences made in the model identified by CT are related to those characteristic of the ground truth model, we derived a number of measures. First, we used the subjective probability of the stimulus, identical to the measure used to assess reaction times. Next, entropy of the predictions is used to characterize the uncertainty associated with

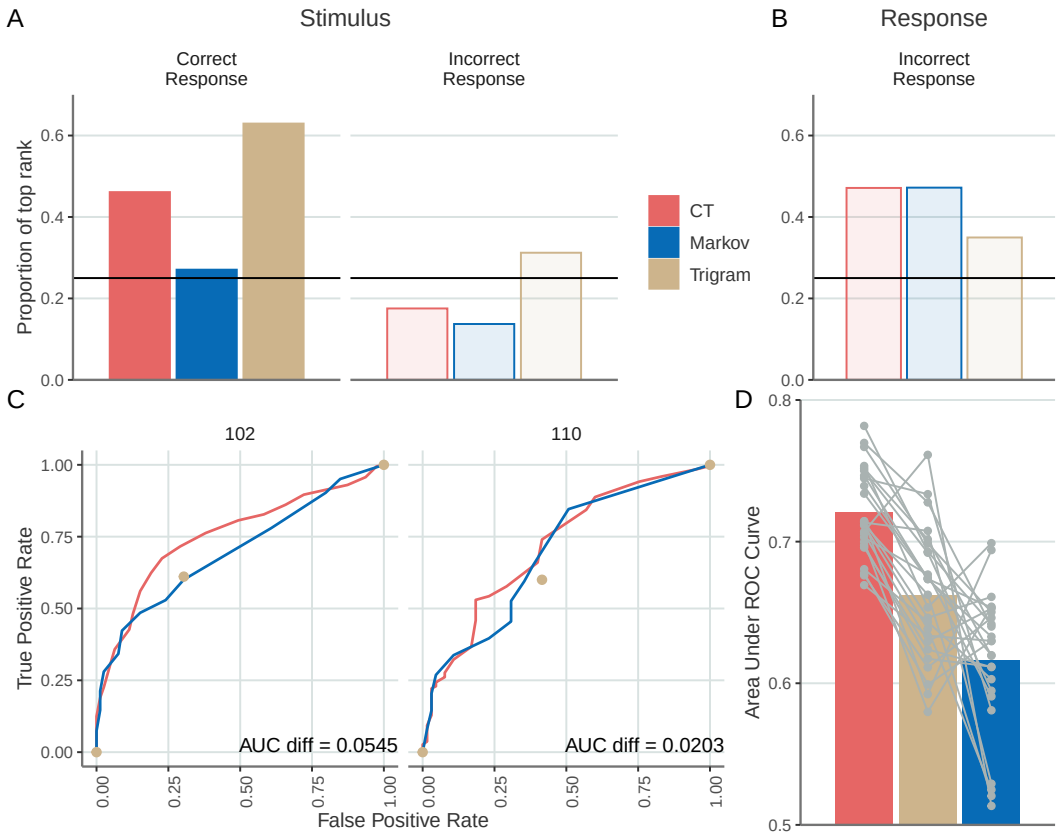


Figure 5: Choice prediction by the different models. Models are trained on response times for correct key presses on Day 8 and tested on both correct and error trials the same day. **A**, Proportion of trials where the model ranked the upcoming stimulus first. For correct trials, the Markov model has only marginal preference for the stimulus. For incorrect trials, the trigram model falsely predicts the stimulus in more than a quarter of the time. **B**, Proportion of trials where the model ranked the button pressed by the participant first. For incorrect responses, all models display a preference towards the actually pressed key over alternatives. **C**, ROC curves for two example participants based on the subjective probabilities of upcoming stimuli (held-out dataset). Since the trigram only renders trials into three groups, ROC consists of two points. Area under the ROC curve difference shows performance difference of CT and Markov models in predicting error trials. **D**, Area under ROC curve. Grey dots show individuals, bars show means. Trigram model area calculated by area under linear interpolation between points.

the stimulus predictions (Fig. 6A). Finally, both models track uncertainty over the latent state of the system continually, which we can measure both before observing a particular stimulus and right after (Fig. 6B). Entropy of state prior characterises the former uncertainty while entropy of state posterior characterises the latter. We calculated these measures for each trial and contrasted them between pattern and random trials both for early models inferred by CT (Day 1) and for models inferred after extensive training (Day 8) (Fig. 6C-F). The ground truth model describes the way stimuli are generated and therefore has no free parameters. In order to make this model reasonably flexible, we introduced two parameters, which correspond to basic uncertainty in model parameters: the noise level in the dynamics and the noise level in the emission distributions. Changes in these noise parameters merely rescale the measures but leaves the sign of the difference unchanged. The parameters we chose were those that resulted in measures on the same scale as for the CT model. None of the measures derived from the early-training CT models were significantly different from zero. Measures of the late-training CT models showed similar tendencies as the ground truth model for the subjective probability of the stimulus, state prior and state posterior entropies. It was only the entropy of the predictions which was qualitatively different from the ground truth model. Taken together, trial-by-trial regressors inferred from response time data provide insights into important quantities of an ideal observer model and have the potential to track these quantities along the training of individuals.

Discussion

In this paper we have built on the idea of Cognitive Tomography, which aims to reverse engineer internal models by separating the task-general component from the task-specific component of a generative model of behaviour, to infer high-dimensional dynamical internal models. Key to our approach was the combination of non-parametric Bayesian methods that allow discovering flexible latent variable models, with probabilistic programming, which allows efficient inference in probabilistic models. The proposed approach yields a tool, which has a number of appealing properties for studying how acquired knowledge about a specific domain affects momentary decisions of biological agents. 1, We used iHMM, a dynamical probabilistic model that can naturally accommodate rich inter-trial dependencies, a characteristic prior of humans[14,15,25]. 2, iHMM has the additional benefit of being able to capture arbitrarily complex statistical structure but not increasing the complexity of the

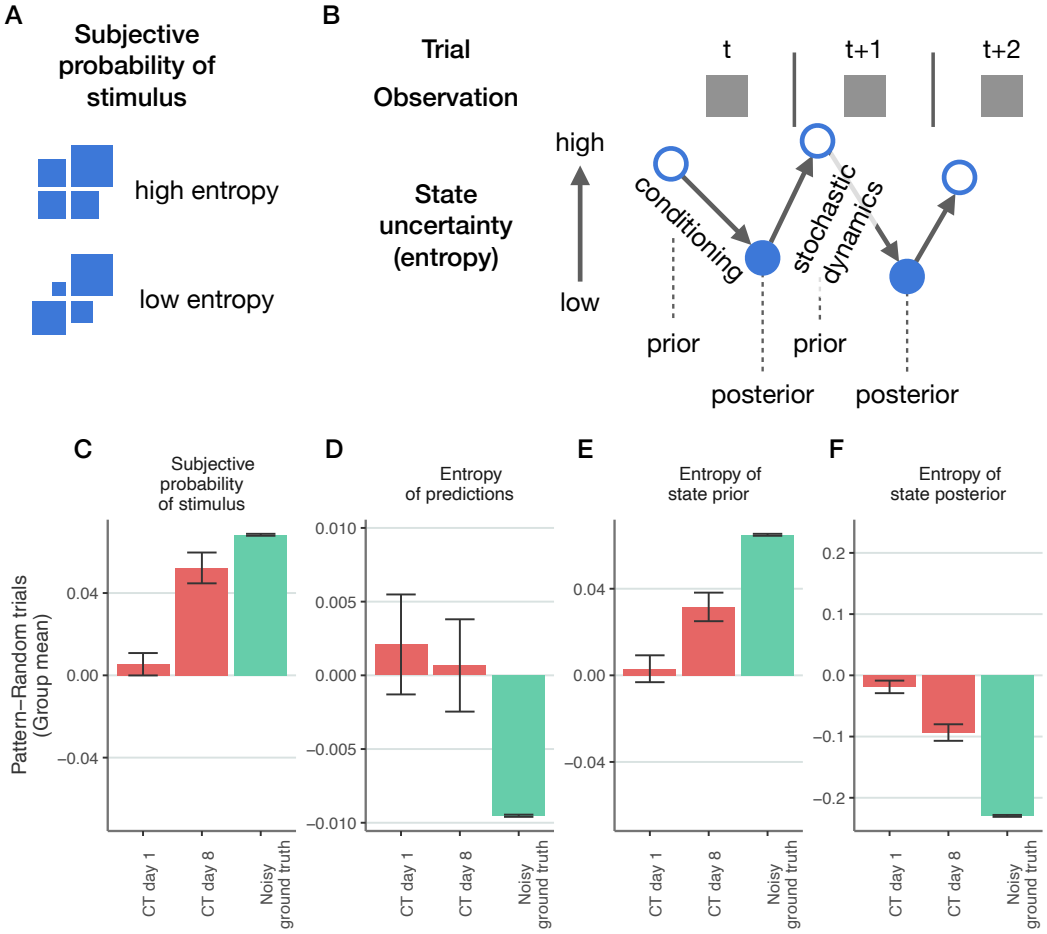


Figure 6: Comparison of the functioning of CT and the ground truth model. **A**, Illustration of low entropy (*top*) and high entropy (*bottom*) predictive distributions of upcoming stimuli. Size of the square corresponds to the probability of a particular stimulus. **B**, Illustration of latent state entropy dynamics across three consecutive trials. Entropy of the latent state distribution calculated based on previous observations (and thus earlier state estimates) but without the knowledge of the current stimulus (entropy of state prior, *open circles*), and the entropy of the latent state distribution based on both the earlier and the current observations (entropy of the state posterior, *filled circles*). **C**–**F** We assessed the predictive probability of upcoming stimulus (**C**), the entropy of the predictive probability distribution (**D**), the entropy of state prior (**E**), and entropy of state posterior (**F**) in both CT (*red bars* and the ground truth model (*green bar*) on individual trials of a stimulus sequence. These four model characteristics were calculated for pattern and random trials separately and the difference between pattern and random trials is shown. CT was evaluated on Day 1 and Day 8 of the training as well. For the ground truth model, it is the deviation from zero and direction of deviation that are the relevant quantities since the scale is determined by two noise parameters of the model and is therefore arbitrary. Bars denote mean differences between pattern and random trial means across trials and across models inferred for different participants, and error bars show 2 s.e.m.

	Model space	Best model in model space
Ground Truth	$p(y_{t+1}\|y_1, y_2, ..., y_t)$	8-state HMM deterministic latent dynamics stochastic observations
Markov	$p(y_{t+1}\|y_t)$	Uniform
Trigram	$p(y_{t+1}\|y_t, y_{t-1})$	(0.625, 0.125, 0.125, 0.125) "high" & "low"
CT	$\sum_{s_t, s_{t+1}} p(y_{t+1}\|s_{t+1}) \cdot p(s_{t+1}\|s_t) \cdot p(s_t\|y_1, y_2, ..., y_t)$	Ground truth

Table 1: Table of models and the maximum likelihood parameter sets for the stimuli in our experiment. The ground truth (the true generative model of the task) can be formalized as an 8-state HMM with states Pattern1, Random1, Pattern2, Random2, Pattern3, Random3, Pattern4, Random4 where the pattern states produce the corresponding sequence element with probability 1 and all the random states produce any of the four observations with equal probability independently. The Markov model (where predictions are produced by conditioning only on the previous observation) fits the task observations best it when predicts all observations with equal probability, since the marginal probabilities of any one stimulus is equal regardless what the previous observation was, because every other trial is random. The trigram model produces a "high triplet" prediction, where the next stimulus is the successor of the stimulus two trials ago in the pattern sequence. All alternatives have equal probability of 0.125. Note that the exact probabilities in this case are not relevant since the trials are categorized into two groups (high and low) and therefore the parameters of the response time model and these probabilities are underspecified. The iHMM model produces prediction for the next stimulus via filtering. A latent state of the sequence is estimated from previous observations using a Hidden Markov Model. This flexible model space includes the ground truth model as well as the Markov model.

model more than necessary[26,27]. 3, We could predict response times of human participants on a trial-by-trial basis. 4, Complex individualized internal models could be learned from behavioural data. 5, We evaluate individual learning curves, thus revealing diverging priors for interpreting the environment. Importantly, we have validated that the statistical structure posited by Cognitive Tomography reflects key characteristics of internal models: specificity to environmental statistics and generality over tasks. For this, we dissociated the contributions of the two main components of the learned model: the task model was replaced for predicting erroneous responses of participants, and adaptation to stimulus statistics was demonstrated by differential recruitment of the internal model under different stimulus statistics. Cognitive Tomography revealed transient models at individuals that were nurtured temporarily only to be abandoned later during training. Individualized internal models helped us to identify internal models that deviate from the 'expert model' and reflect a much simpler structure than that underlying the stimulus, thus indicating strong personalized inductive biases during the acquisition of complex stimulus statistics.

Learning in general is an ill-defined, under-determined problem. Learning requires inductive biases formulated as priors in Bayesian models to efficiently support the acquisition

of models underlying the data[28,29]. The nature of such inductive biases is a fundamental question which concerns both cognitive science and neuroscience, even machine learning[28,30,31]. These inductive biases determine what we can learn. Variance in inductive biases across individuals leads to different learning strategies, and can be tightly linked to the neural representations of the models. The CT approach helps identifying individualized internal models from behavioral data thus allowing the assessment of individual learning strategies. Key to this approach is data efficiency: the experimenter needs to infer internal models from limited data. Bayesian inference that exploits structured models offers exactly this data efficiency while maintaining flexibility to accommodate potentially complex, high-dimensional models[32]. The presented model builds on the original CT analysis performed on faces[13] but differs in a number of fundamental ways. We sought to infer the evolution of the internal model for statistics new to participants. In contrast to the earlier formulation using a 2-dimensional latent space and a static model, inference of a dynamical and potentially high-dimensional model yield a much richer insight into the working of the internal model acquired by humans. Using a structured internal model allows the direct testing of the model against alternatives, in particular the limited computational-capacity Markov model and the ground truth model allows the assessment of crucial characteristics of the model. A well-structured internal model can be used to make arbitrary domain-related inferences within the same model. Based on this, we can decompose the complex inference problem into separately meaningful sub-parts which can be reused in tangential inference problems to serve multiple goals. By showing that the same variables can be used for multiple tasks, it is reasonable to look for signatures of these quantities in neural representations. A possible alternative formalization of this problem could be using POMDPs[33], where internal reward structure and the subjective belief of the effect of the participant's actions are jointly inferred with the internal model. However, in our experiment, the reward and action models have simple structures and hence the problem simplifies to a sequence learning problem. Instead, here we focus on inferring rich internal model structures as well as having an approximate Bayesian estimate instead of point estimates as in [33].

Our model produces moment by moment regressors for (potentially unobserved) variables that are cognitively relevant. Earlier work considered neural correlates of hidden state representations in the orbitofrontal cortex of humans[34] but the internal model was not inferred, rather assumed to be fully known. CT provides an opportunity to design regressors for individualised and potentially changing internal models. In particular, the model differentiates

between objective and subjective uncertainties, characteristics relevant to relate cognitive variables to neural responses[35–38]. The former is akin to a dice-throw, uncertainty about future outcomes which may not be reduced with more information. The latter is uncertainty arising from ambiguity and lack of information about the true current state of the environment. We showed that uncertainties exhibited by a trained individual's internal model show similar patterns in these characteristics as the ground truth model which generated the stimuli, which promises that uncertainties inferred at intermediate stages of learning are meaningful.

Recently, major efforts have been devoted to learning structured models of complex data both in machine learning and in cognitive science[4,7,39,40]. These problems are as diverse as learning to learn[4,41], casual learning[42], learning flexible representational structures[39], visual learning[43]. When applied to human data to reverse engineer the internal models harnessed by humans, past efforts fall into two major categories. 1, Complex (multidimensional) models are inferred from data and fitted to across-participant averaged data[7,44,45], ignoring individual differences. 2, Simple (low dimensional) models are used to predict performance on a participant-by-participant manner, thus resulting in subjective internal models[14,46]. The contribution of this paper is twofold: 1, We exploit recent advances in machine learning to solve the reverse-engineering problem in a setting where complex internal models with high-dimensional latent spaces are required; 2, We contribute to the problem of finding new useful inductive biases by enabling direct access to the internal model learned by individuals.

A widely studied approach to link response time to quantities relevant to task execution is the drift diffusion model, DDM[47]. In its most basic form evidence is stochastically accumulated as time passes such that the rate of accumulation is proportional to the information gained by extended exposure to a stimuli, until evidence reaches a bound where decision is made. Through a compact set of parameters DDM can explain a range of behavioural phenomena, such as decisions under variations in perceptual variables, attentional effects on decision making, the contribution of memory processes to decision making, decision making under time pressure[48–51], and neuronal activity was also shown to display strong correlation with model variables[52,53]. Both LATER and DDM have the potential to incorporate variables relevant to make decisions under uncertainty and the marginal distributions predicted by the two models are comparable. Our choice to use the LATER model was motivated by two major factors. First, LATER is formulated with explicit representation of subjective predictive probability by mapping it onto a single variable of the model. This setting promises that subjective probability can be independently inferred from available

data and the internal model influences a single parameter of the model. As a consequence, subjective probability is formally disentangled from other parameters affecting response times and associated uncertainty can be captured with Bayesian inference. In case of distributing the effect of subjective probability among more than one parameters (starting point, slope, variance) the joint inference of subjective probability with other parameters affecting response times results in correlated distributions. Consequently, maximum likelihood inference, or any other point estimations, the preferred method to fit DDM, will have large uncertainty over the true parameters due to interactions between other variables. Furthermore, this uncertainty remains unnoticed as there is usually no estimation of this uncertainty, only point estimates. Second, trials are usually sorted based on the design of the experiment into more and less predictable trials (with notable exceptions[14]). This leads to a misalignment between the true subjective probabilities of a naive participant and the experimenter's assumptions. Assuming full knowledge of the task and therefore assuming an impeccable internal model in more complex tasks, however, implies that potential variance in the acquired internal models across subjects will be captured in variances in parameters characteristic of the response time model rather than those of the internal model. DDM is considered to be an algorithmic-level model[54] of choices[25], which is indeed useful for linking choice behaviour to neuronal responses[55]. The appeal of the Bayesian description offered by the normative framework used here is that it can accommodate a flexible class of internal models, without the need to adopt algorithmic constraints. Similar algorithmic-level models of behaviour that is based on the flexible and complex internal models yielded by Cognitive Tomography are not available and will be the subject of future research.

Acknowledgements

This research was supported by the National Brain Research Program (project 2017-1.2.1-NKP-2017-00002, D.N., G.O.); Hungarian Scientific Research Fund (NKFIH-OTKA K K125343, G.O.; NKFIH-OTKA K 128016, D.N., NKFIH-OTKA PD 124148, K.J.); Janos Bolyai Research Fellowship of the Hungarian Academy of Sciences (K.J.); IDEXLYON Fellowship of the University of Lyon as part of the Programme Investissements d'Avenir (ANR-16-IDEX-0005) (D.N). B.T. was supported by scholarship by Budapest University of Technology and Economics as well as by Mozaik Education Ltd. (Szeged, Hungary). Resources for the computational analysis were generously provided by the Wigner Data Center.

Author contributions

B.T., D.G.N. and G.O. designed the computational model. M.K., K.J., D.N. designed the experiment. M.K. recorded the experimental sessions. B.T. implemented the models and carried out analyses. B.T., G.O., D.G.N., K.J., D.N. interpreted results. B.T. and G.O. wrote the manuscript and the supplementary information. D.G.N., K.J., D.N. commented on the manuscript at all stages. M.K. wrote report on data collection.

References

[1] Xu, F. & Tenenbaum, J. B. Word learning as Bayesian inference. *Psychological Review* **114**, 245–272 (2007).

[2] Griffiths, T. L. & Tenenbaum, J. B. Theory-Based Causal Induction. *Psychological Review* **116**, 661–716 (2009).

[3] Kemp, C. & Tenenbaum, J. B. Structured Statistical Models of Inductive Reasoning. *Psychological Review* **116**, 20–58 (2009).

[4] Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015). URL `http://cims.nyu.edu/{~}brenden/LakeEtAl2015Science.pdf`.

[5] Niv, Y. Learning task-state representations. *Nature Neuroscience* **22**, 1554–1553 (2019).

[6] Maheu, M., Meyniel, F. & Dehaene, S. Rational arbitration between statistics and rules in human sequence learning. *bioRxiv* (2020).

[7] Battaglia, P. W., Hamrick, J. B. & Tenenbaum, J. B. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences* **110**, 18327–32 (2013). URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3831455{&}tool=pmcentrez{&}rendertype=abstract. arXiv:1404.2263v1`.

[8] Love, B. C., Medin, D. L. & Gureckis, T. M. SUSTAIN: a network model of category learning. *Psychological Review* **111**, 309–332 (2004).

[9] Gershman, S. J. & Niv, Y. Perceptual estimation obeys Occam's razor. *Frontiers in Psychology* **4**, 623 (2013).

[10] Nagy, D. G., Török, B. & Orbán, G. Semantic Compression of Episodic Memories. In *Proceedings of the 40th Conference of the Cognitive Science Society*, 2138–2143 (2018).

[11] Estes, W. K. The problem of inference from curves based on group data. *Psychological bulletin* **53**, 134–140 (1956).

[12] Lesaint, F., Sigaud, O., Flagel, S. B., Robinson, T. E. & Khamassi, M. Modelling Individual Differences in the Form of Pavlovian Conditioned Approach Responses: A Dual Learning Systems Approach with Factored Representations. *PLoS Computational Biology* **10**, e1003466 (2014).

[13] Houlsby, N. M. *et al.* Cognitive Tomography Reveals Complex, Task-Independent Mental Representations. *Current Biology* **23**, 2169–2175 (2013). URL `http://linkinghub.elsevier.com/retrieve/pii/S0960982213011287`.

[14] Glaze, C. M., Kable, J. W. & Gold, J. I. Normative evidence accumulation in unpredictable environments. *eLife* **4**, e08825 (2015).

[15] Urai, A. E., Braun, A. & Donner, T. H. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications* **8**, 14637 (2017). URL `www.nature.com/naturecommunications`.

[16] Braun, A., Urai, A. E. & Donner, T. H. Adaptive History Biases Result from Confidence-Weighted Accumulation of past Choices. *Journal of Neuroscience* **38**, 2418–2429 (2018). URL `http://www.jneurosci.org/content/jneuro/38/10/2418.full.pdf`.

[17] Mathys, C. D. *et al.* Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience* **8**, 825 (2014). URL `http://journal.frontiersin.org/article/10.3389/fnhum.2014.00825/abstract`.

[18] Gael, J. V., Saatci, Y., Teh, Y. W. & Ghahramani, Z. Beam Sampling for the Infinite Hidden Markov Model. *Proceedings of the 25th international conference on Machine learning* 1088–1095 (2008). URL `http://mlg.eng.cam.ac.uk/zoubin/papers/icml2008ihmm.pdf`.

[19] Carpenter, R. & Williams, M. Neural computation of log likelihood in control of saccadic eye movements. *Nature* **377**, 59–62 (1995).

[20] Howard, J. H. & Howard, D. V. Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging* **12**, 634–656 (1997).

[21] Noorani, I. & Carpenter, R. The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews* **64**, 229–251 (2016). URL `https://ac.els-cdn.com/S0149763415301226/1-s2.0-S0149763415301226-main.pdf?{_}tid=d9daf898-2f42-4545-ab0e-8078b038399b{&}acdnat=1530110539{_}1265d3278c42a97caf3bf48597ca4a9b`.

[22] Carpenter, B. *et al.* Stan: A probabilistic programming language. *Journal of Statistical Software* **76** (2017).

[23] Song, S., Howard, J. H. & Howard, D. V. Implicit probabilistic sequence learning is independent of explicit awareness. *Learning & Memory* **14**, 167–176 (2007).

[24] Kóbor, A. *et al.* Tracking the implicit acquisition of nonadjacent transitional probabilities by ERPs. *Memory & Cognition* **47**, 1546–1566 (2019).

[25] Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M. & Donner, T. H. Confirmation Bias through Selective Overweighting of Choice-Consistent Evidence. *Current Biology* **28**, 3128–3135 (2018).

[26] Gershman, S. J. & Blei, D. M. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology* **56**, 1–12 (2012). URL `http://dx.doi.org/10.1016/j.jmp.2011.08.004. 1106.2697`.

[27] MacKay, D. J. C. *Information theory, inference and learning algorithms* (Cambridge University Press, 2003).

[28] Griffiths, T. L., Chater, N., Kemp, C., Perfors, A. & Tenenbaum, J. B. Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences* **17**, 354–367 (2010).

[29] Mitchell, T. M. The Need for Biases in Learning Generalizations. *Readings in Machine Learning* 184–191 (1980). URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.5466`.

[30] Richards, B. A. *et al.* A deep learning framework for neuroscience. *Nature Neuroscience* **22**, 1761–1770 (2019). URL `http://dx.doi.org/10.1038/s41593-019-0520-2`.

[31] Botvinick, M. *et al.* Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences* **23**, 408–422 (2019).

[32] Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521**, 452–459 (2015). URL `https://www.nature.com/articles/nature14541.pdf`.

[33] Wu, Z., Schrater, P., arXiv, X. P. a. p. & Senn, W. Inverse Rational Control: Inferring What You Think from How You Forage. *arXiv* 1805.09864 (2018).

[34] Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* **91**, 1402–1412 (2016).

[35] Barthelmé, S. & Mamassian, P. Evaluation of objective uncertainty in the visual system. *PLoS Computational Biology* **5**, e1000504 (2009).

[36] Bach, D. R. & Dolan, R. J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience* **13**, 572–586 (2012).

[37] Michael, E., de Gardelle, V., Nevado-Holgado, A. & Summerfield, C. Unreliable evidence: 2 sources of uncertainty during perceptual choice. *Cerebral Cortex* **25**, 935–947 (2015).

[38] Pouget, A., Drugowitsch, J. & Kepecs, A. Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience* **19**, 366–374 (2016).

[39] Kemp, C. & Tenenbaum, J. B. The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10687–10692 (2008). URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2492756{&}tool=pmcentrez{&}rendertype=abstract`. `arXiv:1408.1149`.

[40] Saxe, A. M., McClelland, J. L. & Ganguli, S. A mathematical theory of semantic development in deep neural networks. *Proceedings of the National Academy of Sciences* **116**, 11537–11546 (2019).

[41] Braun, D. A., Mehring, C. & Wolpert, D. M. Structure learning in action. *Behavioural brain research* **206**, 157–165 (2010).

[42] Goodman, N. D., Ullman, T. D. & Tenenbaum, J. B. Learning a theory of causality. *Psychological Review* **118**, 110–119 (2011). URL `http://doi.apa.org/getdoi.cfm?doi=10.1037/a0021336`.

[43] Austerweil, J. L. & Griffiths, T. L. A nonparametric Bayesian framework for constructing flexible feature representations. *Psychological Review* **120**, 817–851 (2013).

[44] Orbán, G., Fiser, J., Aslin, R. N. & Lengyel, M. Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2745–2750 (2008).

[45] Griffiths, T. L. & Tenenbaum, J. B. Optimal predictions in everyday cognition. *Psychological Science* **17**, 767–773 (2006).

[46] Gold, J. I. & Stocker, A. A. Visual Decision-Making in an Uncertain and Dynamic World. *Annual Review of Vision Science* **3**, 227–250 (2017). URL `http://www.annualreviews.org/doi/10.1146/annurev-vision-111815-114511`.

[47] Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion Decision Model: Current Issues and History (2016).

[48] Drugowitsch, J., Moreno-Bote, R. N., Churchland, A. K., Shadlen, M. N. & Pouget, A. The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience* **32**, 3612–3628 (2012).

[49] Nosofsky, R. M., Little, D. R., Donkin, C. & Fific, M. Short-Term Memory Scanning Viewed as Exemplar-Based Categorization. *Psychological Review* **118**, 280–315 (2011).

[50] Palmer, J., Huk, A. C. & Shadlen, M. N. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision* **5**, 376–404 (2005).

[51] Smith, P. L. & Ratcliff, R. An Integrated Theory of Attention and Decision Making in Visual Signal Detection. *Psychological Review* **116**, 283–317 (2009).

[52] Gold, J. I. & Shadlen, M. N. Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences* **5**, 10–16 (2001).

[53] Hanes, D. P. & Schall, J. D. Neural control of voluntary movement initiation. *Science* **274**, 427–430 (1996).

[54] Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Henry Holt and Co., Inc., USA, 1982).

[55] Gold, J. I. & Shadlen, M. N. The Neural Basis of Decision Making. *Annual Review of Neuroscience* **30**, 535–574 (2007).

Methods

Contents

1 Experiment: Materials and methods **31**

1.1 Participants 31

1.2 Tasks 31

1.3 Procedure 32

2 Background **33**

2.1 Models for Sequential Prediction 33

2.2 Hidden Markov Model 33

3 Cognitive Tomography **34**

3.1 Doubly Bayesian Model 34

3.2 Infinite Hidden Markov Model 35

3.2.1 Internal model of participants 36

3.2.2 Participant learning the model 37

3.2.3 Inference 39

3.3 Ideal Observer Model 40

3.3.1 Internal Model Components 40

3.3.2 Filtering 40

3.4 Response Time Model 41

3.5 Validation on synthetic datasets 41

4 Alternative models **42**

4.1 Markov model 42

4.1.1 Internal model of participants 42

4.1.2 Inference 42

4.1.3 Relation to sequential motor effects 44

4.1.4 Relation to HMM 44

4.2 Trigram model 44

4.2.1 Internal model of participants 44
4.2.2 Inference 45

5 Model Comparison **45**
5.1 Train and Test Datasets 45

6 Statistical Methods **46**
6.1 Error prediction 46

7 Supplementary Figures **46**

1 Experiment: Materials and methods

1.1 Participants

Twenty-five individuals (22 females and 3 males) aged between 18 and 22 (M_{Age} = 20.4 years, SD_{Age} = 1.0 years) took part in the experiment (we recruited 32 participants, but only 26 completed the experiment; we omitted one further participant because of a system error which resulted in partial loss of their experiment data). They were university students ($M_{Years\ of\ education}$ = 13.3 years, $SD_{education}$ = 1.0 years) from Budapest, Hungary. None of the participants reported history of developmental, psychiatric, neurological or sleep disorders, and they had normal or corrected-to-normal vision. They performed in the normal range on standard neuropsychological tests of short-term and working memory (Digit span task: $M = 6.48$, $SD = 1.15$, Counting span task: $M = 3.76$, $SD = 0.99$)[1]. Before the assessment, all participants gave signed informed consent and received course credit for participation. The study was approved by the Institutional Review Board of Eötvös Loránd University, Hungary.

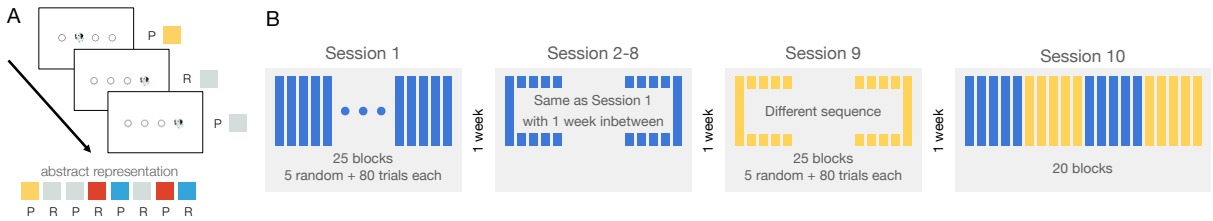
1.2 Tasks

Alternating Serial Reaction Time (ASRT) Task Learning was measured by the ASRT task [2,3]. In this task, a stimulus (a dog's head) appeared in one of four horizontally arranged empty circles on the screen and participants were asked to press the corresponding button as quickly and accurately as they could when the stimulus occurred. The computer was equipped with a keyboard with four heightened keys (Z, C, B, M on a QWERTY keyboard), each corresponding to a circle in a horizontal arrangement. Participants were asked to respond to the stimuli using their middle- and index fingers bimanually. The stimulus remained on the

screen until the participant pressed the correct button. The next stimulus appeared after a 120 ms response-to-stimulus-interval (RSI). The task was presented in blocks of 85 stimuli: unbeknownst to the participants, after the first five warm-up trials consisting of random stimuli, an 8-element alternating sequence was presented ten times (e.g., 2r4r3r1r, where each number represents one of the four circles on the screen and r represents a randomly selected circle out of the four possible ones).

1.3 Procedure

There were ten sessions in the experiment, with one-week delay between the consecutive sessions. Participants performed the ASRT task with the same sequence in the first eight sessions, then an interfering sequence was introduced in Session 9, and both (original and interfering) sequences were tested in Session 10 (see 1). Participants were not given any information about the regularity that was embedded in the task in any of the sessions [3]. They were informed that the main aim of the study was to test how extended practice affected performance on a simple reaction time task. Therefore, we emphasized performing the task as accurately and as fast as they could. Between blocks, the participants received feedback about their average accuracy and reaction time presented on the screen, and then they had a rest period of between 10 and 20 s before starting the next block. On Days 1-9, the ASRT consisted of 25 blocks. One block took about 1-1.5 min, therefore the task took approximately 30 min. For each participant, one of the six unique permutations of the four possible ASRT sequence stimuli was selected in a pseudo-random manner[2–4]. The ASRT task was performed with the same pattern sequence in Sessions 1-8. In Session 9, the ASRT was performed with a new interfering pattern sequence. In Session 10, participants performed 20 blocks of the ASRT task switching between the pattern sequences of Sessions 1-8 and Session 9 every five blocks. In Session 10, the task took approximately 24 min. After performing the ASRT task in Session 10, we tested the amount of explicit knowledge the participants acquired about the task with a short questionnaire. This short questionnaire[3,5] included two questions: "Have you noticed anything special regarding the task?" and "Have you noticed some regularity in the sequence of stimuli?". The participants did not discover the true probabilistic sequence structure.



Extended Data Figure 1: *A* Experimental stimuli and abstract representation used in the paper. *B* Design of the experiment. The experiment consisted of ten sessions, separated by a one-week delay. On Days 1-8, participants performed the ASRT task with sequence 1 throughout 25 blocks (5 epochs) each sessions. On Day 9, an interfering sequence (sequence 2) was introduced. Both sequences were tested on Day 10 with blocks of 5 alternating.

2 Background

2.1 Models for Sequential Prediction

The experimental stimuli form a sequence of discrete observations in discrete time, $\{Y_t\}_{t=1}^{T}$. The task is therefore to predict the upcoming stimulus conditioned on the history of observations:

$$P(Y_{T+1}|Y_1, Y_2, ..., Y_T) \tag{1}$$

In practical terms, learning a model for this temporal prediction task requires imposing a structure over these conditional distributions. Without structural assumptions, there is no statistical dependence among different histories, that is, there is no generalisation from history to future observations.

In the following section we introduce a computational model, the Hidden Markov Model, which can provide a general language for solutions of this problem. It can express arbitrarily complex models given sufficiently large amounts of data. In order to remain as general as possible, we will consider a model space (infinite Hidden Markov Models [6]) which can model all the possible distributions in equation 1. Moreover, we would like to achieve this while being able to express inductive biases in this language which are useful for constraining the possible models in the limited data case.

2.2 Hidden Markov Model

Formally, a Hidden Markov Model comprises of a sequence of hidden states $\{S_t\}_{t=1}^{\infty}$ and a sequence of observations $\{Y_t\}_{t=1}^{\infty}$. In this work we take both the latent states and the observations to be discrete, that is $S_t, Y_t \in \mathbb{N}$ sequence of hidden (latent) states constitute a discrete Markov-chain with transition probabilities $\pi_{ij} = P(S_{t+1} = j|S_t = i)$. In a Markov-chain, the sequence element S_t is conditionally independent of the history conditioned on the

previous state and the transition probabilities:

$$S_t \perp (S_1, S_2, ..., S_{t-2}) \mid S_{t-1}, \pi$$

At (discrete) time t, observation Y_t is governed by the latent state S_t. The observations are generated independently and identically, conditioned on the (latent) state:
$P(Y_t = y \mid S_t) = \phi_{s,y}$ and $Y_t \perp (Y_1, Y_2, ..., Y_{t-1}, S_1, S_2, ..., S_{t-1}) \mid S_t, \phi$

Importantly, since the latent state can incorporate arbitrary information (identical observations at different time-points can correspond to different states), assuming arbitrarily many latent states, we get a completely general solution for the prediction problem in 1. With an adequate prior (e.g. the Hierarchical Dirichlet Process [7] we can learn such structures efficiently [6]). In practical terms this means the assumed number (e.g. posterior mean number) of different states increases with the length of the observation sequence. Loosely: until proven otherwise, a simpler structure is assumed.

3 Cognitive Tomography

We construct a model of behaviour which consists of two parts:

1. An internal model maintained by the participant, which formalizes how latent states assumed to underlie observations evolve and how these states are linked to observations.

2. A model relating the prediction of participants' internal model to their responses (response time model).

3.1 Doubly Bayesian Model

Due to the uncertainty of the participants about the true model and actual state of the stimuli and to the uncertainty of the experimenter about the model maintained by participants and about the actual state of this internal model, the problem can be described as doubly Bayesian. We do Bayesian inference over an internal representation of individuals who themselves do Bayesian inference. Elements of the experimenter's model are introduced in following sections.

Prediction of response times can be described by the following algorithm:

1. We take posterior samples from the behavioural model which consists of parameters of the internal model and the response time model conditioned on data from ten consecutive blocks of trials, where:

   (a) all stimuli, and

(b) response times (with incorrect trials', first five random trials' response times, and response times smaller than 180 msec in each block removed)[1]

are included.

2. For each of the 60 posterior model samples we compute predicted response times by:

   (a) filtering the belief over the latent state over the entire sequence

   (b) produce subjective probabilities for each trial

   (c) produce response time prediction (MAP estimate conditioned on the subjective probability and the response time parameters of the model sample)

   Then we marginalize (i.e. average) over the response time predictions of model samples.

3. We evaluate model performance by computing the R^2 explained variance measure of the predicted response times on the response times of the test dataset.

Note: since actual beliefs depend on past beliefs, one can think of the belief sequence as the path of a light-ray in a large dimensional fog (representing the state uncertainty). During inference, we have a noisy measurement of the light-ray in different points of time and we would like to reconstruct the best explanation of the observation sequence (response times) in terms of a hidden path. As for prediction, the model produces response time predictions for the entire stimulus sequence with no further feedback of response times (i.e. estimated internal beliefs are not updated based on what response time the participant produced on given trials).

3.2 Infinite Hidden Markov Model

The infinite Hidden Markov Model is a non-parametric extension of the Hidden Markov Model, assuming infinitely many states. There is a hierarchical prior imposed over the state transition matrix and the so-called emission distributions relating the latent (hidden) states to observations (Fig. 2).

The hierarchical prior we used is exactly the one defined in [6]. We extended their implementation of their model to a doubly Bayesian behavioural model including the response time.

[1] According to the original formulation in [8], fast response times come from an alternative distribution. We cut off the fast response times (as in [9]) at the fixed 180 msec value. However, we did not fit the cut-off time parameter.

3.2.1 Internal model of participants

A participant is assumed to learn a probabilistic model of the sequence which is formalized as an infinite Hidden Markov Model. At (discrete) time t, observation Y_t is governed by a latent (not directly observable) state S_t. The states $\{S_t\}_{t=1,2,...}$ constitute a Markov-chain, which means the following:

$$p(S_t|S_1, S_2, ..., S_{t-1}) = p(S_t|S_{t-1})$$

That is, the state S_{t-1} holds all information about past regarding the possible evolution of system. In other terms, conditioning on state S_{t-1} renders S_t and all previous states $S_1, S_2, ..., S_{t-2}$ statistically independent.

The observation Y_t at time t is independent of all other observations, conditioned on the latent state S_t (and the model parameters). That is, once the state of the system is decided, the actual previous observations are independent of Y_t.

The parameters governing the state transitions are aggregated in the parameter matrix π:

$$\pi_{i,j} = p(S_t = j|S_{t-1} = i)\ \forall t$$

The observation distributions are given by the parameter matrix ϕ:

$$\phi_{i,k} = p(Y_t = k|S_t = i)$$

At any given time during the task, we assume the participant had estimated the parameters π and ϕ and uses these (point estimates) to do exact filtering over the sequence of observations. That is, in each trial they use the evidence provided by the current stimulus to update their belief over the latent state of the sequence. When doing computations with the participant's internal model, we hold the internal model fixed within shorter time-scales of the task (e.g. one session). The participant represents their belief about the current latent state of the system by a posterior distribution, updated by each incoming observation, while always conditioning on their current estimates $\hat{\pi}$ and $\hat{\phi}$ of π and ϕ respectively.

$$
\begin{aligned}
\hat{s}_t := p(S_t = s_t | Y_1 = y_1, Y_2 = y_2, ..., y_t) \propto & \\
& \propto p(y_t|s_t)p(s_t|y_1, y_2, ..., y_{t-1}) \\
& = \sum_{s_{t-1}} \hat{\phi}_{s_t, y_t} p(s_t|s_{t-1}) p(s_{t-1}|y_1, ..., y_{t-1}) \\
& = \sum_{s_{t-1}} \hat{\phi}_{s_t, y_t} \hat{\pi}_{s_{t-1}, s_t} \hat{s}_{t-1}
\end{aligned}
$$

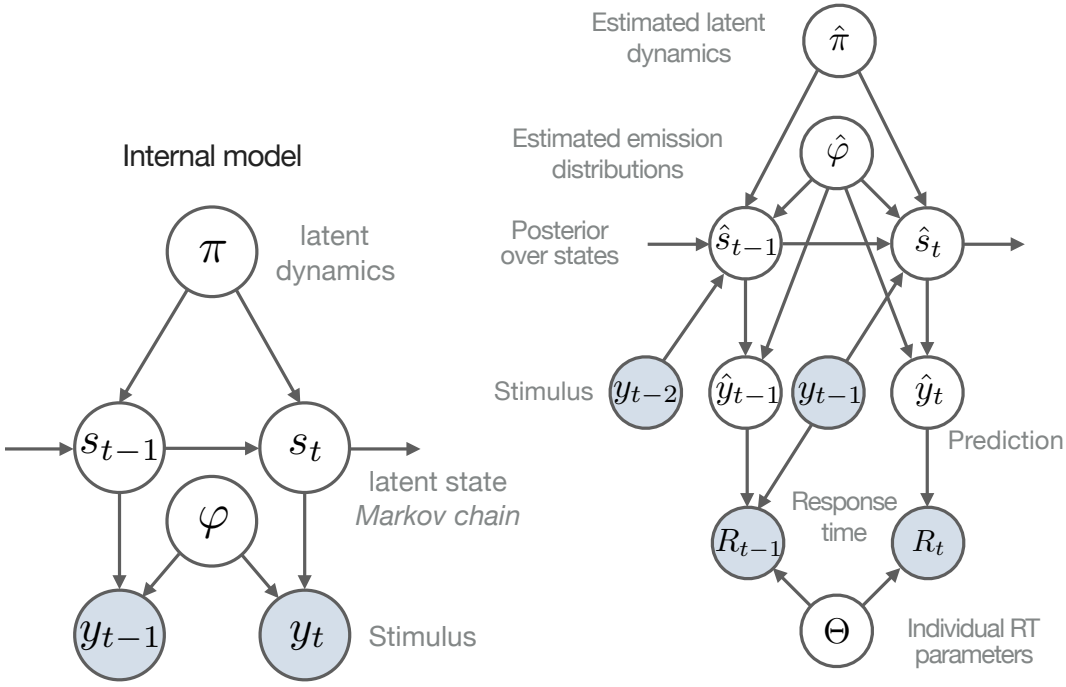
For filtering, stimuli of all trials (including initial random trials at the beginning of each block and stimuli in trials where participant hit the wrong key initially) are used. That is, even if response times are not considered when doing inference over the participant's internal model, the participant is assumed to update their internal beliefs based on the stimulus shown.

Prediction of the next stimulus is computed by marginalizing over the latent state posterior distribution:

$$
\begin{aligned}
p(y_{t+1}|y_1, y_2, ..., y_t) = & \\
& = \sum_{s_{t+1}} p(y_{t+1}|s_{t+1}) p(s_{t+1}|y_1, y_2, ..., y_t) \\
& = \sum_{s_{t+1}} \hat{\phi}_{s_{t+1} y_{t+1}} p(s_{t+1}|y_1, y_2, ..., y_t) \\
& = \sum_{s_t, s_{t+1}} \hat{\phi}_{s_{t+1} y_{t+1}} p(s_{t+1}|s_t) p(s_t|y_1, y_2, ..., y_t) \\
& = \sum_{s_t, s_{t+1}} \hat{\phi}_{s_{t+1} y_{t+1}} \hat{\pi}_{s_t, s_{t+1}} \hat{s}_t
\end{aligned}
$$

3.2.2 Participant learning the model

Throughout the execution of the task, the internal model of the participants is continually updating. We do not directly model the computation of the participants that estimates the current π and ϕ parameters. That is, within a given train or test dataset (10 consecutive blocks) we hold π and ϕ fixed. We do allow, however, for these estimates of π and ϕ to change between sessions.



Extended Data Figure 2: Graphical representation of internal model and generative model of behaviour. Left: Internal model, generative model of the sequence assumed by the participant. Right: generative model of behaviour.

Variable	Notation	Within train/test	Within session, between train-test	Between sessions	Between participants
State transition distribution	$\hat{\pi}$	No	No	Yes	Yes
Observation distribution	$\hat{\phi}$	No	No	Yes	Yes
State belief	$\hat{s}_t$	Yes	Yes	Yes	Yes
Response time parameters	τ_0, μ, σ	No	No	No	No
Prior of observation distribution	H	No	No	No	No
Hierarchical prior over state transitions	α, β	No	No	No	No

Table 1: Summary of when model parameters are allowed to change.

Variable	Prior
State transition distribution	$\hat{\pi}_i \sim \text{Dirichlet}(\alpha_0/K, ..., \alpha_0/K, \alpha_0/K \cdot \epsilon)$
Observation distribution	$\hat{\phi} \sim \text{Dirichlet}(0.8, 0.8, 0.8, 0.8)$
Response time parameters	$\tau_0 \sim \Gamma(1, 10)$ $\mu \sim \Gamma(1, 0.1)$ $\sigma \sim \Gamma(1, 0.01)$
Hierarchical prior over state transitions	$\alpha = 1.3$ $\gamma = 3.8$

Table 2: Parameter priors. Values of the hierarchical prior over state transitions taken from [6]

3.2.3 Inference

We do Approximate Bayesian Inference using a custom sampling method that mixes steps of a Hamiltonian Monte Carlo (HMC) and a Gibbs sampler which samples a slicing parameter (see [6]). The priors used in the model are listed in Table 2.

In order to handle the infinitely many possible states, we use a modified version of the slice sampling method described in [6]. In the original beam sampling algorithm, the authors sample the latent state sequence and make use of the slicing variable to constrain the set of used states to a finite set. They sample the latent sequence and the slicing variables in an alternating fashion. In our case we do not sample latent state sequences, instead, we have to estimate the subjective belief sequence over the latent states. In this latter case, the posterior belief is infinite dimensional and we use slicing to approximate this infinite-dimensional computation with a finite one. At each sampling step, we only look at the latent state belief distribution's $1 - \epsilon$ support where ϵ is sampled from $\text{Uniform}(0.02, 0.2)$.

Four independently and randomly initialised Markov Chains were sampled with 1600 steps of the slice sampling (outer Gibbs-sampling chain) and 30 NUTS steps were taken inbetween the slice sampling steps each time. Samples from the second half of each chain were used to check if estimates of response time parameter means and confidence intervals were identical. For prediction, last 60 unique samples were used from each chain because prediction performance saturates at this number of samples.

3.3 Ideal Observer Model

3.3.1 Internal Model Components

We formalise the ideal observer the following way: at any given point of the experiment, the ideal observer entertains an internal dynamical model comprising of two parts: latent dynamics (the transition probabilities between latent states) and an observational model (conditional distributions of observations conditioned on the latent state).

3.3.2 Filtering

In order to produce predictions for the upcoming observation, conditioning on a fixed model, the ideal observer solves the filtering problem:

$$
\begin{aligned}
& P(Y_t|Y_1, Y_2, ..., Y_{t-1}, \Pi, \Phi) = \\
= & \sum_{s_t=1}^{\infty} P(Y_t|S_t = s_t) \cdot P(S_t = s_t|Y_1, Y_2, ..., Y_{t-1}) \\
= & \sum_{s_t=1}^{\infty} \Phi_{s_t,y_t} \cdot P(S_t = s_t|Y_1, Y_2, ..., Y_{t-1}) \\
= & \sum_{s_t=1}^{\infty} \sum_{s_{t-1}=1}^{\infty} \Phi_{s_t,y_t} \cdot P(S_t = s_t|S_{t-1} = s_{t-1}) \cdot \\
& P(S_{t-1} = s_{t-1}|Y_1, Y_2, ..., Y_{t-1}) \\
= & \sum_{s_t=1}^{\infty} \sum_{s_{t-1}=1}^{\infty} \Phi_{s_t,y_t} \cdot \Pi_{s_{t-1},s_t} \cdot P(S_{t-1} = s_{t-1}|Y_1, Y_2, ..., Y_{t-1})
\end{aligned}
$$

The term filtering is used because as we deduced, the relevant quantity is $P(S_{t-1}|Y_1, Y_2, ..., Y_{t-1})$ which can be filtered through our observations. We carry on this quantity and can calculate it for the next time-step using our model parameters and the observation Y_t.

Importantly, instead of sampling one possible latent trajectory, we have to marginalise over these latent sequences to obtain our prediction for the upcoming stimulus. That is, our prediction is the aggregate of the predictions many possible latent pasts. We combine the predictions of 'had these been the sequence of causes of my past experiences, I should see this' for all possible hypothesised latent cause sequences.

3.4 Response Time Model

In order to connect the predictions of the internal model to measured behaviour, we need to employ a generative model of response times in the form of a conditional probability distribution conditioned on the subjective predicted probability of the upcoming stimulus. To achieve this, we employ the reaction time model of [8] , which in its original formulation states that the majority of saccadic response times come from a reciprocal Normal distribution.

Further studies suggest choice response time distribution should have a similar form[10,11]. However, in other formulations, there is no explicit dependence of the distribution of the RT in a single trial depending on the subjective predicted probability, hence those models are inadequate for our purposes. The generative model for correct response times (LATER model[8]) is:

$$r_n \sim \text{Normal}(\mu, \sigma)$$

$$RT_n = \frac{\theta_0 - \log(p_n)}{r_n}$$

where p_n is the subjective probability (output of the internal model) corresponding to the actual upcoming stimulus and μ, σ, θ_0 are the parameters characterising an individual's response time model. These parameters jointly describe the mean and variance of the response times. Note that in our experiment these parameters comprise all idiosyncratic effects at hand, namely the individual's state, their response times' sensitivity to subjective predicted probabilities, the effects of instruction influencing speed-accuracy trade-off.

The response time parameters are jointly inferred along with the internal representations (dynamical model, observation distribution, latent state inference).

3.5 Validation on synthetic datasets

In order to validate our behavioural model as well as our inference method, we looked at how well we can recover subjective probabilities on a synthetic dataset. We used the algorithm in [6] on synthetic ASRT data to infer a first set of three different internal models from different levels of exposure. These models represent internal models of different synthetic participants (Fig. 3A). We take these models as the ground truth for our synthetic experiment. We trained one model on 640, 1280 and 2400 trials of ASRT stimuli. We then generated response times from the generative model with three parameter settings for each of τ_0, μ, and σ resulting in a total of $3^3 = 27$ different synthetic response time sequences. The resulting response

time distribution's variance is influenced by all four factors – the subjective probabilities (which depends on the internal model) and the three response time parameters. The standard deviation of the response times is an appropriate measure since this can be also computed for data obtained from human participants. We generated the response times for 10 ASRT blocks (the same number we used for inference on human data). Standard deviation of the resulting response times (*symbol colours* on Extended Data Fig. 3B) arise from the interaction of all parameters. Different combinations of the response time parameters resulting in the same standard deviation are marked by identical symbols. Then, we used the CT inference method to generate a second set of (posterior) internal model samples. We computed the same model performance measure as for human data (response time prediction performance) and also the performances for subjective probability predictions (Extended Data Fig. 3B). The results show that the prediction performance of the subjective probabilities exceeds that of the individual response times. Also, as seen on Extended Data Fig. 3C, standard deviation of human participant's response times are considerably below the threshold where our CT inference method in the synthetic data experiment was successful.

4 Alternative models

4.1 Markov model

4.1.1 Internal model of participants

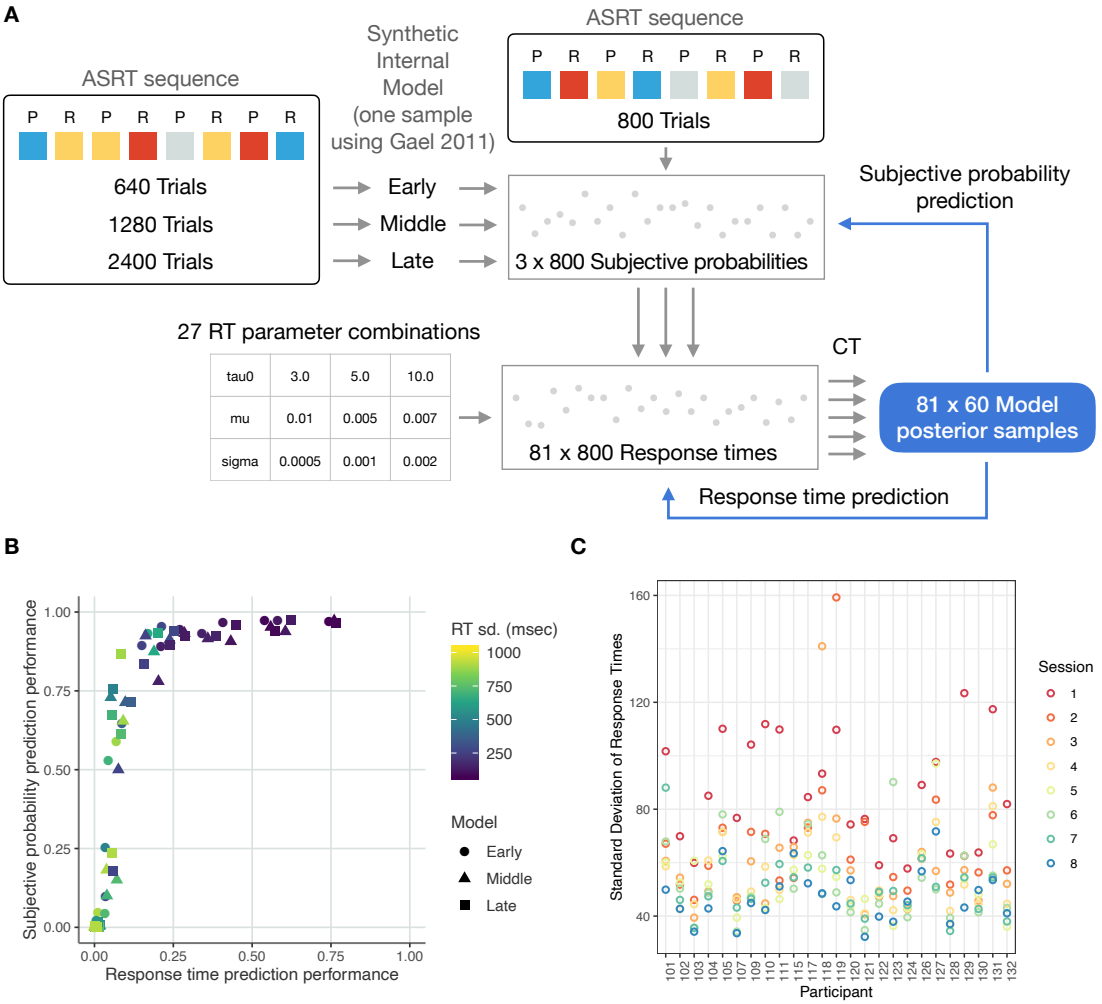
According to this model, the participants assume that the sequence of observations constitute a Markov-chain. That is, for the sequence of observations y_t, we have

$$p(y_t|y_1, y_2, ..., y_{t-1}) = p(y_t|y_{t-1}) \quad \forall t$$

The above equation states that the next observation is independent of all previous observations given the previous observation. This is equivalent to saying that all information (besides parameters governing the sequence) about the state of the sequence is included in the previous observation.

4.1.2 Inference

We use the same parameter priors for the response time model as for the iHMM model and the prior for transition probabilities $\pi_i \sim \text{Dirichlet}(\alpha_0/K, \alpha_0/K, \alpha_0/K, \alpha_0/K)$, where K is the number of states, in this case 4.



Extended Data Figure 3: Synthetic data experiment. *A* We first sampled three versions of synthetic internal models using the original iHMM inference method in [6]. The internal models of the synthetic participants differ in their experience (as how many ASRT trials they had seen) – resulting in an "early", "middle" and "late" model. Then, we generated subjective probability values for each model on a new set of ASRT stimuli (holding the pattern sequence intact). We ran our inference method for 81 synthetic datasets with different parameter settings (*symbols with different colors and shapes*). We use the same number of response times as with the human participants to recover the internal models. *B* Results of our synthetic data experiment. *Symbol colours* correspond to the response time standard deviation. The result shows that while the response time prediction may be at a lower level, the latent predictive probabilities can still be inferred with relatively high accuracy. This shows the inference method can recover the latent structure from a generated response time sequence. *C* Standard deviations of response times of individuals in the first eight experimental sessions.

Four independently and randomly initialised Markov Chains were sampled with 1600 steps taken with the NUTS sampler in STAN. Samples from the second half of each chain were used to check if response time parameter estimates' means and confidence intervals were identical. For prediction, last 60 samples were used from each chain.

4.1.3 Relation to sequential motor effects

The prediction a Markov internal model makes can be rephrased as "how quick a participant will respond to the next stimulus j if the current stimulus is i". In our paradigm, the mapping between the stimuli and the required responses are fixed. For this reason, predictions of a Markov model are indistinguishable from predictions made by a "sequential motor bias" model positing a dependency of the response time on which successive key strokes are made (i.e. which finger the participant uses for the current trial and which one for the previous trial).

Our investigation implies that the Markov models inferred are a superposition of sequential motor effects and an internal model that is Markovian (see Results for our argument).

4.1.4 Relation to HMM

Note that Markov models are subset of Hidden Markov Models. We can always write a Markov model as an HMM if we have a matching number of observation values and latent state values and each observation is unique to a state.

This is particularly important since for an HMM for which the above condition holds, there is an equivalent Markov chain that describes the exact same sequence structure. This is the reason why we term some of the internal models identified by our iHMM method "Markov-like", since they are closely approximated by an actual Markov model.

4.2 Trigram model

The model we describe here is also referred to as 'triplet model' in previous works using the ASRT paradigm. We use the term trigram since it is more commonly used in a sequential prediction modelling context.

4.2.1 Internal model of participants

The model, established in prior literature, sorts trials into High probability and Low probability triplets. This is equivalent to assuming that the participant uses a two-back (or trigram)

model for prediction, predicting the most-likely stimulus conditioning on the previous two observations. Due to the ground truth generative model of the task

4.2.2 Inference

The trigram model has no parameters fitted. Performance is evaluated by the R^2 measure between the response times and the binary variable (high vs low trials) provided by the trigram model.

5 Model Comparison

Since not all models considered are Bayesian (i.e. provide an explicit marginal log-likelihood for the response times), we chose to compare models based on explained variance of response times on a test set. Each model produces response time predictions for each trial and each individual separately. When evaluating on a given test set, in order to control for a shift in mean not related to the inherent structure of the response times, we use R^2 as our performance metric. That is equivalent to assuming that the actual observed response times come from a linear model with the predicted response time as mean and an additive homoscedastic (equal variance irrespective of predicted response time) normal noise term.

The R^2 values were calculated separately for each individual's trials.

5.1 Train and Test Datasets

For the reason described in the above paragraph, for each day (out of 10) of the experiment, out of the 25 blocks each day, we selected blocks 11-20 as a training dataset and blocks 1-10 as test datasets. The main reason for this choice is that on each day in the initial few blocks participants may be engaged in a warm-up phenomenon which fundamentally alters their behaviour in the task. If we use the first 10 blocks as test data, the performance metric may be influenced by 10-30% depending on how many blocks include altered behaviour. However, if we used this part as training data, the whole internal model inference would shift fundamentally, since our inference algorithm assumes a fixed model the entirety of the 10 blocks.

During model inference (train dataset) and performance evaluation (test dataset) the first five random trials and all incorrect response trials' response times are not considered.

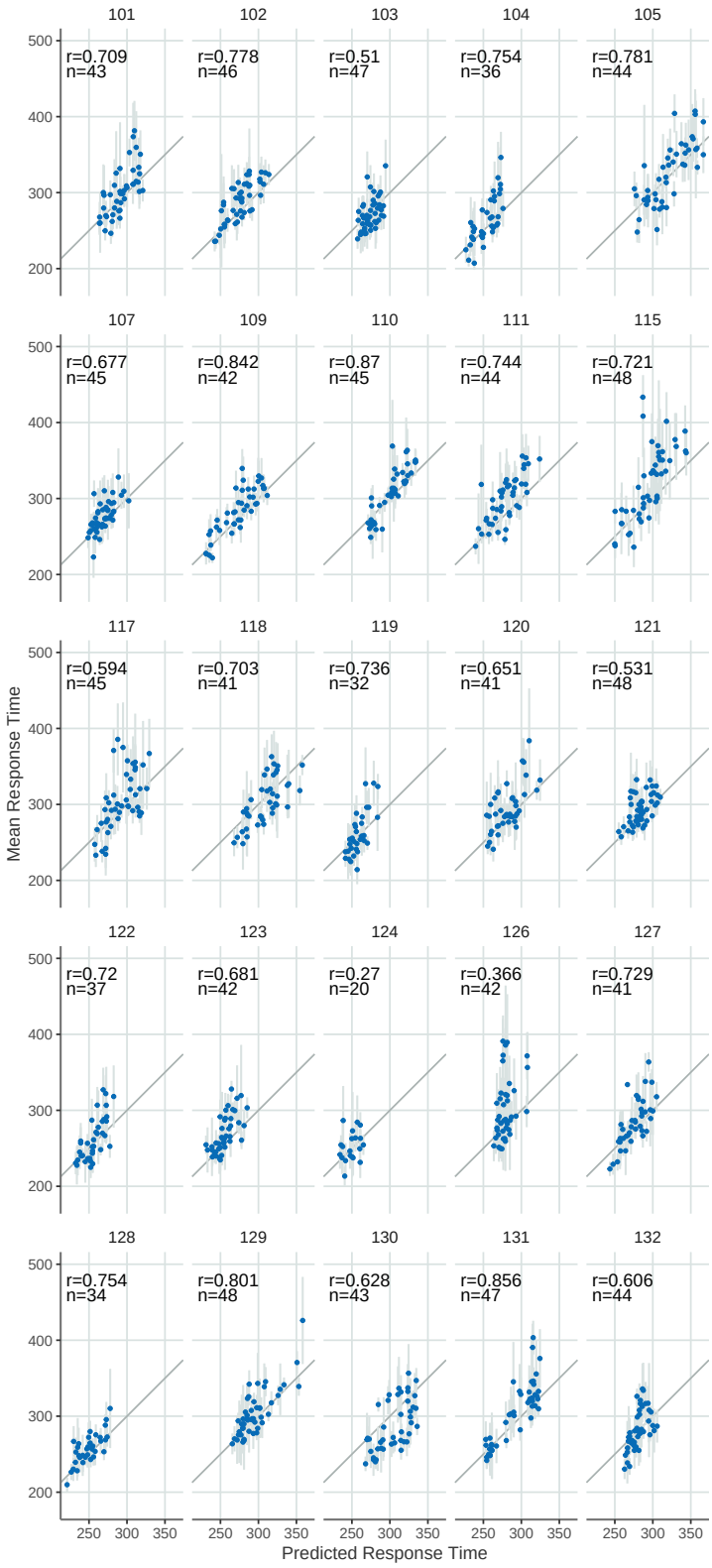
6 Statistical Methods

Normality was not checked prior to t-test comparisons. All reported correlations were computed using Pearson's correlation. T-tests are paired sample tests whenever there is a within-subject comparison. All binomial tests are one-sided. For effect sizes we calculated Cohen's d using the lsr R package.

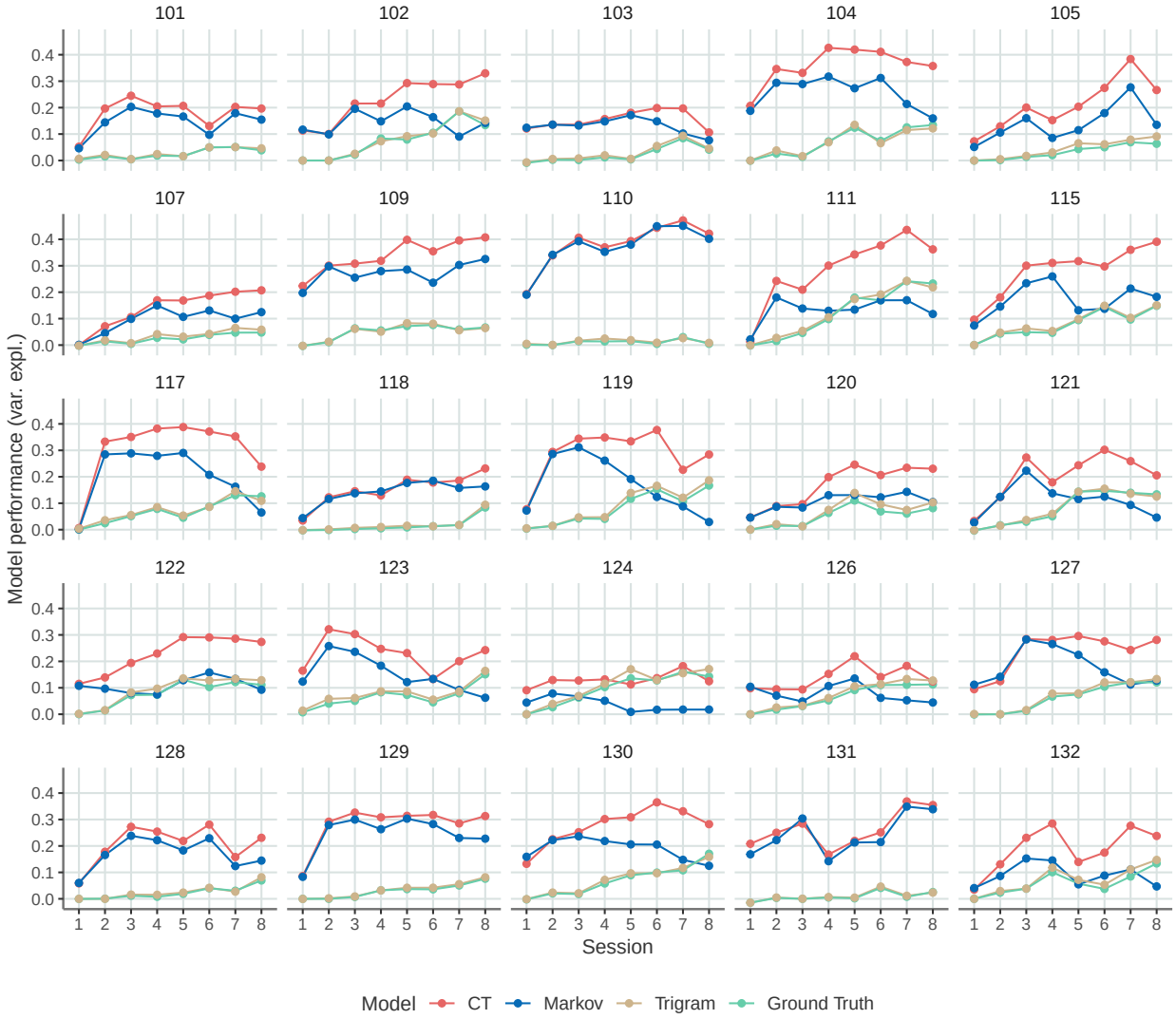
6.1 Error prediction

Just as with response time prediction, the model outputs (for each posterior model sample) a subjective probability estimate for each one of the four possible stimuli for each trial. Then, we take mean over these probability estimates over last 60 unique samples of each chain. We decided on using 60 samples since model performances saturate at this number. On Fig. 5 we compute the rank among the four probability estimates of the stimulus and the choice in correct and incorrect trials. Then, based on the subjective probability estimates of the actual occurring stimulus, we plot the receiver operating characteristic curve for predicting whether a given trial will result in an error. This is done by moving a threshold value from 0 to 1 and predicting correct trial if the subjective probability of the upcoming stimulus is above the threshold and an erroneous trial otherwise. The trigram model has two points (other than the $(0, 0)$ and $(1, 1)$ points). This is because the trigram model predicts 0.25 probability for the all stimuli for the first two trials in each block and 0.625 probability for the more high probability trigram element in all other trials and 0.125 for the other stimuli.

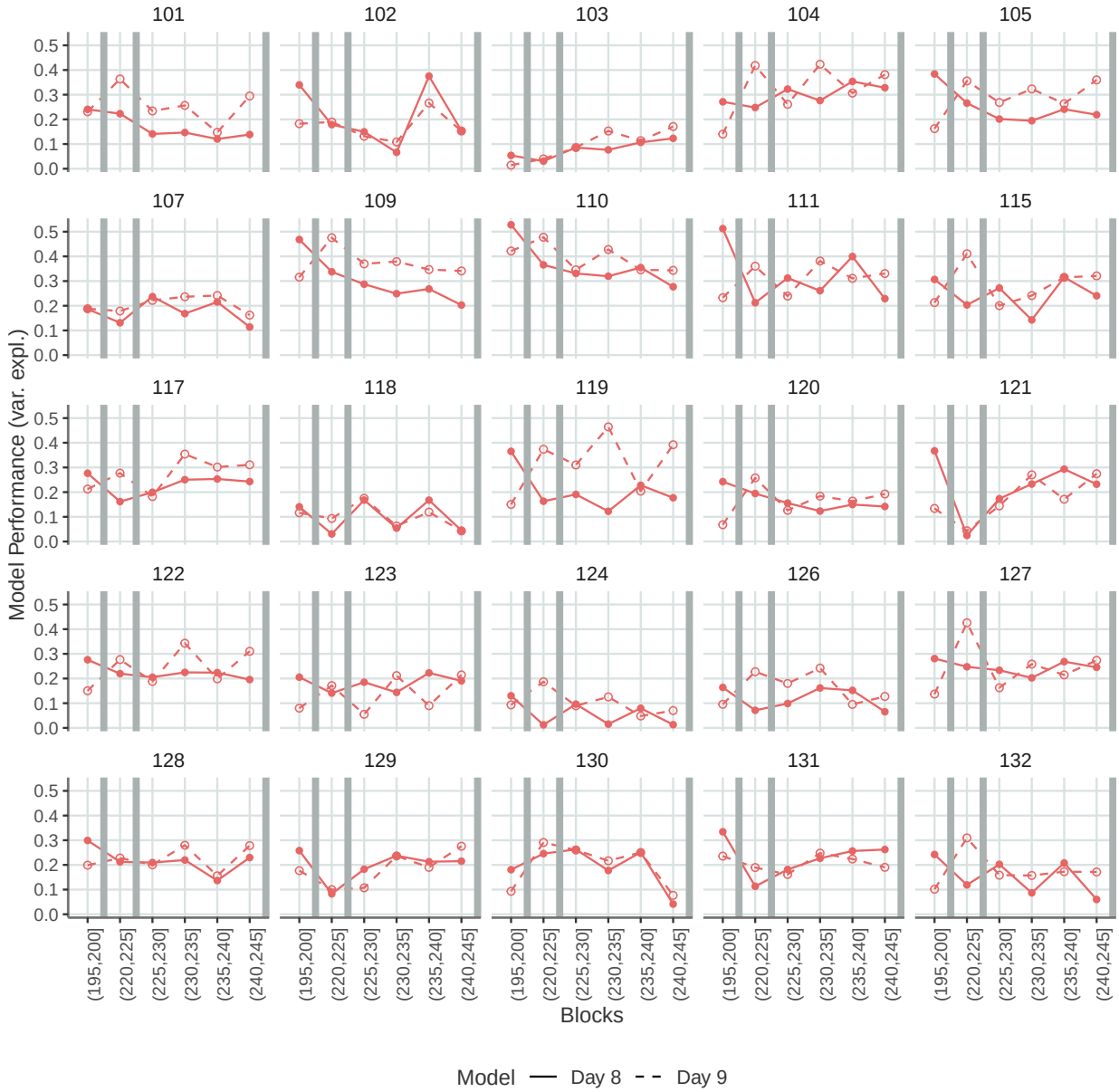
7 Supplementary Figures



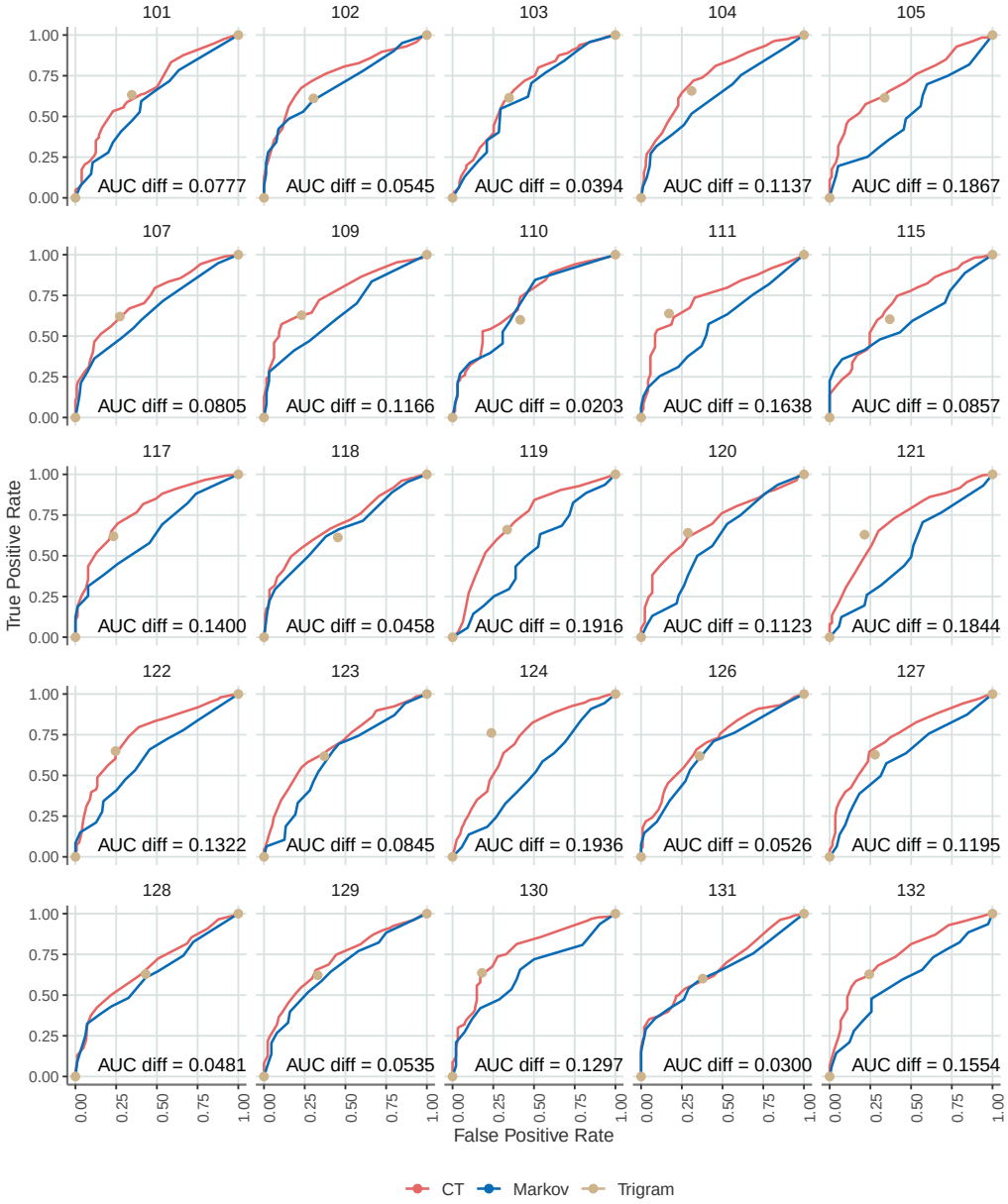
Extended Data Figure 4: Predicted response time means vs. measured response time means on the test set on Day 8 of the experiment grouped by three element sequences for each participant separately. Only those sequences were included which had at least 5 measured correct response times in order to limit the standard error over the measured response time mean. Error bars show 2 s.e.m.



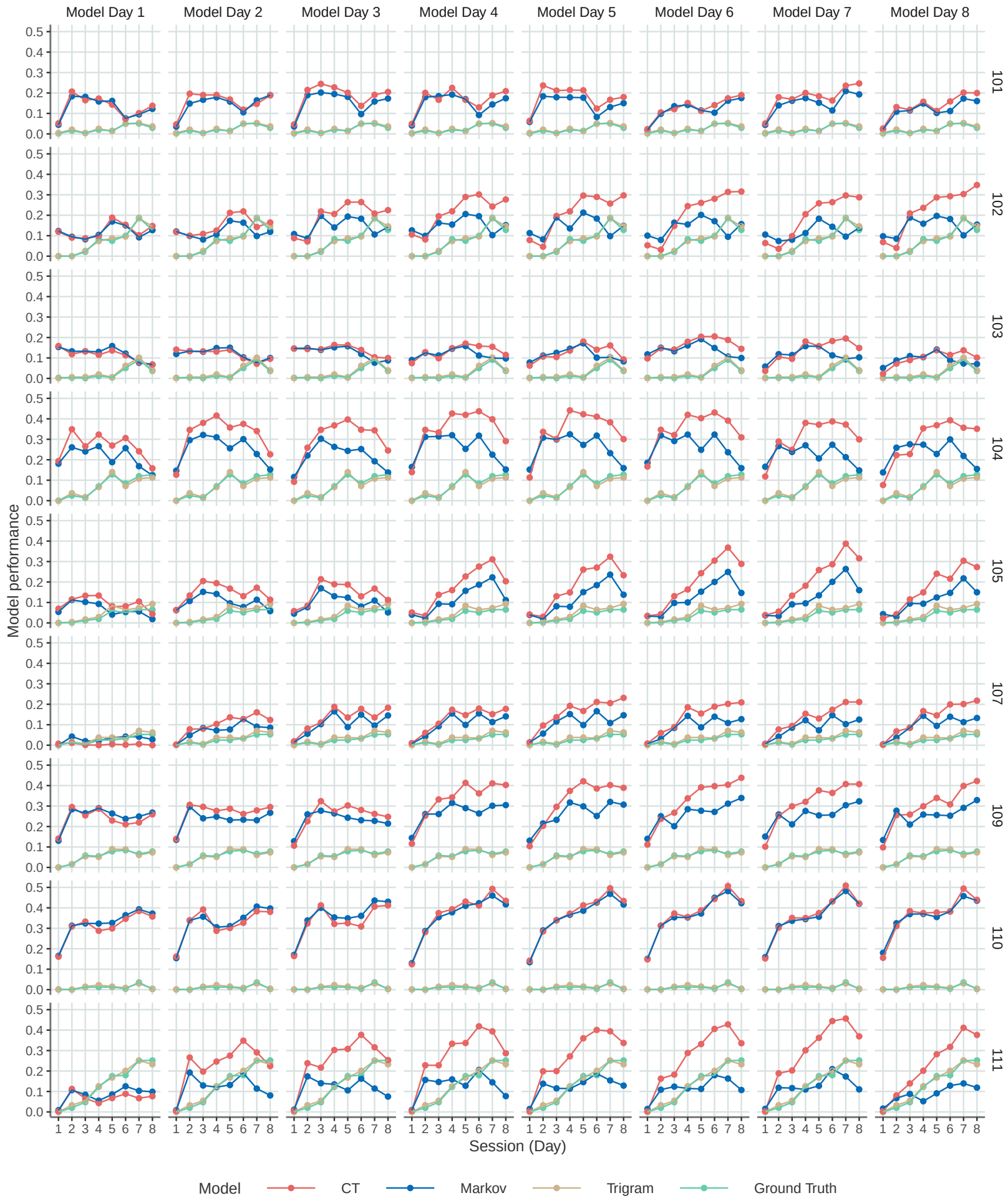
Extended Data Figure 5: Model performances for all models and all participants individually.

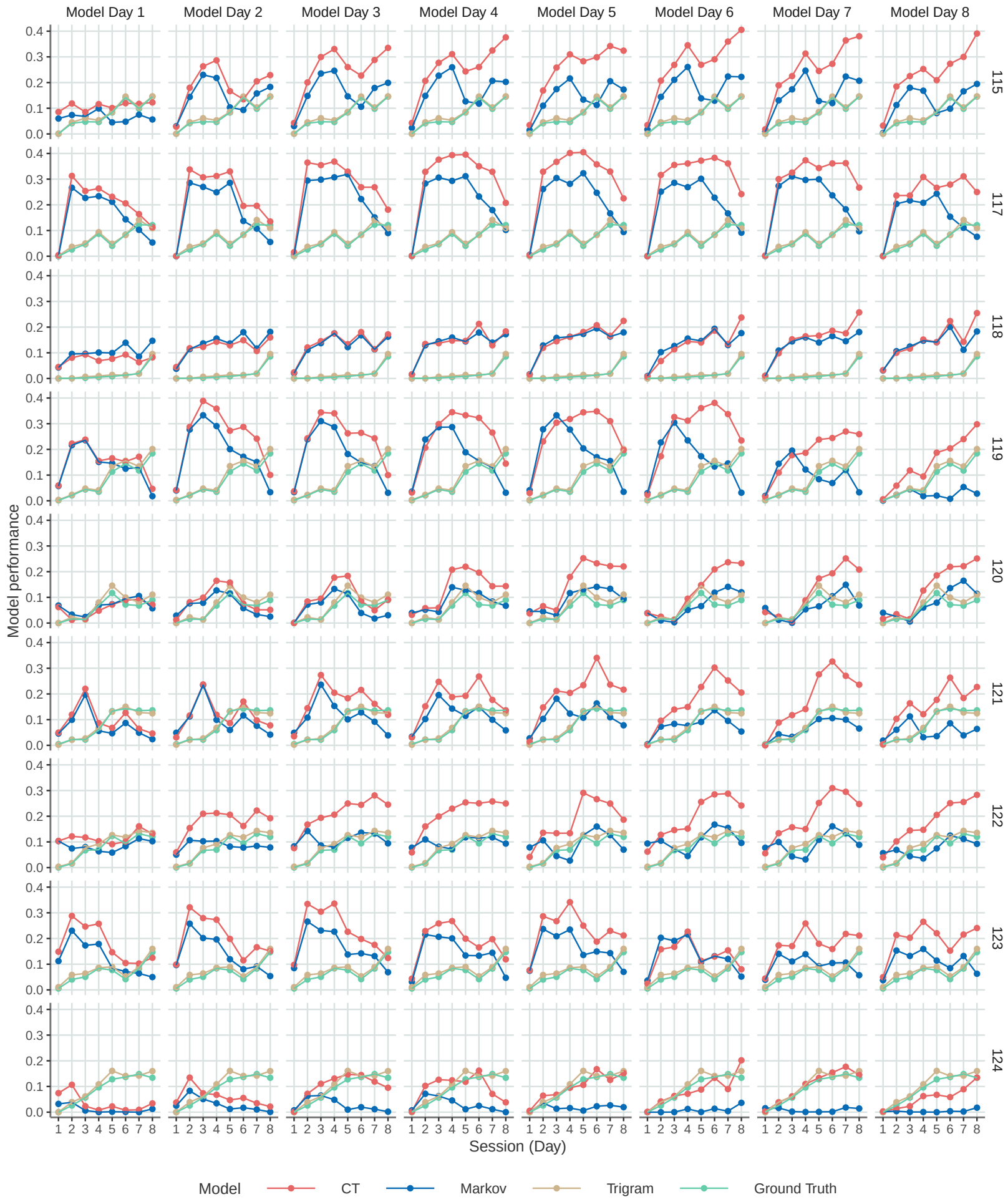


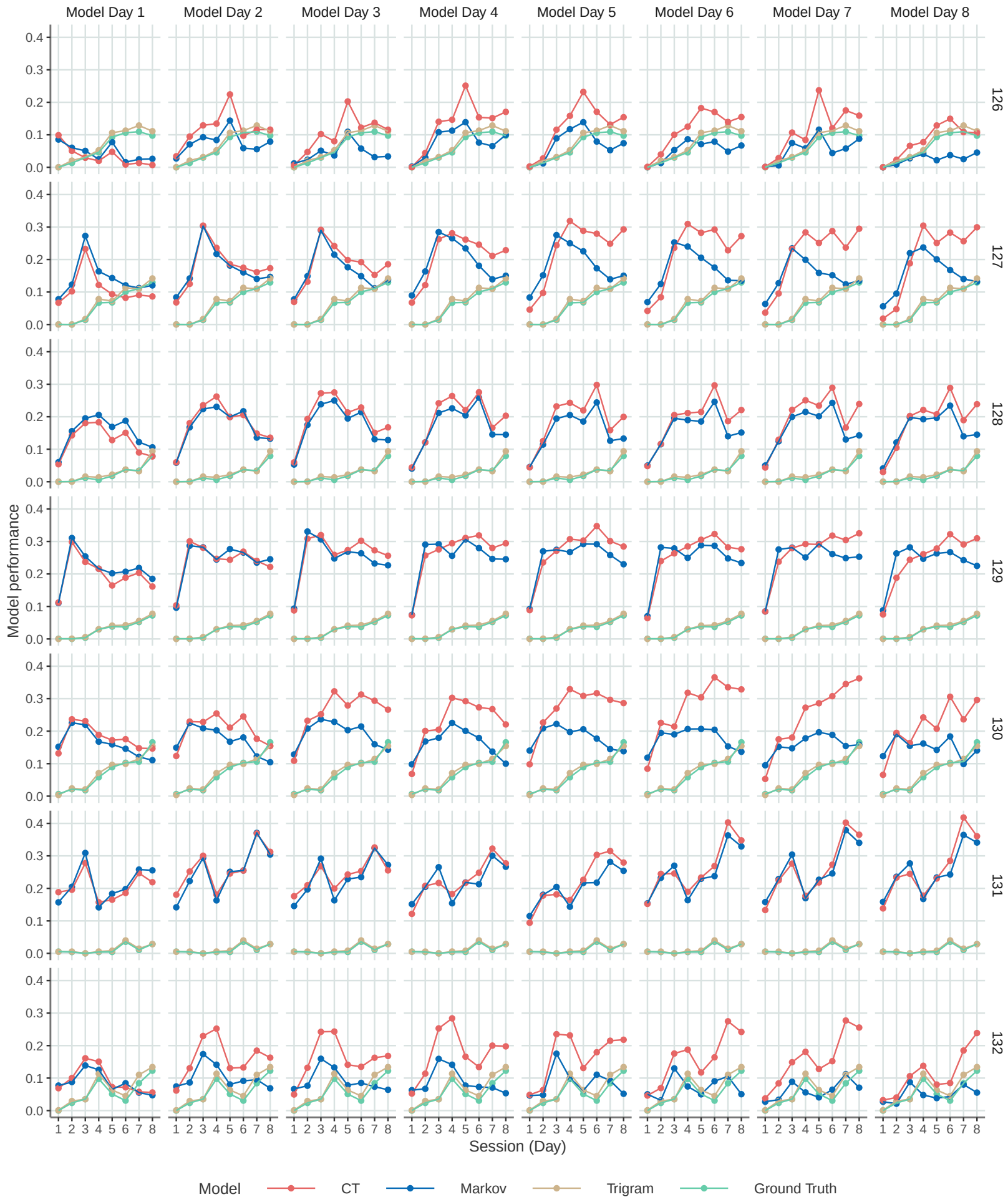
Extended Data Figure 6: Model performances of CT models trained on Day 8 and Day 9 for each individual.



Extended Data Figure 7: Predicting when errors will occur for each participant individually.


Model Day 1
Model Day 2
Model Day 3
Model Day 4
Model Day 5
Model Day 6
Model Day 7
Model Day 8
101
102
103
104
105
107
109
110
111
Model performance
Session (Day)
Model
CT
Markov
Trigram
Ground Truth


Model Day 1
Model Day 2
Model Day 3
Model Day 4
Model Day 5
Model Day 6
Model Day 7
Model Day 8
115
117
118
119
120
121
122
123
124
Model performance
0.0
0.1
0.2
0.3
0.4
1 2 3 4 5 6 7 8
Session (Day)
Model
CT
Markov
Trigram
Ground Truth


Model Day 1
Model Day 2
Model Day 3
Model Day 4
Model Day 5
Model Day 6
Model Day 7
Model Day 8
126
127
128
129
130
131
132
Model performance
0.0
0.1
0.2
0.3
0.4
1 2 3 4 5 6 7 8
Session (Day)
Model
CT
Markov
Trigram
Ground Truth

Extended Data Figure 8: Performances of models trained on different days for each individual. Performances are calculated with RTs above mean + 3.s.d. cut off. This is done to decrease variance of performance. For some participants, the CT model and Markov model diverges, which indicates these individuals infer some non-Markovian structure.

References

[1] Janacsek, K. & Nemeth, D. Implicit sequence learning and working memory: Correlated or complicated? *Cortex* **49**, 2001–2006 (2013). URL `http://dx.doi.org/10.1016/j.cortex.2013.02.012`.

[2] Howard, J. H. & Howard, D. V. Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging* **12**, 634–656 (1997).

[3] Nemeth, D. *et al.* Sleep has no critical role in implicit motor sequence learning in young and old adults. *Experimental Brain Research* **201**, 351–358 (2010).

[4] Kóbor, A., Janacsek, K., Takács, Á. & Nemeth, D. Statistical learning leads to persistent memory: Evidence for one-year consolidation OPEN. *Scientific Reports* **7**, 760 (2017). URL `www.nature.com/scientificreports/`.

[5] Song, S., Howard, J. H. & Howard, D. V. Implicit probabilistic sequence learning is independent of explicit awareness. *Learning and Memory* **14**, 167–176 (2007).

[6] Gael, J. V., Saatci, Y., Teh, Y. W. & Ghahramani, Z. Beam Sampling for the Infinite Hidden Markov Model. *Proceedings of the 25th international conference on Machine learning* 1088–1095 (2008). URL `http://mlg.eng.cam.ac.uk/zoubin/papers/icml2008ihmm.pdf`.

[7] Teh, Y. W., Jordan, M. I., Beal, M. & Blei, D. Hierarchical dirichlet processes. *Journal of the American Statistical ...* 1–41 (2006). URL `http://pubs.amstat.org/doi/abs/10.1198/016214506000000302`. `arXiv:1011.1669v3`.

[8] Carpenter, R. & Williams, M. Neural computation of log likelihood in control of saccadic eye movements. *Nature* **377**, 59–62 (1995).

[9] Kim, T. D., Kabir, M. & Gold, J. I. Behavioral/Cognitive Coupled Decision Processes Update and Maintain Saccadic Priors in a Dynamic Environment. *Journal of Neuroscience* **37**, 3632–3645 (2017). URL `http://www.jneurosci.org/content/jneuro/37/13/3632.full.pdf`.

[10] Brown, S. D. & Heathcote, A. The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology* **57**, 153–178 (2008). URL https://ac.els-cdn.com/S0010028507000722/1-s2.0-S0010028507000722-main.pdf?{_}tid=cf1f6f8c-5935-4868-ad21-fd0cdd65fc01{&}acdnat=1530108438{_}ee4a0bbf5bcbc4d08a41ebba6deeeb81.

[11] Harris, C. M., Waddington, J., Biscione, V. & Manzi, S. Manual choice reaction times in the rate-domain. *Frontiers in Human Neuroscience* **8** (2014). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4051275/pdf/fnhum-08-00418.pdf.