Neuroscience

Using recurrent neural network to estimate irreducible stochasticity in human choice-behavior

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint version 1
January 11, 2024 (this version)

Sent for peer review
September 29, 2023

Posted to preprint server
June 11, 2023

Yoav Ger, Moni Shahar, Nitzan Shahar

Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel • TAD - Center of AI & Data Science, Tel Aviv University, Tel Aviv, Israel • School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel

Abstract

Theoretical computational models are widely used to describe latent cognitive processes. However, these models do not equally explain data across participants, with some individuals showing a bigger predictive gap than others. In the current study, we examined the use of theory-independent models, specifically recurrent neural networks (RNN), to classify the source of a predictive gap in the observed data of a single individual. This approach aims to identify whether the low predictability of behavioral data is mainly due to noisy decision-making or miss-specification of the theoretical model. First, we used computer simulation in the context of reinforcement learning to demonstrate that RNNs can be used to identify model miss-specification in simulated agents with varying degrees of behavioral noise. Specifically, both prediction performance and the number of RNN training epochs (i.e., the point of early stopping) can be used to estimate the amount of stochasticity in the data. Second, we applied our approach to an empirical dataset where the actions of low IQ participants, compared with high IQ participants, showed lower predictability by a well-known theoretical model (i.e., Daw's hybrid model for the two-step task). Both the predictive gap and the point of early stopping of the RNN suggested that model miss-specification is similar across individuals. This led us to a provisional conclusion that low IQ subjects are mostly noisier compared to their high IQ peers, rather than being more miss-specified by the theoretical model. We discuss the implications and limitations of this approach, considering the growing literature in both theoretical and data-driven computational modeling in decision-making science.

eLife assessment

In this study, Ger and colleagues present a **valuable** new technique that uses recurrent neural networks to distinguish between model misspecification and behavioral stochasticity when interpreting cognitive-behavioral model fits. Evidence for the usefulness of this technique, which is currently based primarily on a relatively simple toy problem, is considered **incomplete** but could be improved via comparisons to existing approaches and/or applications to other problems. This technique addresses a long-standing problem that is likely to be of interest to researchers pushing the limits of cognitive computational modeling.

Introduction

Humans' behavior is thought to arise from a set of multidimensional, complex, and latent cognitive processes. Theoretical computational models allow researchers to handle this complexity by putting forward a set of mathematical descriptions that map assumed latent cognitive processes to observed behavior (Daw et al., 2011b; Smith and Ratcliff, 2004). By fitting a theoretical model (also sometimes termed 'symbolic' models) to the observed behavior, the researcher is able to draw conclusions regarding the estimation and architecture of the latent cognitive processes (Wilson and Collins, 2019). While this approach has already led to substantial scientific findings (Montague et al., 2012; Rescorla, 1972), it still faces a major challenge since the true underlying model is always left unknown (Box, 1979). Specifically, theoretical computational models usually stem from strict theoretical assumptions that can differ from the true cognitive mechanism that led to the observed behavior, a problem termed 'model misspecifications' (Beck et al., 2012; Nassar and Frank, 2016). Moreover, the issue of model misspecification is even more critical when considering that most studies in the field focus on group differences and therefore might neglect individual variation in the expressed models (Stephan et al., 2009; Rigoux et al., 2014). In the current study, we address the fundamental methodological problem of theoretical model misspecification by using well-established theory-independent models, a family of highly flexible models that require no prior theoretical specification.

Imagine that you are working on a theoretical question that you have addressed well by assembling a computational model that, after much effort, can replicate and mimic empirical observations. Still, some data will not be accurately predicted by the model, leaving an open question regarding the possibility that better specification of the model (i.e., changing some mechanisms, adding additional processes) would better capture the pattern in your observations. Currently, to address the issue of model misspecification, methodology in theoretical computational science calls for researchers to perform a model comparison analysis (Wilson and Collins, 2019). Here, the researcher is encouraged to put forward a set of different candidate models and quantify which model is most plausible given the observed data (Palminteri et al., 2017). Yet, even after a rigorous process of model comparison, the best-fitting model will still have what seems like room for improvement. This leaves an unanswered question regarding the possibility that yet another different model that was not described by the researcher might better explain and predict the observations (Eckstein et al., 2021; McElreath, 2020).

The distance in terms of variance between observed and predicted model behavior is typically termed the *predictive gap*. The predictive gap between individuals' behavior and model prediction can be attributed to two factors: The first is model *misspecification*, as mentioned above. Here, the individual's true behavior-generating model is different from the one suggested by the researcher (Beck et al., 2012; Nassar and Frank, 2016). The second factor is *stochasticity*, which refers to true noise or the natural randomness in human behavior (Faisal et al., 2008; Findling et al., 2019). The underlying assumption is that some variance in behavior is unpredictable and therefore represents irreducible variance. The assumption of irreducible variance is well accepted across scientific disciplines (Griffiths and Schroeter, 2018), and it suggests that even Laplace's demon (Gleick, 2011), a metaphorical demon that is assumed to have access to all mechanisms and processes in the universe, would not be able to produce perfect predictions. Both model misspecification and irreducible noise influence the predictive gap, yet identifying the contribution of each factor may lead to different implications. A predictive gap attributed mostly to model misspecification suggests that the researcher needs to increase the space of candidate models to further reduce the gap. However, a predictive gap attributed mostly to irreducible noise suggests that such improvement is mostly out of reach.

More recently, data-driven approaches based on neural networks have emerged as an alternative modeling paradigm in cognitive research (Dezfouli et al., 2019b; Song et al., 2021). These networks, which are models trained to learn directly from data rather than relying on theoretical assumptions about human behavior, have been shown to surpass typical theoretical models in pure action prediction (Dezfouli et al., 2019b; Song et al., 2021). The key property of these models is that a priori, the models' parameters do not directly map to some underlying cognitive process. Instead, the free parameters (i.e., weights) are iteratively adjusted during training to improve the network's predictive capability for a desired objective function (LeCun et al., 2015). Furthermore, unlike classical theoretical models, neural networks are overparameterized models that include a large number of learnable free parameters. This property allows the network high flexibility and the ability to approximate a wide range of functions (Siegelmann and Sontag, 1992; Hornik et al., 1989), including functions believed to arise from human cognition (Barak, 2017). For example, Dezfouli et al. (2019b) trained a recurrent neural network (RNN) to predict future human actions in a two-armed bandit task. Their study indicated that the RNN is capable of surpassing baseline RL models in action prediction. In another work, Fintz et al. (2022) trained an RNN model to predict future human actions in a four-armed bandit task. They showed that the RNN was able to capture atypical behaviors such as choice alternations and win-shift-lose-stay, which common cognitive models failed to capture. Finally, Song et al. (2021) trained an RNN model to predict the choices of humans in a reinforcement learning task that included a rich space of possible states and actions. They showed that the RNN was able to outperform the best-known cognitive model.

However, the high flexibility of neural networks also comes with the disadvantage that these networks are often considered black box models. Despite attempts to interpret the networks' latent space (Dezfouli et al., 2019a), it is still not clear how to efficiently interpret behavior using them, which is a major goal in cognitive research (Hasson et al., 2020; Yarkoni and Westfall, 2017). In this study, we aim to leverage the network's high flexibility and predictive capability to address the problem of identifying misspecification in theoretical computational models, using two different estimates:

1. **Predictive performance.** We assume that, for a fixed dataset, the flexibility of neural networks would enable them to capture the mapping between independent variables and dependent variables, up to a point where the remaining predictive gap is primarily due to noise rather than model misspecification (see **Fig. 1a**). Therefore, we hypothesize that across different true generative models, neural networks will reach a predictive gap that closely resembles the predictive gap that remains when the true generative model is known. As a result, theory-independent models can serve as an upper benchmark against which the new data can be predicted based on previous observations. If, for some individuals, the theoretical model is severely misspecified, we expect to observe a significant improvement in our ability to predict unseen data using theory-independent models. However, if an individual is simply exhibiting a noisier pattern of behavior, we anticipate little to no improvement when using theory-independent models.

2. **Point of early stopping.** Neural networks are prone to overfitting, where they fit too closely to the training data and fail to generalize to new datasets (LeCun et al., 2015). To mitigate overfitting, a common technique called early stopping is used (Bishop and Nasrabadi, 2006). In early stopping, the network is trained for multiple epochs, and the predictive performance on a validation set is monitored. Initially, the performance improves as the number of epochs increases, but there comes a point where the network starts learning patterns that reflect noise rather than true underlying patterns in the data. The point of early stopping is determined as the epoch at which the best prediction is achieved. We propose that the point of early stopping largely reflects the amount of noise in the data. With a fixed dataset and network size, an earlier stopping point indicates noisier data (see **Fig. 1b**). The rationale behind this hypothesis is that, under certain realistic assumptions, the probability of stopping at an earlier epoch is higher for a sequence with more noise. This is because the noisy sequence has a lower ratio of true signal to noise, resulting in less information learned by the algorithm over the course of training. Using the point of early stopping to estimate irreducible noise has advantages, including the potential to require fewer data compared to using network predictive accuracy to estimate model misspecification. Estimating the upper bound of the prediction accuracy of the true data generation model using predictive accuracy typically necessitates three sets of data per individual (training, validation, and test). This approach may also require initial auxiliary data for pre-training the recurrent neural network (RNN) to achieve optimal results. However, estimating the point of early stopping can be accomplished with only two sets of data per individual (training and validation). Additionally, this approach may benefit from initializing the network with random weights instead of a pre-trained RNN, as it allows for estimating the maximum possible variance in the optimal epoch estimate among individuals.

In the present study, our aim is to estimate model miss-specification in individual participants. We propose a method that utilizes theory-independent models, specifically recurrent neural networks (RNNs), which possess high flexibility in learning complex features from data without manual engineering. In Study 1, we conducted simulations involving three groups, each consisting of N = 100 simulated agents performing a two-step reinforcement learning decision task. Each group of agents followed a specific data-generating model, and the agents differed in the level of true noise present in their actions. We assumed three hypothetical labs, with each lab having knowledge of only one data-generating model, thus miss-specifying two-thirds of the agents. We demonstrated that the predictive performance of a pre-trained RNN using three data sets per agent (training, validation, test) could be used to classify whether an agent was miss-specified by the lab's theoretical model. Additionally, we showed that the number of optimal training epochs for an RNN with random weights and fewer data (training and validation, without a test set) could serve as an estimate for comparing the amount of noise in agent data, given a fixed dataset and RNN architecture. Next, in Study 2, we analyzed an empirical dataset with N = 54 participants who completed three sessions of a two-step decision task in a lab setting. We examined the fit of a well-known theoretical model (Daw's hybrid model) to participants with different IQ levels. We found that the theoretical model showed a systematically poorer fit to participants with low IQ compared to those with high IQ. By utilizing RNN predictive performance and optimal epochs estimates, we classified the source of the low theoretical model fit in low IQ participants. Our findings converged to suggest that the percentage of model miss-specification did not differ between low and high IQ individuals. Instead, low IQ individuals exhibited noisier decision-making compared to their high IQ counterparts. We propose that theory-independent models, such as RNNs, can be valuable in classifying model miss-specification. However, we acknowledge the limitations of our approach and discuss potential directions for future research.

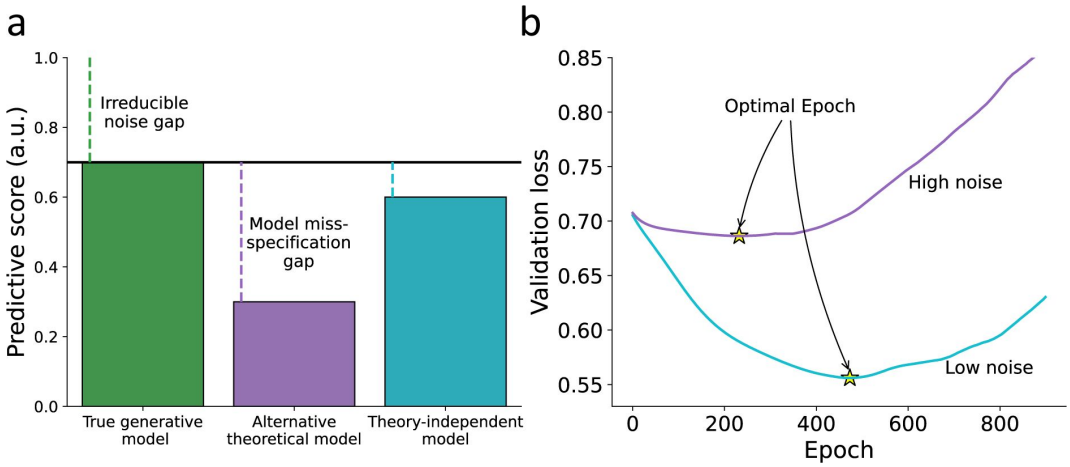

Figure 1

Hypothetical illustration for using theory-independent models to explore the fit of theoretical models to human behavior.

(a) Predictive performance - we illustrate the predictive gap for three hypothetical models: First, the true data-generating model (i.e., forever unknown to the researcher; green) where the remaining gap is due to an irreducible noise component in the individual's behavior. Second, a hypothetical alternative theoretical model (e.g., specified by a researcher; purple), where the remaining predictive gap reflects both irreducible noise and model misspecification. Finally, a hypothetical theory-independent model (turquoise) is assumed to reflect a predictive gap that is mainly due to model misspecification compared with the alternative model, yet does not provide a clear theoretical interpretation. We, therefore, assume that theory-independent models can be used to inform researchers about the amount of improvement that can be further gained by assembling additional theoretical models. **(b)** Point of early stopping - when training a network, we can examine its performance against a validation set as a function of the number of epochs used for training the parameters (x-axis). The point of early stopping reflects the maximum number of training epochs with the best predictive performance, just before the network starts to overfit the data (i.e., learn patterns that are due to noise; indicated by a yellow star). Here, we illustrate two hypothetical learning curves reflecting the point of early stopping for low-noise (purple line) and high-noise (blue line) datasets. Specifically, we illustrate the notion that the point of early stopping can reflect the amount of noise in the data, so a lower point of early stopping reflects noisier data (considering a fixed number of observations for the two datasets and network size).

Study 1

We begin by applying our method to simulated data, where different types of artificial agents governed by distinct generating models performed a two-step task (Daw et al., 2011a). This simulation allowed us to evaluate our approach using artificial data, where we knew the true underlying data-generating model and the level of true noise for each agent. Our main objective was to assess the ability of theory-independent models to inform us about the factors contributing to a poorer fit of a theoretical model, specifically model miss-specification or irreducible noise. To accomplish this, we introduced three theoretical models (Hybrid model-based/model-free, Habit, and K-dominant hand) and two theory-independent models (logistic regression and RNN). We generated data from the three theoretical models and fitted all five models to the data from each respective theoretical model. Subsequently, we utilized the predictive capability of the RNN and logistic regression models, as well as the number of optimal training epochs for the RNN, to estimate and distinguish between model miss-specification and stochastic behavior.

Method

Two-step-task. To test our hypothesis, we employed an exemplar two-step reinforcement learning task, which has been widely utilized in computational modeling studies (see **Fig. 2a**; Daw et al. (2011a)). Each trial of the task consisted of two stages. In the first stage, participants made a choice between two actions, labeled as action A and action B. These actions probabilistically led to one of two possible second-stage states. Action A predominantly led to State-I in the second stage, while action B primarily led to State-II (common transition). However, there was also a minority of trials where the actions led to the opposite states, meaning action A led to State-II and action B led to State-I (rare transition). The probabilities of common and rare transitions were set at 0.7 and 0.3, respectively, and remained constant throughout the task. In the second stage, participants again made a choice between two additional actions and received a binary reward of $0 (no reward) or $1 (reward). The reward probability varied randomly across trials (see **Fig. 2b**).

Theoretical models. We consider three different theoretical models that generate choice behavior in the two-step task.

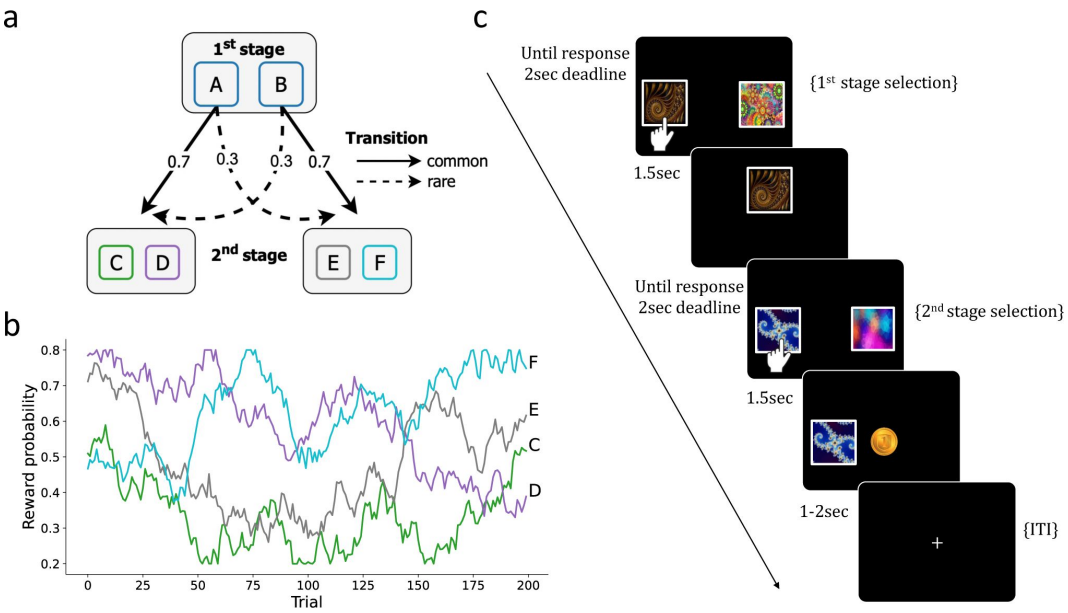

Figure 2

Two-step task.

(a) At the first stage, participants choose between two options (A or B) that probabilistically lead to one of two second-stage states, with a fixed transition probability of 70% (“common”) or 30% (“rare”). In the second stage, participants choose between two options (C/D or E/F) to obtain rewards. **(b)** The reward probability for each second-stage choice is determined by a random walk drift. **(c)** An example of a trial sequence in the two-step task.

1. **Hybrid model.** The first model we considered was the hybrid model originally suggested by Daw et al. (2011a), which assumes that agents' choice behavior is determined by a weighted combination of both model-based (MB) and model-free (MF) RL algorithms (Sutton and Barto, 2018). The contribution of each algorithm is modulated by the free parameter weight w. The weighted value of each first-stage action is calculated according to,

$$Q^{net}(s_1, a_1) = w \cdot Q^{MB}(s_1, a_1) + (1 - w) \cdot Q^{MF}(s_1, a_1), \quad (1)$$

where $Q^{MF}(s_1, a_1)$ is the model-free action value in the first stage updated trial-by-trial and $Q^{MB}(s_1, a_1)$ is the mentally calculated model-based estimation. The MF Q-values were initiated at zero and updated according to action and reward history as follows,

$$Q^{MF}(s_1, a_1) = Q^{MF}(s_1, a_1) + \alpha_1 \cdot \left(Q^{MF}(s_2, a_2) - Q^{MF}(s_1, a_1)\right) + \alpha_1 \cdot \lambda \cdot \left(r_t - Q^{MF}(s_2, a_2)\right), \quad (2)$$

$$Q^{MF}(s_2, a_2) = Q^{MF}(s_2, a_2) + \alpha_2 \cdot \left(r_t - Q^{MF}(s_2, a_2)\right), \quad (3)$$

where α_1, α_2 are the learning rates free parameters of the first and second stage respectively, λ is the discount factor parameter, and $r_t \in [0, 1]$ is the received reward. The model-based action value of the first stage $Q^{MB}(s_1, a_1)$ is calculated according to,

$$Q^{MB}(s_1, a_1) = \sum_{s_2 \in S} P(s_2 | s_1, a_1) \max_{a_2 \in A} Q^{MB}(s_2, a_2), \quad (4)$$

where $P(s_2 | s_1, a_1)$ is the true transition probability of reaching the second-stage state s_2 by performing action a_1 at the first stage. S denotes the second-stage states and A is the set containing the actions available at each second-stage state respectively. Note, that at the second stage model-based (MB) values, $Q^{MB}(s_2, a_2)$ of each action a_2 performed at the second-stage state s_2, are identical to the corresponding model-free values, namely, $Q^{MB}(s_2, a_2) = Q^{MF}(s_2, a_2)$. We used a SoftMax choice rule for both the first and second choices. The SoftMax choice rule transforms the action values to distribution over the action according to,

$$P(s_1, a_1) = \frac{\exp\left[\beta_1 \cdot Q^{net}(s_1, a_1)\right]}{\sum_{\tilde{a} \in A} \exp\left[\beta_1 \cdot Q^{net}(s_1, \tilde{a})\right]}, \quad (5)$$

$$P(s_2, a_2) = \frac{\exp\left[\beta_2 \cdot Q^{MF}(s_2, a_2)\right]}{\sum_{\tilde{a} \in A} \exp\left[\beta_2 \cdot Q^{MF}(s_2, \tilde{a})\right]}, \quad (6)$$

where β_1 and β_2 are inverse-temperature parameters of the first and second stages respectively. This parameter controls the level of stochasticity in the action selection. Where $\beta \to 0$ corresponds to random choices and $\beta \to \infty$ for deterministically choosing the highest value action. Overall, the model includes six free parameters $\theta_{hybrid} = \{w, \alpha_1, \alpha_2, \lambda, \beta_1, \beta_2\}$

2. **Habit model.** According to a habit model, agents tend to repeat previously taken actions regardless of their outcome (Miller et al., 2019). Here, agents' choices are influenced only by previous actions, independent of any transition or reward history. The model keeps track of past action selection in the form of habit strengths, denoted as H, and are similar in spirit to the Q-values described in the previous hybrid model. These values are initiated at 0.5 and updated on both stages of the task according to,

$$H(s_1,\mathbf{a}) = \begin{cases} H(s_1,a_i) + \alpha_1 \cdot \left(1 - H(s_1,a_i)\right) \\ H(s_1,a_j) + \alpha_1 \cdot \left(0 - H(s_1,a_j)\right) \end{cases}, \quad (7)$$

$$H(s_2,\mathbf{a}) = \begin{cases} H(s_2,a_i) + \alpha_2 \cdot \left(1 - H(s_2,a_i)\right) \\ H(s_2,a_j) + \alpha_2 \cdot \left(0 - H(s_2,a_j)\right) \end{cases}, \quad (8)$$

where α_1, and α_2 are the learning rates free parameters of the first and second stages respectively. $H(s, a)$ is the habit strengths matrix of all possible states, and a is the vector of all actions available in the corresponding state. We denote the action selected in the current trial as a_i and the unchosen action as a_j. The updating rule increases the value of the action selected (a_i) toward 1 and the other unselected action (a_j) toward 0. Like the hybrid model, we used SoftMax choice rule for both the first and second choices,

$$P(s_1,a_1) = \frac{\exp\left[\beta_1 \cdot H\left(s_1,a_1\right)\right]}{\sum_{\tilde{a}\in A}\exp\left[\beta_1 \cdot H\left(s_1,\tilde{a}\right)\right]}, \quad (9)$$

$$P(s_2,a_2) = \frac{\exp\left[\beta_2 \cdot H\left(s_2,a_2\right)\right]}{\sum_{\tilde{a}\in A}\exp\left[\beta_2 \cdot H\left(s_2,\tilde{a}\right)\right]}, \quad (10)$$

Overall, the model includes four free parameters $\theta_{habit} = \{\alpha_1, \alpha_2, \beta_1, \beta_2\}$

3. **K dominated-hand model (K-DH).** In this model, an action is selected in a random fashion with some bias towards one action over the other (e.g., due to hand dominancy; note that in the simulation we assigned each action to the same response-key, thus a bias towards one response-key translates directly to a tendency to choose one action over the alternative). In this model, the agent is assumed to change the preference for the dominant hand across a sequence of K actions. For example, in $K = 2$ the agent will have two probabilities (defined by two fixed parameters; p_1, p_2) representing the preference for the dominant hand as a function of position in the two trials' sequence. The K-dominant hand model is therefore designed to generate a random sequence of choices (e.g., A, B, A, B …) with some systematic repetitions. Here, we included agents with a fixed $K = 2$ for the first stage and $K = 1$ for each of the two-second stages, which resulted in a similar number of free parameters as other models in this work. Specifically, the probability of selecting action a_i at each state of the task is given by,

$$p(s_1,a_i) = p_{t \mod 2}, \quad (11)$$

$$p(s_2,a_i) = p_2, \quad p(s_3,a_i) = p_3, \quad (12)$$

Where p_0/p_1 are the probabilities of selecting action a_i in the first stage on the even/odd trials, respectively, and p_2/p_3 are the probabilities of selecting action a_i at each second stage state respectively. The probability of selecting the other action at each state is always set to be the complementary probability (i.e., $1 - p(s, a_i)$). Overall, the model includes four free parameters $\theta_{K\text{-}DH} = \{p_0, p_1, p_2, p_3\}$.

Theory-independent models. We used two theory-independent models. Unlike the theoretical models which are used both to generate, fit, and predict choice data, the theory-independent models were used only to fit and predict choice data.

1. **Recurrent neural network (RNN).** The first theory-independent model we considered was a three-layer recurrent neural network architecture. Our RNN consisted of an: (i) input layer, (ii) hidden layer, and (iii) output layer. At each step, the RNN input was the last trial's first-stage action, reward, and transition type. The output at each step is a probability distribution over the current trial's first-stage actions. The hidden layer was based on a single-layer gated recurrent unit (GRU; Cho et al. (2014)) with five hidden units.

$$\mathbf{x}_t = [r_{t-1}, a_{t-1}, t_{t-1}], \tag{13}$$
$$\mathbf{r}_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r), \tag{14}$$
$$\mathbf{z}_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z), \tag{15}$$
$$\hat{h}_t = \phi(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h), \tag{16}$$
$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \hat{\mathbf{h}}_t, \tag{17}$$
$$\mathbf{o}_t = \mathbf{W}_o \mathbf{h}_t + \mathbf{b}_o, \tag{18}$$
$$p(a_t = A) = \frac{e^{\mathbf{o}_t}}{\sum_i e^{\mathbf{o}_i}}, \tag{19}$$

Where $\mathbf{x}_t$ is the input vector and $\mathbf{h}_{t-1}$ is the previous hidden state vector. σ is the logistic sigmoid function, $\odot$ is the Elementwise Hadamard product, φ is the hyperbolic tangent function. r_t represents the reset gate, z_t the update gate, and h_t the candidate activation vectors. $\mathbf{W}_r$, $\mathbf{W}_z$, $\mathbf{W}_h$ are the input connection weight matrices, $\mathbf{U}_r$, $\mathbf{U}_z$, $\mathbf{U}_h$ are the recurrent connection weight matrices, and $\mathbf{b}_r$, $\mathbf{b}_z$, $\mathbf{b}_h$ are the bias vectors. h_t represents the actual activation vector of the unit (current hidden state). h_t is projected to the output layer with full connections via the weight matrix $\mathbf{W}_o$ and bias vector $\mathbf{b}_o$. The output is then transformed with a SoftMax activation function to action probabilities. Overall, this model includes the following free parameters $\theta_{\text{RNN}} = \{\mathbf{W}_r, \mathbf{W}_z, \mathbf{W}_o, \mathbf{U}_r, \mathbf{U}_z, \mathbf{U}_h, \mathbf{b}_r, \mathbf{b}_z, \mathbf{b}_h, \mathbf{W}_o, \mathbf{b}_o\}$. The total number of free parameters is dependent on the size of the hidden layer (192 for 5 hidden units). Importantly, in the supplementary information, we provide additional analysis where we varied the number of hidden units of the hidden GRU layer (2 and 10). We found similar results, suggesting that at least with the current task, our approach is not sensitive to the size of the network (see S2 in the SI)

2. **Logistic Regression (LR).** The second theory-independent model we considered was a logistic regression model, where the probability of taking the first-stage action is determined by a linear combination of the history of past actions, rewards, and transition, up to j trials back.

$$x = \beta_0 + \sum_j \left(\beta_a^j a_{t-j} + \beta_r^j r_{t-j} + \beta_t^j t_{t-j} + \beta_{a\times r}^j a_{t-j} r_{t-j} + \beta_{a\times t}^j a_{t-j} t_{t-j} + \beta_{r\times t}^j r_{t-j} t_{t-j} + \beta_{a\times r\times t}^j a_{t-j} r_{t-j} t_{t-j}\right), \tag{20}$$

$$p(a_t = 1 \mid \beta) = \frac{1}{1 + e^{-x}}, \tag{21}$$

The dependent and independent variables were coded as follows: a_t: 1 for action A and -1 for action B (first stage actions, see **Fig. 2**). a_{t-j}: 0.5 if the first state action j trials back was A and -0.5 otherwise. r_{t-j}: 0.5 if j trials back was rewarded and -0.5 otherwise. t_{t-j}: 0.5 if j trials back transition was common and -0.5 if it was rare. Overall, the model includes (1 + j × 7) free parameters $\theta_{\text{LR}} = \{\beta_0, \beta_a^j, \beta_r^j, \beta_t^j, \beta_{a\times r}^j, \beta_{a\times t}^j, \beta_{r\times t}^j, \beta_{a\times r\times t}^j\}$.

Simulating data and model fitting. First, we simulated choice-behavior of artificial agents on the two-step task (see **Fig. 2**), using the three distinct theoretical data-generating models (Hybrid, Habit, and K-DH). For each model, we simulated 100 agents with their true underlying free parameters sampled from a range of possible options (see **Table S1** in the **SI**). Each agent was simulated for three blocks of 200 trials each. Then, we separately fitted all five models (three theoretical and two theory-independent) to data simulated from all three theoretical models. In the supplementary information, we provide an additional analysis where we varied the lengths of each block (100 and 500 trials) and found similar results suggestive that our method (at least for the current task) is not sensitive to the amount of data (see **S3** in the **SI**).

1. **Theoretical models.** For each agent individually, we sought optimal parameters $\hat{\theta}_m$ under each of the three theoretical models (Hybrid, Habit, and K-DH) with maximum-likelihood estimation using only one of the agents' blocks (training-set). We repeated this procedure with 10 different initial search points and chose the optimal $\hat{\theta}_m$ using a withheld block (validation-set). Finally, the optimal parameters $\hat{\theta}_m$ were used to record the prediction on an additional withheld block (test-set). We used SciPy library (Virtanen et al., 2020) with the minimize function and L-BFGS-B method to extract best-fit parameters.
2. **RNN.** We trained an individual RNN for each agent in a supervised manner with cross-entropy loss (maximum-likelihood) using Pytorch library (Paszke et al., 2019). We trained with Adam optimizer (Kingma and Ba, 2014) with a constant learning rate of 0.001. Crucially, unlike the other models, RNN's high flexibility may lead it to overfit the training data. This overfitting of the training dataset will result in a low generalization, making the model less useful for making predictions on new data. To prevent this, we used early stopping, a commonly used regularization method (Bishop and Nasrabadi, 2006). We therefore first pre-trained the RNN model using a new auxiliary synthetic dataset consisting of 200 agents from all three theoretical models pooled together. For the auxiliary synthetic data, agents were simulated for two blocks of 200 trials each (used as training and validation for generating the RNN pre-training). The pre-trained RNN model was then fine-tuned for each individual on the main synthetic dataset (see Simulating data and model fitting) using early stopping. Specifically, we used the three blocks of each agent as a train, validation (used for training using early stopping), and a test set used to estimate the predictive accuracy that we report in the results section for further analysis. To further allow an accurate and variable possible estimate of early stopping, we recorded the point of early stopping (using training and validation sets) for an individual RNN for each agent that was not pre-trained, and instead initialized with random weights. The point at which we stopped optimizing the network (i.e., the number of training epochs) was then used in the main analysis and referred as *optimal epoch* estimate.
3. **LR.** We fitted a logistic regression model to the artificial data using scikit-learn library (Pedregosa et al., 2011) with L-BFGS-B method, again using only one of the agents' blocks (training-set). The individual coefficient of each agent was estimated using maximum-likelihood. The number of trials back k was determined using a withheld block (validation-set). Then we used the coefficient to predict the choices of a withheld block (test-set).

$$nLP_i^m = \frac{-\sum_{b=1}^{B}\sum_{t=1}^{T}\log p(a_t^i|\hat{\theta}_m)}{|B|}, \tag{22}$$

Model selection. We compare the different models by their predictive performance of left-out data of the first-stage choice. Hence, we adopted a cross-validation (CV) approach. Specifically, at each CV round, for each agent, and under each model, only one of the blocks (training-set) was used to estimate optimal parameters ($\hat{\theta}_m$). Then the optimal parameters were used to evaluate the predictions on a withheld block (test-set). We averaged across all withheld blocks (three in total) to obtain a single predictive score we denote as nLP_i^m (Negative log-probability; lower is better),

Where *m* denotes the fitted model, *i* the agent index, *B* the total number of blocks, *T* the total number of trials of the corresponding block, $\hat{\theta}_m$ denotes the optimal parameters under model *m* (maximum-likelihood parameters of the validation-set), and $\log p(a_t^i \mid \hat{\theta}_m)$ is the log-probability that the optimal parameters under model *m* assign to the first-stage action a_t of the test-set.

$$noise = 1 - \frac{\beta_1^i - \beta^{min}}{\beta^{max} - \beta^{min}}, \tag{23}$$

Noise estimates. For each agent we quantified the amount of true noise (stochasticity) the agent holds in the first stage action, using his true underlying parameters. Specifically, within each group of agents that shared the same theoretical model we performed a min-max scaling. For the hybrid agents we used the agent's true inverse-temperature parameter of the first stage β_1 to calculate the amount of true noise as fellow,

Where β_1^i is the true first stage inverse-temperature of agent i and β^{min}/β^{max} is the true minimal/maximal inverse-temperature overall hybrid agents. We took the 1 – min-max scaling so that the hybrid agent with the minimal amount of noise will take a value of 0 and the agent with the maximal amount of noise will take the value of 1. For the habit agents, we performed the same calculation, using the true inverse-temperature parameter β_1 of the first stage.

For the K-DH agent, we used the true p_0 and p_1 parameters to quantify the amount of true noise. We measured the distance (in absolute value) between the true p_0/p_1 parameters to 0.5 (i.e., completely random response policy) and summed these resulting distances,

$$\delta = |p_0 - 0.5| + |p_1 - 0.5|, \tag{24}$$

and then performed a min-max scaling as follows,

$$noise = 1 - \frac{\delta^i - \delta^{min}}{\delta^{max} - \delta^{min}}, \tag{25}$$

That is, a K-DH agent that his true p_0, p_1 = 0 (i.e., determinedly choose one of the actions at each trial) will take the value 0. Conversely, a K-DH agent that his true p_0, p_1 = 0.5 (i.e., choose randomly at each trial) will take the value 1.

Results

Classification of model miss-specification using predictive performance. Our first aim was to assess the ability of theory-independent models to predict agent choice data across each theoretical model. For each agent in the artificial dataset and for each cross-validation round, we used one block (training-set) to estimate optimal parameters (separately for each model). Then, the optimal parameters were used to predict the agent's first-stage choice data on a withheld block (test-set). We repeated this procedure for three rounds and averaged the predictive score over all withheld blocks (i.e., nLP_i^m; see Eq. 22). Namely, for each agent we obtained 5 predictive scores, corresponding to the 5 models mentioned above (Hybrid, Habit, K-DH, RNN, LR). We found that across all theoretical models, RNN achieved the second-best predictive score, second only to the true generative model (see **Fig. 3a**). This suggests that RNN is flexible enough to approximate a wide range of different behavioral models. Moreover, we found that the Logistic regression model achieved a poorer predictive score of agent choice data, implying that the model is less expressive compared with RNN.

Next, we performed a hypothetical experiment in which three hypothetical labs estimated individual performance using a theoretical model. Each hypothetical lab was assumed to be aware of only one data-generating model, namely, 'Hybrid-lab', 'Habit-lab', and 'K-DH-lab'. The notion here is that we illustrate a hypothetical situation where unbeknownst to the research lab, subjects are performing the task under different models. For example, the hypothetical Hybrid-lab is assumed to be aware only of the hybrid model and wrongly assumes that all subjects acted according to the Hybrid model. In such a case, all agents with a true generative model of Habit and K-DH models will be misspecified. However, we assumed that the lab has no knowledge regarding these alternative models, and thus will fit the hybrid model to all agents. We were interested to test the extent to which theory-independent models (RNN, LR) can help such a lab to classify if a

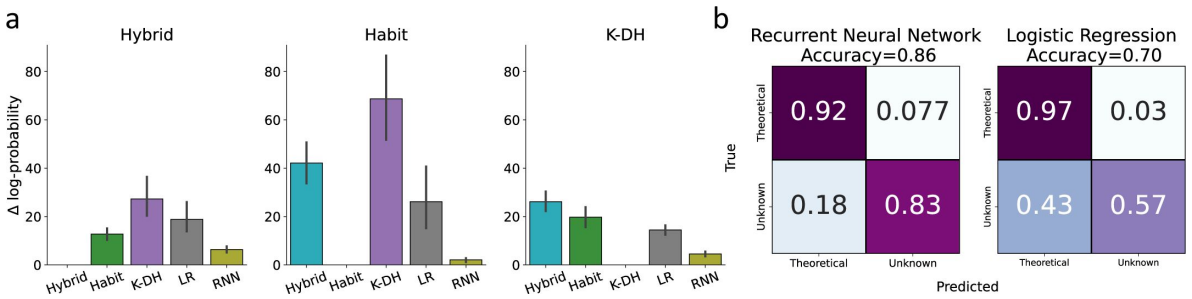

Figure 3

Classification of model-misspecification using the predictive performance of theory-independent models.

Here, we tested our ability to identify model misspecification of theoretical models using RNN and Logistic regression. **(a)** Across three theoretical models (Hybrid, Habit, K-DH) we simulated 100 agents and predicted agents' actions using the same three theoretical models, RNN and logistic regression. We present on the y-axis differences in negative log-probability estimates for each fitted model against the true generative model. For example, the left panel depicts the difference in log-probability estimates across all five models for 100 agents simulated using the hybrid model. As expected, across all three generative theoretical models, RNN achieved the best performance score, second only to the true generative theoretical model (black lines represent 95% CI). **(b)** We calculated a confusion matrix for a hypothetical lab that is familiar only with one theoretical model and uses RNN or logistic regression to try and conclude whether a certain agent shows a high predictive gap due to model miss-specification. Specifically, we simulated data from three theoretical models (Hybrid, Habit, K-DH) and assumed that the hypothetical lab wrongly fitted only one model to all agents. The hypothetical lab then estimated a predictive performance for each agent using RNN and logistic regression and classified each agent as "miss-specified" if the prediction accuracy of the RNN/logistic regression was better compared with the predictive accuracy of the lab's theoretical model. We calculated a confusion matrix where the hypothetical lab classification is contrasted with the true label. Results show good classification by RNN (Accuracy ≈ .86), and logistic regression (Accuracy ≈ .70), with better performance for RNN compared with logistic regression. Overall, these results suggest that RNN/logistic regression can be used to some extent to inform researchers regarding model-misspecification when using theoretical computational modeling.

certain agent might be better explained by another unknown model compared to the hybrid model. For this aim, we performed three classification rounds, where at each round we assumed one hypothetical lab classified each agent into one of two classes: lab theoretical model or unknown alternative model. The classification was done based on the difference between the nLP_i^m scores of the assumed theoretical model and the two theory-independent models of each agent. We averaged across all classification rounds and present two confusion matrices, classification by RNN and by Logistic regression (see **Fig. 3b**). Like before, we found that RNN achieved a higher classification accuracy of 86%, compared to the LR model which reached only 70% accuracy. This finding supports our claim that RNN can signify if a certain subject might be better explained by another unknown model. Importantly, to illustrate the robustness of our results, we provide a supplementary analysis where the exact same analysis was repeated across 100 and 500 observations per agent and found very similar results (see **Fig. S3** in **SI**). Furthermore, to assert that this effect is not due to averaging across all classification rounds in the supplementary information, we provide the confusion matrices for each hypothetical lab (Hybrid, Habit, K-DH). We found that across all hypothetical labs, the RNN achieved a higher true negative rate and a lower true positive rate compared to the LR model (see **S4** in the **SI**).

Using the number of optimal epochs to estimate noise. When fitting an RNN we estimate the number of optimal training epochs that minimize both underfitting and overfitting. We reasoned that for a fixed number of observation and network parameters, the point of early stopping (henceforward, "optimal epochs") should also reflect the amount of noise/information in the behavioral data. Specifically, we hypothesized that the probability of stopping in an earlier epoch is higher for noisier agents since this probability is determined by the ratio between the true signal learned and the noise. To test our hypothesis, we examined the correlations between the number of optimal epochs and the amount of true noise each agent holds (see **Fig. 4a**). As expected, we found that agents with high levels of true noise in their data also showed lower number of optimal epochs (i.e., required less RNN parameter training) compared with less noisy agents [see **Fig. 4a**; linear correlation coefficient r = −.67, −.74, −.62 for the Hybrid, Habit, and K-DH agents respectively; p < .001 across all correlation]. Importantly, we found that this result holds regardless of the agent's true underlying data-generating model, suggesting that the point of early stopping may be used as an index for the amount of true noise each participant holds. We performed the same analysis with different network sizes and different number of observations (i.e., trials) per agent and found very similar results (see **Fig. S2** in **SI**). Therefore, we conclude that the number of optimal epochs reflects the amount of information in the observed behavior. Note, that the number of optimal epochs is not an absolute estimate since on its own it can be influenced by other factors, including the number of observations and the size of the net. However, as long as the number of observations and the size of the network is the same between two datasets, the number of optimal epochs can be used to estimate whether the dataset of one participant is noisier compared with a second dataset.

Study 2

We now present an application of our method to an existing dataset, where humans performed an identical two-step task as reported in Study 1, at three different time points (Kiddle et al., 2018). We predict participants' behavior using a well-established hybrid model (see Hybrid model in Study 1 for further details) and demonstrate that low IQ is associated with a higher predictive gap. We put our method to use and examine whether action prediction of RNN can improve the predictive gap of the theoretical model.

Specifically, we assume that if low IQ participants are more frequently miss-specified by the hybrid model, then RNN will show a greater reduction of the predictive gap for low compared with high IQ individuals (see **Fig. 5a**). However, another possibility is that the higher predictive

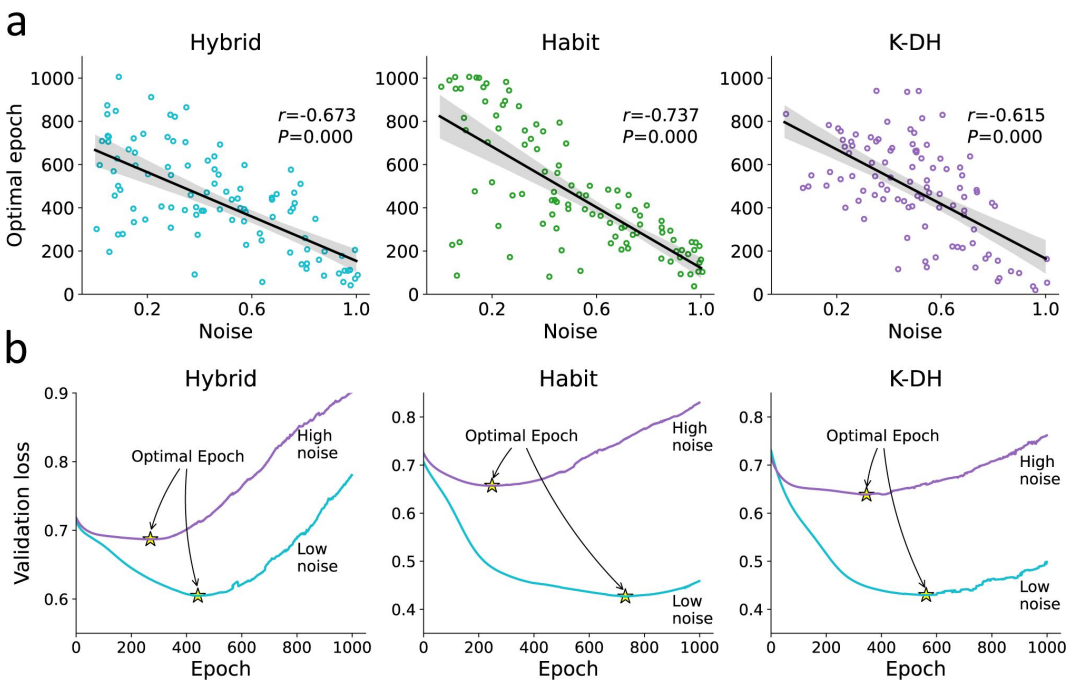

Figure 4

Optimal epoch relation to true noise.

For each agent we iteratively trained the RNN while examining its predictive ability. We then recorded for each agent the number of optimal epochs, which is the number of RNN training iterations that minimized underfitting and overfitting. **(a)** We found a strong association between the number of optimal RNN training epochs (y-axis) and the agent's true noise level. This result demonstrates that the number of optimal RNN epochs can be used as a proxy for the amount of information in a certain dataset (given a fixed number of observations and network size). **(b)** For illustration purposes, we plotted the RNN loss curves for agents with low vs. high levels of noise. Specifically, we present loss curves where for each RNN training epoch (x-axis) we estimate the predictive accuracy using a validation data set. We estimated the loss curves of 300 artificial agents (from three theoretical models) and divided them into two groups according to their true noise (high vs. low). We show that the point of optimal epoch (early stopping; denoted yellow star) was higher to agents with low internal noise.

gap of low compared with high IQ participants in the theoretical hybrid model is mostly due to the noisier behavior of the low IQ participants' part. In that case, RNN should have a similar contribution to the predictive gap, regardless of IQ (see **Fig. 5b**).

Furthermore, we estimate the number of optimal epochs for each individual and examine the correlation with IQ. If low IQ participants' behavior is less predicted by the theoretical model mostly due to them being misspecified, then we should not observe any correlation between the optimal number of epochs and IQ (since across IQ participants are assumed to have a similar amount of information in their behavioral data). However, if the behavior of low IQ participants is noisier compared with high IQ, then we should expect a negative correlation between the number of epochs and IQ.

Methods

Dataset. We studied a previously published dataset taken from NSPN U-Change Cognition Cohort (Kiddle et al., 2018). We focus on a subset of the dataset, of which healthy volunteers (ages 14 to 24 years; N = 54) performed the two-step task at three distinct time points (6 and 18 months after the first measurement). At the first time point (Time-I), each subject performed the task for 121 trials, in the second (Time-II) 121 trials, and at the third (Time-III) 201 trials. In the preprocessing step, we omitted from the analysis trials with RTs below 150ms. The dataset also includes for each subject a Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). Importantly, we also utilized another subset of the NSPN dataset that included N = 515 individuals who performed the two-step task in only two different time points (Time-I: 121 trials, Time-III: 201 trials).

Model fitting. For each subject choice data, we fitted both the theoretical Hybrid model and a theory-independent RNN model. We followed the same procedure presented in Study 1, where at each cross-validation round (three in total), we used each subject three distinct measurements as a train, validation, and test sets respectively.

$$P(s_1,a_1)=\frac{\exp\left[\beta_1\cdot\left(Q^{net}(s_1,a_1)+\kappa_t\left(a\right)\right)\right]}{\sum_{\tilde{a}\in A}\exp\left[\beta_1\cdot\left(Q^{net}(s_1,\tilde{a})+\kappa_t\left(\tilde{a}\right)\right)\right]},\qquad \kappa_t\left(a\right)=\begin{cases}\kappa & \text{if } a=a_{t-1}\\ 0 & \text{otherwise}\end{cases}, \tag{26}$$

Hybrid. To comply with most studies that examined two-step task behavior using a hybrid model (Daw et al., 2011a; Gillan et al., 2016; Shahar et al., 2019), we included to the hybrid model described in Study 1 an additional first-stage choice perseveration parameter,

where $-0.5 < \kappa < 0.5$ is the choice perseveration free-parameter so that $\kappa > 0$ corresponded to selecting the same action as the previous trial and $\kappa < 0$ corresponded to switching to the other action. All other model details were similar to the model presented in Study 1. Overall, the model includes seven free parameters $\theta_{\text{hybrid}} = \{w, \alpha_1, \alpha_2, \lambda, \beta_1, \beta_2, \kappa\}$.

RNN. The RNN architecture was identical to the one described in Study 1 (see Theory-independent models) We first pre-trained the RNN using a subset of the NSPN dataset consisting of N = 515 individuals who performed the two-step task at two different time points (Time-I: 121 trials, Time-III: 201 trials). As in Study 1, we used one block (from each individual) for training and the other block for validation. We then fine-tuned the pre-trained RNN weights to fit the choice behavior of each of the N = 54 individuals using three different blocks for each individual, following the same cross-validation procedure of Study 1. In this procedure, each subject's three distinct measurements were used as training, validation, and test sets, respectively, in three rounds. We were concerned that it might be that the pre-training set (N = 515, two sessions) had a different proportion of high/low IQ participants compared to the main set that we were interested in

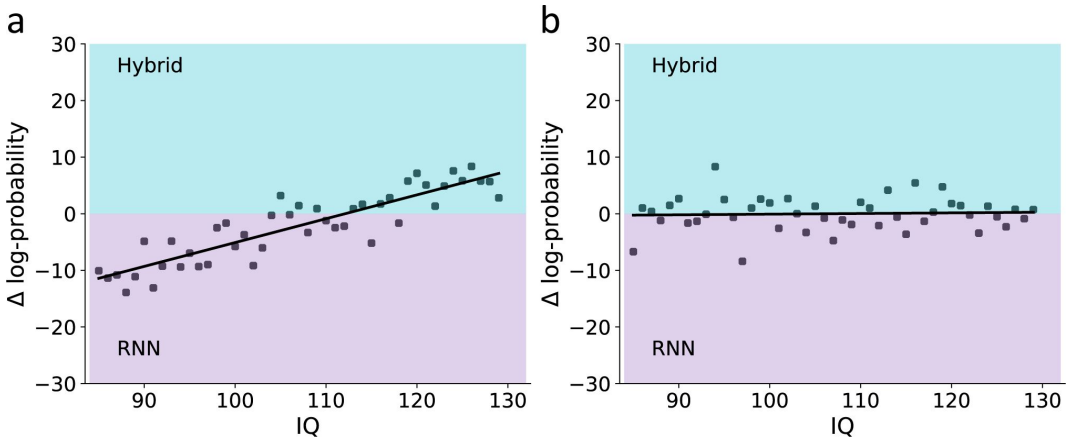

Figure 5

Prediction scenarios for the use of RNN to examine model miss-specification as a function of IQ.

To demonstrate the ability to use RNN to identify model miss-specification, we illustrate two hypothetical associations between IQ and the difference in predictive performance for a theoretical hybrid model vs. RNN. **(a)** Prediction for a scenario where there is a higher frequency of model miss-specification among low vs. high IQ individuals: Here, we assume that low IQ participants are more frequently miss-specified by the theoretical hybrid model compared with high IQ individuals. Therefore, the prediction here is that for low IQ individuals, the RNN will provide higher predictive gap improvement compared with high IQ individuals. **(b)** Prediction for a scenario where there is equal frequency of model miss-specification among low vs. high IQ individuals: Here, we assume that the frequency of model miss-specification is similar across levels of IQ. Therefore, we predict no association between IQ and the change in predictive accuracy for RNN vs. the theoretical hybrid model.

estimating (N = 54, three sessions). However, we tested and found that the distribution of IQ scores was identical for both subsets (M = 112.31, SD = 10.75 for N = 54 and M = 110.83, SD = 10.92 for N = 515; $t(567) = -0.954$, $p = 0.34$).

Results

Predictive performance. We began our investigation by estimating the correlation between IQ and the predictive score of the hybrid model. We found a positive correlation between individuals' IQ and the hybrid predictive score [see **Fig. 6a**; linear correlation coefficient $r = 0.28$, $p > 0.05$, 95% CI (0.014, 0.548)] so that lower IQ was associated with a higher predictive gap. This finding can reflect a systematic miss-specification of individuals with low IQ by the hybrid model. Alternatively, it might be that low IQ is associated with noisy behavior which then led to a higher predictive gap. To address this question, we predicted individuals' actions using RNN and examined whether the association between individuals' predictive gap and IQ is attenuated when using RNN. If the association between IQ and predictive gap of the theoretical hybrid model is due to higher rates of model misspecification for low compared with high IQ individuals, then RNN would attenuate this association. Specifically, RNN should be able to overcome the miss-specification issue the hybrid model might have, thus leading to a similar predictive gap across IQ levels. However, if the association between IQ and predictive gap of the theoretical hybrid model is mostly due to noisy behavior in low compared with high IQ, then we should observe the same correlation between predictive gap and IQ for both the theoretical hybrid model and RNN (see **Fig. 6b**, linear correlation coefficient $r = 0.07$, $p = 0.58$).

To examine whether RNN's predictive accuracy shows an attenuated association with IQ compared to the theoretical hybrid predictive accuracy, we estimated the paired interaction of model-type (RNN vs. Hybrid) and IQ on predictive accuracy estimates. Specifically, we fitted a hierarchical Bayesian regression model, where the dependent variable was the individual's predictive score (measured in negative log-probability) that was predicted using the individual's IQ score, model type (coded as -1 for hybrid and 1 for RNN), and their paired interaction. We found a main effect for IQ, such that individuals with higher IQ had higher log-probability scores (see **Fig. S5a** in the **SI**; posterior median = 0.614; 95%$_{HDI}$ between 0.022 and 1.208; probability of direction (pd) 97.9%). Moreover, we also found a main effect for Model-Type, such that the RNN model obtained on average a higher predictive score than the Hybrid (see **Fig. S5b** in the **SI**; posterior median = 9.577; 95%$_{HDI}$ between 6.94 and 12.2; probability of direction (pd) 100%).

Importantly, we did not find support for an interaction effect of IQ × Model-type on individuals' predictive scores, suggesting that the association between IQ and predictive accuracy was similar for the RNN and hybrid models (see **Fig. 6c**; posterior median = -0.11; 95% HDI between -0.35 and 0.13; probability of direction (pd) 81.1%). This result suggests that RNN did not improve the predictive accuracy of the hybrid model more for low vs high IQ individuals. Specifically, the lack of interaction between model-type and IQ can be taken as evidence that the frequency of model miss-specification is not different in low vs. high IQ individuals. To further assert our finding, we now turn to examine the association between IQ and the number of optimal RNN epochs as a proxy for the amount of noise in the individual's data.

Optimal epoch relation with IQ. The comparison of the predictive performance of the hybrid vs. RNN models suggested that low IQ individuals showed noisier behavior compared to their high IQ peers. To further validate our finding that lower IQ individuals' choice data is noisier (rather than more frequently miss-specified), we examined the number of optimal RNN training epochs for each participant. Here, we trained an individual RNN model for each subject initialized with random weights. We then examined the relationship between an individual's IQ score and their corresponding number of optimal RNN training epochs (i.e., the point at which we stopped training the RNN to avoid overfitting). We found a significant positive correlation, such that

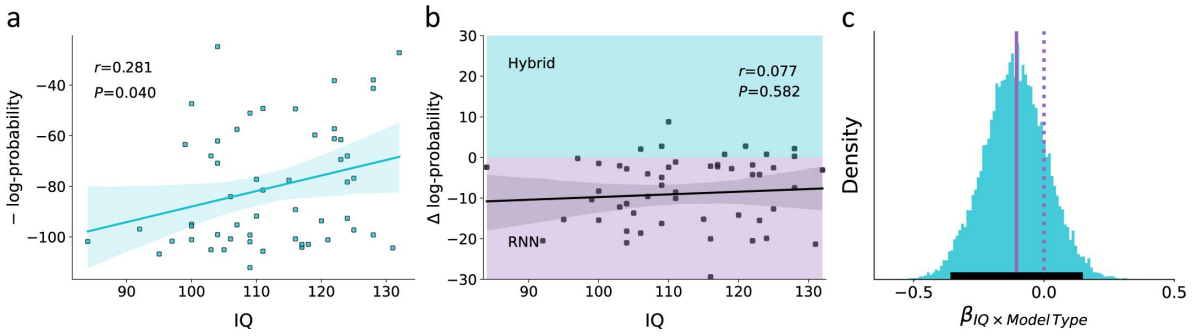

Figure 6

Empirical association of the predictive score for the theoretical hybrid model, RNN, and IQ.

(a) Association between IQ and the predictive score obtained from a theoretical hybrid model, showing that low IQ individuals are associated with lower predictive accuracy (shaded area corresponds to the 95% confidence interval for the regression line). We assumed that if this association is mostly due to higher rates of model miss-specification in low compared with high IQ individuals, then RNN should overcome this difficulty and show similar predictive accuracy across both models. **(b)** Difference in the predictive score for the theoretical hybrid model vs. RNN. Across different IQ scores, we found no significant difference in the predictive score of one of the models over the other. This finding suggests that the lower predictive score of low IQ individuals is not due to systematic model miss-specification, but rather due to noisier behavior (shaded turquoise/purple areas signify better predictive score of the Hybrid/RNN models, respectively; shaded dark area corresponds to the 95% confidence interval for the regression line). **(c)** Posterior distribution for the interaction term from a Bayesian regression, suggesting no effect for the paired interaction of model-type (hybrid vs. RNN) and IQ on the model's predictive score. This null effect suggests that the association between IQ and predictive accuracy was similar for both hybrid and RNN models (soiled/dashed purple lines indicate median/zero, respectively; lower solid black line indicates $95\%_{HDI}$). These results suggest that the higher predictive gap of the theoretical model for low compared with high IQ individuals is mostly due to individual differences in the levels of behavioral noise rather than systematic model miss-specification.

individuals with higher IQ had a higher number of optimal training epochs [see **Fig. 7a**; linear correlation coefficient $r = 0.33$, $p < 0.05$, $95\%_{CI}$ (0.072, 0.597)]. To further illustrate this finding, we divided participants into two groups according to their IQ score – low/high (using the median as the threshold). We then recorded the prediction of the validation data of each participant throughout the training procedure and averaged within each of the two groups (see **Fig. 7b**). We found that the optimal epoch of the high IQ group was significantly greater ($M = 332.80$, $SD = 178.06$) than that of the low IQ group ($M = 227.34$, $SD = 206.90$; $t(52) = 2.011$, $p < 0.05$). These findings suggest noisier behavior for low compared to high IQ individuals.

Discussion

Developing a theoretical computational model requires the researcher to state theoretical assumptions regarding the investigated process, assumptions that might differ from the true data generating process, and thus exposed to the problem of model miss-specification. Here, we sought to construct a method that tackles this problem in the context of theoretical computational models of human decision-making behavior. Taking advantage of the high flexibility of theory-independent models (i.e., RNN), that are not theoretically constrained by assumptions of behavior, we proposed a method to indicate if a predictive gap observed in a single agent choice data is mostly due to model miss-specification or rather an inherent stochasticity in the behavior being modeled. We further suggest that the number of training epochs, used to reach an optimal balance between under and overfitting for RNN can be used as a relative estimate of noise in the individual behavior.

In Study 1, we validated our approach using simulated data from artificial agents generated by different theoretical models in a two-step multi-armed bandit task. By comparing the predictive performance of the RNN model with each theoretical model, we demonstrated the RNN's ability to determine whether a predictive gap observed in an individual agent is primarily due to model miss-specification or inherent stochasticity in their behavior. Additionally, we showed that the point of early stopping in the RNN training process is strongly associated with the level of true irreducible stochasticity in an agent's behavior. A higher number of optimal training epochs is indicative of less noise in the agent's behavior, given a fixed amount of data and network size. One practical advantage of using the point of early stopping is that it can be estimated using only two sets of data (training and validation), unlike predictive accuracy, which requires an additional test data set.

Next, in Study 2, we applied these methods to an empirical dataset of human participants performing the same two-step task. We found that Daw's hybrid model (Daw et al., 2011a), a well-known theoretical model, exhibited a consistently poorer fit to participants with low IQ compared to those with high IQ. To investigate the source of this poorer fit, we individually fitted an RNN model to each participant and compared the predictive performance of both the hybrid model and the RNN model, as well as the point of early stopping for each participant. We observed that the RNN model showed a similar reduction in behavior prediction for both low and high IQ individuals, consistent with the hybrid model. Additionally, the RNN's optimal epoch estimates were systematically higher for high IQ individuals. These findings provide preliminary evidence suggesting that the behavior of low IQ participants is characterized by greater inherent noise, rather than being more mis-specified by the hybrid model, when compared to their high IQ counterparts.

The use of theory-independent neural network models fitted directly to behavioral data in decision-making tasks has garnered significant attention in recent years (Dezfouli et al., 2019a,b; Fintz et al., 2022; Peterson et al., 2021; Song et al., 2021). Previous research has highlighted the advantages of this approach by demonstrating that recurrent neural networks (RNNs) are capable of capturing behavioral features that traditional theoretical models struggle to

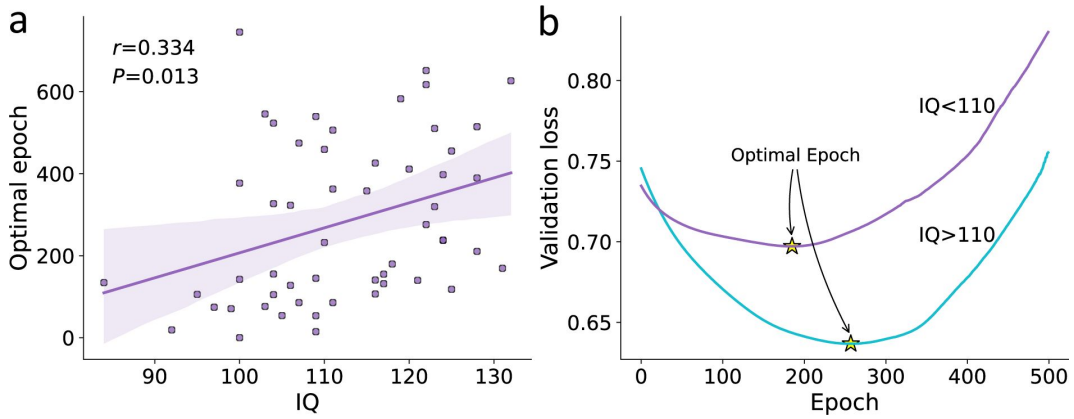

Figure 7

Association between IQ and individual's optimal epoch estimates.

(a) Relation between the number of optimal RNN training epochs (y-axis) and IQ score (x-axis; shaded area corresponds to the 95% confidence interval for the regression line). We found that IQ significantly correlates with individuals' optimal RNN epoch estimates, such that lower IQ participants required fewer RNN training epochs to reach the optimal training point. **(b)** Loss curve of validation data averaged over the low IQ group (IQ < 110; purple) and high IQ group (IQ > 110; turquoise). The point of optimal epoch (early stopping) is denoted by a yellow star. We found that the optimal epoch for the high IQ group was significantly higher than that for the low IQ group. Overall, when combined with our simulation study demonstrating the association between the number of optimal RNN epochs and true noise (for a fixed number of observations and network size; see **Fig. 4**), these results suggest noisier decision-making for low IQ individuals compared to high IQ individuals. This finding, along with the results obtained from the RNN predictive accuracy (see **Fig. 6**), suggests that low IQ individuals are not more frequently miss-specified compared to their high IQ peers.

replicate. In a groundbreaking study by Dezfouli et al. (2019b), it was shown that an RNN model fitted to participants performing a two-armed bandit task outperformed a typical reinforcement learning (RL) model in pure action prediction. This finding has been extended to more complex decision-making tasks such as the four-armed bandit (Fintz et al., 2022) and build-your-own-stimulus tasks (Song et al., 2021). While most studies have primarily focused on the high predictive performance of RNN models, there are also examples of leveraging RNNs to gain theoretical insights. For instance, Dezfouli, Ashtiani, et al. (2019) developed an autoencoder model utilizing an RNN encoder to map each participant's behavior into a low-dimensional latent space, which facilitated the study of individual differences and variation in decision-making strategies. In another study, Song et al. (2021) proposed an RNN model trained with an embedding specific to each participant, showing that these embeddings were effective in capturing differences in various cognitive variables. Thus, previous research has demonstrated that RNNs, unconstrained by specific theoretical assumptions, have the ability to capture and predict human behavior, resulting in remarkably low predictive gaps.

We believe that our work contributes to the growing body of research in the field, making four key contributions. Firstly, through simulation, we have validated the assertion that RNNs can achieve nearly optimal fit and approximate various behavioral models without the need for prior theoretical specification. Secondly, we introduce a novel application of the point of early stopping in RNN training. Traditionally utilized to prevent overfitting, we demonstrate its utility as a unique individual measure of the amount of inherent noise present in both artificial and human behavioral choice data. Unlike predictive accuracy, which requires group-level pre-training and three data sets per individual (training, validation, and test), the point of early stopping only necessitates two data sets (training and validation) and can be employed without empirical RNN pre-training. This allows for a more flexible estimate of stopping points between individuals. Thirdly, we present an application of the RNN model to human choice data in the context of a multi-stage two-step decision task, expanding upon previous research that has predominantly focused on single-stage tasks. Lastly, we utilize the RNN model to investigate the relationship between model miss-specification and individuals' intelligence scores.

Our work also addresses the question of inherent randomness or irreducible stochasticity in human behavior. The question of irreducible stochasticity has broader implications not only for cognitive science but also as a fundamental question in general science. We provide provisional evidence suggesting that individuals with lower intelligence scores exhibit noisier decision patterns that cannot be reduced by any model. However, several questions remain unanswered, such as identifying the specific stages of the decision-making process that underlie this behavioral variability. In our current work, we primarily focus on noise that arises from the deliberation process. This aligns with previous studies that have proposed the concept of "decision acuity" (Moutoussis et al., 2021), which represents a dimensional structure in core decision-making strongly related to the inverse temperature parameter (β). It suggests that decision variability originates from differences in reward sensitivities. However, another line of research has proposed the idea of "computation noise," which refers to random noise inherent to the learning process that corrupts the value updating of each action (Findling et al., 2019; Findling and Wyart, 2021). Further studies are needed to elucidate the extent to which these two factors contribute to the noisier decision patterns observed in individuals with lower intelligence. Another open question pertains to identifying the neural correlates and mechanisms underlying this variability. Nevertheless, the significant advantage of the current method is that researchers can use the number of optimal RNN epochs to estimate the amount of noise in observations without specifying a theoretical mechanism. This capability enhances the interpretation and utilization of theoretical models.

Several limitations should be considered in our proposed approach. The first limitation is that the empirical dataset used in our study consisted of data collected from human participants at a single time point, serving as the training set for our RNN. The test set data, collected with a time interval

of approximately 6 and 18 months, introduced the possibility of changes in participants' decision-making strategies over time. In our analysis, we neglected any possible changes in participants' decision-making strategies during that time, changes that may lead to poorer generalization performance of our approach. Thus, further studies are needed to eliminate such possible explanations. Second, the interpretation of the number of optimal epochs as an estimate of the amount of information/noise in the individual data is relative and can only be interpreted at this point against other datasets. While we found that the association between the optimal number of epochs and true noise is robust across network size and the number of observations (see **Fig. S2**), the interpretation of this estimate can be made only when two datasets are estimated with the same number of observations and network size.

To sum up, we show that both the predictive gap and the point of early stopping of neural networks can be used to estimate whether a certain predictive gap found for a theoretical model is mostly due to model miss-specification or irreducible noise. We hope this work would lead to further studies exploring the utilization of neural networks to enhance theoretical computational models.

Code availability

The code is publicly available via: *https://github.com/yoavger/using_rnn_to_estimate_irreducible_stochasticity*

Acknowledgements

The study was funded by the Israel Ministry of Science and Technology, and the Israeli Science Foundation (Grant 2536/20 awarded to N.S.).

Supplementary Information

S1. Simulation specification

In study 1 we simulated 300 agents performing the two-step task (Daw et al., 2011) for three blocks of 200 trials each. For each agent, his model parameters were sampled randomly according to Table S1.

Model	Hybrid Daw et al. (2011)	Habit Miller et al. (2019)	K-DH
Parameter range	$w \sim U(0,1)$ $\alpha_{1,2} \sim U(0,1)$ $\beta_{1,2} \sim U(0,10)$ $\lambda \sim U(0,1)$	$\alpha_{1,2} \sim U(0,1)$ $\beta_{1,2} \sim U(0,10)$	$p_{0,1,2,3} \sim U(0,1)$
# Agents	100	100	100
# trials	200	200	200
# blocks	3	3	3

Table S1

Simulation specification for Study 1.

S2. Varying the size of the network hidden layer and the number of simulated trials

To test how the network's hidden size and the number of simulated trials affect the relationship between the number of optimal epochs and the true noise each agent holds, we conducted a repeated analysis similar to study 1 in the main text. In this analysis, we simulated agents with varying trial lengths (100, 200, and 500 trials per block) and trained RNN models with different hidden sizes (2, 5, and 10 cells in the GRU layer). For each combination, we simulated 100 agents from each of the three theoretical models (Hybrid, Habit, and K-DH), resulting in a total of 300 agents. Subsequently, we fitted an RNN model to the behavior of each agent, training it with only one block and using a withheld block to determine the optimal epoch for each agent. Across all combinations, we observed a strong relationship between the number of optimal epochs and the true noise held by each agent. Specifically, agents with a high level of true noise were better explained by an RNN model trained with a small number of epochs, while agents with a low level of true noise were better explained by an RNN model trained with more epochs (see **Fig. S2**).

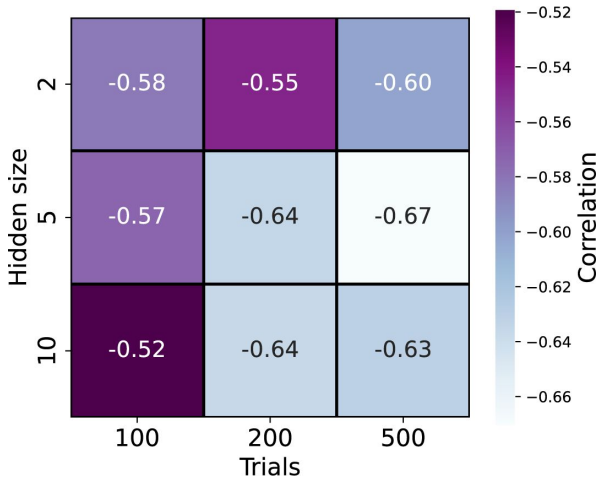

Figure S2

Correlation matrix between optimal epoch and the true noise.

Each square denotes the Pearson correlation between the number of optimal epochs and the true noise each agent holds. For each square, the correlation was calculated using agents that were simulated for 100, 200, or 500 trials (X-axis) and the RNN model hidden layer consists of 2, 5, or 10 cells in the hidden GRU layer (Y-axis). We found, across all combinations a negative relation such that longer training was most beneficial in predicting the action of agents with low levels of true noise.

S3. Sensitivity to the lengths of simulated trials

To evaluate the sensitivity of our proposed method to the number of simulated trials, we conducted a repeated analysis similar to study 1 in the main text (see Classification of model misspecification using predictive performance in the main text). Specifically, we generated a comparable artificial dataset to that of study 1, with the only difference being that we varied the number of simulated trials within each block (100 and 500 trials). Importantly, the size of the hidden GRU layer in the RNN was held constant at 5 cells across all conditions. Across all lengths of simulated trials, we observed a similar overall result to the main text. In particular, regardless of the theoretical models used, the RNN achieved the second-best predictive score, surpassed only by the true generative model (see **Fig. S3**, left panel). Additionally, the RNN model exhibited higher classification accuracy compared to the LR model (see **Fig. S3**, right panel). These findings suggest that our method is relatively insensitive to the number of trials used in the simulations.

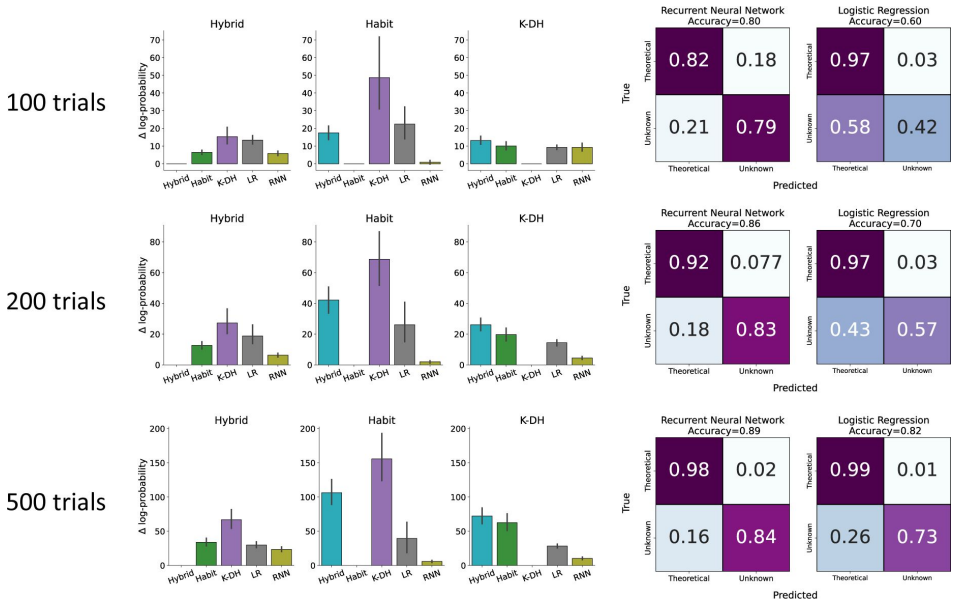

Figure S3

Action prediction and agent classification for different lengths of simulated trials.

The difference in the negative log-probability obtained by each model to each group of agents (left) and Classification of each agent (right) for various simulated trial lengths [100, 200 (**Fig. 3** in the main text), 500]. Overall RNN outperformed the LR model in both predictive score and classification accuracy.

S4. Classification of model-misspecification within each theoretical model

In the main text, we presented the average results obtained from all classification rounds of the hypothetical experiment (see **Fig. 3b** in the main text). In this section, we provide the confusion matrices for each round, which display the classification performance of each hypothetical lab (see **Fig. S4**). Across all theoretical models, we observed that the RNN achieved a higher true negative rate (bottom right square) compared to logistic regression. Specifically, the RNN exhibited a better ability to correctly identify cases where an agent was better explained by another unknown model. On the other hand, logistic regression demonstrated a higher true positive rate (upper right square) compared to the RNN. However, when considering overall classification accuracy, the RNN model outperformed logistic regression. This finding provides further support for our claim that the RNN model offers greater expressiveness compared to logistic regression.

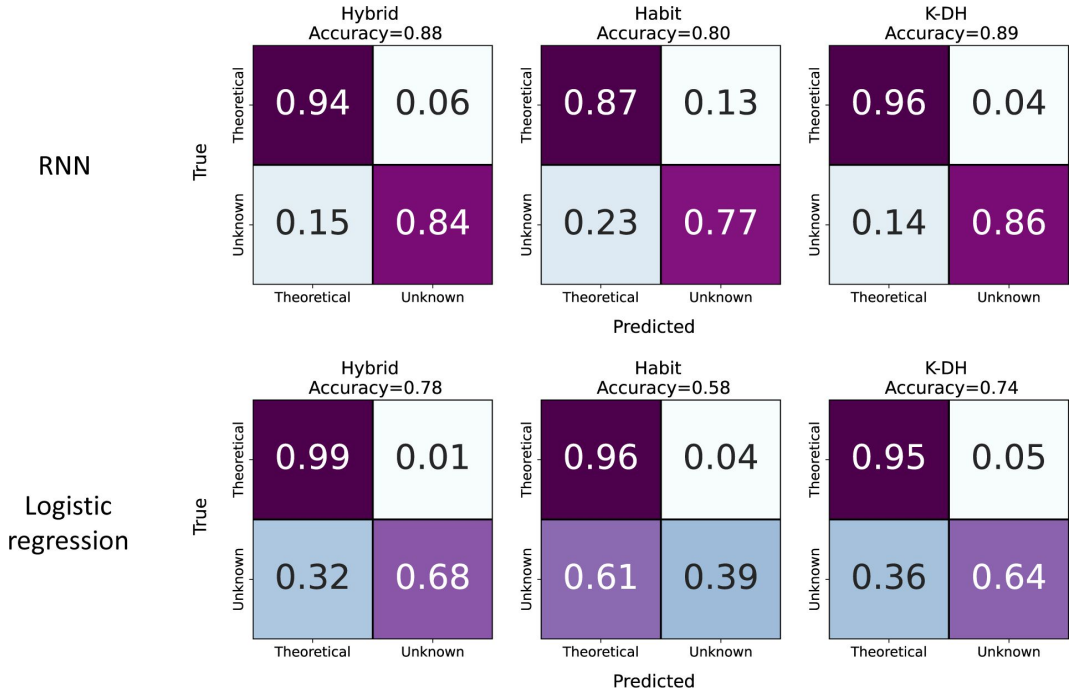

Figure S4

Agent classification of each hypothetical lab.

Classification by RNN (top) and Classification by LR (bottom). Across all theoretical models (Hybrid, Habit, and K-DH), RNN achieved a higher True negative rate (bottom right square) compared to Logistic regression. In addition, although RNN achieved a lower True positive rate (upper left square) compared to Logistic regression, the overall classification accuracy of the RNN model is higher.

S5. Bayesian regression analysis

We provide two additional posterior distributions of the Bayesian regression coefficient, as discussed in Study 2 (see Predictive performance in the main text). Firstly, we observed a main effect for IQ, indicating that individuals with higher IQ scores had higher log-probability scores (see **Fig. S5a**; posterior median=0.614; 95%$_{HDI}$ between 0.022 and 1.208; probability of direction (pd) 97.9%). Secondly, We also found a main effect for Model-Type, such that the RNN model obtains on averaged higher predictive score than the Hybrid (**Fig. S5b**, posterior median=9.577; 95%$_{HDI}$ between 6.94 and 12.2; probability of direction (pd) 100%).

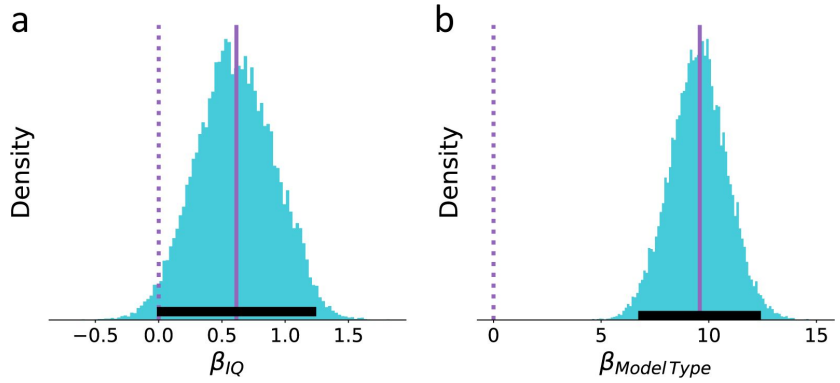

Figure S5

Posterior distribution for the regression coefficient.

(a) Individual's IQ score and **(b)** model-type on predictive Log-probability scores (soiled/dashed purple lines indicate median/zero, respectively; lower solid black line indicates 95%$_{HDI}$). We found two main effects of individual's IQ and model-type on predictive Log-probability scores. Such that, low IQ participant's choices are less predictable than high IQ participant's and that on average the RNN model reaches a better predictive score than the Hybrid model.

References

Barak O. (2017) **Recurrent neural networks as versatile tools of neuroscience research** *Current opinion in neurobiology* **46**:1–6

Beck J. M., Ma W. J., Pitkow X., Latham P. E., Pouget A. (2012) **Not noisy, just wrong: the role of suboptimal inference in behavioral variability** *Neuron* **74**:30–39

Bishop C. M., Nasrabadi N. M. (2006) **Pattern recognition and machine learning**

Box G. E. (1979) **Robustness in the strategy of scientific model building** *Robustness in statistics* :201–236

Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. (2014) **Learning phrase representations using rnn encoder-decoder for statistical machine translation** *arXiv preprint arXiv:1406.1078*

Daw N. D., Gershman S. J., Seymour B., Dayan P., Dolan R. J. (2011) **Model-based influences on humans' choices and striatal prediction errors** *Neuron* **69**:1204–1215

Daw N. D. (2011) **Trial-by-trial data analysis using computational models** *Decision making, affect, and learning: Attention and performance XXIII* **23**

Dezfouli A., Ashtiani H., Ghattas O., Nock R., Dayan P., Ong C. S. (2019) **Disentangled behavioural representations** *Advances in neural information processing systems* **32**

Dezfouli A., Griffiths K., Ramos F., Dayan P., Balleine B. W. (2019) **Models that learn how humans learn: the case of decision-making and its disorders** *PLoS computational biology* **15**

Eckstein M. K., Wilbrecht L., Collins A. G. (2021) **What do reinforcement learning models measure? interpreting model parameters in cognition and neuroscience** *Current opinion in behavioral sciences* **41**:128–137

Faisal A. A., Selen L. P., Wolpert D. M. (2008) **Noise in the nervous system** *Nature reviews neuroscience* **9**:292–303

Findling C., Wyart V. (2021) **Computation noise in human learning and decision-making: origin, impact, function** *Current Opinion in Behavioral Sciences* **38**:124–132

Findling C., Skvortsova V., Dromnelle R., Palminteri S., Wyart V. (2019) **Computational noise in reward-guided learning drives behavioral variability in volatile environments** *Nature neuroscience* **22**:2066–2077

Fintz M., Osadchy M., Hertz U. (2022) **Using deep learning to predict human decisions and using cognitive models to explain deep learning models** *Scientific reports* **12**

Gillan C. M., Kosinski M., Whelan R., Phelps E. A., Daw N. D. (2016) **Characterizing a psychiatric symptom dimension related to deficits in goal-directed control** *elife* **5**

Gleick J. (2011) **The information: A history, a theory, a flood**

Griffiths D. J., Schroeter D. F. (2018) **Introduction to quantum mechanics**

Hasson U., Nastase S. A., Goldstein A. (2020) **Direct fit to nature: an evolutionary perspective on biological and artificial neural networks** *Neuron* **105**:416–434

Hornik K., Stinchcombe M., White H. (1989) **Multilayer feedforward networks are universal approximators** *Neural networks* **2**:359–366

Kiddle B., Inkster B., Prabhu G., Moutoussis M., Whitaker K. J., Bullmore E. T., Dolan R. J., Fonagy P., Goodyer I. M., Jones P. B. (2018) **Cohort profile: the nspn 2400 cohort: a developmental sample supporting the wellcome trust neuroscience in psychiatry network** *International journal of epidemiology* **47**:18–19

Kingma D. P., Ba J. (2014) **Adam: A method for stochastic optimization** *arXiv preprint arXiv:1412.6980*

LeCun Y., Bengio Y., Hinton G. (2015) **Deep learning** *nature* **521**:436–444

McElreath R. (2020) **Statistical rethinking: A Bayesian course with examples in R and Stan**

Miller K. J., Shenhav A., Ludvig E. A. (2019) **Habits without values** *Psychological review* **126**

Montague P. R., Dolan R. J., Friston K. J., Dayan P. (2012) **Computational psychiatry** *Trends in cognitive sciences* **16**:72–80

Moutoussis M., Garzón B., Neufeld S., Bach D. R., Rigoli F., Goodyer I., Bullmore E., Fonagy P., Jones P., Hauser T. (2021) **Decision-making ability, psychopathology, and brain connectivity** *Neuron* **109**:2025–2040

Nassar M. R., Frank M. J. (2016) **Taming the beast: extracting generalizable knowledge from computational models of cognition** *Current opinion in behavioral sciences* **11**:49–54

Palminteri S., Wyart V., Koechlin E. (2017) **The importance of falsification in computational cognitive modeling** *Trends in cognitive sciences* **21**:425–433

Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., Killeen T., Lin Z., Gimelshein N., Antiga L. (2019) **Pytorch: An imperative style, high-performance deep learning library** *Advances in neural information processing systems* **32**

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V. (2011) **Scikit-learn: Machine learning in python** *the Journal of machine Learning research* **12**:2825–2830

Peterson J. C., Bourgin D. D., Agrawal M., Reichman D., Griffiths T. L. (2021) **Using large-scale experiments and machine learning to discover theories of human decision-making** *Science* **372**:1209–1214

Rescorla R. A. (1972) **A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement** *Classical conditioning, Current research and theory* **2**:64–69

Rigoux L., Stephan K. E., Friston K. J., Daunizeau J. (2014) **Bayesian model selection for group studies—revisited** *Neuroimage* **84**:971–985

Shahar N., Moran R., Hauser T. U., Kievit R. A., McNamee D., Moutoussis M., Consortium N., Dolan R. J. (2019) **Credit assignment to state-independent task representations and its relationship with model-based decision making** *Proceedings of the National Academy of Sciences* **116**:15871–15876

Siegelmann H. T., Sontag E. D. (1992) **H. T. Siegelmann and E. D. Sontag. On the computational power of neural nets. In Proceedings of the fifth annual workshop on Computational learning theory, pages 440–449, 1992.** :440–449

Smith P. L., Ratcliff R. (2004) **Psychology and neurobiology of simple decisions** *Trends in neurosciences* **27**:161–168

Song M., Niv Y., Cai M. (2021) **Using recurrent neural networks to understand human reward learning** *Proceedings of the Annual Meeting of the Cognitive Science Society* **43**:1388–1394

Stephan K. E., Penny W. D., Daunizeau J., Moran R. J., Friston K. J. (2009) **Bayesian model selection for group studies** *Neuroimage* **46**:1004–1017

Sutton R. S., Barto A. G. (2018) **Reinforcement learning: An introduction**

Virtanen P., Gommers R., Oliphant T. E., Haberland M., Reddy T., Cournapeau D., Burovski E., Peterson P., Weckesser W., Bright J. (2020) **Scipy 1.0: fundamental algorithms for scientific computing in python** *Nature methods* **17**:261–272

Wechsler D. (1999) **Wechsler abbreviated scale of intelligence**

Wilson R. C., Collins A. G. (2019) **Ten simple rules for the computational modeling of behavioral data** *Elife* **8**

Yarkoni T., Westfall J. (2017) **Choosing prediction over explanation in psychology: Lessons from machine learning** *Perspectives on Psychological Science* **12**:1100–1122

Article and author information

Yoav Ger
Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
For correspondence: yoavger@mail.tau.ac.il

Moni Shahar
TAD - Center of AI & Data Science, Tel Aviv University, Tel Aviv, Israel

Nitzan Shahar
Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel, School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel

Copyright

source are credited.

Editors

Reviewing Editor
Joshua Gold
University of Pennsylvania, Philadelphia, United States of America

Senior Editor
Joshua Gold
University of Pennsylvania, Philadelphia, United States of America

Reviewer #1 (Public Review):

Summary:
Ger and colleagues address an issue that often impedes computational modeling: the inherent ambiguity between stochasticity in behavior and structural mismatch between the assumed and true model. They propose a solution to use RNNs to estimate the ceiling on explainable variation within a behavioral dataset. With this information in hand, it is possible to determine the extent to which "worse fits" result from behavioral stochasticity versus failures of the cognitive model to capture nuances in behavior (model misspecification). The authors demonstrate the efficacy of the approach in a synthetic toy problem and then use the method to show that poorer model fits to 2-step data in participants with low IQ are actually due to an increase in inherent stochasticity, rather than systemic mismatch between model and behavior.

Strengths:
Overall I found the ideas conveyed in the paper interesting and the paper to be extremely clear and well-written. The method itself is clever and intuitive and I believe it could be useful in certain circumstances, particularly ones where the sources of structure in behavioral data are unknown. In general, the support for the method is clear and compelling. The flexibility of the method also means that it can be applied to different types of behavioral data - without any hypotheses about the exact behavioral features that might be present in a given task.

Weaknesses:
That said, I have some concerns with the manuscript in its current form, largely related to the applicability of the proposed methods for problems of importance in computational cognitive neuroscience. This concern stems from the fact that the toy problem explored in the manuscript is somewhat simple, and the theoretical problem addressed in it could have been identified through other means (for example through the use of posterior predictive checking for model validation), and the actual behavioral data analyzed were interpreted as a null result (failure to reject that the behavioral stochasticity hypothesis), rather than actual identification of model-misspecification. I expand on these primary concerns and raise several smaller points below.

A primary question I have about this work is whether the method described would actually provide any advantage for real cognitive modeling problems beyond what is typically done to minimize the chance of model misspecification (in particular, post-predictive checking). The toy problem examined in the manuscript is pretty extreme (two of the three synthetic agents are very far from what a human would do on the task, and the models deviate from one another to a degree that detecting the difference should not be difficult for any method). The issue posed in the toy data would easily be identified by following good modeling practices, which include using posterior predictive checking over summary measures to identify model insufficiencies, which in turn would call for the need for a broader set of models (See Wilson

& Collins 2019). Thus, I am left wondering whether this method could actually identify model misspecification in real world data, particularly in situations where standard posterior predictive checking would fall short. The conclusions from the main empirical data set rest largely on a null result, and the utility of a method for detecting model misspecification seems like it should depend on its ability to detect its presence, not just its absence, in real data.

Beyond the question of its advantage above and beyond data- and hypothesis-informed methods for identifying model misspecification, I am also concerned that if the method does identify a model-insufficiency, then you still would need to use these other methods in order to understand what aspect of behavior deviated from model predictions in order to design a better model. In general, it seems that the authors should be clear that this is a tool that might be helpful in some situations, but that it will need to be used in combination with other well-described modeling techniques (posterior predictive checking for model validation and guiding cognitive model extensions to capture unexplained features of the data). A general stylistic concern I have with this manuscript is that it presents and characterizes a new tool to help with cognitive computational modeling, but it does not really adhere to best modeling practices (see Collins & Wilson, eLife), which involve looking at data to identify core behavioral features and simulating data from best-fitting models to confirm that these features are reproduced. One could take away from this paper that you would be better off fitting a neural network to your behavioral data rather than carefully comparing the predictions of your cognitive model to your actual data, but I think that would be a highly misleading takeaway since summary measures of behavior would just as easily have diagnosed the model misspecification in the toy problem, and have the added advantage that they provide information about which cognitive processes are missing in such cases.

As a more minor point, it is also worth noting that this method could not distinguish behavioral stochasticity from the deterministic structure that is not repeated across training/test sets (for example, because a specific sequence is present in the training set but not the test set). This should be included in the discussion of method limitations. It was also not entirely clear to me whether the method could be applied to real behavioral data without extensive pretraining (on >500 participants) which would certainly limit its applicability for standard cases.

The authors focus on model misspecification, but in reality, all of our models are misspecified to some degree since the true process-generating behavior almost certainly deviates from our simple models (ie. as George Box is frequently quoted, "all models are wrong, but some of them are useful"). It would be useful to have some more nuanced discussion of situations in which misspecification is and is not problematic.

- https://doi.org/10.7554/eLife.90082.1.sa1

Reviewer #2 (Public Review):

SUMMARY:
In this manuscript, Ger and colleagues propose two complementary analytical methods aimed at quantifying the model misspecification and irreducible stochasticity in human choice behavior. The first method involves fitting recurrent neural networks (RNNs) and theoretical models to human choices and interpreting the better performance of RNNs as providing evidence of the misspecifications of theoretical models. The second method involves estimating the number of training iterations for which the fitted RNN achieves the best prediction of human choice behavior in a separate, validation data set, following an approach known as "early stopping". This number is then interpreted as a proxy for the amount of explainable variability in behavior, such that fewer iterations (earlier stopping) correspond to a higher amount of irreducible stochasticity in the data. The authors validate

the two methods using simulations of choice behavior in a two-stage task, where the simulated behavior is generated by different known models. Finally, the authors use their approach in a real data set of human choices in the two-stage task, concluding that low-IQ subjects exhibit greater levels of stochasticity than high-IQ subjects.

STRENGTHS:
The manuscript explores an extremely important topic to scientists interested in characterizing human decision-making. While it is generally acknowledged that any computational model of behavior will be limited in its ability to describe a particular data set, one should hope to understand whether these limitations arise due to model misspecification or due to irreducible stochasticity in the data. Evidence for the former suggests that better models ought to exist; evidence for the latter suggests they might not.

To address this important topic, the authors elaborate carefully on the rationale of their proposed approach. They describe a variety of simulations - for which the ground truth models and the amount of behavioral stochasticity are known - to validate their approaches. This enables the reader to understand the benefits (and limitations) of these approaches when applied to the two-stage task, a task paradigm commonly used in the field. Through a set of convincing analyses, the authors demonstrate that their approach is capable of identifying situations where an alternative, untested computational model can outperform the set of tested models, before applying these techniques to a realistic data set.

WEAKNESSES:
The most significant weakness is that the paper rests on the implicit assumption that the fitted RNNs explain as much variance as possible, an assumption that is likely incorrect and which can result in incorrect conclusions. While in low-dimensional tasks RNNs can predict behavior as well as the data-generating models, this is not *always* the case, and the paper itself illustrates (in Figure 3) several cases where the fitted RNNs fall short of the ground-truth model. In such cases, we cannot conclude that a subject exhibiting a relatively poor RNN fit necessarily has a relatively high degree of behavioral stochasticity. Instead, it is at least conceivable that this subject's behavior is generated precisely (i.e., with low noise) by an alternative model that is poorly fit by an RNN - e.g., a model with long-term sequential dependencies, which RNNs are known to have difficulties in capturing.

These situations could lead to incorrect conclusions for both of the proposed methods. First, the model misspecification analysis might show equal predictive performance for a particular theoretical model and for the RNN. While a scientist might be inclined to conclude that the theoretical model explains the maximum amount of explainable variance and therefore that no better model should exist, the scenario in the previous paragraph suggests that a superior model might nonetheless exist. Second, in the early-stopping analysis, a particular subject may achieve optimal validation performance with fewer epochs than another, leading the scientist to conclude that this subject exhibits higher behavioral noise. However, as before, this could again result from the fact that this subject's behavior is produced with little noise by a different model. Admittedly, the existence of such scenarios *in principle* does not mean that such scenarios are common, and the conclusions drawn in the paper are likely appropriate for the particular examples analyzed. However, it is much less obvious that the RNNs will provide optimal fits in other types of tasks, particularly those with more complex rules and long-term sequential dependencies, and in such scenarios, an ill-advised scientist might end up drawing incorrect conclusions from the application of the proposed approaches.

In addition to this general limitation, the paper also makes a few additional claims that are not fully supported by the provided evidence. For example, Figure 4 highlights the relationship between the optimal epochs and agent noise. Yet, it is nonetheless possible that the optimal epoch is influenced by model parameters other than inverse temperature (e.g., learning rate). This could again lead to invalid conclusions, such as concluding that low-IQ is

associated with optimal epoch when an alternative account might be that low-IQ is associated with low learning rate, which in turn is associated with optimal epoch. Yet additional factors such as the deep double-descent (Nakkiran et al., ICLR 2020) can also influence the optimal epoch value as computed by the authors.

An additional issue is that Figure 4 reports an association between optimal epoch and noise, but noise is normalized by the true minimal/maximal inverse-temperature of hybrid agents (Eq. 23). It is thus possible that the relationship does not hold for more extreme values of inverse-temperature such as beta=0 (extremely noisy behavior) or beta=inf (deterministic behavior), two important special cases that should be incorporated in the current study. Finally, even taking the association in Figure 4 at face value, there are potential issues with inferring noise from the optimal epoch when their correlation is only r~=0.7. As shown in the figures, upon finding a very low optimal epoch for a particular subject, one might be compelled to infer high amounts of noise, even though several agents may exhibit a low optimal epoch despite having very little noise.

APPRAISAL AND DISCUSSION:
Overall, the authors propose a novel method that aims to solve an important problem, but whose generality might be limited only to special cases. In the future, it would be beneficial to test the proposed approach in a broader setting, including simulations of different tasks, different model classes, different model parameters, and different amounts of behavioral noise. Nonetheless, even without such additional work, the proposed methods are likely to be used by cognitive scientists and neuroscientists interested in assessing the quality and limits of their behavioral models.

- https://doi.org/10.7554/eLife.90082.1.sa0