A generalized method for dynamic noise inference in modeling sequential decision-making

Jing-Jing Li (jl3676@berkeley.edu)
Helen Wills Neuroscience Institute
University of California, Berkeley

Chengchun Shi (c.shi7@lse.ac.uk)
Department of Statistics
London School of Economics and Political Science

Lexin Li (lexinli@berkeley.edu)
Department of Biostatistics and Epidemiology
Helen Wills Neuroscience Institute
University of California, Berkeley

Anne G.E. Collins (annecollins@berkeley.edu)
Department of Psychology
Helen Wills Neuroscience Institute
University of California, Berkeley

Abstract

Computational cognitive modeling is an important tool for understanding the processes that support human and animal decision-making. Choice data in sequential decision-making tasks are inherently noisy, and separating noise from signal can improve the quality of computational modeling. Currently, most models assume that noise is constant, or *static*, typically by including a parameter (e.g., uniform ε) to estimate the noise level. However, this assumption is not guaranteed to hold – for example, an agent can lapse into an inattentive phase for a series of trials in the middle of otherwise low-noise performance. Assuming that noise is static could bias parameter and model identification. Here, we propose a new method to *dynamically* infer noise in choice behavior, under a model assumption that agents can transition between two discrete latent states (for example, attentive and noisy). Using four empirical datasets with diverse behavioral and modeling features, we demonstrate that our method improves model fit and that it can be easily incorporated into existing fitting procedures, including maximum likelihood estimation and hierarchical Bayesian modeling.

Keywords: computational modeling; decision-making; noise; human behavior; animal behavior; reinforcement learning

Introduction

Computational modeling has helped cognitive scientists, psychologists, and neuroscientists translate theories into quantitative predictions, and to better explain behavior with mathematical equations (Palminteri, Wyart, & Koechlin, 2017; Wilson & Collins, 2019). Computational modeling is particularly useful for explaining choice behavior in decision-making tasks – it reveals links between participants' observable choices and putative latent internal variables such as objective or subjective value (Tversky & Kahneman, 1992), strength of evidence (Bitzer, Park, Blankenburg, & Kiebel, 2014), and history of past outcomes (Dayan & Niv, 2008). This link between internal latent variables and choices is made via a policy: the probability of making a choice among multiple options based on past and current information.

An important feature of choice behavior produced by biological agents is the inherent noise and variance, which can be attributed to multiple sources including inattention (Esterman & Rothlein, 2019; Warm, Parasuraman, & Matthews, 2008), stochastic exploration (Wilson, Geana, White, Ludvig, & Cohen, 2014), and internal computation noise (Findling & Wyart, 2021). Choice randomization can be adaptive, as it encourages exploration and can come close to optimal performance if implemented correctly (Chapelle & Li, 2011; Thompson, 1933; Wang & Wilson, 2018).

Computational cognitive models often focus on noiseless information processing over internal latent variables – for example, in reinforcement learning, how the choice values are updated with each outcome (Daw & Tobler, 2014). A common approach to model behavioral noise is to include simple parameterized noise into the model's policy (Wilson & Collins, 2019). For example, a greedy policy can be "softened" by a logistic or softmax function with an inverse temperature parameter, β, such that choices among more similar options are noisier than choices among more different ones. Another approach is to use an ε-greedy policy, where the noise level parameter, ε, blends a uniform decision noise into a greedy policy. This approach is motivated by a different intuition: that lapses in choice patterns can happen independently of the specific internal values used to make decisions. Multiple noise processes can be used jointly in a model when appropriate (Collins & Frank, 2012).

Failure to account for a noisy choice process in modeling could lead to under- or over-emphasis of certain data points, and thus inappropriate conclusions (Nassar & Frank, 2016; Schaaf, Jepma, Visser, & Huizenga, 2019). However, commonly used policies with noisy decision processes share strong assumptions. In particular, they assume that the level of noise in the policy is fixed, or "static," over the duration of the experiment. This assumption could hold for some sources of noise, such as computation noise, but many other sources are not guaranteed to generate consistent levels of noise. For instance, a subject might disengage during some periods of the experiment, but not others. How much subjects explore through choice randomization could also vary over time. Therefore, such models with static noise inference might fail to capture the variance in noise level, which can impact the quality of computational modeling.

To resolve this issue, we introduce a dynamic noise inference method that infers the likelihood of noise in choice behavior trial-by-trial, allowing it to vary over time. Our method makes looser assumptions than the static noise inference method, making it suitable to solve a broader range of problems (Fig. 1). Specifically, our dynamic noise policy models the presence of random noise as the result of switching between two latent states that correspond to two different types of noise policy (e.g., a fully inattentive, random policy vs. a standard, attentive, softmax policy). We assume that a hidden Markov process governs transitions between the two

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

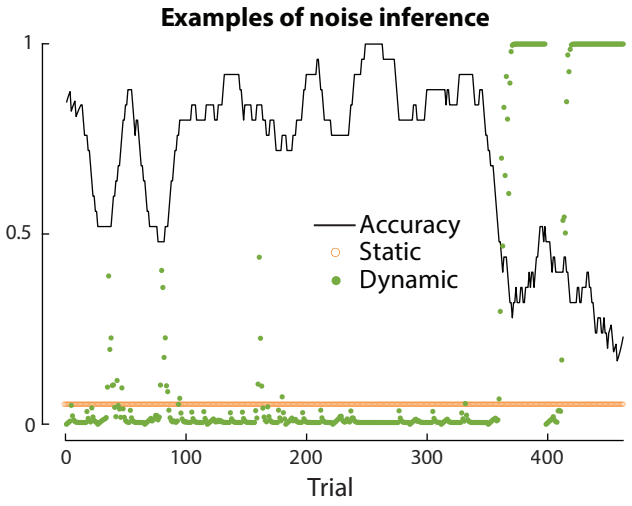

Figure 1: Examples of static and dynamic noise inference fitted to behavioral data (black line). In static noise inference, the likelihood of the random policy, or noise level (orange circles), is modeled as a constant parameter ε. In dynamic noise inference, it is estimated by the inferred probability of being in the random state $\lambda(0)$ (green dots), which varies according to choice behavior.

latent states. This model captures the assumption that noise levels in decision making tend to be temporally autocorrelated, which may be a reflection of an evolved expectation of temporally autocorrelated environments (Group et al., 2014).

We show that this dynamic noise policy can be inferred trial-by-trial in sequential decision-making. On each trial, the model infers the probability of the agent being in each latent state using observation, choice, and reward data. It estimates the choice probability as a weighted average of decisions generated by the random policy and the learned policy, which is then used to estimate likelihood. Therefore, dynamic noise inference can be incorporated into any decision-making model with analytical likelihood. Model parameters can be estimated using procedures that optimize the likelihood or its posterior distribution, including maximum likelihood estimation and hierarchical Bayesian methods (Piray, Dezfouli, Heskes, Frank, & Daw, 2019). Fig. 1 illustrates examples of static and dynamic noise inference on human choice behavioral data from Eckstein et al. (2022). The likelihoods of a noisy policy inferred by the static and dynamic methods are shown in conjunction with choice accuracy. In this example, choice accuracy drops steeply to a random level around Trial 350, indicating a higher likelihood of noisy latent state contamination. This change is captured by dynamic noise inference but not the static method.

Although dynamic noise inference can be applied to modeling any sequential choice behavior, in this work, we focus on illustrating it on reinforcement learning models (Daw & Tobler, 2014). In the following sections, we will outline the dynamic noise inference method, prove its mathematical validity, and compare and contrast it against the static noise inference method theoretically and empirically. We show that using the dynamic noise policy can improve modeling in four distinct datasets, across three experimental paradigms and models, and two species.

Methods

In a sequential decision-making task, the data collected include observation-action pairs (o_t, a_t) over the learning trajectory for time $t = 1, 2, ..., T$. In a reinforcement learning task, reward r_t is additionally collected. We assume that choices are generated by a Markov decision process (Puterman, 2014). The decision-making model leads to a policy $\pi(a|o)$ that the agent uses to choose between discrete actions given the observation. The policy may include noise mechanisms, such as using the softmax function for action selection, and it is conditional on the model's latent variables and parameters (e.g., learned values and learning rates for reinforcement learning models). We describe two extensions of such a decision model: the static noise inference method that implements the classic ε-mechanism (Nassar & Frank, 2016) and the new dynamic noise inference method. The parameters, θ, of both extended models can be optimized by maximizing the likelihood of the data given the model, denoted as $\mathcal{L}(\theta)$. Below, we focus on the model policies; all other model equations (such as reinforcement learning value updates) are taken from the published models and not reported here.

Static noise inference

Static noise policies assume that decision noise is at a constant level ε throughout the learning trajectory. At any time t, from the set of available actions A, the agent samples an action uniformly at random (with probability ε) or based on its learned policy (with probability $1-\varepsilon$). Static noise inference can be incorporated into likelihood estimation according to Algorithm 1. Thus, any model that can be fitted with likelihood-based procedures can incorporate static noise into its policy.

Algorithm 1: Static noise likelihood computation

Initialize $L(\theta) = 0$;
for $t = 1, 2, ..., T$ **do**
  Calculate the action probability $\pi_t(a_t|o_t)$;
  $L(\theta) \leftarrow L(\theta) + \log[\varepsilon \cdot \frac{1}{|A|} + (1-\varepsilon) \cdot \pi_t(a_t|o_t)]$;
  Update the policy with (o_t, a_t, r_t).
end

Dynamic noise inference

The dynamic noise inference method models decision noise by assuming that the agent is in one of two latent states at any given time: the *random state* in which the agent chooses actions uniformly at random or the *policy state* in which decisions are made according to the true model policy. The transitions between both states are governed by two parameters:

T_0^1 and T_1^0, the probabilities of transitioning from the random state to the policy state and vice versa. From these transition probabilities, we can calculate the stay probability for each latent state: $1-T_0^1$ for the random state and $1-T_1^0$ for the policy state.

The state is composed of an observation o_t, often encoding the stimulus, and unobserved, latent variables including the learned policy and h_t, where $h_t \in \{0,1\}$ indicates whether the agent is in the random state or policy state at time t. It is further assumed that r_t and o_t are conditionally independent of the latent states up to time t given the observed data history, since rewards and future observations in behavioral experiments do not depend on subjects' unobserved mental states.

Our goal is to maximize the following log-likelihood:

$$\begin{aligned}\mathcal{L}(\theta) &= \sum_{t=1}^{T} \log \mathbb{P}(a_t|o_t,\bar{o}_{t-1};\theta) \\ &= \sum_{t=1}^{T} \log \mathbb{P}\Big(\sum_i \mathbb{P}(a_t|o_t,h_t=i;\theta)\mathbb{P}(h_t=i|\bar{o}_{t-1};\theta)\Big)\end{aligned}$$

where $\bar{o}_{t-1}$ denotes the observation-action-reward triplets up to time $t-1$. Let $\lambda_t(i) = \mathbb{P}(h_t = i|\bar{o}_{t-1};\theta)$ be the prediction probability of being in the latent state $i \in \{0,1\}$ at time t. We have

$$\lambda_t(i) = \sum_j \mathbb{P}(h_t = i|h_{t-1}=j,\bar{o}_{t-1};\theta)\mathbb{P}(h_{t-1}=j|\bar{o}_{t-1};\theta),$$

where $j \in \{0,1\}$ and

$$\mathbb{P}(h_{t-1}=j|\bar{o}_{t-1};\theta) = \frac{\mathbb{P}(h_{t-1}=j,a_{t-1},r_{t-1}|o_{t-1},\bar{o}_{t-2};\theta)}{\sum_k \mathbb{P}(h_{t-1}=k,a_{t-1},r_{t-1}|o_{t-1},\bar{o}_{t-2};\theta)}.$$

Notice that for any given k, each term in the denominator, including the nominator with $j=k$, is equal to $\mathbb{P}(r_{t-1}|o_{t-1},a_{t-1},h_{t-1}=k,\bar{o}_{t-2};\theta) \times \mathbb{P}(a_{t-1},h_{t-1}=k|o_{t-1},\bar{o}_{t-2};\theta)$, the first term of which is independent of h_{t-1} and is canceled out between the nominator and denominator. Thus, the above ratio is equal to

$$\frac{\mathbb{P}(a_{t-1}|h_{t-1}=j,o_{t-1},\bar{o}_{t-2};\theta)\mathbb{P}(h_{t-1}=j|\bar{o}_{t-2};\theta)}{\sum_k \mathbb{P}(a_{t-1}|h_{t-1}=k,o_{t-1},\bar{o}_{t-2};\theta)\mathbb{P}(h_{t-1}=k|\bar{o}_{t-2};\theta)}.$$

We can now compute $\lambda_t(i)$ by plugging the above expression into its calculation, which then allows us to calculate $\mathcal{L}(\theta)$. The probabilities needed to infer $\lambda_t(i)$ and $\mathcal{L}(\theta)$ can be iteratively updated according to Algorithm 2 over the learning trajectory. These calculations can be easily incorporated into fitting procedures based on optimizing the model's likelihood, including maximum likelihood estimation and hierarchical Bayesian modeling.

Note that static noise inference is a special case of dynamic noise inference with $T_1^0 = \varepsilon$ and $T_0^1 = 1-\varepsilon$, such that the probabilities of staying in the random state and the policy state are ε and $1-\varepsilon$, respectively. That is, the probability of transitioning into each latent state is independent of the current state. Thus, with optimal parameters, likelihood estimates made with dynamic noise inference must be no worse than estimates made with static noise inference, despite one additional parameter.

Algorithm 2: Dynamic noise inference likelihood computation

Initialize $L(\theta) = 0$ and $\lambda_0(i)$ for $i \in \{0,1\}$;
for $t = 1,2,...,T$ **do**
  Calculate the action probability $\pi_t(a_t|o_t)$;
  $l_t(\theta) = \log[\frac{1}{|A|} \cdot \lambda_{t-1}(0) + \pi_t(a_t|o_t) \cdot \lambda_{t-1}(1)]$;
  $L(\theta) \leftarrow L(\theta) + l_t(\theta)$;
  $\lambda_t(h) \leftarrow \frac{\frac{1}{|A|} \cdot \lambda_{t-1}(0) \cdot T_0^h + \pi_t(a_t|o_t) \cdot \lambda_{t-1}(1) \cdot T_1^h}{\exp(l_t(\theta))}$;
  Update the policy with (o_t, a_t, r_t).
end

Model evaluation

To investigate whether our new dynamic noise inference policy improves modeling in decision-making tasks, we fitted models with static and dynamic noise inference on four empirical datasets. The Dynamic Foraging dataset consists of mice choice behavior on a two-armed bandit task with changing reward probabilities (Grossman et al., 2022). The RLWM dataset contains human behavioral data on a task testing the interactions between reinforcement learning and working memory (Collins, 2018). To further demonstrate our method on diverse tasks, models, and populations, we included two more human behavioral datasets with different versions of the 2-step task, which differentiates model-based and model-free reinforcement learning, tested on adult and developmental populations (Kool, Cushman, & Gershman, 2016; Nussenbaum, Scheuplein, Phaneuf, Evans, & Hartley, 2020). For each dataset, we added both noise inference mechanisms to the best performing model reported in the paper, except that on the RLWM dataset, we used an improved model based on later work (Master et al., 2020). For simplicity, all models were fitted using maximum likelihood estimation on the individual level (Wilson & Collins, 2019) using the MATLAB global optimization toolbox and fmincon function.

To evaluate model fit, we calculated the difference in the Akaike information criterion (AIC) between static and dynamic noise inference model fit (Akaike, 1974). One-tailed Wilcoxon signed-rank tests were performed to test the alternative hypothesis that dynamic noise inference yields lower AIC than static noise inference. Tests using the protected exceedance probability (pxp) led to the same conclusions (Rigoux, Stephan, Friston, & Daunizeau, 2014). We verified that this model comparison procedure was adequate for model identification on simulated data. We validated the models by simulating choice behavior using fitted parameters for 1,000 times and comparing it to empirical data.

Furthermore, we assessed the robustness of the dynamic noise inference models via the recovery of model parameters and the prediction probability. Specifically, we recovered the transition probability parameters, T_0^1 and T_1^0, and the prediction probability of the policy state, $\lambda(1)$, by re-fitting the models on data simulated using fitted parameters. We also verified that other model parameters were recoverable. Simulations were performed once per subject for parameter recovery and 1,000 times per subject for prediction probability recovery. Error bars represent the standard error of the mean.

Results

On the Dynamic Foraging dataset (Grossman et al., 2022), the model with dynamic noise inference significantly outperformed its static counterpart (ΔAIC =-7.77; p = 0.00027; pxp = 0.963 for the dynamic model) at the group level (Fig. 2A left). Individually, dynamic noise inference improved model fit for 29 out of 48 mice (Fig. 2A right). Fit improvement was substantially higher for some animals than others.

Mice were inferred to be in the policy state 89% of the time on average, as estimated by the $\lambda(1)$ distribution simulated using fitted parameters (Fig. 2B right). We verified that this inferred latent state was interpretable using simulations. Indeed, the recovered values of the prediction probability of the policy state, $\lambda(1)$, were similar to the true values overall, with slightly higher recovered values for true values between 0.1 and 0.9, which accounted for a minority of the data (Fig. 2B left). The parameter governing the transition probability from the policy state to the random state, T_1^0, recovered well with most recovered values close to true values (Fig. 2C left). The transition probability parameter from the random state to the policy state, T_0^1, recovered well for most animals, with a few outliers whose recovered values were much higher than the true values (Fig. 2C right). This likely happened when T_1^0 was very low, such that the animals were rarely inferred to be in the random state, under-powering the recovery of T_0^1.

In model validation, both models successfully reproduced key behavioral patterns (Fig. 2D). In behavioral data and model simulations, the learning curves after sharper reward probability changes (i.e., high-low to low-high transitions) were steeper than those after smaller reward probability changes (i.e., medium-low to low-high transitions).

On the RLWM dataset (Collins, 2018), the dynamic noise inference model fitted better than the static noise inference model on average, but not significantly so (ΔAIC = -1.50, p = 0.749; Fig. 3A left). The numerical difference was driven

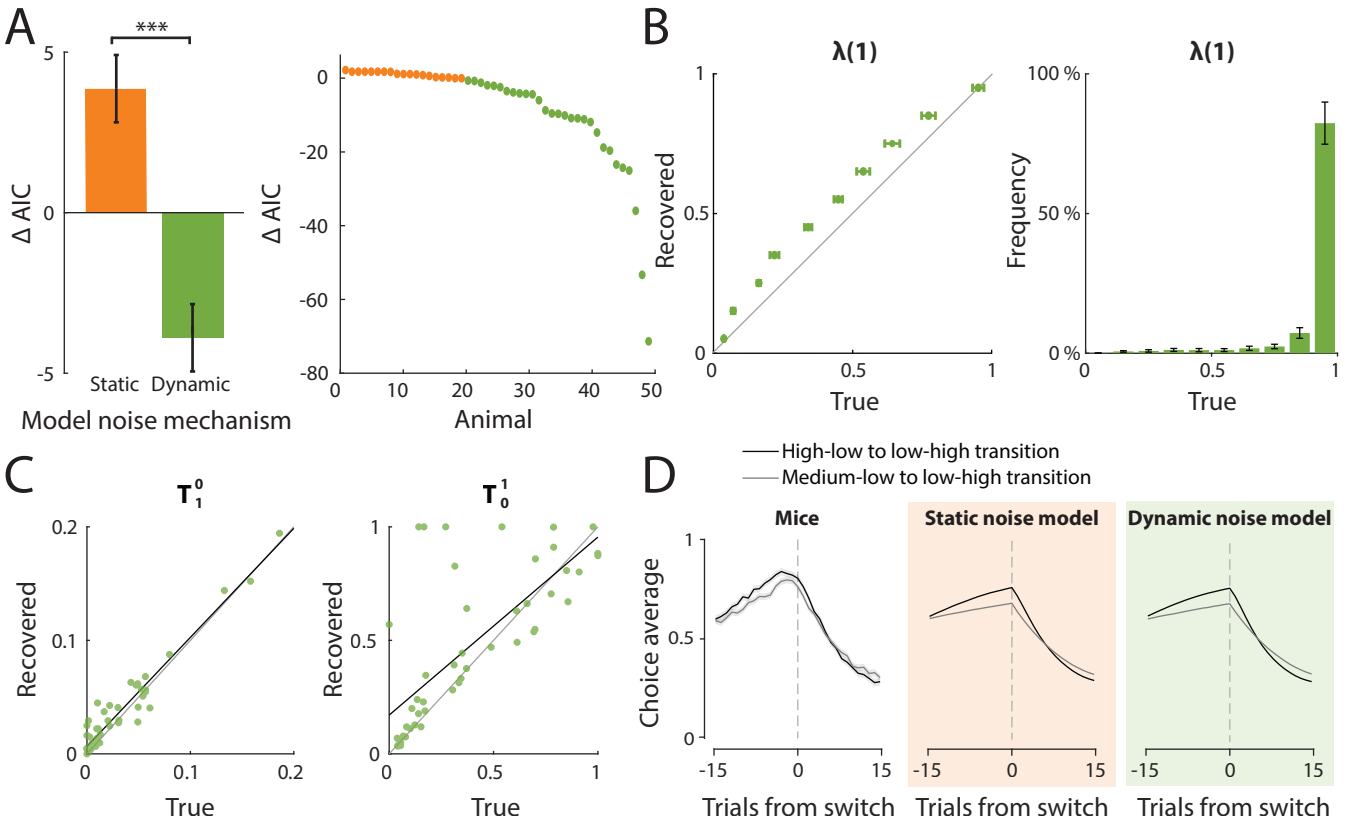

Figure 2: Model evaluation on mice Dynamic Foraging data (Grossman et al., 2022). A) The differences in AIC between models with static and dynamic noise inference at the group (left) and individual (right) levels. On the right, the colors indicate the winning models (orange: static; green: dynamic). Simulations show that the B) prediction probability and C) latent state transition probability parameters are recoverable. D) Both models reproduce key behavioral patterns.

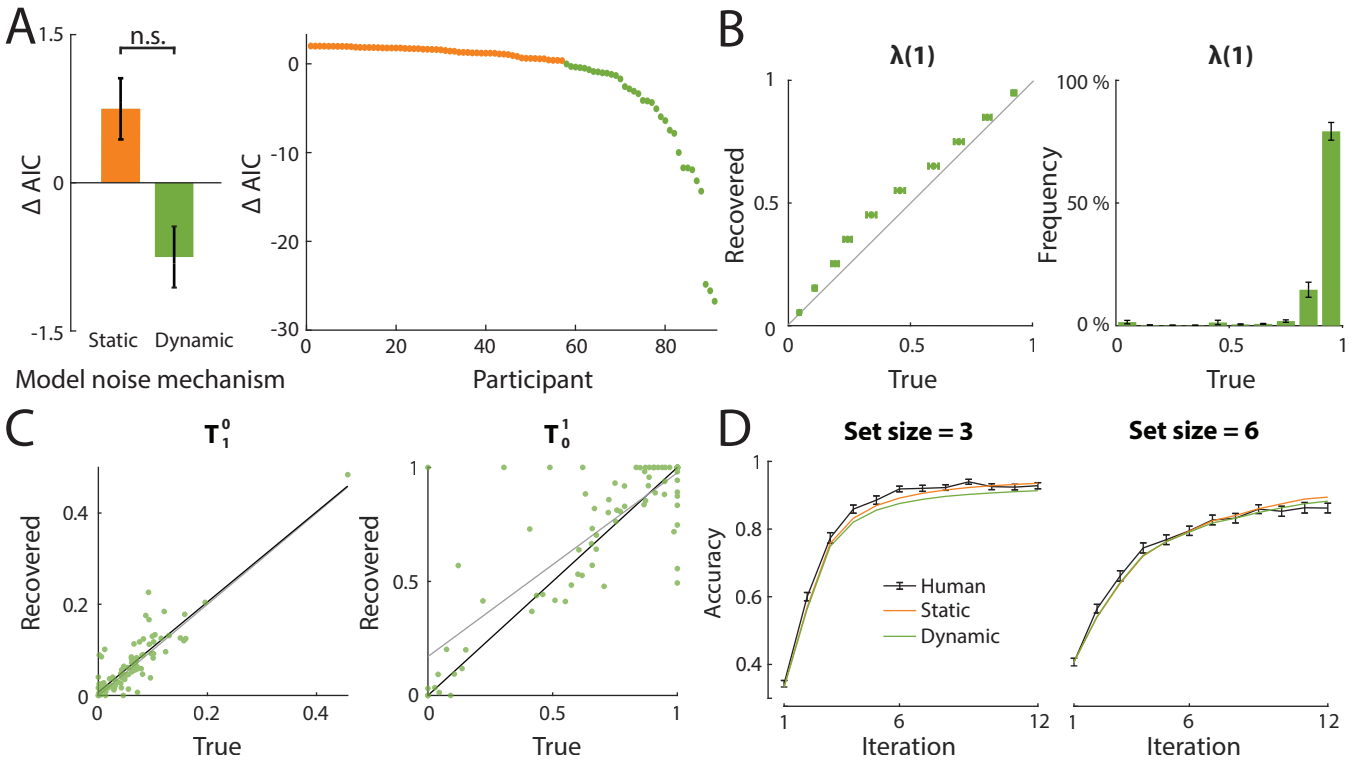

Figure 3: Model evaluation on human RLWM data (Collins, 2018). A) The differences in AIC between models with static and dynamic noise inference at the group (left) and individual (right) levels. Simulations show that the B) prediction probability and C) latent state transition probability parameters are recoverable. D) Both models reproduce key behavioral patterns.

by 34 of 91 subjects for whom dynamic noise inference improved fit substantially (Fig. 3A right).

Participants were estimated to be in the policy state 91% of the time based on simulations using the fitted parameters (Fig. 3B right). Similarly to the Dynamic Foraging dataset, the prediction probability and latent state transition probability parameters recovered well overall, with small positive biases in $\lambda(1)$ recovery and some outliers in T_0^1 recovery (Fig. 3B, C). The distribution of simulated true $\lambda(1)$ was shaped similarly to that of the Dynamic Foraging dataset, with predominantly values higher than 0.9. In model validation, both models showed human-like behavioral patterns (Fig. 3D). Specifically, initial learning was faster and choice accuracy converged to a higher value for set size of 3 than 6.

For both 2-step datasets, the model improved fit on the group level on average (Fig. 4), though this improvement was not statistically significant (top: p=0.853, bottom: p=0.422). Subject-level ΔAIC followed similar patterns to those of Dynamic Foraging and RLWM.

On all datasets, we verified that the addition of the noise inference parameters did not impair the identifiability of other model parameters.

Discussion

Our results show that dynamic noise inference can improve model fit across diverse species, populations, tasks, and models, qualifying it as a better alternative to static noise inference. While fit did not always improve at the group level, the improvements for some individuals were substantial. Dynamic noise inference is effective, versatile, and easy to be incorporated into existing model fitting procedures such as maximum likelihood estimation and hierarchical Bayesian modeling. Additionally, although the present work only considers the softmax policy, the method can be applied to various other decision policies including Thompson sampling and the upper confidence bound algorithm.

Dynamic noise inference assumes that making choices randomly and according to the learned policy are distinct, binary latent states. Biologically, this assumption aligns with an established literature on how norepinephrine modulates attention, a major contributor to varying noise levels: the phasic or tonic mode of activity of the noradrenergic locus coeruleus system closely correlates to good or poor task performance (Aston-Jones, Rajkowski, & Cohen, 1999; Berridge & Waterhouse, 2003). It is worth noting that the binary assumption of the latent states may not always be accurate. Nonetheless, it is a less strict assumption than that of static noise inference, which additionally assumes that the probability of transitioning into each latent state is independent of the current state. Thus, although dynamic noise inference may be limited by its binary latent state assumption, it is still more suitable to solve a broader range of problems than static noise inference.

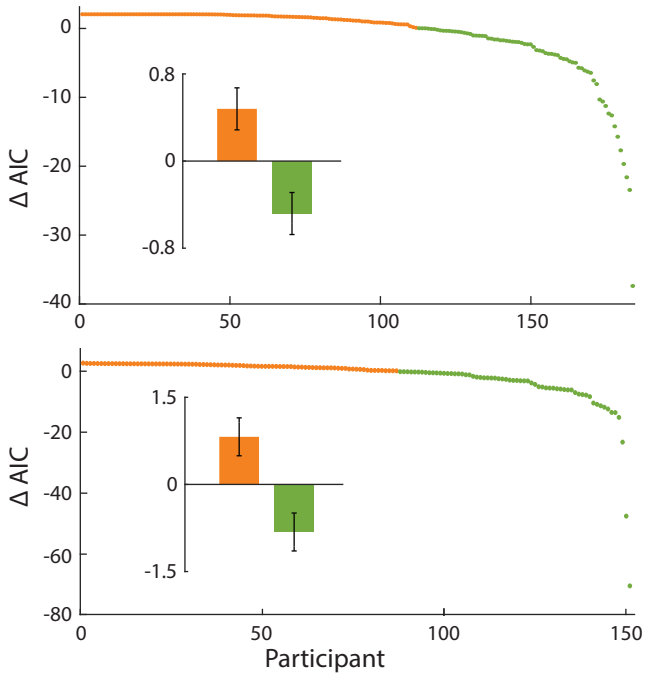

Figure 4: Model fit on human 2-step data (Kool et al., 2016; Nussenbaum et al., 2020). The differences in AIC between models with static and dynamic noise inference at the individual and group levels.

Dynamic noise inference takes effect by identifying periods of choice behavior that are better explained by random noise than the learned policy. These noisy segments of data can be caused by various factors including extended attentional lapses of the subjects. Their likelihoods are lower-bounded by that of the random policy, which limits their impacts on the estimation of the overall likelihood and model parameters. Thus, dynamic noise inference can mitigate the effects of noisy periods of data on model fitting, which could help improve data efficiency by preventing the exclusion of entire sessions or subjects due to noise in a subset of the data.

We compared model fit using the AIC metric to show that dynamic noise inference improves fit at the group level and that non-trivial proportions of subjects benefit substantially from it (Fig. 2A, 3A, 4). Although dynamic noise inference did not significantly improve fit on some of the datasets at the group level, it could still be beneficial, particularly for relatively noisy datasets (e.g., animal and online behavioral data) and capturing individual differences. In our future work, we will further characterize these benefits in detail.

We showed that the dynamic noise inference model is robust through the recovery of latent state transition probabilities and the prediction probability, despite small overestimation of $\lambda(1)$ and outliers in T_0^1 recovery. Note that both observations were caused by skewed distributions of fitted latent state transition probability parameters (i.e., low T_1^0 and high T_0^1; Fig. 2C, 3C). These transition probabilities led to imbalanced latent state frequencies: the policy state was much more frequent than the random state. This limited possible transitions out of the random state, which provided insufficient data for the accurate estimation of T_0^1. Additionally, the higher frequency of the policy state than the random state caused imbalanced $\lambda(1)$ frequencies favoring extremely high values, which may have led to a positive bias in $\lambda(1)$ estimation (Fig. 2B & 3B). However, it is important to keep in mind that the latent state transition probability parameters are likely to be distributed similarly on other empirical datasets, assuming humans and animals make choices according to their learned policies most of the time. Therefore, these observations are likely to persist in further applications of dynamic noise inference to model human and animal choice behavior.

Although small differences in model validation were observed on both datasets, there is insufficient evidence for qualitative distinctions between the behavior simulated by static and dynamic noise inference models. Future work should further validate dynamic noise inference experimentally, for example, by comparing inferred prediction probabilities to an independent measure of attention, such as biometrics or reaction time, and testing whether inferred latent states capture this measure.

Having shown that dynamic noise inference may benefit model fit, the next questions are: *how* does this improvement of fit manifest, and does it impact the insights we can derive from computational modeling? Our future work will focus on systematically characterizing the benefits of dynamic noise inference through further theoretical and empirical analyses. For example, we will demonstrate that dynamic noise inference can improve parameter fit, especially in the presence of long periods of noisy behavior (e.g., lapses).

Our work is related to other recent work identifying discrete latent policy states (Ashwood et al., 2022). Here, our assumptions, which allow two distinct latent noise states, are more complex than those of most classic policy models (e.g., ε-greedy). However, they are simpler than assumptions made by other approaches with more latent states and more complex policies (Ashwood et al., 2022). The intermediate complexity level of our assumptions allows us to capture changing noise levels while keeping simple likelihood estimations accessible, and should thus make dynamic noise inference a more versatile tool.

In summary, our dynamic noise inference method promises potential improvements over the static noise inference method currently used in the modeling literature of decision-making behavior. Dynamic noise inference enables us to capture different degrees of task-engagement in different task periods, limiting contamination of model-fitting by noisy periods, without requiring ad-hoc data curating. Based on the empirical evaluation of the method reported in the current work, we expect that dynamic noise inference in modeling choice behavior will strengthen modeling in many decision-making paradigms, while keeping additional model complexity and assumptions minimal.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Ashwood, Z. C., Roy, N. A., Stone, I. R., Laboratory, I. B., Urai, A. E., Churchland, A. K., ... Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, *25*(2), 201–212.

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological psychiatry*, *46*(9), 1309–1320.

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain research reviews*, *42*(1), 33–84.

Bitzer, S., Park, H., Blankenburg, F., & Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a bayesian model. *Frontiers in human neuroscience*, *8*, 102.

Chapelle, O., & Li, L. (2011). An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, *24*.

Collins, A. G. (2018). The tortoise and the hare: Interactions between reinforcement learning and working memory. *Journal of cognitive neuroscience*, *30*(10), 1422–1432.

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035.

Daw, N. D., & Tobler, P. N. (2014). Value learning through reinforcement: the basics of dopamine and reinforcement learning. In *Neuroeconomics* (pp. 283–298). Elsevier.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology*, *18*(2), 185–196.

Eckstein, M. K., Master, S. L., Xia, L., Dahl, R. E., Wilbrecht, L., & Collins, A. G. (2022). The interpretation of computational model parameters depends on the context. *Elife*, *11*, e75474.

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current opinion in psychology*, *29*, 174–180.

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

Grossman, C. D., Bari, B. A., & Cohen, J. Y. (2022). Serotonin neurons modulate learning rate through uncertainty. *Current Biology*, *32*(3), 586–599.

Group, T. M. A. D., Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E., ... McNamara, J. M. (2014). The evolution of decision rules in complex environments. *Trends in cognitive sciences*, *18*(3), 153–161.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off? *PLoS computational biology*, *12*(8), e1005090.

Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., & Collins, A. G. (2020). Disentangling the systems contributing to changes in learning during adolescence. *Developmental cognitive neuroscience*, *41*, 100732.

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.

Nussenbaum, K., Scheuplein, M., Phaneuf, C. V., Evans, M. D., & Hartley, C. A. (2020). Moving developmental research online: comparing in-lab and web-based studies of model-based reinforcement learning. *Collabra: Psychology*, *6*(1).

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in cognitive sciences*, *21*(6), 425–433.

Piray, P., Dezfouli, A., Heskes, T., Frank, M. J., & Daw, N. D. (2019). Hierarchical bayesian inference for concurrent model fitting and comparison for group studies. *PLoS computational biology*, *15*(6), e1007043.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

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

Schaaf, J. V., Jepma, M., Visser, I., & Huizenga, H. M. (2019). A hierarchical bayesian approach to assess learning and guessing strategies in reinforcement learning. *Journal of Mathematical Psychology*, *93*, 102276.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*(3-4), 285–294.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

Wang, S., & Wilson, R. (2018). Any way the brain blows? the nature of decision noise in random exploration.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human factors*, *50*(3), 433–441.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, *8*, e49547.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074.