A Domain-general Strategy for Hidden-state Inference in Humans and Noisy Networks

Junseok K Lee[1,2], Valentin Wyart[1,2,3]

[1]Laboratoire de Neurosciences Cognitives et Computationnelles, Institut National de la Santé et de la Recherche Médicale (Inserm), Paris, France,
[2]Département d'Études Cognitives, École Normale Supérieure, Université Paris Sciences et Lettres (PSL University), Paris, France
[3]Institut du Psychotraumatisme de l'Enfant et de l'Adolescent, Conseil Départemental Yvelines et Hauts-de-Seine, Versailles, France

European Research Council
Established by the European Commission
Inserm
La science pour la santé
From science to health

Abstract

Understanding the hidden (latent) states and structures that generate observations of the world is a fundamental aspect of cognition, wherein humans demonstrate exceptional proficiency in the ability to apply similar cognitive strategies across superficially dissimilar contexts sharing the same latent structure. While previous efforts to understand the computational bases of these cognitive strategies through cognitive modeling have largely focused on single contexts, here we take a novel approach which combines two tasks with the same reversal structure. Through cognitive modeling, we first show that humans use the same noisy hidden-state inference strategy across these superficially dissimilar tasks that require reversal learning. Then, using recurrent neural networks (RNNs) featuring either exact or noisy computations trained on the same tasks, we show that noisy RNNs – like humans – benefit from reusing the same latent representations for solving the two tasks. Together, our findings underscore the significance of computation noise in constraining the use of mental resources, shedding light on its potential functional role in cognition.

Behavioral task (for humans and neural networks)

Q1: ***How do people perform inference in contexts of differing modalities?***

Common reward/stimulus generation

reward-guided (**bandit**)

bandit-specific

- stimuli are **numerical values** (rewards)
- stimuli are **controllable** outcomes of actions
- rewards are drawn from one of two distributions whose mean value changes depending on the current hidden-state

common in both tasks

- one set of generative statistics produces a *unique stimulus trajectory* used for both rewards and colors

⇒ **the two tasks are externally different in every way but can be solved by the same Bayes-optimal strategy**

perceptual categorization (**fairy**)

fairy-specific

- stimuli are **colors** ranging from green to red
- stimuli are **uncontrollable** observations
- stimuli are sampled from the distribution corresponding to the current hidden-state

Results from human participants

A1: ***Individuals demonstrate similar & stable behavior between the task contexts.***

Behavior (N = 149)

- at the group-level, participants show similar performance between the two tasks

[median split on PC1 from PCA on above measures]

- stable between-participant gradient of performance

Cognitive modeling

Q2: ***What computational mechanism dvrives this behavior?***

Computational model

modular description

parameter description

sensory
- stimulus compression a — *captures distortions in stimuli perception*
- stimulus bias ω — *captures biases in stimuli perception*

inference
- hazard rate h — *captures perceived volatility*
- prior compression γ — *captures categorization of beliefs*
- inference noise σ — *captures imprecision of evidence integration*

selection
- stimulus evidence bias η — *captures bias towards recent stimuli*

A2: parameter fits (N = 149)

- sensory and selection parameters show group-level differences and do not correlate across contexts (evidence of task differences at the level of these modules)
- inference parameters show no group-level differences and correlate across tasks (evidence of context-general inference)

test-retest stability (N = 106)

- test-retest gives plausible upper bounds on cross-context stability
  - inference noise is highly stable

Q3: ***Inference noise a general cognitive parameter?***

Cognitive ability and model parameters

- lower inference noise (more precise inference) is the **sole** predictor of higher cognitive ability in both conditions and sessions

only inference noise inversely correlated with higher accuracy in both conditions

Structure of cognitive ability and inference noise

alternative models

- link between cognitive ability and inference noise can be
  - h1: both general and context-specific
  - h2: only general
  - h3: only context-specific

model recovery

- the three hypotheses regarding the structure of cognitive ability and inference noise are differentiable

the hierarchical structure of inference noise best explains the data

model comparison

A3: ***inference noise is a general cognitive parameter, but could be modulated in a context-dependent way[1]***

Recurrent neural networks

Q4: ***What role does computation noise have on cross-context inference?***

Network architecture

Network types

exact computation
all computation is precisely done given some set of weights and bias

noisy computation
noise is added to the linear combination of previous state and current input at the hidden layer distorting network inference

Objectives

unconstrained inference
task accuracy is maximized

constrained inference
task accuracy is maximized while also minimizing a penalty cost on aligned dimensions of hidden layer activations

Training procedure

For each epoch, a batch of novel blocks are generated

RNN performs the tasks

Loss calculated for back-propagation

Constraint mechanism

a network naturally settles to an aligned state

the network is now less aligned after training with penalties

RNN performance

(50 network instantiations)

under free optimization...
- all exact RNNs reach Bayes-optimal performance
- noisy RNNs reach Bayes-optimal performance if sufficiently large hidden layer size or if noise sufficiently low

under alignment constraints...
- exact RNNs still reach Bayes-optimal performance
- noisy RNNs fail to reach Bayes-optimal performance
- penalty of alignment on accuracy increases with lower hidden unit size and higher noise

under free optimization...
- RNNs with more noise tend to be less aligned
- RNNs with smaller hidden layers align more

under alignment constraints...
- exact RNNs can fully misalign and achieve optimal performance
- noisy RNNs show lower performance with increasing misalignment
  - effect is exacerbated with higher levels of noise and lower resource size

⇒ **noisy RNNs achieve good performance between different contexts through the alignment of resources for inference**

A4: ***Shared inference may be a natural result of the combination of resource constraints and limited precision of neural representations***
- computation noise also associated with cognitive advantages such as zero-shot adaptability to unencountered adversity in neural networks[2]

Q5: ***Is the alignment of cognitive resources linked to transfer and few-shot learning in humans?***

A5: ***To be determined...***

[1] Lee, J. K., Rouault, M., & Wyart, V. (2023). Adaptive tuning of human learning and choice variability to unexpected uncertainty. Science Advances, 9(13), eadd0501.
[2] Findling, C., & Wyart, V. (2021). Computation noise in human learning and decision-making: origin, impact, function. Current Opinion in Behavioral Sciences, 38, 124-132.

The 7th annual conference on Cognitive Computational Neuroscience
Massachusetts Institute of Technology
Boston, Massachusetts, USA August 6 - 9, 2024