Interpretation of Individual Differences in Cognitive Computational Neuroscience

Jessica V. Schaaf (Jessica.Schaaf@radboudumc.nl)
Donders Institute, Nijmegen, The Netherlands

Steven Miletić (Steven@miletic.nl)
Leiden University, Leiden, The Netherlands

Anna C.K. van Duijvenvoorde (A.C.K.van.Duijvenvoorde@fsw.leidenuniv.nl)
Leiden University, Leiden, The Netherlands

Hilde M. Huizenga (H.M.Huizenga@uva.nl)
University of Amsterdam, Amsterdam The Netherlands

Abstract

Computational neuroscience offers a valuable opportunity to understand the neural mechanisms underlying behavior. Suppose that you fit a computational model to behavioral data to generate an individual-specific prediction error regressor. You in turn use this regressor to model activity in a brain region of interest. What do individual differences in the resulting regression weights mean? Typically researchers interpret these individual differences as differences in neural coding. Yet, in five scenarios, we illustrate through simulations that such individual differences may stem from other factors. By acknowledging these alternative interpretations of individual differences, and by openly sharing reproducible code, we aim to advance the understanding and interpretation of individual differences in computational neuroscience.

Keywords: computational neuroscience; individual differences; interpretation; reinforcement learning

Individual Differences in Computational Neuroscience

Studying individual differences by means of computational neuroscience has two major advantages: it makes theories on the origins of individual differences in neural coding explicit (Guest & Martin, 2021) and it allows researchers to investigate individual differences in neural coding of latent variables underlying cognitive processes (Hartley & Somerville, 2015). Many computational neuroscience studies use a so-called latent-input approach (Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017). In this approach, a computational model is fit to the behavioral data, after which a latent variable derived from this model (e.g., a prediction error) is entered as a predictor in the functional Magnetic Resonance Imaging (fMRI) analysis. Individual differences in the resulting regression weights are typically interpreted as individual differences in neural coding, and subsequently related to factors like age or socio-economic status. As we illustrate in five scenarios, the interpretation of individual differences based on these regression weights is anything but straightforward (see also Lebreton, Bavard, Daunizeau, & Palminteri, 2019).

Take a simple reinforcement-learning task in which participants repeatedly choose between two stimuli (e.g., a chair and a clock). After making a choice, the participant experiences an outcome (e.g., winning or losing a dollar) allowing them to gradually learn the value of each stimulus. This learned value V then subsequently guides choice behavior in the next trial. Formally, this process can be described as follows (Rescorla & Wagner, 1972): $V_{chair,i,t+1} = V_{chair,i,t} + \alpha_i \cdot PE_{i,t}$, where the prediction error (PE) is the difference between the observed outcome (O) and the value of the chosen option, $PE_{i,t} = O_{i,t} - V_{chair,i,t}$, and the learning rate (α) indicates how fast participants update values based on prediction errors.

The individual-specific prediction error variable can then be included as a first-level regressor in the fMRI General Linear Model (GLM) after convolution with a hemodynamic response function. That is, schematically, $neuralsignal_{i,t} = \phi_i \cdot PE_{i,t} + \varepsilon_{i,t}$, where ϕ is the neural coding parameter and ε refers to noise.

Scenario 1: Lack of Individual Differences in Neural Coding

The first scenario computational neuroscientists may encounter is that they find individual differences in learning rates, but no individual differences in ϕ. A perhaps intuitive interpretation suggests that if there are individual differences in the neural signal, these must be reflected in the ϕ_i parameter. However, individual differences can be present in the neural signal but absent in ϕ_i. This is because the neural coding parameter ϕ_i captures the relative size of the neural signal to the size of the prediction error. As such, even in the absence of individual differences in the neural coding parameter, differences in learning rates lead to differences in prediction error variance which in turn leads to differences in neural signal variance. This scenario should thus be interpreted as no individual differences in neural coding on top of those observed in behavior rather than no such differences in neural response.

Scenario 2: Spurious Individual Differences in Neural Coding due to Neglected Individual Differences in the Duration of the Neural Response

A second scenario, also highlighted by Mumford et al. (2024), occurs when computational neuroscientists overlook individ-

ual differences in the duration of the neural response. Neuroscientists typically assume a constant duration in the fMRI GLM (Grinband, Wager, Lindquist, Ferrera, & Hirsch, 2008). Yet, a response lasting twice as long generates a neural signal nearly identical to that of a response twice as strong (Mumford et al., 2024). As such, when there are individual differences in the duration of prediction-error related responses, these appear as individual differences in the strength of prediction-error coding (i.e., in ϕ_i) in the fMRI GLM when duration is unaccounted for.

Scenario 3: Spurious Individual Differences in Neural Coding due to Neglected Individual Differences in Outcome Sensitivity

We now turn to a scenario in which the computational model is inadequately specified. This third scenario computational neuroscientists may encounter, concerns individual differences in outcome sensitivity (not to be confused with inverse temperatures as discussed in the next scenario). Specifically, it may be that some participants are more sensitive to outcomes (e.g., winning or losing a dollar) than others (Pedersen, Frank, & Biele, 2017). A straightforward way to implement outcome sensitivity in the computational model is to introduce an individual-specific outcome sensitivity parameter (Huys, Pizzagalli, Bogdan, & Dayan, 2013). This outcome sensitivity parameter γ_i weighs the observed outcome: $PE_{i,t} = \gamma_i \cdot O_{i,t} - V_{chair,i,t}$. From this adjusted equation, it becomes clear that the outcome sensitivity parameter γ_i influences prediction errors. Simulations (shown in Figure 1A) confirm that outcome sensitivity affects prediction error variance. As a result, if a researcher fails to adequately model individual differences in outcome sensitivity, they will introduce spurious individual differences in the neural coding parameter ϕ_i (see Figure 1C).

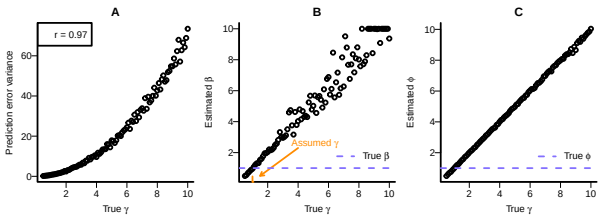

Figure 1: The effect of outcome sensitivity on prediction error variance (A), estimated inverse temperature in case of model misspecification (B), and the estimated neural coding parameter in case of model misspecification (C).

Scenario 4: No Spurious Individual Differences in Neural Coding due to Individual Differences in Inverse Temperature

Up until now, all scenarios have focused on the learning part of the reinforcement learning process. Yet, learned values are used to generate choices. This choice part of the reinforcement learning process can be formally expressed as follows, $Pr(choice_{i,t} = chair) = 1/(1 + e^{-\beta_i(V_{i,t,chair} - V_{i,t,clock})})$, in which the choice probability is determined by the difference between the values of the two stimuli. The higher the inverse temperature parameter β_i, the more choices are guided by the difference in the values of the two stimuli. The inverse temperature does not weigh prediction errors. Accordingly, we observe no systematic relationship between β and prediction error variance (Figure 2A). While β can be accurately estimated from the behavioral data (Figure 2B), it does not influence the estimated neural coding parameter ϕ (Figure 2C).

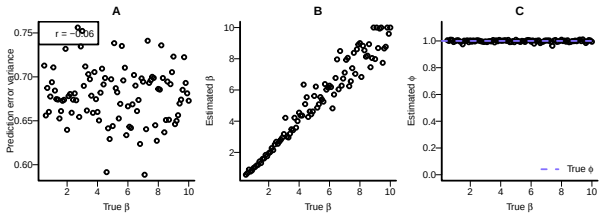

Figure 2: The effect of inverse temperature settings on prediction error variance (A), the estimated inverse temperature (B), and the estimated neural coding parameter (C).

Scenario 5: Spurious Individual Differences in Neural Coding due to Neglected Individual Differences in the Computational Model

In scenario 3, we discussed how inadequate specification of the computational model, that is, neglecting individual differences in outcome sensitivity, may induce spurious individual differences in the neural coding parameter. This fifth scenario, illustrated in Figure 3 shows how misspecification of the type of learning rate cause spurious individual differences in the neural coding parameter respectively.

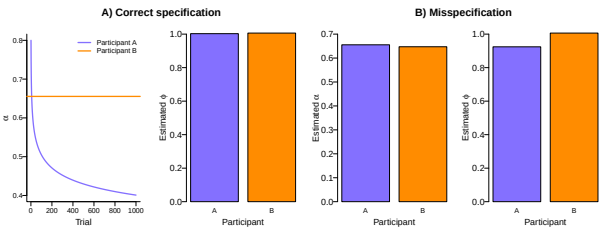

Figure 3: When participants truly use different type of learning rates α but have the same neural coding parameter ϕ (A), ignoring individual differences in the type of α causes spurious individual differences in ϕ (B).

Concluding Remarks

A computational neuroscience approach to studying the origins of individual differences has gained increased popularity. We here addressed potential challenges in the interpreta-

tion of individual differences from computational neuroscience studies by presenting five scenarios that computational neuroscientists may encounter. Although we illustrate the scenarios in a reinforcement-learning context, they arguably generalize to other computational neuroscience fields adopting a latent input approach (scenarios 1-2), other choice paradigms (scenario 4), and other models including parameters that affect regressor variance (scenarios 3 and 5). As such, our results widely aid the understanding and interpretation of individual differences in computational neuroscience.

References

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fmri designs. *NeuroImage*, *43*(3), 509-520.

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, *16*(4), 789-802.

Hartley, C. A., & Somerville, L. H. (2015). The neuroscience of adolescent decision-making. *Current Opinion in Behavioral Sciences*, *5*, 108-115.

Huys, Q. J. M., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*(1), 1-16.

Lebreton, M., Bavard, S., Daunizeau, J., & Palminteri, S. (2019). Assessing inter-individual differences with task-related functional neuroimaging. *Nature Human Behaviour*, *3*, 897-905.

Mumford, J. A., Bissett, P. G., Jones, H. M., S, S., Rios, J. A. H., & Poldrack, R. A. (2024). The response time paradox in functional magnetic resonance imaging analyses. *Nature Human Behaviour*, *8*(2), 349-360.

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin and Review*, *24*(4), 1234-1251.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning: Current research and theory.*

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65-79.