Helmholtz Institute for Human-Centered AI
Helmholtz Munich

Submission of our article titled "Model-based exploration strategies can be measured convergently but are not associated with questionnaire assessments and psychiatric diagnostic criteria" to Scientific Reports

Dear Dr. Marszalek,

We would like you to consider our manuscript entitled "Model-based exploration strategies can be measured convergently but are not associated with questionnaire assessments and psychiatric diagnostic criteria" for publication as an Article in *Scientific Reports*. We study the psychometric properties of three common exploration-exploitation tasks. More specifically, we investigate whether computational model parameters extracted from these tasks are temporally stable across a 6-week period, whether they measure converging latent constructs and whether they relate to self-report measures of everyday exploration. Our findings shed light on important shortcomings in contemporary exploration measures and on limitations in the generalizability of these task-based measures. To circumvent these problems, we present a concrete suggestion of how the measurement of exploration behavior can be made more stable over time and de-correlated from task-specific artifacts (i.e., task-specific strategies and parameter correlations induced during model fitting). Thereby, our study extends previous research published in *Nature Communications* (Anvari et al. 2023) by (a) moving from a model-free measurement of exploration to a theoretically informed model-based measurement, (b) testing the divergent validity from a performance-based, latent, construct, (c) testing the external validity, and (d) evaluating criterion validity using this model-based approach. Our study also shows that previous results reported in *Nature Human Behaviour* (Schurr et al. 2024, Fan et al. 2023), building on task-specific measurements of exploration, have to be interpreted carefully.

More specifically, we focus on three well-established few-armed bandit tasks: The Horizon task (Wilson et al. 2014), the Two-armed bandit task (Gershman 2018) and the Restless bandit task (Daw et al. 2004). We find that most model parameters lack acceptable temporal stability, they do not correlate across tasks and also do not correlate with self-reported measures of exploration. By unifying the modeling approaches across tasks and removing a highly correlated model parameter in the Two-armed bandit, we strongly improved the convergence across tasks and were able to form latent factors of the two main exploration strategies (value-guided exploration and directed exploration). These latent factors were however still not related to self-reported measures of exploration. They did however correlate strongly with measures of working memory capacity, suggesting that exploration strategies are affected by general cognitive capacity. We were further not able to replicate previously observed associations between traits of anxiety and depression with exploration behavior (e.g., Fan et al., 2023) when moving from a task-based perspective to a perspective of converging evidence from several tasks. Taken together, these results discourage the use of model parameters from a single task for studying individual differences and show ways to derive more generalizable constructs of exploration behavior. This article should be exciting to anyone interested in exploration behavior and information seeking more broadly but also to an audience interested in computational psychiatry, and in how task-based measures can be used to study individual differences in psychological traits.

If it is helpful when selecting reviewers, qualified individuals include Farid Anvari (expert in the measurement of individual differences using cognitive tasks), Roey Schurr or Maria Waltmann (both experts in the computational modeling of exploration behavior), Haoxue Fan or Anahit Mkritchian (both experts in computational psychiatry), Kou Murayama (expert in theories of motivation and curiosity), and Robert Wilson (expert in modeling intermediate difficulty).

This article is rich with implications, and communicates them in a readily digestible format. For these reasons, we believe that our article would make a good contribution to *Scientific Reports*.

Thank you for your consideration, and we are looking forward to hearing from you.

Kind regards,
Kristin Witte, Mirko Thalmann, Eric Schulz