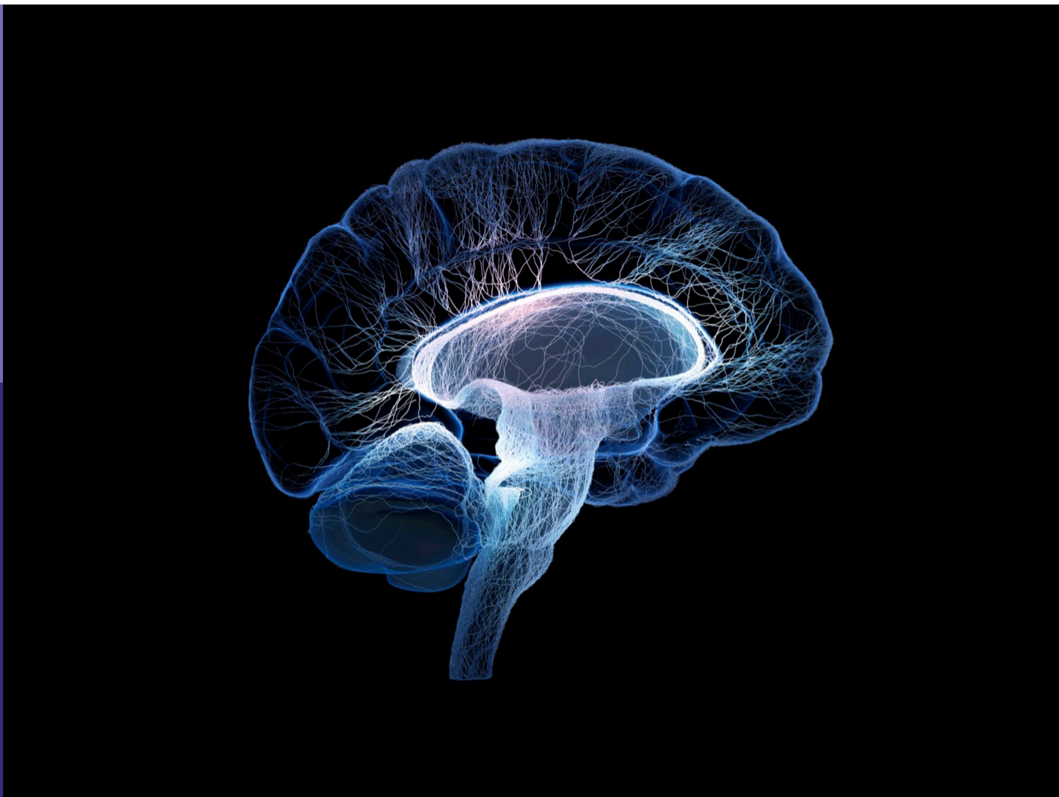
frontiers | Research Topics

Novel applications of Bayesian and other models in translational neuroscience

Edited by
Reza Rastmanesh, Jacob Raber, Edward W. Hsu
and Benjamin R. Pittman-Polletta

Published in
Frontiers in Neuroscience
Frontiers in Computational Neuroscience



ISSN 1664-8714
ISBN 978-2-8325-4882-0
DOI 10.3389/978-2-8325-4882-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Novel applications of Bayesian and other models in translational neuroscience

Topic editors

Reza Rastmanesh — American Physical Society, United States

Jacob Raber — Oregon Health and Science University, United States

Edward W. Hsu — The University of Utah, United States

Benjamin R. Pittman-Polletta — Boston University, United States

Citation

Rastmanesh, R., Raber, J., Hsu, E. W., Pittman-Polletta, B. R., eds. (2024). *Novel applications of Bayesian and other models in translational neuroscience*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4882-0

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 29 April 2024
DOI 10.3389/fnins.2024.1373633

Editorial: Novel applications of Bayesian and other models in translational neuroscience

Jacob Raber[1]*, Benjamin R. Pittman-Polletta[2] and Reza Rastmanesh[3]*

[1]Department of Behavioral Neuroscience, Neurology, and Radiation Medicine, Division of Neuroscience, Oregon National Primate Research Center, Oregon Health and Science University, Portland, OR, United States, [2]Department of Mathematics and Statistics, Boston University, Boston, MA, United States, [3]American Physical Society, College Park, MD, United States

KEYWORDS

translational neuroscience, Bayesian and other models, data analysis, probabilistic, Bayesian analytics

OPEN ACCESS

EDITED AND REVIEWED BY
Guo-Yuan Yang,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Jacob Raber
✉ raberj@ohsu.edu
Reza Rastmanesh
✉ r.rastmanesh@gmail.com

RECEIVED 20 January 2024
ACCEPTED 15 April 2024
PUBLISHED 29 April 2024

CITATION
Raber J, Pittman-Polletta BR and Rastmanesh R (2024) Editorial: Novel applications of Bayesian and other models in translational neuroscience.
Front. Neurosci. 18:1373633.
doi: 10.3389/fnins.2024.1373633

Editorial on the Research Topic
Novel applications of Bayesian and other models in translational neuroscience

The task of both the brain and the neuroscientist is to reason about large numbers of variables that are both mutually interdependent and uncertain (i.e., probabilistic). This partly explains why statistical models - and Bayesian models in particular – have been increasingly prominent in both theoretical accounts of brain function and methodologies for analyzing neural data. Bayes' theorem specifies the optimal way to combine prior beliefs with data in probabilistic inference,[1] offering a powerful tool for reasoning under uncertainty (van Amersfoort et al., 2020). Within the framework of *Bayesian networks*, the values (or rather probability distributions) of multiple variables interrelated through a network of conditional dependencies can be calculated from observational data by successive applications of Bayes' theorem. Bayesian networks can be used as statistical models for a large and general class of dynamical phenomena, and can be constructed using expert knowledge or learned from data through the process of *structure learning*. Recent theories of brain function suggest that perception, cognition, and action can all be fruitfully understood as forms of Bayesian inference, in which an internal generative model of the world is inverted to fit sensory data. This internal generative model can be formalized as a Bayesian network that is dynamic and *hierarchically deep* – i.e., composed of multiple levels of (increasingly abstract) explanatory variables evolving in time. Inversion of this network is believed to be implemented via *predictive processing*, in which brain activity principally encodes the difference between model-generated predictions and sensory data, i.e., *prediction errors*. In perception, the model is changed to match the sensory data, while in action, the sensory data is changed to match the model through so-called *active inference*.

[1] Bayes' theorem states that the *conditional probability* of some occurrence A given observed data B, $P(A\|B)$, is proportional to the product of the *prior probability* of the event, $P(A)$, and the *likelihood* of the observation given the event, $P(B\|A)$. We can think of the conditional distribution $P(B\|A)$ as a *generative model* of the data, which we *invert* to calculate the *posterior probability* $P(A\|B)$.

In perhaps the farthest-reaching formulation of these hypotheses, the free-energy principle, the brain accomplishes Bayesian inference by performing a gradient descent on free energy. This ensures that the accuracy of the internal model (and its predictions) increases, while its complexity decreases (Bruineberg et al., 2016).

However, while Bayesian, predictive, and statistical models have been proposed as qualitative and quantitative models and tools for basic research, the applications of these models to translational neuroscience have been understudied and underreported. Exceptions include variational Bayesian mixed-effects inference, which has been successfully tested for use in classification studies (Brodersen et al., 2013), and a recently-published multi-task Bayesian compressive sensing approach to simultaneously estimate the full posterior of the CSA-ODF and diffusion-weighted volumes from multi-shell HARDI acquisitions. This Research Topic collects further research applying Bayesian and statistical tools, techniques, and theories to the prediction or anticipation of brain function in humans and animal models under physiological and pathological conditions.

Many of the studies in this Topic employ Bayesian networks (BNs) to analyze and make predictions about neurophysiological data. In Fan et al., structure learning is applied to create a predictive model for ischemic stroke (IS) by discovering a BN linking risk factors to IS in patients with dilated cardiomyopathy (DCM). As Fan et al. point out, a major advantage of BNs is their utility in classifying imbalanced datasets, a common challenge in real-world data. In Carvalho do Nascimento et al., techniques from structure learning for BNs are applied to the discovery of functional connectivity networks in the domain of interpersonal neural synchronization (INS). The proposed two-step network estimation method allows inference of the time-varying probabilistic dependencies between brain regions both within and between subjects. Carvalho do Nascimento et al. demonstrate the utility of their method in the analysis of fNIRS hyperscanning data recorded simultaneously from violinists playing a duet, confirming that one player was leading the other. In Chen, techniques from structure learning are applied to create a data fusion method, called Bayesian Multisource Data Integration, to model the interactions among data sources (i.e., imaging modalities) and behavioral variables. The proposed method constructs a Bayesian network model associating features in each data source with behavioral outcome variables. The generated Bayesian network is transparent and easy to understand. It can be used to understand how behavioral changes depend on features in each data source, and to identify which features synergistically contribute to behavioral outcomes, which are redundant, and which are uninformative.

Thome et al. take the use of Bayesian statistical models for data analysis a step further. They propose a novel use for interpretable latent variable models. These models probabilistically link behavioral observations to an underlying latent process, and have increasingly been used to draw inferences about cognition from observed behavior. The latent process usually connects experimental variables to cognitive computation. While such models provide important insights into the latent processes generating behavior, one important aspect has often been overlooked. They may also be used to generate precise and falsifiable behavioral predictions as a function of the modeled experimental variables. In doing so, they pinpoint how experimental conditions must be designed to elicit desired behavior and generate adaptive experiments. These ideas are exemplified on the process of delay discounting (DD). After inferring DD models from behavior on a typical DD task, the models are leveraged to generate a second adaptive DD task, which elicits 9 graded behavioral discounting probabilities across participants. Models are then validated and contrasted to competing models in the field by assessing the out-of-sample prediction error. They also report evidence for inter-individual differences with respect to the most suitable models underlying behavior. Finally, they outline how to adapt the proposed method to the investigation of other cognitive processes including reinforcement learning.

Priorelli and Stoianov further the application of Bayesian network models of the brain, presenting a normative computational theory of how the brain may support visually-guided goal-directed actions in dynamically changing environments. This theory extends active inference, a theory of cortical processing according to which the brain maintains beliefs over the environmental state, and motor control signals try to fulfill the corresponding sensory predictions. The authors propose that the neural circuitry in the Posterior Parietal Cortex (PPC) compute flexible intentions (Duarte-Carvajalino et al., 2014)—or motor plans from a belief over targets—to dynamically generate goal-directed actions, and develop a computational formalization of this process. A proof-of-concept agent embodying visual and proprioceptive sensors and an actuated upper limb was tested on target-reaching tasks. The agent behaved correctly under various conditions, including static and dynamic targets, different sensory feedbacks, sensory precisions, intention gains, and movement policies; limit conditions were individuated, too. Active inference driven by dynamic and flexible intentions can thus support goal-directed behavior in constantly changing environments, and the PPC might putatively host its core intention mechanism. More broadly, the study provides a normative computational basis for research on goal-directed behavior in end-to-end settings and further advances mechanistic theories of active biological systems.

Mezzetti et al. apply Bayesian models to the analysis of psychometric data, extending their use of generalized linear mixed models (GLMM) and two-level methods in a Bayesian framework. This allows them to apply *a priori* knowledge from the literature and from previous experiments to estimation of psychometric functions, reducing the uncertainty of the parameters through the combination of prior knowledge and the experimental data. Evaluating uncertainties between and within participants through posterior distributions, Mezzetti et al. use a special type of Bayesian model, the power prior distribution, to modulate the weight of the prior, constructed from a first set of data, and use it to fit a second one. Their models estimated the probability distributions of the parameters of interest conveying information about the effects of the experimental variables and their uncertainty, as well as the reliability of individual participants.

The work collected in this Topic also includes translational applications of more general statistical models and approaches. Floyrac et al. used auditory evoked potentials recorded non-invasively during an oddball paradigm in a cohort of 29 post-cardiac arrest anoxic comatose patients to predict return to consciousness and good neurological outcomes. By extracting features from the standard and the deviant auditory stimulations independently and using machine learning to cluster patients within the two-dimensional space determined by these features, they were able to predict patients' neurological outcomes with a sensitivity of 0.83 and an accuracy of 0.90, even when using data only from one electrode. Ren et al. constructed a diagnostic model for cognitive impairment, a common disorder in patients with epilepsy, using the clinical and the phase locking value functional connectivity features of the electroencephalogram (EEG). Yoshiiwa et al., motivated by electroencephalographic studies of working memory demonstrating cortical activity and oscillatory representations without clarifying how the stored information is retained in the brain, measured scalp electroencephalography data while participants performed a modified n-back working memory task. They then calculated the current intensities from the estimated cortical currents by introducing a statistical map generated using Neurosynth as prior information. Their results indicate that the representation of executive control over memory retention may be mediated through both persistent neural activity and oscillatory representations in the beta and gamma bands over multiple cortical regions that contribute to visual working memory functions. Yazawa et al. created an arterially perfused *in situ* brainstem and spinal cord preparation that allowed them to investigate functional interactions in the CNS from the neonatal to adult period, bypassing the technical limitations on the spatial and temporal scope of *in vitro* neonatal rodent spinal cord preparations imposed by low oxygen tension in deep tissues. Using their novel preparation, they explored whether the absence of interferon regulatory factor 8 (IRF8) – which affects behavior and modulates Alzheimer's disease progression in a mouse model – influences the development of lumbar central pattern generator (CPG) networks in mice of all ages. Finally, Mount et al. explored how autism spectrum disorder (ASD) risk genes influence neural circuit computation during behavior by performing large-scale cellular calcium imaging from hundreds of individual CA1 neurons simultaneously in transgenic mice with total knockout of the X-linked ASD-risk gene *NEXMIF* (neurite extension and migration factor). As *NEXMIF* knockout in mice led to profound learning and memory deficits, they examined the CA1 network during voluntary locomotion, a fundamental component of spatial memory. They found that in wild-type mice the CA1 network desynchronizes during locomotion, consistent with increased network information coding during active behavior. Upon *NEXMIF* knockout, the CA1 network is over-synchronized regardless of behavioral state and fails to desynchronize during locomotion, highlighting how perturbations in ASD-implicated genes create abnormal network synchronization that could contribute to ASD-related behaviors.

In conclusion, it is our hope that the work collected in this Topic will serve as a basis for future studies exploring the potential application of Bayesian and other models in Translational Neuroscience.

Author contributions

JR: Conceptualization, Writing—original draft, Writing—review & editing. BP-P: Conceptualization, Writing—original draft, Writing—review & editing. RR: Conceptualization, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by R21 AG079158, U19 AT010829, RG-2206-39711, U01 NS123658, and CPCP-2023-002 (JR). BP-P was funded by NIH grants 1R21MH127483-01-A1 and 1P50MH132642-01.

Conflict of interest

RR was employed by American Physical Society.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Brodersen, K. H., Daunizeau, J., Mathys, C., Chumbley, J. R., Buhmann, J. M., and Stephan, K. E. (2013). Variational Bayesian mixed-effects inference for classification studies. *Neuroimage* 76, 345–361.doi: 10.1016/j.neuroimage.2013.03.008

Bruineberg, J., Kiverstein, J., and Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese* 195, 2417–2444. doi: 10.1007/s11229-016-1239-1

Duarte-Carvajalino, J. M., Lenglet, C., Xu, J., Yacoub, E., Ugurbil, K., Moeller, S., et al. (2014). Estimation of the CSA-ODF using Bayesian compressed sensing of multi-shell HARDI. *Magn. Reson. Med.* 72, 1471–1485. doi: 10.1002/mrm.25046

van Amersfoort, J., Smith, L., Teh, Y., and Gal, Y. (2020). Uncertainty Estimation Using a Single Deep Deterministic Neural Network. arXiv:2003. 02037

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fnins.2022.1044680

Bayesian multisource data integration for explainable brain-behavior analysis

Rong Chen*

Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD, United States

OPEN ACCESS

EDITED BY
Reza Rastmanesh,
American Physical Society,
United States

REVIEWED BY
Karl Friston,
University College London,
United Kingdom
Victorita Neacsu,
University College London,
United Kingdom

*CORRESPONDENCE
Rong Chen
rchen@som.umaryland.edu

SPECIALTY SECTION
This article was submitted to Translational Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 14 September 2022
ACCEPTED 12 October 2022
PUBLISHED 28 October 2022

CITATION
Chen R (2022) Bayesian multisource data integration for explainable brain-behavior analysis.
Front. Neurosci. 16:1044680.
doi: 10.3389/fnins.2022.1044680

Different data sources can provide complementary information. Moving from a simple approach based on using one data source at a time to a systems approach that integrates multiple data sources provides an opportunity to understand complex brain disorders or cognitive processes. We propose a data fusion method, called Bayesian Multisource Data Integration, to model the interactions among data sources and behavioral variables. The proposed method generates representations from data sources and uses Bayesian network modeling to associate representations with behavioral variables. The generated Bayesian network is transparent and easy to understand. Bayesian inference is used to understand how the perturbation of representation is related to behavioral changes. The proposed method was assessed on the simulated data and data from the Adolescent Brain Cognitive Development study. For the Adolescent Brain Cognitive Development study, we found diffusion tensor imaging and resting-state functional magnetic resonance imaging were synergistic in understanding the fluid intelligence composite and the total score composite in healthy youth (9–11 years of age).

KEYWORDS

Bayesian network, brain-behavior analysis, explainable AI, Bayesian inference, data fusion

1. Introduction

A central topic in neuroscience is understanding the association between the brain and behavior in normal and diseased states. Neuroimaging provides a non-invasive tool to study brain structure and function *in vivo* and is a powerful tool for brain-behavior analysis. A brain characterization framework is referred to as a data source ("source" here means the source or cause of a particular data feature). A data source can be an imaging method such as resting-state functional magnetic resonance imaging (fMRI); or it can be a kind of feature from an imaging method, for example, structural MRI can generate four data sources: volume, thickness, surface, and curvature. Most existing neuroimaging studies focus on a single data source. Many brain disorders are complex diseases. It's highly unlikely that one source will be able to fully capture the brain disorder. Different sources can provide complementary information. Moving from a simple approach based on using one source at a time to a systems approach that integrates multiple sources provides an opportunity to identify composite neuroimaging biomarkers for brain disorders.

Explainable AI (XAI) aims to develop AI algorithms in which the processes of action (e.g., predictions or recommendations) can be easily understood by users. Explainable models enable users to understand and appropriately trust the developed models. Interpreting the decision-making process of models in the biomedical domain is especially important.

We propose a method, called Bayesian Multisource Data Integration (BAMDI), to model the interactions among data sources and behavioral variables. BAMDI generates a representation from a data source and associates the representation with behavioral variables. The generated representation is referred to as embedding. The embedding is a set of vectors. Each vector is referred to as a factor. BAMDI has the following features. First, it centers on brain-behavior analysis. Many data integration methods focus on generating shared representation and cannot answer the question of how cross-source interactions are related to the behavior (Geenjaar et al., 2021; Zhang et al., 2022). In contrast, BAMDI represents interactions among different sources and behavioral variables as a Bayesian network. Brain-behavior analysis is the core of BAMDI. Second, BAMDI is an XAI method. Unlike some black-box methods, the Bayesian network generated by BAMDI is transparent and easy to understand. We use Bayesian inference to understand how the perturbation of a factor is related to the behavioral change.

Various Bayesian fusion methods for neuroimaging data have been proposed. Wei et al. developed a Bayesian fusion method to provide informative (empirical) neuronal priors—derived from dynamic causal modeling of electroencephalogram data—for subsequent dynamic causal modeling of fMRI data (Wei et al., 2020). Kang et al. proposed a Bayesian hierarchical spatiotemporal model to combine diffusion tensor imaging (DTI) and fMRI data (Kang et al., 2017). This method uses DTI-based structural connectivity to construct an informative prior for functional connectivity estimation. A parametric Bayesian multi-task learning based approach is developed to fuse univariate trajectories of neuroimaging features across subjects (Aksman et al., 2019). This Bayesian method fuses neuroimaging data across subjects, instead of modalities. Different from the above methods, the proposed method centers on modeling the interactions among data sources and behavioral variables with Bayesian network modeling, an XAI method.

In what follows, we first describe the overall design of BAMDI and its constituent modules. Following this, we applied BAMDI to simulated data to establish face validity. In other words, to ensure that the proposed scheme can recover the known brain-behavior mappings used to generate synthetic data. After this, we applied BAMDI to empirical data—from a publicly available databank—to characterize the relationship between MRI data from children, and their behavioral phenotypes as assessed with a battery of standard neurocognitive instruments.

2. Methods

2.1. Background

One of the foundations of BAMDI is Bayesian network modeling (Pearl, 1988; Koller and Friedman, 2009). A Bayesian network $\mathcal{B} = \{\mathcal{G}, \Theta\}$ is a probabilistic graphical model, where $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is a directed acyclic graph. A node X in $\mathcal{V}$ is a random variable in the problem domain. $\mathcal{E}$ is the edge set. A parent node of X is a node from which there exists a directed edge to X. The parent set of X is denoted by $pa(X)$. The local distribution is the conditional distribution $P(X|pa(X))$. The full specification of local distribution is the parameterization of the network. Θ is the set of parameters. The joint distribution can be represented compactly: $P(\mathcal{V}) = \prod_i P(X_i|pa(X_i))$. In BAMDI, we adopt the discrete Bayesian network representation and all nodes are discrete variables because the discrete Bayesian network can represent any kind of distribution among discrete variables and has high representation power. In a discrete Bayesian network, $P(X_i|pa(X_i))$ is a conditional probability table. For node X_i, the conditional probability $\theta_{ijk} = P[X_i = k|pa(X_i) = \mathbf{j}]$ is the probability that node X_i assumes state k when the parent set of X_i assumes state $\mathbf{j}$. If X_i has no parents, then θ_{ijk} is the marginal distribution of X_i. $\Theta = \{\theta_{ijk}\}$ is the parameters of discrete Bayesian network.

Bayesian network structure learning aims to learn $\mathcal{G}$. Bayesian network parameter learning is the process to estimate Θ. Score-based structure learning methods use a score that reflects how well the data support the structure and search for a structure that can optimize the fitness score. For discrete Bayesian networks, a widely used score is the Bayesian Dirichlet equivalent uniform (BDeu) score (Heckerman et al., 1995).

Bayesian network inference performs queries about probability distribution once some evidence about variables is available. The task of inference is to compute $P(\mathbf{Y}|\mathbf{X} = \mathbf{x})$, the posterior distribution of the query variables $\mathbf{Y}$, conditioned on $\mathbf{X} = \mathbf{x}$. In this paper, we use the algorithm in Lauritzen and Spiegelhalter (1988) to solve the inference problem.

2.2. Bayesian multisource data integration

The basic idea of BAMDI is as follows. In our data generation model, we imagine that there exist various brain states that generate a variety of neuroimaging data features. For example, being in one state or another state determines the pattern of functional connectivity in regional resting-state fMRI time courses. To model brain-behavior relationships, we assume that brain states (i.e., "factors") cause a particular behavioral disposition that is reflected in behavioral measures or scores. That is, the brain states are the parent nodes of behavioral states which can be measured by behavioral variables.

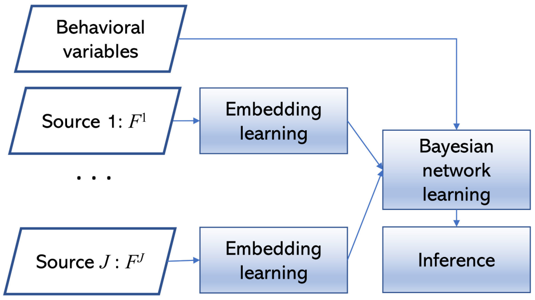


FIGURE 1
The BAMDI algorithm.

There can be many different kinds of brain states that may, or may not, interact in causing a particular behavioral state. Similarly, a particular behavioral state can be caused by one or more brain states. The problem then is to identify the brain-behavior associations in terms of the structure of a Bayesian network. This is accomplished using Bayesian network structure learning, following the identification of brain states using a clustering algorithm.

BAMDI learns a Bayesian network $\mathcal{B}$ from the observed data $\mathbf{D}$. It includes these main modules: embedding learning, Bayesian network learning, and inference. The algorithm is depicted in Figure 1. For source j, the feature set $\mathbf{F}^j$ is a vector with dimension $|\mathbf{F}^j|$, where $|\mathbf{F}^j|$ is the cardinality of $\mathbf{F}^j$. For a study with I subjects, the observed data $\mathbf{S}^j$ is an $I \times |\mathbf{F}^j|$ data matrix. For a study with J data sources and K behavioral variables, the whole dataset includes $\{\mathbf{F}^1, \ldots, \mathbf{F}^J\}$ and the associated behavioral variables $\mathbf{B} = \{B_1, \ldots, B_K\}$.

The first module is embedding learning. For each data source, we use graph-based clustering to generate an embedding. For $\mathbf{S}^j$, we group subjects into clusters. We normalize variables in $\mathbf{F}^j$ to zero-mean and unit variance. For subjects i_1 and i_2, we calculate the Euclidean distance d_{i_1,i_2} and obtain the similarity score as $1/(1 + d_{i_1,i_2})$. For a study with I subjects, this step generates an $I \times I$ similarity matrix that can be treated as a weighted graph. Then we use the multi-level modularity optimization algorithm (Blondel et al., 2008) to detect community structures in the weighted graph. The number of communities is determined by the algorithm. If subjects i_1 and i_2 belong to the same community, they are in the same cluster. Clustering generates a partition of the subject space. We convert this categorical variable into the embedding with one-hot encoding. Each cluster is associated with a binary variable that represents whether a subject belongs to the cluster (0—no, 1—yes). We use C_l^j to denote the l^{th} factor of the embedding for source j. $\mathbf{C}^j = \{C_1^j, \ldots, C_L^j\}$. For example, if the clustering algorithm generates 5 clusters, then the embedding contains 5 binary factors.

The second module is Bayesian network learning. We construct a Bayesian network $\mathcal{B}$ to describe interactions among $\{\mathbf{C}^1, \ldots, \mathbf{C}^J, \mathbf{B}\}$. We use Bayesian network classifier with inverse-tree structure (BNCIT) to solve this problem (Chen and Herskovits, 2005a,b). BNCIT is an efficient Bayesian network learning algorithm. In BNCIT, the parent set of a node in $\mathbf{B}$ is a subset of $\{\mathbf{C}^1, \ldots, \mathbf{C}^J\}$. There are no edges from $\mathbf{B}$ to $\{\mathbf{C}^1, \ldots, \mathbf{C}^J\}$. We adopt this kind of Bayesian network structure because we focus on studying how the embedding will affect behavioral variables. For a node X in $\mathbf{B}$, we search for a subset $\mathbf{C}^s$ of $\{\mathbf{C}^1, \ldots, \mathbf{C}^J\}$ which can maximize the BDeu score for structure $\mathbf{C}^s \rightarrow X$. That is, the parent set of X is determined by $\mathbf{C}^* = argmax_{\mathbf{C}^s}\, BDeu(\mathbf{C}^s \rightarrow X)$. This search process runs in a node-by-node fashion. After structure learning, the parameters are estimated by the maximum a posteriori method.

The inference module centers on explaining the generated model. The Bayesian network structure reveals important brain-behavior patterns. If the parent set of a behavioral variable includes factors from different data sources, then these sources are synergistic regarding this behavioral variable. If two behavioral variables have shared parent nodes, then these two behavioral variables have a shared brain mechanism. If the factors from a specific data source j are not associated with any behavioral variables, then source j provides little information about behaviors or source j is redundant.

A factor is a binary variable. We use two scores, divergence and mode change, to quantify how the change of factor C's state influences the marginal distribution of behavioral variable B by comparing $P(B|C = 0)$ and $P(B|C = 1)$. Both $P(B|C = 0)$ and $P(B|C = 1)$ are discrete probability distributions. We calculate the Jensen–Shannon divergence which is a symmetrized and smoothed version of the Kullback–Leibler divergence (Lin, 1991). For distributions p and q, the Kullback–Leibler divergence is defined as $D_{KL}(p\|q) = \sum p \log \frac{p}{q}$. The Jensen–Shannon divergence is defined as $D_{KL}(p\|m) + D_{KL}(q\|m)$, where $m = (p+q)/2$ and $D_{KL}(p\|m)$ is the Kullback–Leibler divergence between p and m. The Jensen–Shannon divergence is between 0 (identical) and 1 (maximally different) when the base 2 logarithm is used. For mode change, if the mode of $P(B|C = 0)$ is different from that of $P(B|C = 1)$, the value of this score is 1; otherwise, it is 0.

3. Results

3.1. Simulated data

We generated simulated data with three data sources (M1, M2, and M3) and four behavioral variables (BV_1, BV_2, BV_3, BV_4). Sources M1, M2, and M3 included 10, 10, and 30 variables, respectively. Source M1 included 2 clusters: samples 1–50 and 151–200 were sampled from a multivariate Gaussian distribution with mean $= \{3, \ldots, 3\}$ and samples 51–150 were

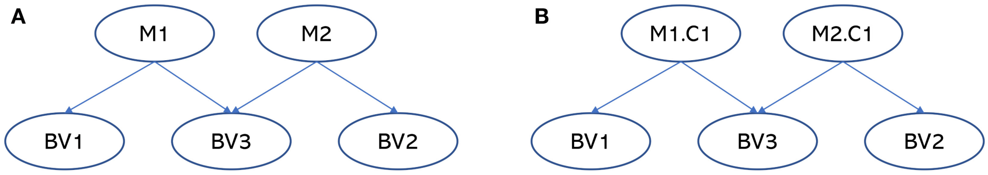


FIGURE 2
The Bayesian networks for the simulated data. **(A)** Is the ground-truth Bayesian network model to generate the simulated data and **(B)** is the Bayesian network generated by BAMDI. In the ground-truth model, BV_1 is associated with M_1, BV_2 is associated with M_2, and BV_3 is associated with both M_1 and M_2. In the model generated by BAMDI, $M1.C1$ is factor 1 from source M1. $M2.C1$ is factor 1 from source M2. Other factors were not associated with any behavioral variables and were not shown in the figure. The model generated by BAMDI matches the ground-truth model perfectly.

sampled from a multivariate Gaussian distribution with mean $= \{8, \ldots, 8\}$. Source M2 included 2 clusters: samples 1–150 were sampled from a multivariate Gaussian distribution with mean $= \{15, \ldots, 15\}$ and samples 151–200 were sampled from a multivariate Gaussian distribution with mean $= \{18, \ldots, 18\}$. For source M3, all samples (1–200) were generated from a multivariate Gaussian distribution with mean $= \{2, \ldots, 2\}$.

Let M_1 be a categorical variable to represent the cluster structure of source M1. $M_1 = 0$ for samples 1–50 and 151–200 and $M_1 = 1$ for samples 51–150. $M_2 = 0$ for samples 1–150 and $M_2 = 1$ for samples 151–200. BV_1 was a noisy version of M_1 with flipping noise 0.1. BV_2 was a noisy version of M_2 with flipping noise 0.1. BV_3 was a noisy version of $[M_1 \text{ OR } M_2]$. BV_4 was randomly sampled from $\{0, 1\}$ and was not associated with M_1 or M_2. M_3 and BV_4 were isolated variables. M_3 was not associated with any behavioral variables and BV_4 was not associated with any sources. We included them to assess whether BAMDI can handle isolated sources and behavioral variables.

BAMDI detected two, two, and four clusters for sources M1, M2, and M3, respectively. There were eight factors in the generated embedding (two of them from M1, two of them from M2, and four of them from M3). Figure 2 is the generated Bayesian network. In this figure, $M1.C1$ is factor 1 from source M1. $M2.C1$ is factor 1 from source M2. Among these factors, two of them ($M1.C1$ and $M2.C1$) were associated with some behavioral variables. Other factors were not associated with any behavioral variables and were not shown in the figure. BV_4 was not associated with any factors and was not shown in the figure. There are important brain-behavior patterns that can be elucidated from the Bayesian network. First, the Bayesian network revealed that BV_1 was associated with source M1, BV_2 was associated with source M2, and BV_3 was associated with sources M1 and M2. This is expected. Second, BV_1 and BV_3 had a shared brain mechanism because $M1.C1$ was a common parent node. BV_2 and BV_3 had a shared brain mechanism because $M2.C1$ was a common parent node. Third, sources M1 and M2 were synergistic regarding BV_3 because $M1.C1$ and $M2.C1$ were jointly predictive of BV_3.

3.2. The Adolescent Brain Cognitive Development study

In this experiment, participant data were obtained from the baseline Adolescent Brain Cognitive Development (ABCD) study (release 3.0). 11875 youth (baseline 9–11 years of age) were recruited. Written informed consents were obtained from all parents. All children provided assent to a research protocol approved by the institutional review board at each study site. Details of ABCD MRI acquisition and sequence parameters are in Casey et al. (2018).

Our analysis included these MRI modalities: DTI and resting-state fMRI (rs-fMRI). For DTI, the ABCD database provides a variable for imaging quality. Low quality DTI data were excluded from our analysis. For DTI, standard measures related to white matter microstructural tissue properties were calculated. We used Fractional Anisotropy (FA) which is a measure of the degree of anisotropic water diffusion within a region. FA was averaged across voxels within the Destrieux region-of-interest (ROI) of sub-adjacent white matter. This process generated 148 features (2 hemispheres × 74 regions). The average measures for white matter voxels in the left hemisphere, right hemisphere, and whole brain were also calculated to represent global effects. There were 151 DTI-derived features. To remove batch effects, we used the ComBat algorithm (Fortin et al., 2018) to harmonize these DTI features.

Head motion is a major problem in rs-fMRI and leads to spurious findings. For a 4D rs-fMRI volume, the ABCD database provides information about the total number of frames and the number of frames with low motion. We generated a quality score for motion that was defined as the number of frames with low motion divided by the total number of frames. The quality score was used as an indicator of the overall motion level. We selected subjects with at least half of the frames without excessive head motion (the quality score of motion > 0.5). We excluded subjects with incomplete data (those with missing values).

For rs-fMRI, the imaging-derived features were correlation between distributed networks of brain regions (Marek et al., 2019). Thirteen brain networks were detected, including auditory network ("ad"), cingulo-opercular network ("cgc"), cingulo-parietal network ("ca"), default network ("dt"), dorsal attention network ("dla"), frontoparietal network ("fo"), "none" network ("n"), retrosplenial temporal network ("rspltp"), sensorimotor hand network ("smh"), sensorimotor mouth network ("smm"), salience network ("sa"), ventral attention network ("vta"), and visual network ("vs") (Gordon et al., 2017). Notice that these brain networks comprised ROIs with positive correlations, which means that the average signal reflects the activity of the network. Each network was treated as a node. Functional connectivity between node A and node B was measured by calculating the correlation coefficient between the average signal of A and that of B. There were 78 rs-fMRI-derived features. Each feature represented functional connectivity between a brain network pair.

In the ABCD study, the NIH Toolbox cognition measures were used to assess child cognition (Luciana et al., 2018). The seven cognitive tasks in the NIH Toolbox included the dimensional change card sort task to assess cognitive flexibility ("cardsort"), list sorting working memory task to assess working memory ("list"), picture sequence memory task to assess episodic memory ("picture"), pattern comparison processing speed task to assess processing speed ("pattern"), picture vocabulary task to measure vocabulary comprehension ("picvocab"), oral reading recognition task to measure language/reading decoding ("reading"), and the flanker task to assess attention and inhibition ("flanker"). The neurocognitive battery was administrated using an iPad with one-on-one monitoring by a research assistant. The total time for administration was about 35 min. Based on the seven task scores, three composite scores were generated: a total score composite ("totalcomp"), a crystallized intelligence composite ("cryst"), and a fluid intelligence composite ("fluidcomp"). The age-corrected total score composite has a mean of 100

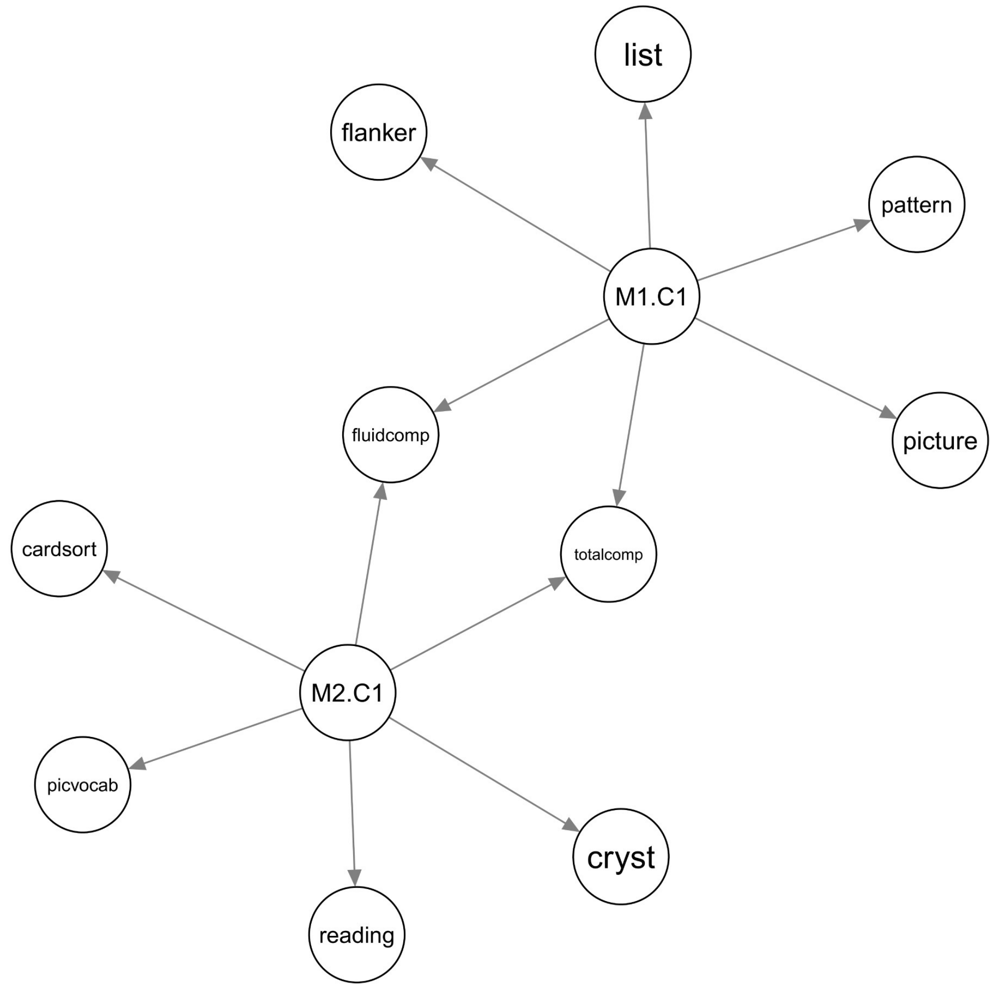


FIGURE 3
The Bayesian network for the ABCD study. Source M1 is DTI and source M2 is rs-fMRI. *M1.C1* is factor 1 from DTI. *M2.C1* is factor 1 from rs-fMRI. Other factors were not associated with any behavioral variables and were not shown in the figure.

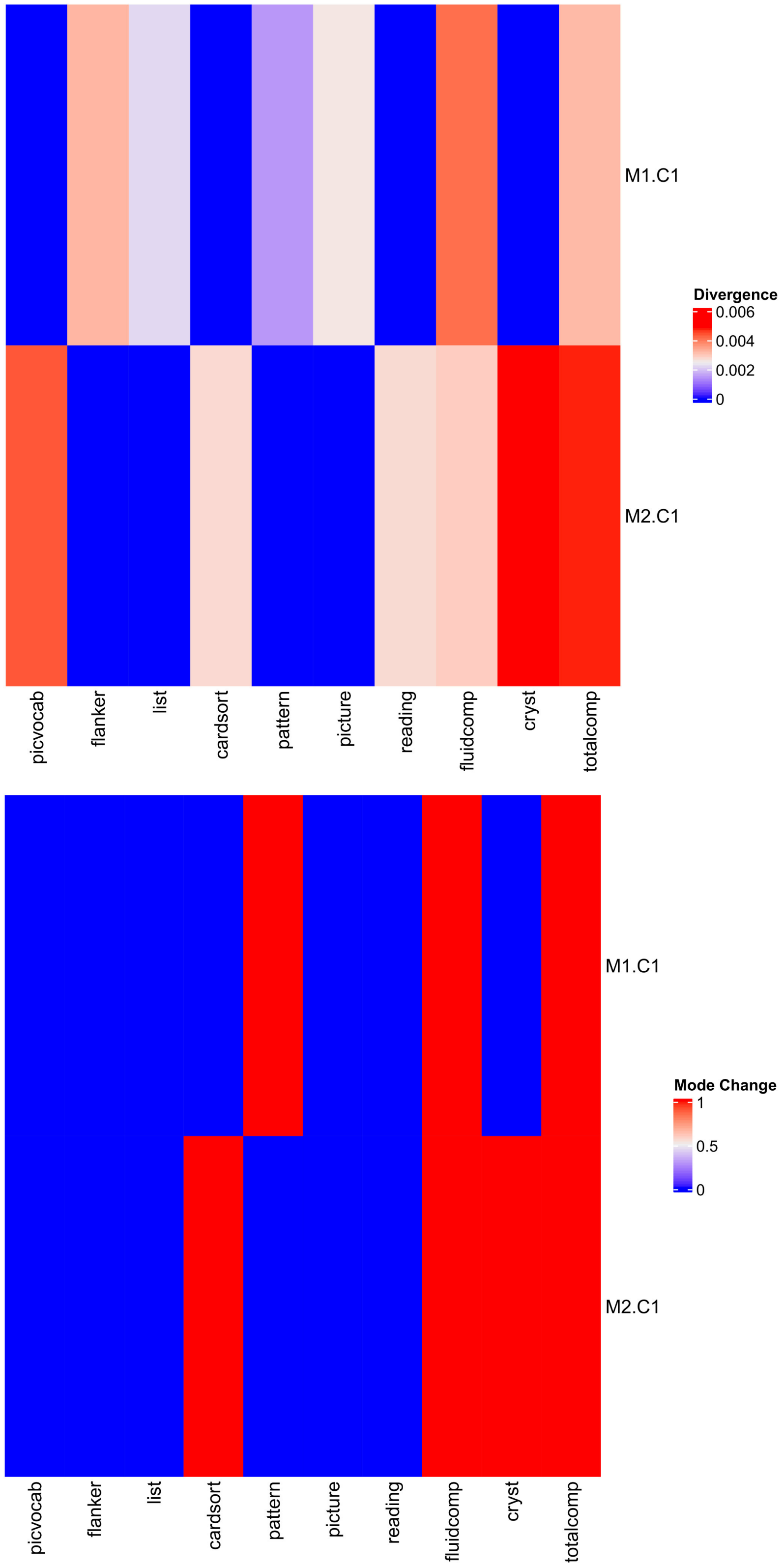


FIGURE 4
The divergence and mode change score for the ABCD study.

and a standard deviation of 15. For measures of cognition, higher scores represented better cognitive ability. The age-corrected scores were used as the behavioral variables in this study. These behavioral variables were binarized based on the sample median.

For DTI (source 1), BAMDI generated two factors. For rs-fMRI (source 2), BAMDI generated three factors. Among these five factors, two of them were associated with behavioral variables (Figure 3). DTI and rs-fMRI were synergistic regarding the fluid intelligence composite and the total score composite. The list sorting, flanker, picture sequence memory, and pattern comparison processing speed tasks were associated with DTI. The dimensional change card sort, picture vocabulary, oral reading recognition tasks, and crystallized intelligence composite were associated with rs-fMRI.

The divergence and mode change score are depicted in Figure 4. $M1.C1$ (factor 1 from DTI) had high divergence and high mode change score for the fluid intelligence composite and total score composite. That is, the change of $M1.C1$ changed the posterior marginal distribution of the fluid intelligence composite and total score composite. $M2.C1$ (factor 1 from rs-fMRI) had high divergence and high mode change score for the fluid intelligence composite, total score composite, and crystallized intelligence composite. That is, the change of $M2.C1$ changed the posterior marginal distribution of the fluid intelligence composite, total score composite, and crystallized intelligence composite.

To annotate important factors, we detected imaging markers to characterize factors. For a factor C^j from source j, we performed analysis of variance (ANOVA) with an imaging feature F^j as the dependent variable and C^j as the independent variable. Then we ranked imaging features based on the effect size and selected the top 10% features as the imaging markers. The imaging markers are shown in Figure 5. For DTI, the factor $M1.C1$ represented a subtype that had lower FA in the whole brain, right hemisphere, left superior frontal gyrus, left supramarginal gyrus, left superior parietal lobule, left precuneus, left lateral aspect of the superior temporal gyrus, right superior frontal gyrus, right angular gyrus, right supramarginal gyrus, right lateral aspect of the superior temporal gyrus, right central sulcus, right intraparietal sulcus and transverse parietal sulci, and right superior temporal sulcus. For rs-fMRI, the factor $M2.C1$ represented a subtype that had higher functional connectivity between the default network and auditory network, frontoparietal network and auditory network, "none" network and auditory network, sensorimotor hand network and frontoparietal network, and lower functional connectivity between visual network and auditory network, visual network and cingulo-opercular network, visual network and sensorimotor hand network, and visual network and ventral attention network.

4. Discussion

Data fusion is important for the understanding of inter-dependencies and relations across heterogeneous types of data. We propose a data fusion method called BAMDI to model the interactions among data sources and behavioral variables. The generated Bayesian network describes brain-behavior relationships. It is explainable: (1) the structure of Bayesian network reveals important brain-behavior patterns such as source synergy; (2) the divergence and mode change score assess how the change of factor affects the marginal distribution of behavioral variables.

We assessed the performance of BAMDI in two studies: simulated data and the ABCD study. For the simulated data, BAMDI correctly detected the brain-behavior patterns including BV_3 is a noisy version of $[M_1 \text{ OR } M_2]$. For the ABCD study, the two data sources, DTI and rs-fMRI, were synergistic regarding the fluid intelligence composite and the total score composite. The change of $M1.C1$, a DTI-derived factor that was characterized by lower FA in many regions, changed the posterior marginal distribution of the fluid intelligence composite and total score composite. The change of $M2.C1$, a rs-fMRI derived factor characterized by hyper-connectivity related to the auditory network and hypo-connectivity related to the visual network, changed the posterior marginal distribution of the fluid intelligence composite, total score composite, and crystallized intelligence composite.

Data integration methods can be classified into three different categories: early integration, intermediate integration, and late integration. Early integration focuses on combining data before applying a learning algorithm. An example of early integration is learning a common latent representation. Intermediate integration produces a joint model learned from different sources simultaneously. Late integration methods model different sources separately, then combines the outputs. BAMDI is a late integration method. BAMDI is also related to collective learning. Collective learning (Chen et al., 2004) is a machine learning framework to learn a model from multiple and diverse datasets by stage-wise learning (local learning and cross learning). Under this framework, the embedding learning step in BAMDI is local learning and the Bayesian network learning step in BAMDI is cross learning.

One of the limitations of BAMDI is that it requires discrete behavioral variables. Some behavioral variables such as disease diagnosis (normal controls or Alzheimer's disease) are naturally discrete; while others may be continuous. For continuous behavioral variables, we need to discretize them and this discretization process may cause a loss of information. We could extend BAMDI to handle continuous behavioral variables. In this extension, we adopt the conditional Gaussian Bayesian network representation and the local distribution $P(X|pa(X))$ is a Gaussian mixture. This will be the focus of our future work.

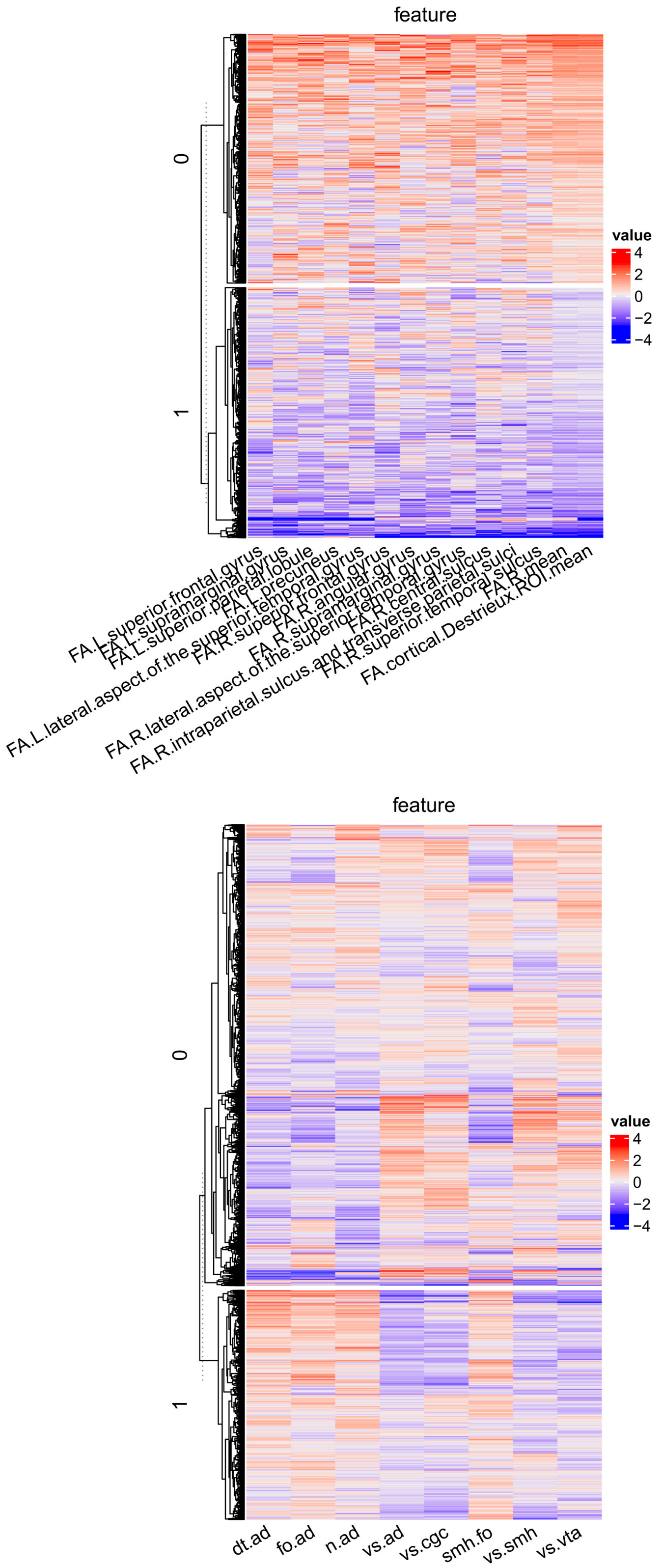


FIGURE 5
The imaging markers for DTI and rs-fMRI based factors.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://abcdstudy.org/.

Ethics statement

The ABCD study was approved by the ABCD Site Ethics Committee. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/principal-investigators/. The patients/participants provided their written informed consent to participate in this study.

Author contributions

RC designed the study, implemented the algorithm, conducted the experiments, and wrote the manuscript.

Funding

This work was partially supported by the NIH NINDS and the BRAIN Initiative (R01NS110421) and NIH NIDA (UG3DA053802).

Acknowledgments

ABCD acknowledgment: data used in the preparation of this article were obtained from the ABCD Study (https://abcdstudy.org/), and are held in the NIMH Data Archive. This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9–10 and follow them over 10 yr into early adulthood. The ABCD Study was supported by the National Institutes of Health (NIH) and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners/. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/principal-investigators/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. The ABCD data repository grows and changes over time.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

References

Aksman, L. M., Scelsi, M. A., Marquand, A. F., Alexander, D. C., Ourselin, S., Altmann, A., and ADNI (2019). Modeling longitudinal imaging biomarkers with parametric Bayesian multi-task learning. *Hum. Brain Mapp.* 40, 3982–4000. doi: 10.1002/hbm.24682

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008, P10008. doi: 10.1088/1742-5468/2008/10/P10008

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., et al. (2018). The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* 32, 43–54. doi: 10.1016/j.dcn.2018.03.001

Chen, R., and Herskovits, E. H. (2005a). "A Bayesian network classifier with inverse tree structure for voxelwise magnetic resonance image analysis," in *Proceeding of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining - KDD '05* (New York, NY: ACM Press), 4. doi: 10.1145/1081870.1081875

Chen, R., and Herskovits, E. H. (2005b). Graphical-model-based morphometric analysis. *IEEE Trans. Med. Imaging* 24, 1237–1248. doi: 10.1109/TMI.2005.854305

Chen, R., Sivakumar, K., and Kargupta, H. (2004). Collective mining of Bayesian networks from distributed heterogeneous data. *Knowledge Inform. Syst.* 6, 164–187. doi: 10.1007/s10115-003-0107-8

Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., et al. (2018). Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* 167, 104–120. doi: 10.1016/j.neuroimage.2017.11.024

Geenjaar, E., Lewis, N., Fu, Z., Venkatdas, R., Plis, S., and Calhoun, V. (2021). "Fusing multimodal neuroimaging data with a variational autoencoder," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE), 3630–3633. doi: 10.1109/EMBC46164.2021.9630806

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., et al. (2017). Precision functional mapping of individual human brains. *Neuron* 95, 791–807. doi: 10.1016/j.neuron.2017.07.011

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Mach. Learn.* 20, 197–243. doi: 10.1007/BF00994016

Kang, H., Ombao, H., Fonnesbeck, C., Ding, Z., and Morgan, V. L. (2017). A Bayesian double fusion model for resting-state brain connectivity using joint functional and structural data. *Brain Connect.* 7, 219–227. doi: 10.1089/brain.2016.0447

Koller, D., and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques.* Cambridge, MA: MIT Press.

Lauritzen, S. L., and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. R. Stat. Soc. Ser. B* 50, 157–194. doi: 10.1111/j.2517-6161.1988.tb01721.x

Lin, J. (1991). Divergence measures based on the shannon entropy. *IEEE Trans. Inform. Theory* 37, 145–151. doi: 10.1109/18.61115

Luciana, M., Bjork, J., Nagel, B., Barch, D., Gonzalez, R., Nixon, S., et al. (2018). Adolescent neurocognitive development and impacts of substance use: overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Dev. Cogn. Neurosci.* 32, 67–79. doi: 10.1016/j.dcn.2018.02.006

Marek, S., Tervo-Clemmens, B., Nielsen, A. N., Wheelock, M. D., Miller, R. L., Laumann, T. O., et al. (2019). Identifying reproducible individual differences in childhood functional brain networks: an ABCD study. *Dev. Cogn. Neurosci.* 40, 100706. doi: 10.1016/j.dcn.2019.100706

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann. doi: 10.1016/B978-0-08-051489-5.50008-4

Wei, H., Jafarian, A., Zeidman, P., Litvak, V., Razi, A., Hu, D., et al. (2020). Bayesian fusion and multimodal DCM for EEG and fMRI. *Neuroimage* 211, 116595. doi: 10.1016/j.neuroimage.2020.116595

Zhang, J., Wang, H., Zhao, Y., Guo, L., and Du, L. (2022). Identification of multimodal brain imaging association via a parameter decomposition based sparse multi-view canonical correlation analysis method. *BMC Bioinformatics* 23(Suppl 3), 128. doi: 10.1186/s12859-022-04669-z

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 November 2022
DOI 10.3389/fnins.2022.1043922

Check for updates

OPEN ACCESS

EDITED BY
Reza Rastmanesh,
The Nutrition Society, United Kingdom

REVIEWED BY
Karl Olof Lovblad,
Hôpitaux Universitaires de Genève,
Switzerland
Qinqin Liu,
The Second Affiliated Hospital
of Harbin Medical University, China

*CORRESPONDENCE
Guang-Zhi Liu
guangzhi2002@hotmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Translational Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 14 September 2022
ACCEPTED 25 October 2022
PUBLISHED 09 November 2022

CITATION
Fan Z-X, Wang C-B, Fang L-B, Ma L, Niu T-T, Wang Z-Y, Lu J-F, Yuan B-Y and Liu G-Z (2022) Risk factors and a Bayesian network model to predict ischemic stroke in patients with dilated cardiomyopathy. *Front. Neurosci.* 16:1043922. doi: 10.3389/fnins.2022.1043922

Risk factors and a Bayesian network model to predict ischemic stroke in patients with dilated cardiomyopathy

Ze-Xin Fan[1]†, Chao-Bin Wang[2]†, Li-Bo Fang[3], Lin Ma[1], Tian-Tong Niu[1], Ze-Yi Wang[1], Jian-Feng Lu[1], Bo-Yi Yuan[1] and Guang-Zhi Liu[1]*

[1]Department of Neurology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China,
[2]Department of Neurology, Beijing Fangshan District Liangxiang Hospital, Beijing, China,
[3]Department of Neurology, Beijing Fuxing Hospital, Capital Medical University, Beijing, China

Objective: This study aimed to identify risk factors and create a predictive model for ischemic stroke (IS) in patients with dilated cardiomyopathy (DCM) using the Bayesian network (BN) approach.

Materials and methods: We collected clinical data of 634 patients with DCM treated at three referral management centers in Beijing between 2016 and 2021, including 127 with and 507 without IS. The patients were randomly divided into training (441 cases) and test (193 cases) sets at a ratio of 7:3. A BN model was established using the Tabu search algorithm with the training set data and verified with the test set data. The BN and logistic regression models were compared using the area under the receiver operating characteristic curve (AUC).

Results: Multivariate logistic regression analysis showed that hypertension, hyperlipidemia, atrial fibrillation/flutter, estimated glomerular filtration rate (eGFR), and intracardiac thrombosis were associated with IS. The BN model found that hyperlipidemia, atrial fibrillation (AF) or atrial flutter, eGFR, and intracardiac thrombosis were closely associated with IS. Compared to the logistic regression model, the BN model for IS performed better or equally well in the training and test sets, with respective accuracies of 83.7 and 85.5%, AUC of 0.763 [95% confidence interval (CI), 0.708–0.818] and 0.822 (95% CI, 0.748–0.896), sensitivities of 20.2 and 44.2%, and specificities of 98.3 and 97.3%.

Conclusion: Hypertension, hyperlipidemia, AF or atrial flutter, low eGFR, and intracardiac thrombosis were good predictors of IS in patients with DCM. The BN model was superior to the traditional logistic regression model in predicting IS in patients with DCM and is, therefore, more suitable for early IS detection and diagnosis, and could help prevent the occurrence and recurrence of IS in this patient cohort.

KEYWORDS

Bayesian network, stroke, dilated cardiomyopathy, prediction model, risk factor

Introduction

Dilated cardiomyopathy (DCM) is a myocardial disease characterized by left ventricular (LV) dilation and systolic dysfunction in the absence of coronary artery disease or abnormal loading conditions sufficient to produce LV impairment (Elliott, 2000). DCM most frequently occurs in younger adults, and its most common clinical manifestations include congestive heart failure, sudden death, arrhythmias, and thromboembolic events (Japp et al., 2016). Ischemic stroke (IS) is a catastrophic thromboembolic complication of DCM, reported in several case reports and case series (Spengos and Vemmos, 2010; Jeon et al., 2012; Kawano et al., 2014; Zhdanova et al., 2016; Li et al., 2017). Thus,early identification of IS in patients with DCM is important because it can improve clinical outcomes and reduce medical costs. So far, many prediction models have been proposed to estimate the probability of developing stroke under certain conditions [e.g., nonvalvular atrial fibrillation (AF), transient ischaemic attack (TIA)], such as the Framingham score (D'Agostino et al., 2008), ABCD (2) score (Johnston et al., 2007), and CHA2DS2-VASc score (Lip et al., 2010). Of them, the most commonly used models is the Framingham Stroke Risk Profile,which was created using Cox proportional hazards regression modeling of Framingham Study data to identify factors that were most predictive of the 10-year probability of stroke.

In general, traditional logistic regression requires independent variables that are uncorrelated with each other, but in practice, the factors affecting the occurrence of IS are not independent and may interact with each other to form a complex relationship network. Unlike logistic regression, Bayesian network (BN) can well reflect the potential relationship and relationship strength between variables by constructing directed acyclic graph and conditional probability table (Park et al., 2018). In addition, increasing evidence has confirmed successful application of BN in medical diagnosis, expert systems, statistical decision making, learning, and prediction (Agrahari et al., 2018; Zhang et al., 2019). However, an agreed set of guidelines or reports on developing predictive models for IS in DCM cohorts are currently unavailable. Hence, there is a great need for further work toward constructing highly predictive models for early IS detection and diagnosis. This study established and compared traditional logistic regression and BN predictive models for IS occurrence using known risk factors.

Materials and methods

Patients and data collection

We selected 634 patients with DCM treated at three referral management centers between January 2016 and August 2021, mainly because Beijing Anzhen Hospital is one of the largest national centers for cardiovascular disease. The following inclusion criteria were used: (i) age ≥ 18 years; (ii) diagnosis of DCM following the European Society of Cardiology proposal which is based on systolic dysfunction and LV dilatation confirmed by echocardiography or cardiac magnetic resonance imaging and after excluding abnormal loading conditions or coronary artery disease (Pinto et al., 2016). The exclusion criteria were as follows: (i) patients with ischemic cardiomyopathy, rheumatic heart disease, arrhythmogenic cardiomyopathy, congenital heart disease, pulmonary heart disease, drug-induced cardiomyopathy, hypertensive heart disease, perinatal cardiomyopathy, valvular heart disease, and alcoholic cardiomyopathy; (ii) patients with missing clinical data. IS was diagnosed based on medical history, clinical examination, and cranial magnetic resonance imaging and magnetic resonance angiography scan results and confirmed by two attending neurologists.

Data collected at the first hospital admission, including demographic information, medical history, comorbidities, echocardiography, electrocardiogram, and laboratory tests, were collected from the electronic medical records. For patients with multiple admissions due to recurrent stroke, the data of the first admission were used in this study. This study followed the principles of the Declaration of Helsinki.

Abbreviations: AUC, area under the receiver operating characteristic curve; BN, Bayesian network; CI, confidence interval; DCM, dilated cardiomyopathy; eGFR, estimated glomerular filtration rate; IS, ischemic stroke; LV, left ventricular; LVEF, left ventricular ejection fraction; OR, odds ratio.

As Harrell (2015) stated, when developing a prediction model for dichotomous outcomes, the sample size should be at least 10 times the independent variable. In our research, 9 independent variables were finally included in multivariate analysis, and then the number of samples in each group should be at least 90. In fact, the number of cases of DCM with IS or without IS was 127 and 507, respectively, thus the sample size was enough to develop the prediction model.

Quality control

The data extraction process from the medical records was standardized, and the investigators familiarized themselves with it before starting data retrieval for this study. Data entry followed a double-entry method. If discrepancies were found during the review process, the medical records were consulted, and the data were corrected.

Data processing for predictive variables

Before building the predictive model, the collected data are preprocessed based on previous literatures. According to the studies by Li (Li et al., 2017) and Sharma (Sharma et al., 2000), AF and intracardiac thrombus are common risk factors for IS, as well-known risk factor for embolic complications (Orenes-Piñero et al., 2017). Hence, in this study, AF and intracardiac thrombus is used as risk factors for IS. Apart from these two variables, Deng (Deng et al., 2019) and Fukui (Fukui et al., 2017) also reported that lower estimated glomerular filtration rate (eGFR) was related to IS risk, with their predictive validity being well-verified. Thus, five basic characteristics (sex, age, AF, intracardiac thrombus and eGFR) of participants are ascertained. Additionally, according to biostatistics literature (Rosner, 2016), data will lose its measure of confidence if its missing value ratio > 30%. Therefore, for our study, some instances were removed from the dataset if they had more than 6 missing attributes (6 of 18). These missing attributes normally result from time conflicts and failures in the tests. Finally, a total of 26 instances were utilized as the primary dataset.

Logistic regression was utilized to screen for possible IS-related factors and evaluate assess their associated risk intensities. Logistic regression models were then applied to predict the IS, splitting the data into training and testing sets at a ratio of 7:3 using the random number table method. The training dataset was used to fit the prediction model (to "train" the algorithm), and then the model was utilized to predict the variable of interest from the test dataset. Similarly, a BN model of the IS-related risk factors in patients with DCM was established by a Tabu search algorithm using the training dataset. The test dataset was used to assess the models' accuracy. Before establishing the BN model, all IS-related factors were quantified and coded (**Supplementary Table 1** in **Supplementary material 1**).

Bayesian networks

As a probabilistic graphical model, the BN uses directed acyclic graphs to describe the probabilistic relationships between variables (Liao et al., 2017). The directed acyclic graph nodes stand for random variables $U = \{X_i, \ldots, X_n\}$, and the directed edges (E) stand for the probabilistic dependency relations

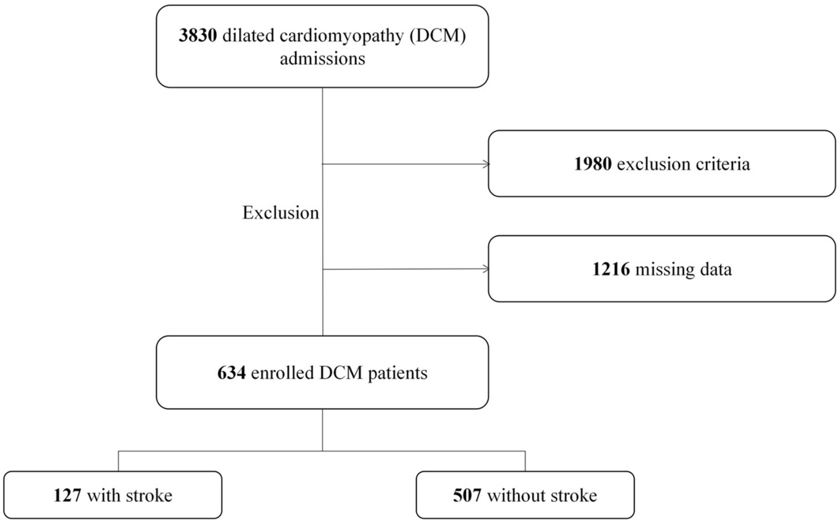

3830 dilated cardiomyopathy (DCM) admissions
Exclusion
1980 exclusion criteria
1216 missing data
634 enrolled DCM patients
127 with stroke
507 without stroke

FIGURE 1
Flowchart describing the screening of patients with dilated cardiomyopathy (DCM).

between the variables. If a directional arc from X_1 to X_2 is seen, we can infer that X_1 causes X_2; thus, X_1 and X_2 are usually defined as the parent and child, respectively. Each node has a conditional probability distribution table representing the parent node's state. The BN is a representation of the joint probability distributions of random variables $X = \{X_1, \ldots, X_n\}$; thus, a probability expression can be obtained:

$$\begin{aligned} P(X_1, ..., X_n) &= P(X_1)P(X_2|X_1)...P(X_n|X_1, X_2, ...X_{n-1}) \\ &= \prod_{1}^{n} P(X_i|\pi(X_i)) \end{aligned}$$

where $\pi(X_i)$ represents the collection of the parents of X_i; $\pi(X_i) \subseteq \{X_1 \ldots, X_{i-1}\}$ (Zhang et al., 2019).

In the present study, the collected dataset was utilized to construct a BN model for predicting the occurrence of IS. We extracted from the patient data 26 random variables for each instance. We initially filtered the nodes using logistic regression, in order to avoid including too many nodes and adding excessive complexity to the network structure. We then established the optimal model on the basis of Tabu search algorithm (Zhang et al., 2019).

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows, Version 23.0 (IBM Corp., Armonk, NY, USA). Continuous variables are presented as mean ± standard deviation or median (interquartile range). Categorical variables are expressed as numbers and percentages. Normally distributed data were analyzed using the Student's *t*-test (hematocrit, hemoglobin), and non-normally distributed data were analyzed using the Mann-Whitney *U* test [age, systolic blood pressure, leukocyte, platelet, eGFR, serum sodium (Na+), high-sensitivity C-reactive protein (Hs-CRP), D-dimer, left ventricular end-diastolic diameter, left ventricular ejection fraction, left atrium diameter, pulmonary arterial pressure]. Categorical variables were analyzed using the chi-squared test (male, smoking, drinking, hyperuricemia, hypertension, hyperlipidemia, diabetes, AF or atrial flutter, cardiac function, left bundle branch block, mitral regurgitation, and intracardiac thrombosis). Binary logistic regression analysis assessed the variables associated with DCM-related IS. Variables demonstrating an association with the outcome at a level of < 0.05 in univariate analysis were candidates for further multivariate analysis. Receiver operating characteristic analysis assessed the predictive models, and their areas under the curve (AUCs) were calculated. Furthermore, Delong test was applied to test the statistical significance of the difference between the AUC values. Hosmer–Lemeshow test and calibration plots were used to assess the calibration of each model. Statistical significance was set at $P < 0.05$. RStudio software, Version 4.2.0,[1] was employed for structural learning of the BN and parameter estimation using the maximum likelihood estimation method. The BNs' topology and conditional probability distribution tables were drawn using the Netica32 software (Norsys Software Corp., Vancouver, BC, Canada).

Results

Patients selection

Among the 3,830 patients diagnosed with DCM, 3,196 were excluded because of secondary cardiomyopathy etiologies or missing data. Finally, 634 eligible cases, including 127 with and 507 without IS were included in the study (**Figure 1**).

Risk factors for ischemic stroke

Multiple variables, including basic characteristics, stroke risk factors, echocardiography findings [i.e., left ventricular end-diastolic diameter, LV ejection fraction (LVEF), and left atrium diameter], electrocardiogram, and laboratory results, were compared between patients with and without IS (**Table 1**). Of the 26 variables, nine were associated with IS by univariate logistic regression: hypertension [odds ratio (OR), 1.561; 95% confidence interval (CI), 1.068–2.282; $P = 0.022$], hyperlipidemia (OR, 1.548; 95% CI, 1.018–2.354; $P = 0.041$), AF or atrial flutter (OR, 1.754; 95% CI, 1.159–2.655; $P = 0.008$), eGFR (OR, 0.980; 95% CI, 0.971–0.988; $P < 0.001$), serum sodium (OR, 0.915; 95% CI, 0.865–0.968; $P = 0.002$), Hs-CRP (OR, 1.029; 95% CI, 1.010–1.048; $P = 0.002$), D-dimer (OR, 1.000; 95% CI, 1.000–1.001; $P = 0.015$), cardiac function (classes III and IV; OR, 1.720; 95% CI, 1.093–2.706; $P = 0.019$), and intracardiac thrombosis (OR, 5.682; 95% CI, 3.130–10.315; $P < 0.001$).

The following five significant variables were retained in the final multivariate logistic regression model after performing a backward stepwise variable selection: hypertension (OR, 1.531; 95% CI, 1.004–2.334; $P = 0.048$), hyperlipidemia (OR, 1.723; 95% CI, 1.088–2.729; $P = 0.020$), atrial fibrillation/flutter (OR, 1.597; 95% CI, 1.017–2.507; $P = 0.042$), eGFR (OR, 0.986; 95% CI, 0.977–0.995; P = 0.003), and intracardiac thrombosis (OR, 5.417; 95% CI, 2.849–10.300; $P < 0.001$; **Table 2**).

Bayesian network structure

The BN model of the IS-related factors consisted of 10 nodes and 13 directed edges. The nodes represented IS,

[1] https://www.rstudio.com/

TABLE 1 Baseline data of patients with dilated cardiomyopathy (DCM).

Variables	DCM with IS (n = 127)	DCM without IS (n = 507)	P-value
Age, years	58 (49, 63)	56 (47, 65)	0.39
Male	96 (75.6%)	377 (74.4%)	0.776
Current smoking	38 (29.9%)	111 (21.9%)	0.056
Current drinking	30 (23.6%)	87 (17.2%)	0.093
Hyperuricemia	30 (23.6%)	87 (17.2%)	0.093
Hypertension	58 (45.7%)	169 (33.3%)	0.01
Hyperlipidemia	43 (33.9%)	126 (24.9%)	0.04
Diabetes	37 (29.1%)	122 (24.1%)	0.238
AF or atrial flutter	46 (36.2%)	125 (24.7%)	0.009
Cardiac function (class III, IV)	98 (77.2%)	336 (66.3%)	0.018
Left bundle branch block	22 (17.3%)	94 (18.5%)	0.751
Mitral regurgitation (moderate to severe)	68 (53.5%)	297 (58.6%)	0.304
Systolic blood pressure (mmHg)	116 (103, 130)	116 (102, 126)	0.372
Leukocyte (10^9/L)	6.7 (6.0, 8.4)	6.8 (5.8, 8.1)	0.551
Hematocrit (%)	41.6 ± 5.5	42.2 ± 5.3	0.552
Platelets (10^9/L)	198 (169, 242)	202 (169, 246)	0.45
Hemoglobin (g/L)	141.5 ± 20.9	144.6 ± 19.3	0.4
eGFR (mL/min/1.73 m^2)	81.1 (59.4, 96.5)	90.2 (72.5, 102.1)	< 0.001
Serum Na + (mmol/L)	138.8 (136.4, 140.9)	139.7 (137.8, 141.2)	0.006
Hs-CRP (mg/L)	3.33 (0.97, 10.19)	1.82 (0.8, 5.8)	0.001
D-dimer (ng/mL)	240 (100, 611)	135 (78, 298)	< 0.001
Echocardiography			
LVEDD	64 (59, 71)	66 (60, 74)	0.064
LVEF	30 (25, 37)	30 (25, 37)	0.769
LAD	45 (40, 50)	45 (40, 50)	0.964
PAD	30 (25,45)	30 (25, 45)	0.45
Intracardiac thrombosis	27 (21.2%)	23 (4.5%)	< 0.001

AF, atrial fibrillation; eGFR, estimated glomerular filtration rate; Hs-CRP, high-sensitivity C-reactive protein; IS, ischemic stroke; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; LAD, left atrium diameter; PAD, pulmonary arterial pressure.

hypertension, hyperlipidemia, AF/atrial flutter, eGFR, serum sodium, high-sensitivity C-reactive protein, D-dimer, cardiac function (class III or IV), and intracardiac thrombosis. Nodes directly linked to IS through complex network relationships included hyperlipidemia, atrial fibrillation/flutter, eGFR, and intracardiac thrombosis; heart failure (cardiac function classes III and IV) was indirectly associated with eGFR and intracardiac thrombosis, and hypertension was either directly or indirectly linked with IS through its association with hyperlipidemia (**Figure 2**). Based on the maximum likelihood estimation, the common variables predicting IS were hypertension, hyperlipidemia, atrial fibrillation/flutter, eGFR, and intracardiac thrombosis (**Table 3**).

Model performance evaluation

Compared with the logistic regression predictive model, the BN model for predicting IS achieved higher or equal scores in the training and test datasets (**Table 4**). The BN model achieved accuracies of 83.7 and 85.5%, AUCs of 0.763 (95% CI, 0.708–0.818) and 0.822 (95% CI, 0.748–0.896), sensitivities of 20.2 and 44.2%, and specificities of 98.3 and 97.3% in the training and test datasets, respectively. The logistic regression predictive model achieved accuracies of 83.0 and 84.5%, AUCs of 0.714 (95% CI, 0.649–0.778) and 0.769 (95% CI, 0.674–0.864), sensitivities of 17.9 and 39.5%, and the same specificities as the BN model (**Figure 3**). However, the Delong test revealed that there were no statistical differences in the AUC values between BN model and logistic regression model in either training datasets or test cohorts (P = 0.199 or P = 0.388). In addition, the calibration plots showed that the predicted probabilities of IS agreed well with the actual observations (**Figure 4**), and the Hosmer–Lemeshow test also demonstrated good calibration for BN model in training sets (P = 0.9999, chi square = 0.462, degree of freedom = 8) and test sets (P > 0.9999, chi square = 0, degree of freedom = 8), as well as for logistic regression model in training sets (P = 0.8234, chi square = 4.359, degree of freedom = 8) and test sets (P = 0.1028, chi square = 13.273, degree of freedom = 8).

Discussion

Generally, disease risk prediction requires a statistical risk factor model (Zhang et al., 2016). The present study used univariate and multivariate logistic regression models to screen the main risk factors for IS in patients with DCM. Subsequently, we constructed a BN model to estimate the conditional probability of each node based on the univariate analysis using the Tabu search algorithm. Our BN analysis suggested that hypertension, hyperlipidemia, AF or atrial flutter, eGFR, and intracardiac thrombosis was directly associated with IS, while cardiac insufficiency (i.e., heart failure) was indirectly linked to IS through eGFR and intracardiac thrombosis. Our findings are consistent with a retrospective case series of cardioembolic strokes with hypertrophic cardiomyopathy (n = 8) or DCM (n = 12), showing that over half of the patients with DCM had reduced LVEF (< 40%), enlarged left ventricular end-diastolic dimension (> 5.6 cm) and left atrium diameter (> 4 cm), and most (60%) had documented sinus rhythm when AF was diagnosed at stroke onset or during follow-up (Li et al., 2017). Together with well-known cardiovascular risk factors,

TABLE 2 Risk factors of ischemic stroke in patients with dilated cardiomyopathy (DCM): Univariate and multivariate binary logistic regression analysis.

Characteristics	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Hyperlipidemia	1.548 (1.018–2.354)	0.041	1.723 (1.088–2.729)	0.020
Hypertension	1.561 (1.068–2.282)	0.022	1.531 (1.004–2.334)	0.048
AF or atrial flutter	1.754 (1.159–2.655)	0.008	1.597 (1.017–2.507)	0.042
eGFR (mL/min/1.73 m^2)	0.980 (0.971–0.988)	< 0.001	0.986 (0.977–0.995)	0.003
Serum sodium [Na](mmol/L)	0.915 (0.865–0.968)	0.002	0.965 (0.905–1.028)	0.267
Hs-CRP (mg/L)	1.029 (1.010–1.048)	0.002	1.014 (0.999–1.030)	0.071
D-dimer (ng/mL)	1.000 (1.000–1.001)	0.015	1.000 (1.0–1.0)	0.249
Cardiac function (class III, IV)	1.720 (1.093–2.706)	0.019	1.205 (0.732–1.981)	0.463
Intracardiac thrombosis	5.682 (3.130–10.315)	< 0.001	5.417 (2.849–10.300)	< 0.001

AF, atrial fibrillation; eGFR, estimated glomerular filtration rate; Hs-CRP, high-sensitivity C-reactive protein.

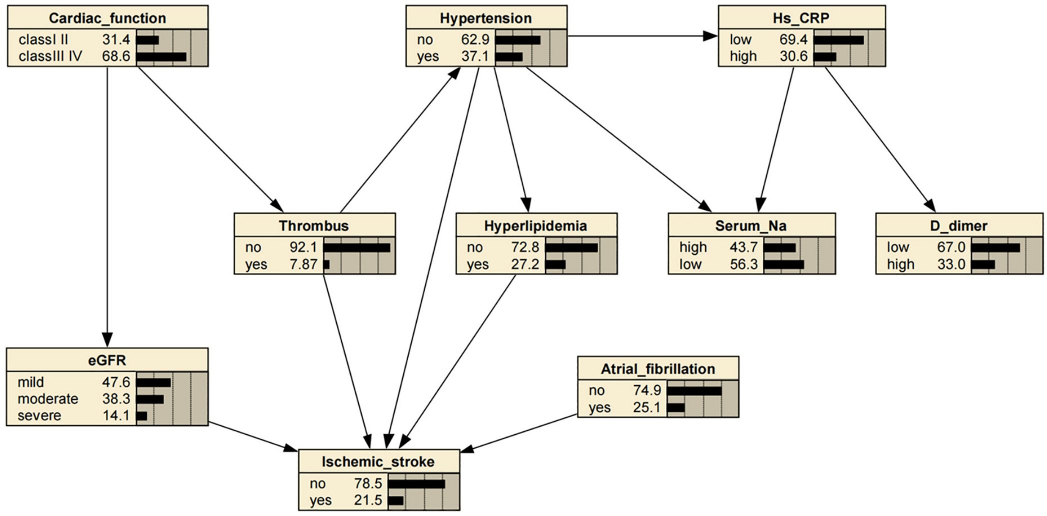


FIGURE 2

Bayesian network (BN) for predicting occurrence of ischemic stroke (IS) in patients with dilated cardiomyopathy (DCM). The BN model used nine variables selected by univariate logistic regression analysis. Estimated glomerular filtration rate (eGFR), high-sensitivity C-reactive protein (Hs−CRP), Serum sodium [Na], and D-dimer levels were defined according to their values. eGFR ml/min/1.73 m^2: mild (≥ 90), moderate (60−90), severe (≤ 60). Hs-CRP levels (mg/L): low (< 5), high (≥ 5). Serum sodium [Na] levels (mmol/L): high (≥ 140), low (< 140). D-dimer levels (ng/ml): low (< 240), high (≥ 240).

such as hypertension and hyperlipidemia (O'Donnell et al., 2010; Wang et al., 2022), these risk factors could prompt or contribute to the formation of intracardiac thrombi, resulting in cardioembolic stroke (Crawford et al., 2004; Li et al., 2017). Moreover, a retrospective cohort study by Deng et al. reported that decreased eGFR (≤ 60 mL/min/1.73 m^2) was associated with IS in patients with DCM (Deng et al., 2019). However, the underlying mechanism remains uncertain; therefore, we can only speculate that decreased eGFR in patients with DCM promotes the formation of thrombi through excessive oxidative stress on the vascular endothelium and activation of the renin-angiotensin system. Nonetheless, more evidence is required to address these issues.

In our study, cardiac insufficiency (i.e., heart failure) was indirectly linked to IS through eGFR and intracardiac thrombosis. This is noteworthy as a study by Kostas et al. revealed that heart failure, as a predictor independent of age, sex, stroke severity, and other stroke-related risk factors, could predict death in patients with stroke (Vemmos et al., 2012). Under pathophysiological conditions, patients with heart failure often have a decreased LVEF and abnormal intracardiac blood flow due to LV systolic dysfunction caused by LV dilation. Furthermore, endothelial dysfunction and changes in blood components (e.g., platelet function) have been observed in some patients with heart failure but normal LVEF, contributing to increased susceptibility

TABLE 3 The conditional probability table of the training set basing on ischemic stroke (IS) as the target node.

eGFR (mL/min/1.73 m^2)	Hypertension	Hyperlipidemia	Intracardiac thrombosis	AF/atrial flutter	Ischemic stroke	
					Yes	No
≥ 90	yes	no	no	no	0.14	0.86
≥ 90	yes	yes	no	no	0.18	0.82
≥ 90	yes	yes	no	yes	0.29	0.71
60–90	yes	yes	yes	no	1.00	0.00
60–90	no	yes	yes	no	1.00	0.00
≥ 90	no	no	no	no	0.15	0.85
≥ 90	no	yes	no	no	0.04	0.96
≥ 90	no	no	no	yes	0.19	0.81
≥ 90	yes	no	no	yes	0.09	0.91
≥ 90	no	yes	no	yes	0.20	0.80
60–90	no	no	no	no	0.02	0.98
60–90	yes	no	no	no	0.33	0.67
60–90	yes	yes	no	no	0.22	0.78
60–90	no	yes	no	no	0.10	0.90
60–90	no	no	no	yes	0.05	0.95
60–90	yes	no	no	yes	0.14	0.86
60–90	yes	yes	no	yes	0.25	0.75
60–90	no	yes	no	yes	0.40	0.60
≤ 60	no	no	no	no	0.29	0.71
≤ 60	yes	no	no	no	0.14	0.86
≤ 60	no	yes	no	no	0.50	0.50
≤ 60	yes	yes	no	no	0.50	0.50
≤ 60	yes	no	no	yes	0.50	0.50
≤ 60	no	no	no	yes	0.12	0.88
≤ 60	no	yes	no	yes	0.50	0.50
≤ 60	yes	yes	no	yes	0.50	0.50
≥ 90	no	no	yes	no	0.38	0.62
≥ 90	no	yes	yes	no	0.00	1.00
≥ 90	yes	no	yes	yes	0.00	1.00
≥ 90	yes	yes	yes	yes	0.00	1.00
60–90	no	no	yes	no	0.20	0.80
60–90	no	no	yes	yes	0.67	0.33
60–90	yes	no	yes	yes	0.00	1.00
60–90	no	yes	yes	yes	1.00	0.00
≤ 60	no	no	yes	no	0.75	0.25
≤ 60	no	no	yes	yes	1.00	0.00

to thromboembolism (Schumacher et al., 2018). Heart failure development might activate the sympathetic nervous system and the renin-angiotensin-aldosterone system, leading to constriction of glomerular afferent arterioles and decreased glomerular filtration rate and renal blood flow due to low cardiac output (Braunwald, 2019). Therefore, further investigation should determine the role of heart failure in the pathogenesis of IS in patients with DCM and whether timely therapy to improve cardiac function could reduce the occurrence of IS.

Bayesian network (BN) models possess certain advantages in the medical domain, including adaptability and strong robustness against missing values (Sheng et al., 2019). As to adaptability, building the BN model can start with limited domain knowledge, which is then simplified or extended by inputting new knowledge to meet various needs. Clinicians can add patients' updated knowledge, letting the BN model automatically adjust the probabilities. As to strong robustness against missing values, the BN model does not need complete knowledge of the topic; it can utilize available knowledge to perform its prediction. The BN model has been used to infer the probability of IS in patients with DCM. As shown in

TABLE 4 The performance of different predictive models.

Model	Accuracy	AUC	Sensitivity	Specificity
Bayesian network (training set)	83.67%	0.763	20.23%	98.32%
Logistic regression (training set)	82.99%	0.714	17.86%	98.32%
Bayesian network (test set)	85.49%	0.822	44.19%	97.33%
Logistic regression (test set)	84.45%	0.769	39.53%	97.33%

AUC, area under the curve.

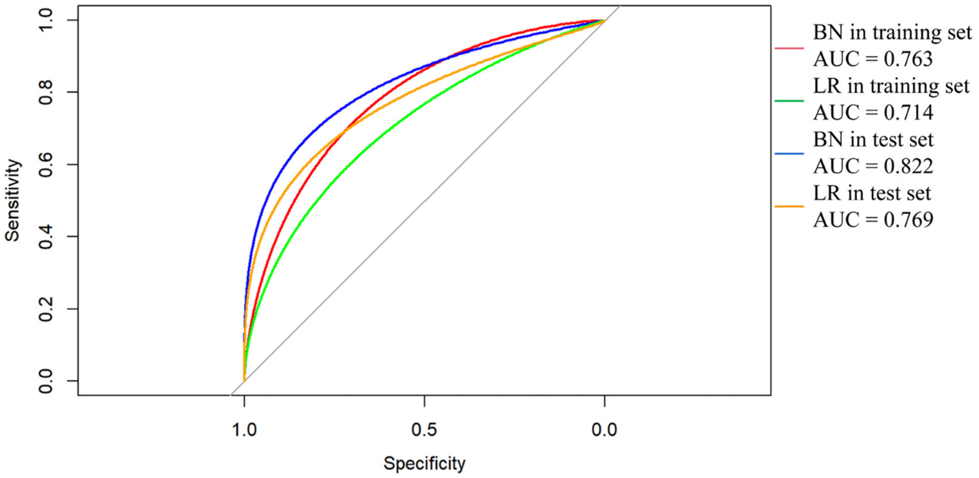


FIGURE 3
Receiver operating characteristic (ROC) curves of Bayesian network (BN) model and logistic regression (LR) model for predicting ischemic stroke (IS) in patients with dilated cardiomyopathy (DCM). The areas under the curve (AUC) of BN model predicting IS was 0.763 (95% CI, 0.708–0.818) and 0.822 (95% CI, 0.748–0.896) in (red line) training and (blue line) test datasets, respectively. The AUC of LR model predicting IS was 0.714 (95% CI, 0.649–0.778) and 0.769 (95% CI, 0.674–0.864) in (green line) training and (orange line) test datasets.

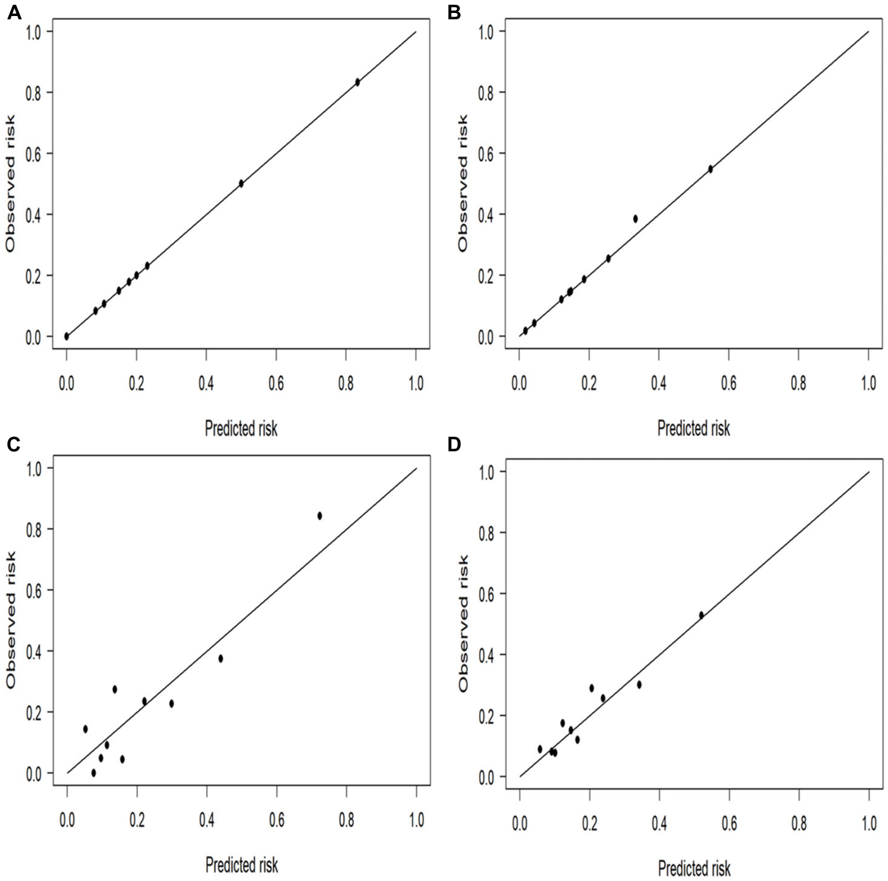


FIGURE 4
Calibration plots for the four prediction models in both cohorts. The perfect prediction should be on the 45-degree line. The calibration plots showed that the predicted risk of ischemic stroke (IS) agreed well with the observed risk, in either Bayesian network model of **(A)** test and **(B)** training datasets, or in logistic regression model of **(C)** test and **(D)** training datasets.

Table 3, patients with hypertension but without hyperlipidemia, abnormal renal function, intracardiac thrombosis, and AF or atrial flatter had a probability of 0.14 for concurrent IS; if the patient had hypertension and hyperlipidemia, the probability was 0.18; if the patients had atrial fibrillation/flatter, hypertension, and hyperlipidemia, the probability increased to 0.29; if the patient's eGFR was 60–90 mL/min/1.73 m^2, with hyperlipidemia, intracardiac thrombosis, but without AF or atrial flatter, the probability was 1.0. Hence, our results substantiated that the BN model based on the Tabu search algorithm had a flexible inference mechanism, making it very helpful for early IS detection and diagnosis in patients with DCM and, more importantly, for preventing the occurrence and recurrence of IS.

Besides its ability to generate an interpretable prediction and reduced uncertainty, BN is a powerful machine learning method to classify imbalanced datasets (Drummond and Holte, 2003; Monsalve-Torra et al., 2016), an important feature because a class imbalance is one of the most important challenges in real-world studies (Maldonado et al., 2014). In our study, calibration was good for both BN model and logistic regression model. Besides, the performance of our proposed BN model was promising and satisfactory in terms of accuracy, AUC, sensitivity, and specificity when compared to the traditional logistic regression model, albeit not statistically significant (e.g., AUC). This is possibly because logistic regression relies on independent variables, but the clinical features of IS and related factors are not independent; complex interaction networks might exist among them. Applied logistic regression models can predict the probability of developing IS until the state of the variables is known; however, in clinical practice, factors utilized for model prediction might be missing, leading to their inability to predict (Lee et al., 2005). In contrast the BN is constructed based on disease-related knowledge, fully mining potential information from the data and revealing the multilevel interactions between multiple factors. Additionally, the BN can outperform the radial basis function and multilayer perceptron in terms of sensitivity (Monsalve-Torra et al., 2016). In contrast, BN achieved a sensitivity of approximately 40% for identifying IS in our study. Three possible reasons for the imperfect sensitivity of our BN model were hypothesized. (i) The used dataset was not complex (contained only 26 attributes). The included attributes were derived from general information, including the subjects' basic characteristics and simple accessory tests, rather than special radiographic data such as brain neuroimaging. The main reason for using such a dataset was to develop a predictive model for IS that can be easily utilized in community clinics or rural hospitals. Hence, special neuroimaging data that might have improved its performance could not be included. (ii) The dataset used was not large (n = 634). The identification accuracy would undoubtedly be increased if a larger dataset was utilized (Wang et al., 2014). (iii) Skewed dataset could impact the model's performance (Watt and Bui, 2008); for example, males comprised 70% of the patients. Therefore, the reliability and validity of the BN model could be improved by using advanced learning algorithms.

In conclusion, our study is the first to propose a BN model to predict IS in patients with DCM, achieving a better performance than the traditional logistic regression model. Hypertension, hyperlipidemia, AF or atrial flutter, lower eGFR, and intracardiac thrombosis were good predictors of IS in our patient cohort. However, this study had some limitations. First, the number of patients with DCM complicated by IS was small. Second, as a retrospective study, clinical and laboratory data (e.g., troponin and B-type natriuretic peptide) were incomplete. Finally, the BN-directed edges reflected probability dependence between variables rather than a causal relationship. Therefore, long-term, multicenter prospective studies should be conducted to gain more insights into the potential causal relationship between the risk factors and IS in patients with DCM, optimize disease prevention strategies, and ultimately improve the long-term survival of patients with DCM.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethical Committee of Beijing Anzhen Hospital, Beijing Fangshan District Liangxiang Hospital, and Beijing Fuxing Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

G-ZL conceived the experiments. Z-XF, C-BW, and L-BF conducted the experiments. LM, T-TN, Z-YW, J-FL, and B-YY analyzed the results. All authors reviewed the manuscript.

Funding

This work was supported by the China National Natural Science Foundation (82071342).

Acknowledgments

We thank all the participants, research assistants, and outcome assessors involved in this study. We also thank Editage (www.editage.com) for English language editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1043922/full#supplementary-material

References

Agrahari, R., Foroushani, A., Docking, T. R., Chang, L., Duns, G., Hudoba, M., et al. (2018). Applications of Bayesian network models in predicting types of hematological malignancies. *Sci. Rep.* 8:6951. doi: 10.1038/s41598-018-24758-5

Braunwald, E. (2019). Diabetes, heart failure, and renal dysfunction: The vicious circles. *Prog. Cardiovasc. Dis.* 62, 298–302. doi: 10.1016/j.pcad.2019.07.003

Crawford, T. C., Smith, W. T. IV, Velazquez, E. J., Taylor, S. M., Jollis, J. G., and Kisslo, J. (2004). Prognostic usefulness of left ventricular thrombus by echocardiography in dilated cardiomyopathy in predicting stroke, transient ischemic attack, and death. *Am. J. Cardiol.* 93, 500–503. doi: 10.1016/j.amjcard.2003.10.056

D'Agostino, R. B. Sr., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., et al. (2008). General cardiovascular risk profile for use in primary care: The Framingham heart study. *Circulation* 117, 743–753. doi: 10.1161/CIRCULATIONAHA.107.699579

Deng, Y., Chen, Z., Hu, L., Xu, Z., Hu, J., Ma, J., et al. (2019). Decreased eGFR is associated with ischemic stroke in patients with dilated cardiomyopathy. *Clin. Appl. Thromb. Hemost.* 25:1076029619866909. doi: 10.1177/1076029619866909

Drummond, C., and Holte, R. C. (2003). "C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling," in *Proceedings of the workshop on learning from imbalanced datasets II*, (Washington, DC: Citeseer).

Elliott, P. (2000). Cardiomyopathy. Diagnosis and management of dilated cardiomyopathy. *Heart* 84, 106–112. doi: 10.1136/heart.84.1.106

Fukui, S., Imazeki, R., Amano, Y., Kudo, Y., Amari, K., Yamamoto, M., et al. (2017). Common and specific risk factors for ischemic stroke in elderly: Differences based on type of ischemic stroke and aging. *J. Neurol. Sci.* 380, 85–91. doi: 10.1016/j.jns.2017.07.001

Harrell, F. E. Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*, 2nd Edn. Berlin: Springer, doi: 10.1007/978-3-319-19425-7

Japp, A. G., Gulati, A., Cook, S. A., Cowie, M. R., and Prasad, S. K. (2016). The diagnosis and evaluation of dilated cardiomyopathy. *J. Am. Coll. Cardiol.* 67, 2996–3010. doi: 10.1016/j.jacc.2016.03.590

Jeon, G. J., Song, B. G., Park, Y. H., Kang, G. H., Chun, W. J., and Oh, J. H. (2012). Acute stroke and limb ischemia secondary to catastrophic massive intracardiac thrombus in a 40-year-old patient with dilated cardiomyopathy. *Cardiol. Res.* 3, 37–40. doi: 10.4021/cr142w

Johnston, S. C., Rothwell, P. M., Nguyen-Huynh, M. N., Giles, M. F., Elkins, J. S., Bernstein, A. L., et al. (2007). Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *Lancet* 369, 283–292. doi: 10.1016/S0140-6736(07)60150-0

Kawano, H., Inatomi, Y., Hirano, T., and Yonehara, T. (2014). Cerebral infarction in both carotid and vertebrobasilar territories associated with a persistent primitive hypoglossal artery with severe dilated cardiomyopathy. *J. Stroke Cerebrovasc. Dis.* 23, 176–178. doi: 10.1016/j.jstrokecerebrovasdis.2012.07.020

Lee, S. M., Abbott, P., and Johantgen, M. (2005). Logistic regression and Bayesian networks to study outcomes using large data sets. *Nurs. Res.* 54, 133–138. doi: 10.1097/00006199-200503000-00009

Li, C. H., Ma, S. K. T., and Chang, R. S. (2017). Cardioembolic stroke and cardiomyopathy: Rhythm is the key. *J. Neurol. Sci.* 380, 172–173. doi: 10.1016/j.jns.2017.07.032

Liao, Y., Xu, B., Wang, J., and Liu, X. (2017). A new method for assessing the risk of infectious disease outbreak. *Sci. Rep.* 7:40084. doi: 10.1038/srep40084

Lip, G. Y., Nieuwlaat, R., Pisters, R., Lane, D. A., and Crijns, H. J. (2010). Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: The euro heart survey on atrial fibrillation. *Chest* 137, 263–272. doi: 10.1378/chest.09-1584

Maldonado, S., Weber, R., and Famili, F. (2014). Feature selection for high-dimensional class-imbalanced data sets using support vector machines. *Inf. Sci.* 286, 228–246. doi: 10.1016/j.ins.2014.07.015

Monsalve-Torra, A., Ruiz-Fernandez, D., Marin-Alonso, O., Soriano-Payá, A., Camacho-Mackenzie, J., and Carreño-Jaimes, M. (2016). Using machine learning methods for predicting inhospital mortality in patients undergoing open repair of abdominal aortic aneurysm. *J. Biomed. Inf.* 62, 195–201. doi: 10.1016/j.jbi.2016.07.007

O'Donnell, M. J., Xavier, D., Liu, L., Zhang, H., Chin, S. L., Rao-Melacini, P., et al. (2010). Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): A case-control study. *Lancet* 376, 112–123. doi: 10.1016/S0140-6736(10)60834-3

Orenes-Piñero, E., Esteve-Pastor, M. A., Valdés, M., Lip, G. Y. H., and Marín, F. (2017). Efficacy of non-vitamin-K antagonist oral anticoagulants for intracardiac thrombi resolution in nonvalvular atrial fibrillation. *Drug Discov. Today* 22, 1565–1571. doi: 10.1016/j.drudis.2017.05.010

Park, E., Chang, H. J., and Nam, H. S. (2018). A Bayesian network model for predicting post-stroke outcomes with available risk factors. *Front. Neurol.* 9:699. doi: 10.3389/fneur.2018.00699

Pinto, Y. M., Elliott, P. M., Arbustini, E., Adler, Y., Anastasakis, A., Böhm, M., et al. (2016). Proposal for a revised definition of dilated cardiomyopathy, hypokinetic non-dilated cardiomyopathy, and its implications for clinical practice: A position statement of the ESC working group on myocardial and pericardial diseases. *Eur. Heart J.* 37, 1850–1858. doi: 10.1093/eurheartj/ehv727

Rosner, B. (2016). *Fundamentals of biostatistics*, 8th Edn. Boston, MA: Cengage Learning.

Schumacher, K., Kornej, J., Shantsila, E., and Lip, G. Y. H. (2018). Heart failure and stroke. *Curr. Heart Fail. Rep.* 15, 287–296. doi: 10.1007/s11897-018-0405-9

Sharma, N. D., McCullough, P. A., Philbin, E. F., and Weaver, W. D. (2000). Left ventricular thrombus and subsequent thromboembolism in patients with severe systolic dysfunction. *Chest* 117, 314–320. doi: 10.1378/chest.117.2.314

Sheng, B., Huang, L., Wang, X., Zhuang, J., Tang, L., Deng, C., et al. (2019). Identification of knee osteoarthritis based on Bayesian network: Pilot study. *JMIR Med. Inform.* 7:e13562. doi: 10.2196/13562

Spengos, K., and Vemmos, K. N. (2010). Etiology and outcome of cardioembolic stroke in young adults in Greece. *Hellenic J. Cardiol.* 51, 127–132.

Vemmos, K., Ntaios, G., Savvari, P., Vemmou, A. M., Koroboki, E., Manios, E., et al. (2012). Stroke aetiology and predictors of outcome in patients with heart failure and acute stroke: A 10-year follow-up study. *Eur. J. Heart Fail.* 14, 211–218. doi: 10.1093/eurjhf/hfr172

Wang, C., Du, Z., Ye, N., Shi, C., Liu, S., Geng, D., et al. (2022). Hyperlipidemia and hypertension have synergistic interaction on ischemic stroke: Insights from a general population survey in China. *BMC Cardiovasc. Disord.* 22:47. doi: 10.1186/s12872-022-02491-2

Wang, K. J., Makond, B., and Wang, K. M. (2014). Modeling and predicting the occurrence of brain metastasis from lung cancer by Bayesian network: A case study of Taiwan. *Comput. Biol. Med.* 47, 147–160. doi: 10.1016/j.compbiomed.2014.02.002

Watt, E. W., and Bui, A. A. (2008). Evaluation of a dynamic bayesian belief network to predict osteoarthritic knee pain using data from the osteoarthritis initiative. *AMIA Annu. Symp. Proc.* 6, 788–792.

Zhang, X., Yuan, Z., Ji, J., Li, H., and Xue, F. (2016). Network or regression-based methods for disease discrimination: A comparison study. *BMC Med. Res. Methodol.* 16:100. doi: 10.1186/s12874-016-0207-2

Zhang, Z., Zhang, J., Wei, Z., Ren, H., Song, W., Pan, J., et al. (2019). Application of tabu search-based bayesian networks in exploring related factors of liver cirrhosis complicated with hepatic encephalopathy and disease identification. *Sci. Rep.* 9:6251. doi: 10.1038/s41598-019-42791-w

Zhdanova, S. G., Petrikov, S. S., Ramazanov, G. R., Khamidova, L. T., Aliev, I. S., and Sarkisyan, Z. O. (2016). Dilated cardiomyopathy as a cause of ischemic stroke. *Zh. Nevrol. Psikhiatr. Im. S S Korsakova* 116, 44–47. doi: 10.17116/jnevro20161168244-47

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fnins.2022.1077735

Check for updates

OPEN ACCESS

EDITED BY
Reza Rastmanesh,
The Nutrition Society, United Kingdom

REVIEWED BY
Paul H. E. Tiesinga,
Radboud University, Netherlands
Guangzhi Liu,
Beijing Anzhen Hospital, Capital Medical University, China

*CORRESPONDENCE
Georgia Koppe
✉ georgia.koppe@zi-mannheim.de

SPECIALTY SECTION
This article was submitted to Translational Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 23 October 2022
ACCEPTED 14 December 2022
PUBLISHED 09 January 2023

CITATION
Thome J, Pinger M, Durstewitz D, Sommer WH, Kirsch P and Koppe G (2023) Model-based experimental manipulation of probabilistic behavior in interpretable behavioral latent variable models.
Front. Neurosci. 16:1077735.
doi: 10.3389/fnins.2022.1077735

Model-based experimental manipulation of probabilistic behavior in interpretable behavioral latent variable models

Janine Thome[1,2], Mathieu Pinger[3], Daniel Durstewitz[1,4], Wolfgang H. Sommer[5], Peter Kirsch[3,6] and Georgia Koppe[1,2]*

[1]Department of Theoretical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany, [2]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany, [3]Department of Clinical Psychology, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany, [4]Faculty of Physics and Astronomy, Heidelberg University, Heidelberg, Germany, [5]Institute of Psychopharmacology, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany, [6]Institute of Psychology, Heidelberg University, Heidelberg, Germany

Introduction: Interpretable latent variable models that probabilistically link behavioral observations to an underlying latent process have increasingly been used to draw inferences on cognition from observed behavior. The latent process usually connects experimental variables to cognitive computation. While such models provide important insights into the latent processes generating behavior, one important aspect has often been overlooked. They may also be used to generate precise and falsifiable behavioral predictions as a function of the modeled experimental variables. In doing so, they pinpoint how experimental conditions must be designed to elicit desired behavior and generate adaptive experiments.

Methods: These ideas are exemplified on the process of delay discounting (DD). After inferring DD models from behavior on a typical DD task, the models are leveraged to generate a second adaptive DD task. Experimental trials in this task are designed to elicit 9 graded behavioral discounting probabilities across participants. Models are then validated and contrasted to competing models in the field by assessing the ouf-of-sample prediction error.

Results: The proposed framework induces discounting probabilities on nine levels. In contrast to several alternative models, the applied model exhibits high validity as indicated by a comparably low prediction error. We also report evidence for inter-individual differences with respect to the most suitable models underlying behavior. Finally, we outline how to adapt the proposed method to the investigation of other cognitive processes including reinforcement learning.

Discussion: Inducing graded behavioral frequencies with the proposed framework may help to highly resolve the underlying cognitive construct and associated neuronal substrates.

KEYWORDS

reward discounting, delay discounting, computational models, behavioral model, design optimization, adaptive design, homogenizing behavior, computational psychiatry

Introduction

Behavioral latent variable models which describe an individual's trial-by-trial behavior in terms of well interpretable generative equations have become increasingly popular in neuroscience and psychiatry to quantify the mechanisms underlying cognitive processes involved in decision-making (Durstewitz et al., 2016; Huys et al., 2016). By inferring such models from an individual's choice sequence recorded during an experiment, the underlying cognitive processes which echo in these choices can be mapped onto an often low-dimensional set of interpretable model parameters, the involved sub-functions can be teased apart, and hypotheses directed at the algorithmic principles of the given process may be addressed (e.g., Huys et al., 2013; Collins et al., 2017; Koppe et al., 2017; Thome et al., 2022).

Besides these clear advantages, as of yet, one important aspect of such models has often been overlooked. Since they attempt to fully explain trial-by-trial behavior, these models typically incorporate all relevant factors necessary to describe variations in behavior. This also means that they (implicitly) predict how behavior would change if any of the relevant factors is varied. On the one hand, such model-based predictions can be leveraged to steer or *induce* behavior by manipulating the experiment (by varying one or more of the above-mentioned relevant factors), thus providing a formal recipe to generate adaptive model-based experiments (Thome et al., 2022). On the other hand, by comparing a broad range of these predictions to actual behavioral observations, we obtain a formal framework ideally suited to validate a given model. Here, we therefore build on a previously introduced generic model-based framework to improve the generation of adaptive experiments (Thome et al., 2022), and couple it to a formal behavioral model validation procedure.

We have illustrated the procedure in the context of delay discounting. Delay discounting refers to the tendency of an individual to favor immediate as compared to temporally distant outcomes due to future outcome devaluation. Since individuals differ strongly in their discounting behavior, adaptive tasks which aim at adjusting trials to the individual to induce and measure more homogeneous discounting behavior, have been the means of choice for quite some time (Monterosso et al., 2007; Ripke et al., 2012; Cavagnaro et al., 2016; Koffarnus et al., 2017; Pooseh et al., 2018; Ahn et al., 2020; Knorr et al., 2020). In a typical delay discounting task such as the intertemporal choice task (ICT), participants are faced with a series of choices between a delayed larger and immediate smaller reward (e.g., Mazur, 1987). A common model of behavior in the ICT assumes that choices are probabilistic draws based on internal choice values with a higher likelihood for choices of higher value (e.g., Pine et al., 2009; Prevost et al., 2010; Miedl, 2012; Peters et al., 2012; Ahn et al., 2020). These choice values are computed based on the presented rewards and delays in the experiment, the discounting function, and individual-specific discounting parameters which regulate its behavior. A probabilistic function maps these latent values to probabilities for immediate and delayed choices. By setting the conditional probability for an immediate choice to a given response probability for each unique participant, we can resolve the model equations for a condition that expresses how experimental stimuli need to be selected so that we can expect to observe this response probability. For example, when setting the discounting probability in an ICT to 0.5, this condition will express how to adjust rewards and delays in a given participant to obtain 50% discounted choices. We have recently successfully applied this framework to induce a 0.3, 0.5, and 0.7 discounting probability across individuals (Thome et al., 2022).

On the other hand, manipulating the experimental variables simultaneously renders *predictions* over behavioral response probabilities for a given model. The fields of statistical learning theory and machine learning (ML) instruct us on how to make use of such predictions to objectively assess model validity (Hastie et al., 2009; Koppe et al., 2021). Validity here refers to whether a function – for instance, a statistical model – generalizes well to the population and has a low expected prediction error (PE; Hastie et al., 2009; Koppe et al., 2021). In short, a method or function is valid if we can infer it on a sample and use it to predict new unseen measurements with low error (Hastie et al., 2009). This corresponds well to the psychological perspective on validity by which validity denotes the extent to which evidence and theory justify the interpretation of measurements (Schmidt-Atzert and Amelang, 2021). The appealing part about assessing a PE is that it provides an objective way to assess predictive validity, and may further yield quantitative information on how and where (i.e., in what domain) a method is valid. Here, we thus extend our model-based adaptive approach to (a) predict and induce a wider range of behavioral response probabilities, and (b) use these predictions to perform a formal assessment of the PE.

The advantages of such a procedure are manifold. For one, the approach provides a recipe of how to generate adaptive experiments that ensure similar behavioral probabilities between participants. Effectively, such a procedure reduces behavioral variance within experimental conditions and thereby increases statistical power (Winer, 1971; Mumford, 2012). At the same time, the proposed procedure relocates between-subject variability into the adaptive experimental variables and model parameters, such that this information is preserved and can be systematically studied (Kanai and Rees, 2011; Hedge et al., 2018). Second, by generating experimental conditions which induce graded response probabilities, we may also induce graded intensities of the underlying process, resolving it at a finer level. This is beneficial when linking behavior to, for instance, neuro(physio)logical mechanisms (Dagher et al., 1999; Grinband et al., 2006; Wood et al., 2008; Hare et al., 2009; Ripke et al., 2014; Grosskopf et al., 2021). Finally, the formal assessment of model validity in terms of (out-of-sample) estimates of the PE allows us to compare and select between a class of available or novel models, and evaluate the models on a wide range of the model domain.

The present work illustrates this procedure in the context of monetary reward delay discounting, and expands it to a broader class of cognitive domains. We address the hypothesis whether by applying the proposed approach we can successfully induce (relative) discounting frequencies on a 9-level graded scale (ranging from 0.1 to 0.9) which, to the best of our knowledge, has never been attempted before. We then illustrate how to formally assess (predictive) model validity within this framework by comparing predicted to induced response frequencies, and evaluate several models on a group and single-subject level. Finally, we outline how to adapt the approach to latent variable models which are history dependent as well as response models which are multi-categorical.

Materials and methods

Experimental design

General framework

The key aspect of the proposed framework is to experimentally manipulate the latent process of a latent variable model, and thereby generate precise and falsifiable hypotheses about the data generating process (and associated cognitive functioning) in a systematic manner. These hypotheses (i.e., predictions) are statements about the frequency of observed responses in consequence of the experimental manipulation. Latent variable models which formalize the latent process and its dependencies on experimental variables, and probabilistically link this process to behavior, provide the means for such a manipulation. This is because these models let us track how changes in the experimental variables will affect behavioral probabilities. By making use of this property, we can systematically tune the experiment to generate a given behavioral probability.

The framework proceeds in two experimental runs (see **Figure 1**). The first run (termed 'run A') serves to generate data to infer the models and thus the latent process of interest. The models are then leveraged to generate predictions and associated experimental manipulations which are then assessed in a second run (termed 'run B'). By separating run A and B in this way, we ensure that the trial-generation procedure is not biased by type of model applied, that is, the model is not inferred on trials it has itself selected. Validity of the instrument is measured by comparing these predictions to observations made in run B (**Figure 1B**). We illustrate and evaluate the framework here based on the case where we have a latent variable model with no history dependence combined with a binary response model. A transfer of the proposed approach to latent variable models with history dependence and response models which are multi-categorical is found in the Results section.

Binary response models with no history-dependence

Delay discounting provides a prominent example of a binary choice process in which the latent process does not depend on history (i.e., each choice depends only on current and not previous choice values). In the delay discounting example, run A and run B consist of an ICT. In this task, participants are faced with a series of binary choice trials in which they are asked to choose between an immediate smaller reward and a delayed larger reward (see **Figure 1A**). The collected data set d thus consists of T pairs of *observed* variables $d = \{(x_i, y_i),\ i = 1, 2, \ldots, T\}$, where x_i are predictor vectors of immediate and delayed reward and delay pairs, and y_i are one-dimensional (dichotomous) observed responses of immediate ($y_i = 1$) and delayed ($y_i = 0$) choices. The sequence of observed choices $Y = \{y_1, \ldots, y_T\}$ is modeled as i.i.d. Bernoulli random variables $y_i \sim Bi(1, \mu(x_i))$, for $i = 1, 2, \ldots, T$, where $\mu(x_i)$ is the probability of an immediate choice given the predictor x_i.

The probabilistic latent variable discounting model estimates $\mu(x_i)$ by mapping the observed predictor vectors x_i onto the conditional mean of the distribution of y_i via a latent process f_λ, that is, $\widehat{\mu}(x_i) : = E[y_i|f_\lambda(x_i)]$. f_λ is a discounting function mapping observable predictors x_i onto internally represented *latent* values v_i of these predictors, and

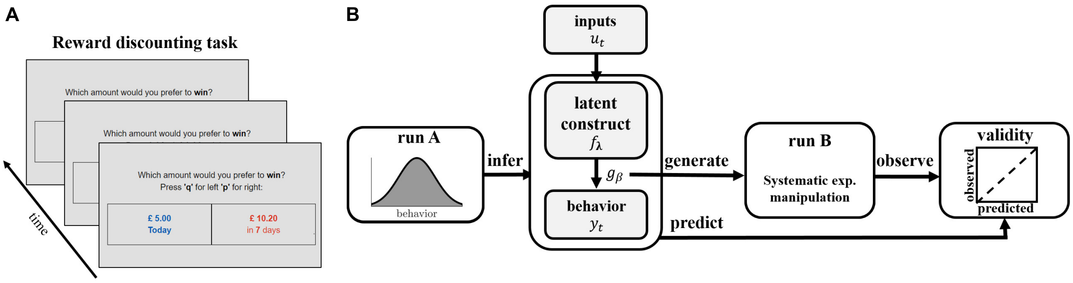


FIGURE 1
Schematic illustration of task and experimental framework. **(A)** Illustration of the reward discounting task. Participants are faced with a series of binary choice trials, in which they are asked to choose between an immediate smaller reward and a delayed larger reward. **(B)** Illustration of experimental protocol. Participants perform run A of the reward discounting task. Latent discounting models f_λ are inferred on each participant's sequence of observed behavioral choices $Y = \{y_1, \ldots, y_T\}$ in run A and used to generate trials of run B. Trials are systematically manipulated by varying experimental inputs u_t to induce discounting frequencies ranging from 0.1 to 0.9, based on the expectation of the probability distribution g_β. Validity is assessed by comparing predicted and observed rel. choice frequencies.

$\widehat{\mu}(x_i)$ maps these values onto the conditional mean of the Bernoulli distribution.

As discounting model f_λ, we chose a hyperbolic function with exponential delay termed 'modified hyperboloid model' in the following (after Mazur, 1987; Rachlin, 2006). The discounting model is a vector valued function mapping two rewards r presented at two delays D (displayed in each trial of the ICT and collected in predictors x_i above) onto internally represented values v for the two associated choices.

$$f_\lambda(r, D) : = (\frac{1}{1 + \kappa \cdot D^s})r \qquad (1)$$

where κ is a discount parameter capturing the individual tendency to discount, and s is a scaling parameter, both $\subset \lambda$. In each trial of the ICT, the discounting model thus maps an immediate reward r_{imm} (presented at 0 delay) and a delayed reward r_{del} presented at delay D_i onto immediate and delayed values v_{imm} and v_{del} for the respective choice. Since the immediate reward has 0 delay, it is equal to its latent value, i.e., $r_{imm} = v_{\text{imm}}$. We will refer to the factor in front of r in the following as the 'discount factor.' We have previously shown that this model performs consistently better at predicting unseen behavior than a number of other models (Thome et al., 2022; see also Estle et al., 2006; Odum et al., 2006; Rachlin, 2006; Rodzon et al., 2011; McKerchar et al., 2013; Cox and Dallery, 2016; Białaszek et al., 2020).

We choose a sigmoid function to map these two latent values onto the conditional mean, that is, onto the probability for selecting the immediate choice option.

$$\widehat{\mu}(x_i) = \frac{1}{1 + e^{\beta(v_{del} - v_{imm})}} \qquad (2)$$

where β is an individual-specific parameter which captures the sensitivity to differences in choice values (see also **Figure 2A**). Eqn. (2) maps differences in values to immediate choice probabilities p_{imm} (see **Figure 2A**), in close analogy to a psychometric function (e.g., Wichmann and Hill, 2001). It provides a condition which permits the systematic manipulation of experimental conditions. By setting p_{imm} to a given probability, we obtain an equation which may be resolved for an observable and tunable experimental variable. For instance, solving Eqn. (2) for the immediate reward r_{imm} by plugging in the model assumptions, we obtain the following condition.

$$r_{imm} = (\frac{1}{1{+}\kappa D^s})r_{del} + \frac{log(\frac{p_{imm}}{1 - p_{imm}})}{\beta} \qquad (3)$$

defined for $0 < p_{imm} < 1$. By inserting a set of fixed delays and delayed rewards, as well as inferred subjective parameters κ and β, Eqn. (3) expresses how to experimentally manipulate the presented immediate reward to obtain a desired immediate choice probability p_{imm} in a given individual (please see **Figures 2A–C** on operating principles of the method). By aiming to construct experimental conditions with similar immediate choice probabilities across participants, we are effectively homogenizing behavior across participants. We make the implicit assumption here that behavior is homogeneous if, within an experimental condition, different participants display similar frequencies, that is, they show similar probabilities, for the available behavioral options.

For the conducted experiment, models inferred on run A were applied according to this framework to generate an ICT with nine experimental conditions inducing graded immediate choice probabilities of $p_{imm} = \{0.1{,}0.2{,}0.3{,}0.4{,}0.5{,}0.6{,}0.7{,}0.8{,}0.9\}$ presented in run B (see **Figure 1**), for simplicity referred to as 'induced frequencies' in the following.

Experimental settings

Trials in run A followed a previously developed protocol optimized to elicit discounting behavior across participants (Thome et al., 2022) and optimized in line with results obtained from a preliminary experiment (see **Supplementary Text 2**).

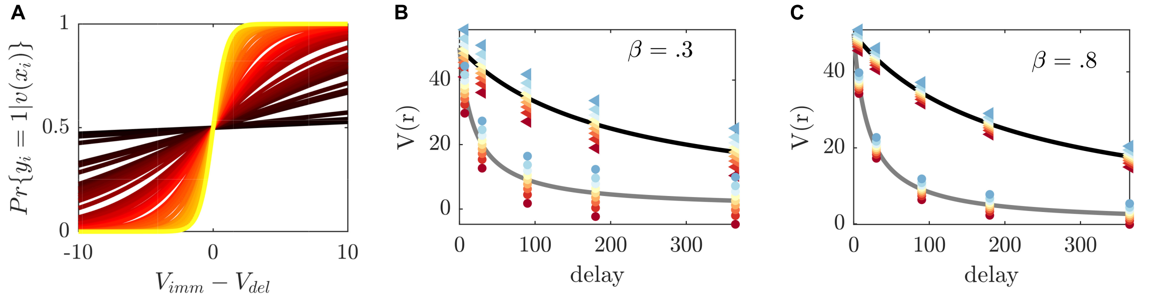


FIGURE 2
Illustration of method principles. **(A)** Immediate choice probability [cf. Eqn. (2)] as a function of the difference between immediate and delayed value for β estimates in our sample (color-coded from largest $\beta = 2$ in yellow to smallest $\beta = 0.01$ in dark red). The indifference point, i.e., the point at which immediate and delayed choice probability (and immediate and delayed choice values) are equal, is at 0. If v_{imm}/v_{del} immediate choice probability is below/above 0.5. β regulates the steepness of the curve and thus the sensitivity toward differences in values. **(B)** Discounted value of a reward of size 50 (y-axis) delayed at different time points (x-axis) and two exemplary κ's ($\kappa = 0.005$ in black and $\kappa = 0.05$ gray). The method's selected immediate rewards at a given delay are displayed as colored dots from blue to red with respect to the induced choice probability from 0.1 to 0.9 (triangles/circles are associated with $\kappa = 0.005/0.05$, respectively). To induce the same probabilities at different delays, the difference between the presented immediate choice values (depicted as colored dots) and delayed values (depicted on the discounting curve) is constant [see Eqn. (2), $v_{del} - v_{imm}$]. For participants with different κ, the reward and value ratios will therefore vary. The left graph depicts selected rewards for a hypothetical $\beta = 0.3$ and **(C)** the right graph for $\beta = 0.8$. β thus regulates the precise difference between immediate and delayed values, with higher β resulting in smaller differences, making the differentiation between the two more difficult.

Rewards and delays varied across trials. Delays were set to $D = \{7, 30, 90, 180, 365\}$ days and delayed rewards to $r_{del} = \{5, 10, 20, 50, 100\}$ £ (UK). Immediate rewards were selected based on the described model guided procedure [Eqn. (3)], chosen to elicit an equal probability for immediate and delayed choices at 4 different population representative discounting parameters. Run A thus consisted of 100 trials (5 delays × 5 delayed rewards × 4 discounting parameters). Trials of run B were generated via Eqn. (3), and the parameters inferred on run A, to induce 9 probability gratings ranging from 0.1 to 0.9. With nine gradings and using the same delays and delayed rewards as in run A, run B consisted of 225 trials.

A few parameter and stimulus constellations could result in immediate rewards smaller than 0, or equal immediate and delayed rewards. To avoid such trials, the delays (and corresponding immediate rewards) were iteratively adjusted in the trial-generating procedure until reaching a minimum of 1 or a maximum of 365 days. If still not resolved, negative immediate rewards were set to 1 penny, while immediate rewards equaling delayed rewards were reduced by 1 penny, respectively. This adaptation could result in a slight deviation of the induced frequencies (see **Figure 3B**, red line). Trials were self-paced, allowing for a maximum decision phase of 10 s, with a 1 s inter-stimulus-interval.

Model inference

Discounting models were inferred on run A and run B separately via maximum likelihood estimation (MLE). Given Bernoulli i.i.d. assumptions, the models' likelihood is given by $p(Y|X,\theta) = \prod_{i=1}^{T} p_\theta(y_i|v(x_i))$, where $p_\theta(y_i|v(x_i))$ is given by Eqn. (2) in case that y_i refers to the immediate choice, and by 1 minus this probability for the delayed choice, respectively. Under inspection of the preliminary experiment (see **Supplementary Text 2**), parameters were constrained to $\beta \epsilon$ [0.001, 2], $s \epsilon$ [0, 1], and $\kappa \epsilon$ [0, 1000], and optimization was performed using a Quasi-Newton algorithm (the limited-memory BFGS algorithm) implemented via the optimize.minimize() function from the SciPy library[1], starting from multiple initial conditions.

Sample and data assessment

Fifty healthy participants (24 males, 25 females, and 1 undefined) participated in the study, recruited via the following website: https://www.prolific.co. Participants were eligible if they were between the age of 18 and 65 with current residency in the United Kingdom (UK) and were reimbursed £7.50 per hour to participate in the study. Please see **Supplementary Tables 1, 3** for more information on the sample.

All participants accessed the study through a link on the Prolific website. They completed a consent form, filled out sociodemographic information, and took part in run A of the experiment. After completing run A, the alcohol use disorder identification questionnaire (AUDIT; Babor et al., 1992) and the short version of the Barratt Impulsiveness Scale (BIS-15; Spinella, 2007) were filled out, immediately followed by run B of the experiment. The whole procedure took 28.4 (±8.39) minutes on average. The study was approved by the local ethics committee of the Medical Faculty Mannheim, University of Heidelberg (2019-633N).

[1] https://scipy.org/citing-scipy/

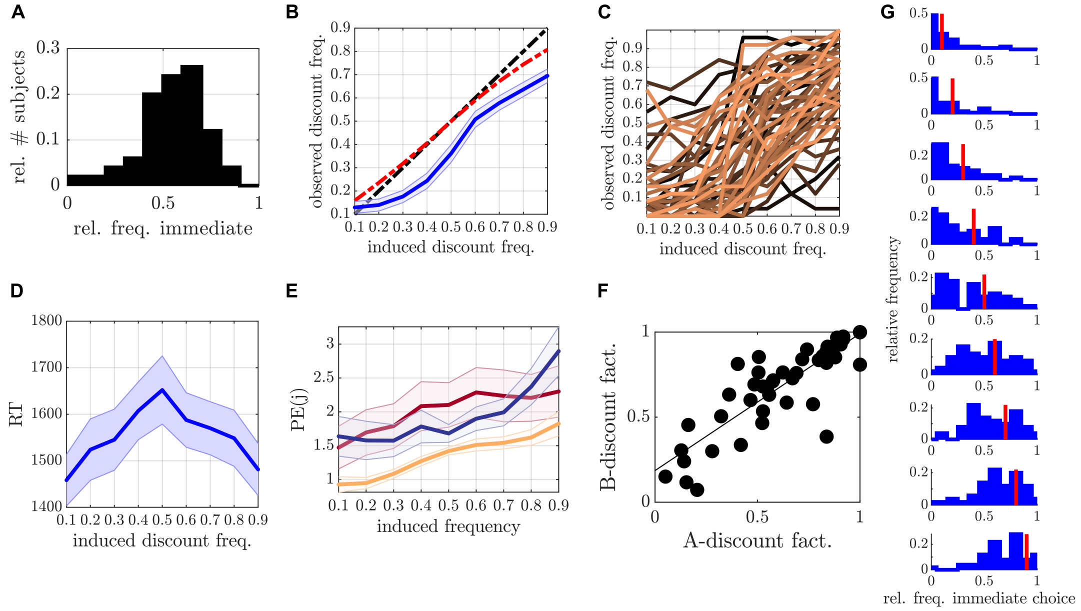


FIGURE 3

(A) Relative frequency of immediate choices in run A. **(B)** Relative frequency of discounted choices (*y*-axis) as a function of model induced frequencies (*x*-axis) averaged over all participants (mean and SEM are displayed in blue). The black dashed line marks the identity, while the red dashed line shows the actual predicted frequencies according to the models. **(C)** Single participant curves. **(D)** Mean and SEM of reaction time (RT) as a function of experimental conditions. **(E)** Mean and SEM of prediction error (PE) as a function of experimental conditions for modified hyperboloid (yellow), hyperbolic (red), and hyperboloid control (blue) models. **(F)** Discount factor in run A (*x*-axis) and run B (*y*-axis) illustrated for delay = 90, indicating a reliable estimate of discounting across runs. **(G)** Histograms of immediate choice behavior for 0.1 (top) to 0.9 (bottom) frequency conditions (conditions indicated by the red line).

Data analysis

Measured variables

We assessed the frequency of discounted choices (that is, choices in favor of the objectively smaller outcome) and median reaction time (RT) across run A, and for all 9 probability gradings (i.e., experimental conditions) in run B, model parameters (i.e., $\beta, \kappa, \ s$) the discount factor(s), as well as total scores of AUDIT and BIS/BAS questionnaires.

Inferential statistics

The agreement between experimentally induced frequencies and observed behavioral frequencies was assessed via a general linear model (GLM) with induced frequencies as linear predictor variables. The hypothesized inverted U-functional relationship between induced frequencies and RT was assessed via a GLM with quadratic induced frequencies as curvilinear predictor variables (hypothesizing higher/lower RT toward more difficult/easy choices) in run B. We report t-statistics on the regression coefficients of these models.

Prediction error assessment

Predictive validity was assessed by approximating the PE using cross-validation (CV). Rooted in statistical learning theory (Hastie et al., 2009; Efron, 2021), the PE quantifies the error made when applying a prediction rule, here the statistical model, to unseen (out-of-sample) data. Assuming that the (x, y) data pairs in the ICT follow an (unknown) joint distribution F, the PE quantifies the error made when drawing a new pair with only the predictor variable x observed and predicting y with $\widehat{\mu}(x)$ (cf. section "General framework") based on the model. Given some loss function $L(y, \widehat{\mu}(x))$ which assesses the deviation between observation and prediction, the PE is assessed as the expected loss under F, i.e., $Err = E_F\{L(y, \widehat{\mu}(x))\}$ (Hastie et al., 2009; Efron and Hastie, 2021). Since this expectation goes over all (x, y) pairs, this integral may not be computed directly, but is in practice often approximated by resampling methods such as CV. Using CV, here we approximate the PE by $\widehat{Err}_{CV} = \frac{1}{T}\sum_{i=1}^{T} L(y_i, \widehat{\mu}(x_i))$, where the summation runs exclusively over data pairs observed in run B and the prediction is based on models inferred on run A [denoted 'PE (run B)'], and vice versa [denoted 'PE (run A)']. As most appropriate for dichotomous data, we employ as error function the binomial deviance (Efron, 2021), given by $L(\mu, \widehat{\mu}) :=2\{\mu log\frac{\mu}{\widehat{\mu}} + (1-\mu)log(\frac{1-\mu}{1-\widehat{\mu}})\}$, such that the PE was assessed as

$$\widehat{Err}_{CV} = \frac{2}{T}\sum_{i=1}^{T}\{y_i log\frac{y_i}{\widehat{\mu}(x_i)} + (1-y_i)log(\frac{1-y_i}{1-\widehat{\mu}(x_i)})\}, \quad (4)$$

and $\widehat{\mu}(x_i) \in [0, 1]$ was truncated to [1e-10, 1 – 1e-10] to avoid infinities.

Since the integral in the PE runs across all possible (x, y) pairs, sampling a broad range of data pairs in run B – as achieved here by including nine experimental levels – should improve the estimation of Err by $\widehat{Err}_{CV}$. It furthermore allows to dissect and examine the PE for different experimental conditions. PEs were computed for each participant (i.e., at fixed parameter values), and averaged over to obtain a population estimate.

Model comparisons

The two experimental runs and PE assessment allows the objective comparison of different models in their prediction ability. Several models which varied in the assumption about the computation of the delayed values (and therefore the latent variable model) were evaluated in terms of PE (see **Supplementary Table 2**). These included the common hyperbolic model (Mazur, 1987; Davison and McCarthy, 1988), the exponential model (Samuelson, 1937), the constant sensitivity (CS) model (Ebert and Prelec, 2007), the modified hyperboloid model used for trial-generation (Mazur, 1987; Rachlin, 2006), the quasi-hyperbolic model (Phelps and Pollak, 1968; Laibson, 1997), the (conventional) hyperboloid model (Loewenstein and Prelec, 1992; Green et al., 1994), the double exponential model (van den Bos and McClure, 2013), and a control model to the modified hyperboloid model with β in Eqn. (2) fixed to 1. Details on these models can be found in the **Supplementary Table 2**. In addition, to investigate the model fit on a single participant level, we counted the number of individuals best described by each model.

Behavioral homogeneity

Reductions in behavioral variation within experimental conditions (i.e., increase in behavioral homogeneity) was tested by comparing variances of immediate choice frequencies via F-Tests across conditions of run B between the experiment and the preliminary data reported in the **Supplementary Text 2**.

Test–retest reliability

Finally, test–retest reliability was assessed by correlating the inferred parameters β, κ, s, as well as the discount factor(s) across runs A and B via Pearson correlation coefficients. Elements greater than 3 scaled median absolute deviation away from the mean were removed for these analyses to avoid spurious correlations.

Results

The experiment is divided into two runs, run A and run B, where trials of run B were generated based on models inferred on single participant behavior in run A. Run B trials were generated such as to induce nine levels of discounting probability, ranging from 0.1 to 0.9. Most of the following results therefore concentrate on analyzing the success of inducing these probabilities in run B.

In run A, we observed an average frequency of discounted choices of 54% (±16%; see **Figure 3A**). Only 4% of our sample showed less than 20% discounted choices, rendering good conditions for model parameters to converge (see also **Supplementary Tables 1, 3** and **Supplementary Text 1** for further information on effects of gender, or associations to subjective measurements and sociodemographic information).

Inferential statistics

In run B, observed discounting frequencies increased with induced frequencies on a group and individual level [group level slope: $T(7) = 13.91$, $p < 0.001$; **Figure 3B**; individual slopes: $T(49) = 16.51$, $p < 0.001$; **Figure 3C**]. On average, the offset and slope parameters obtained from the GLM came close to what was theoretically expected by the models, with an observed average offset of -0.017 [± 0.23; $T(7) = -0.51$, $p = 0.63$] and a slope of 0.80 (±0.34) (where the expected offset and slope lay at 0.07 and 0.84, see **Figure 3B** red line). Median RTs moreover followed an inverse quadratic curve [significance of inverse quadratic predictor within GLM: $T(6) = 7.41$, $p < 0.001$; **Figure 3D**] as hypothesized.

Examining test–retest-reliability, the parameters β, s, and the discount factor (evaluated at $D = 90$) were significantly correlated between runs (β: $r = 0.60$, $p < 0.001$; s 0.38, $p = 0.006$; discount factor: $r = 0.85$, $p < 0.001$; see also **Figure 3F**), but not κ ($r = 0.23$, $p = 0.21$). The lack in reliability of κ was likely due to intercorrelations between κ and s known for this model (run A: $r = -0.47$, $p < 0.001$; run B: $r = -0.4$, $p = 0.005$; see also Thome et al., 2022), which, however, do not affect reliability of the discount factors.

Behavioral homogeneity

To investigate whether the experimental framework was able to reduce variance within the induced experimental conditions, we compared variances within conditions of run B to the preliminary experiment (see **Supplementary Text 2**). All variances were either lower than or similar to those in the preliminary experiment (please also see **Figure 3G** for choice frequency distributions). Significantly lower variances were observed at frequencies 0.3 and 0.7 [0.3: $F(48{,}49) = 2.03$, $p = 0.015$, 0.7: $F(48{,}49) = 1.8$, $p = 0.044$], as well as marginally lower at 0.1, 0.2, 0.6, and 0.9 [0.1: $F(48{,}49) = 1.63$, $p = 0.091$, 0.2: $F(48{,}49) = 1.7$, $p = 0.067$; 0.6: $F(48{,}49) = 1.72$, $p = 0.062$; 0.9: $F(48{,}49) = 1.76$, $p = 0.051$]. Collectively, these results suggest that the induction protocol generated graded behavior which centered (comparatively) narrowly around model predictions.

Prediction error assessment

Corroborating these findings, we observed a comparatively low PE in run B for the applied modified hyperboloid models, that is, a low deviation between observed responses and responses predicted by the models inferred on run A (**Figure 4A** left). Statistically, the PE was lower than for the hyperbolic model ($p = 0.046$), the exponential model ($p = 0.021$), the double exponential model ($p = 0.012$), and the control model ($p < 0.001$) (and marginally lower than for the quasi-hyperbolic model; $p = 0.084$). This was largely consistent with the PE assessed on run A based on models inferred on run B (**Figure 4A** right), although here, the CS model performed significantly worse ($p < 0.001$), and the double-exponential model comparably ($p = 0.173$). These differences in the prediction ability of the evaluated models were not observed when applying the Akaike information criterion (AIC) as an in-sample error estimate of the PE (see **Figure 4B**), suggesting the AIC was less adequate to distinguish between models.

Interestingly, when examining the PE in the different experimental conditions of run B, we observed an increase in PE for higher induction frequencies (**Figure 3E**). Also, on an individual level, not all participants were best described by the hyperboloid models. In fact, we observed a wide spread over all models when counting the number of participants with lowest PE in each model (**Figure 4C**).

Application to other latent variable models

Our framework to generate an adaptive experimental design was described and evaluated here based on the special case where the behavior generating model is characterized by a time-independent latent variable model and a simple binary response variable model. We therefore briefly outline here how to proceed when transferring the proposed framework to other cognitive functions and applications in which (a) the latent variable model is history-dependent (as for instance during learning), or (b) the response variable model is multi-categorical (as in tasks with more than two response options).

History-dependent latent variable model

We will first consider the case in which the computation of values within the latent variable model depends on previous values and is thus history dependent. As a simple example, we assume to be learning values toward two stimuli u_1 and u_2 via a Rescorla–Wagner type model in which our latent variable model f_λ now describes the formation of associative memory traces (=values) in time as a function of the reward prediction error,

$$f_\lambda\left(u_i\right) := \; v_t\left(u_i\right) \; = v_{t-1}\left(u_i\right) + \lambda\left(r_t\left(u_i\right) - v_{t-1}\left(u_i\right)\right) \tag{5}$$

where λ is a learning rate parameter, $r_t\left(u_i\right)$ is a reward or outcome associated with choosing the respective stimulus $u_i, \; i = \{1, 2\}$, at time t, and $v_{t-1}(u_i)$ is its prediction (Rescorla and Wagner, 1972). Eqn. (5) comes down to a recursive relationship in time which we can expand to its initial value:

$$v_t\left(u_i\right) \; = \; (\lambda \sum_{n=0}^{t-2} \gamma^n r_{t-n}(u_i)) + \gamma^{t-1} v_1(u_i) \qquad (6),$$

$$\text{where } \gamma \; : \; = 1 - \lambda$$

For the response model, we select between these two stimuli such that we again end up with a Bernoulli process. To obtain a condition for generating adaptive trials which will induce a desired probability for selecting u_1 at time t, denoted here by p_1 (in analogy to selecting the immediate choice with probability p_{imm}), we need to insert Eqn. (6) into a sigmoid such as in Eqn. (2), and then solve for p_1. If, for simplicity, we further define $c_i := \lambda \sum_{n=1}^{t-2} \gamma^n r_{t-n}(u_i)$ (which collects the history of rewards obtained for selecting stimulus i), we obtain the following trial-generating condition for this history dependent model:

$$r_t(u_1) \; = \; \frac{log(\frac{p_1}{1-p_1})}{\lambda\beta} + r_t\left(u_2\right) + \frac{c_2 - c_1 + \gamma^{t-1} v_1\left(u_2\right) - \gamma^{t-1} v_1(u_1)}{\lambda} \tag{7}$$

Eqn. (7) also makes sense intuitively. If we consider no prior knowledge [i.e., $v_1\left(u_i\right) \; = 0$] and no reward history (i.e., $c_i \; = \; 0$) and aim at generating trials which induce equal response probabilities for both options, we need to equalize the two rewards. If, in contrast, we have a higher initial value for selecting stimulus 2, we will need to add reward to stimulus 1. Finally, if we have already observed multiple rewards, the initial values will lose and reward history (reflected in c_i) will gain importance. Such adaptive approaches may prove particularly suitable to address and control inter-individual variability in memory formation (e.g., Lonsdorf and Merz, 2017), and serve as an effective alternative (or addition) to threshold-based adaptation procedures.

Multi-categorical response model

In the second case, we consider a history-independent latent variable model coupled to a multiple choice response model. In such a case, the response probability p_k of a response $y_k, \; k \; = \; 1, ..., K$, can be modeled in terms of a softmax function as $p_k \; = \; \frac{e^{\beta v_k}}{\sum_i e^{\beta v_i}}$, and the likelihood function will now follow a multinomial distribution. If we want to generate trials which will induce predetermined probabilities p_k for response options y_k with associated values v_k, we therefore obtain the trial generating condition(s)

$$v_k \; = \; \frac{log\left(\sum_{j \neq k}^{K} e^{\beta v_j}\right)}{\beta} + \frac{log(\frac{p_k}{1-p_k})}{\beta}.$$

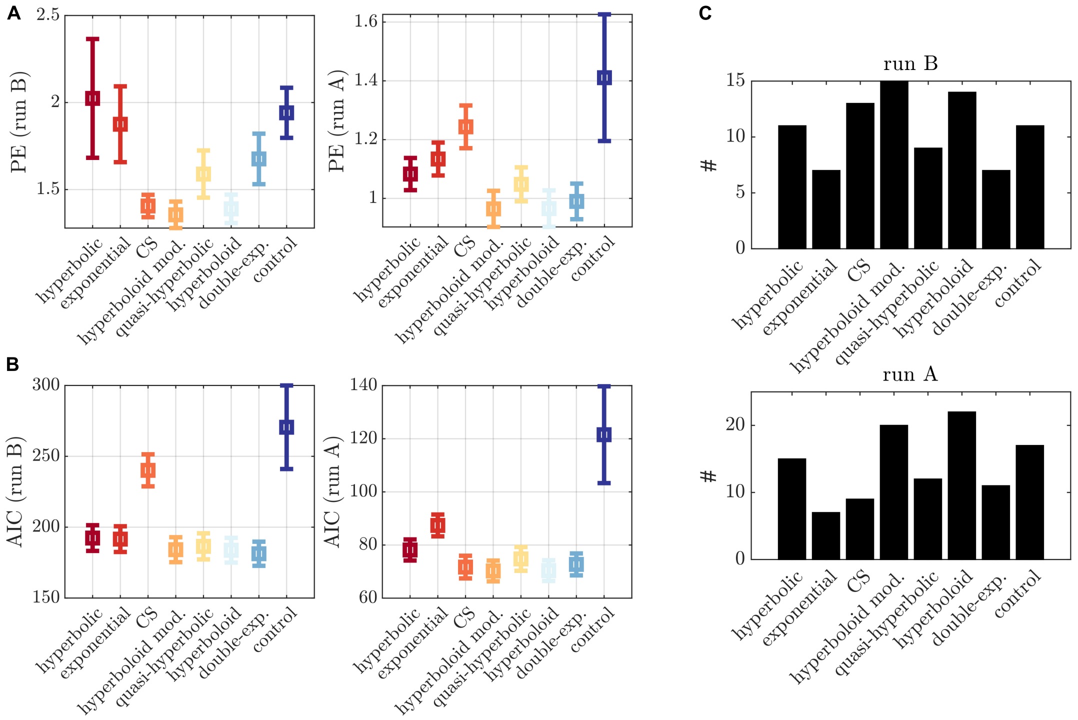


FIGURE 4
Model comparisons. **(A)** Left: PE in run B based on models inferred on run A for different models (*x*-axis). Right: PE in run A based on models inferred on run B for different models (*x*-axis). **(B)** AIC evaluated on run B (left) and run A (right). **(C)** Number of individuals (*y*-axis) with lowest PE for each model (*x*-axis) for run B (top) and run A (bottom), applying a tolerance threshold of 0.01. Cases with multiple minima were counted multiple times.

Since in this multi-categorical case, we aim at controlling the probability of all K options simultaneously, we will also need to solve these K equations simultaneously, for instance, by some form of constrained optimization.

Discussion

A general challenge in psychological and other sciences is that we want to uncover processes that are not directly observable, also termed latent processes. We draw inferences on these processes by observing behavior. In this context, experiments serve to generate conditions, that is, experimental manipulations, which differentially engage the latent process and are hypothesized to manifest in behavioral differences which allows us to study its nature in more detail. To draw an accurate inference on the underlying process based on these manipulations, we need to rely on our experiment and the latent process model being valid.

In the current work, we propose a framework which leverages interpretable probabilistic behavioral latent variable models to guide experimental manipulations and address the assessment of validity. By expressing the process in relation to the experiment and linking it probabilistically to behavior, these models allow us to generate precise and falsifiable hypotheses, i.e., predictions, and tune experimental variables to address these hypotheses (Thome et al., 2022). Predictions in this context are formulated as observable behavioral response probabilities. Assessing the deviation of predictions from out-of-sample observations facilitates the objective assessment of predictive validity. Here we apply the proposed approach to predict and induce graded choice frequencies on an individual participant level.

We illustrate the procedure in the context of delay discounting. Delay discounting is an influential psychological process, related to a variety of different traits such as impulsivity (Keidel et al., 2021), self-control (Levitt et al., 2020), intelligence (Shamosh and Gray, 2008), socio-economic status (Kohler et al., 2022), or personality (Keidel et al., 2021). It measures the tendency of an individual to devalue distant as compared to close future outcomes (Ainslie, 1975; Frederick et al., 2002; da Matta et al., 2012). Overly steep discounting has moreover been used to explain maladaptive behavior in addiction (Rabin and O'Donoghue, 1999; O'Donoghue and Rabin, 2000) and alcohol risk (Kohler et al., 2022), serving as a biomarker for the disease (Story et al., 2014; Bickel, 2015; Bailey et al., 2021; Cheng et al., 2021). Delay discounting is therefore of wide interest to both psychology and psychiatry. The general principle of the proposed framework in the context of delay discounting is that since we can infer probabilistic models which formalize how rewards are discounted as a function of delays and rewards (the cognitive process), then if we know the function (by model inference), we may determine how to vary experimental components so as to influence behavior.

We applied the proposed model-based approach to invoke discounting probabilities ranging from 0.1 to 0.9 in an (online) monetary reward discounting paradigm in a sample of healthy individuals. In line with model predictions, we observed a continuous (mostly linear) average increase in discounting frequencies coinciding with induced frequencies. Analyses of mean RT – as indirect measure of processing time – supported this notion as it followed an inverse U-function with higher RT toward trials with induced frequencies close to 0.5. This is to be expected since trials which induce equal or close to equal probabilities for immediate and delayed options are more difficult and may thus require more processing time (e.g., Ratcliff and Rouder, 1998). We also observed high test–retest reliability for the discount factor, replicating previous findings (Thome et al., 2022).

The model-based framework was successful at significantly reducing between-subject variance within several of the manipulated experimental conditions. This was observed in terms of lower behavioral variability in run B compared to a preliminary experiment with similar settings (see **Supplementary Text 2**). Low variance within experimental conditions is a prerequisite to obtain high power in associated statistical tests (e.g., Winer, 1971). In that sense, the proposed framework may also be seen as a tool which converts inter-subject variability into homogeneous 'treatment conditions,' increasing statistical power of an experimental design (Winer, 1971; Jackson, 2011; Boslaugh, 2012). At the same time, it does not eliminate important between-subject variability *per se* (Hedge et al., 2018; Goodhew and Edwards, 2019). Rather, between-subject variance is systematically relocated and captured in the (interpretable) model parameters and experimental variations. Relationships of this between-subject variance to other variables such as brain mechanisms or societal factors can therefore be explored.

The main strengths of the present framework though are the possibility to induce graded levels of behavior and to formally validate the trial-generating model and related models which reflect variations of a latent process. Assessing graded levels of behavioral probability benefits the resolution of the cognitive process at a fine-grained level. This is because behavioral probabilities reflect the intensity by which a cognitive process is engaged (in this example, the strength of temporal discounting). By studying fine gradings of behavioral probability, we may study the process on a dimensional level from low to high intensity. These intensities may be related, for instance, to neuro(physio)logical recordings to map the finely resolved latent process onto neural mechanisms (e.g., Ripke et al., 2014; Grosskopf et al., 2021, p. 20; Batsikadze et al., 2022). This may be of particular importance to psychiatry, where we aim at slowly moving away from studying psychiatric entities to stratifying patients in terms of dimensional alterations in different functional domains (RDOC; Insel et al., 2010).

The validation of the framework is realized by the implementation of two consecutive experimental runs which permit an estimation of the PE by cross-validation. We exploited this arrangement to validate the employed model by assessing its PE and comparing it to several discounting models proposed in the literature. The closely related hyperboloid models and the constant sensitivity model generated particularly low average PEs whereas the most commonly applied hyperbolic and exponential models performed comparatively poorly (in line with previous observations; Thome et al., 2022).

A modified hyperboloid (control) model with choice parameter β fixed to 1 performed particularly poorly (see **Figure 4A**). This emphasizes the importance of tuning β to the individual participant for a valid behavioral induction protocol. As outlined in the **Supplementary Text 2**, recovering β with high precision comes at the cost of increasing trial numbers. This is in line with a recent study by Pooseh et al. (2018) which performed simulation analyses to illustrate that at least 50–120 iterations are necessary for parameters to converge to their true values even when using an adaptive model-based Bayesian delay discounting framework. It challenges recent methods which infer discounting parameters in very few trials. For instance, Ahn et al. (2020) proposed a method to infer hyperbolic discounting models in less than 10 trials. While the authors demonstrate remarkably high reliability in measuring κ, they acknowledge poor reliability in β. It remains unclear how recovering models on the basis of few trials affects validity of other adaptive model-based designs. Unfortunately, most studies which have developed adaptive designs do not provide direct evidence for model validity, that is, they do not directly report predicted and actually induced response frequencies, (Monterosso et al., 2007; Cavagnaro et al., 2016; Koffarnus et al., 2017; Pooseh et al., 2018; Ahn et al., 2020), making it difficult to draw conclusions on validity of the available methods more generally.

An interesting insight of the present study is that model validity decreased particularly around hard trials, which are the main target of most other adaptive delay discounting methods (e.g., Ripke et al., 2012; Ahn et al., 2020; Knorr et al., 2020), and around larger immediate choice frequencies. One possible explanation for PE increases around hard trials is that the slope of the probability curve is steepest around hard choices (see **Figure 2A** where $v_{imm} = v_{del}$). Small biases in the inference of discounted values (e.g., due to biases in parameter estimates) have the largest effect on changes in immediate choice probabilities, possibly resulting in higher behavioral variability in these conditions. This once more emphasizes the importance of an unbiased valid recovery of model parameters.

While it remains unclear why higher frequencies were associated with a higher PE, the example shows how dissecting the PE may uncover domains at which a method is less valid. In fact, a particular advantage of the proposed adaptive approach is that it allows to systematically perturb the different factors relevant to a choice process, obtain model-based predictions for these perturbations, and then validate the model thereon. For instance, in the present example, one could analogously vary the delay period (rather than immediate reward), and – using the CV approach – formally validate a broad range of the domain the discounting function is defined on. Effectively, we can thereby improve the criterion we aim at predicting to assess predictive validity.

In sum, the model evaluation results illustrate how the framework may be leveraged to select among a set of available models delineating variations of a given process model and identify domains at which a model may fail based on out-of-sample approximations of the PE (Hastie et al., 2009; Efron, 2021; Koppe et al., 2021). In the same way it could be used to identify and validate novel models or, differentiate a given model to alternative models (implicating discriminant validity). Out-of-sample predictions are crucial for validation since in-sample error estimates (as still commonly applied) have repeatedly been shown to be strongly biased (Hurvich and Tsai, 1989; Kuha, 2004; Hastie et al., 2009). In the present context, the AIC did not discriminate well between models whereas the out-of-sample PE did.

The present study separates model inference which is always based on the same constant set of trials in each participant (run A), from model-based prediction and manipulation (performed on run B). This differs from iterative approaches – most often employed in psychophysics to identify a psychometric function – which generate successive trials online based on an underlying (often simple sigmoid) model which is assumed to be true (Leek, 2001; Shen and Richards, 2012; see also Pooseh et al., 2018). Such approaches are ill-suited to compare an applied model to related models since the model-based procedure already biases trial selection and may moreover result in unequal trials and trial numbers per participant. Biased trial selections may likely favor some models over others (Owen et al., 2021; see also pitfalls of successive procedures Leek, 2001; Shen and Richards, 2012).

Historically, psychophysics has originated in aspirations to identify objective 'laws of nature' which map physical objects to sensation, i.e., rules which are thought to apply to everyone such as the Weber–Fechner law or Steven's power law (Weber, 1835; Fechner, 1860; Stevens, 1957). Physical properties of experimental stimuli are therefore also typically directly mapped to detection or discrimination probabilities without an additional subjective transformation in between. Although latent variable models have more recently been applied to detect inter-individual differences (e.g., Taubert et al., 2012; Chakroborty et al., 2021; Owen et al., 2021), they are typically not used to generate adaptive trials (although see Thomas et al., 2021). Evaluating among a larger class of different subjective models has moreover not been of primary concern.

The latter may be specifically relevant to scientific disciplines which focus on uncovering inter-individual differences in (subjectively modulated) cognitive processes such as in the field of psychiatry for instance (e.g., Kanai and Rees, 2011). Here the focus often lies on how individuals (differentially) learn, interpret, or attribute information and how these processes may be subjectively modulated or biased (e.g., Koppe et al., 2017). In the present study, inter-individual differences are also supported by the observation of a high spread in the assignment of different discounting models to the individual participants, indicating different participants may best be described by slightly different ways of assigning subjective value to delayed outcomes (see also Cavagnaro et al., 2016).

While illustrated here on monetary delay discounting, the proposed framework may be adopted to many other contexts. Other popular and widely applied examples of interpretable latent variable models are for instance variants of reinforcement learning (RL) models which formalize the latent 'learning' process (e.g., Durstewitz et al., 2016; Sutton and Barto, 2018), and drift diffusion models which formalize latent evidence accumulation (Ratcliff, 1978). We outline here how to proceed in the case of reinforcement learning where latent variable models are history-dependent, as well as multi-categorical response models, where responses are not simply binary. To further name a few application examples in these contexts: by adapting the design to environmental stimuli, one could study the incentives at which individuals will cease to discount future environmental outcomes with a given probability (i.e., certainty). Translating the paradigm to different cognitive processes such as associative memory, one may aim at adjusting stimuli to homogenize associative memory traces or induce comparable learning speeds, which have been found to be highly heterogeneous (see e.g., Lonsdorf and Merz, 2017). Finally, in experiments of social interaction, one could even conceive of constructing artificial agents that follow individualized model-based behavioral suggestions which aim at inducing cooperative behavior in their interaction partner. This could for instance prove useful when training social skills or reducing negative biases in a therapeutic context.

Conclusion

We propose a generic framework to manipulate and validate experimental conditions based on a specific class of interpretable behavioral latent variable models. These models may be leveraged to generate precise and falsifiable behavioral predictions which may be used to evoke graded and homogeneous choice probabilities. Statistical learning theory formally defines how to assess the degree of agreement between observations and predictions and thus how indicative observations are of the latent process, sometimes referred to as predictive validity (Yarkoni and Westfall, 2017). Assessing validity in terms of PEs in this context has a number of advantages. For one, since the PE may be used to uncover domains at which an instrument may fail to be valid, it may provide insights into how an instrument or model may be improved. Also, a low PE provides evidence for the latent process model itself, as experimental manipulations follow proposed hypotheses. As illustrated earlier, this paves the way to identify novel models, delineate differences to alternative models, or improve current models by model selection. Finally, improving validity in the above mentioned sense should help us homogenize behavior between participants, as a more valid experiment will generate more precise behavioral predictions by which participants may be grouped. The proposed approach can in principle be applied with little adaptation to other cognitive domains including learning and other types of decision making, as also outlined here.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/GKoppe/data_code_repository_gradedDiscounting.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Faculty Mannheim, Heidelberg University (2019-633N). The participants provided their written informed consent to participate in this study.

Author contributions

GK and PK conceptualized the study. GK, JT, MP, PK, and WS contributed to the design of the study. MP compiled the online experiment and collected the data. GK and JT performed the statistical analyses and wrote the manuscript. DD, GK, JT, MP, PK, and WS contributed to reading, revising, and approving the submitted manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the German Research Foundation (DFG) within the collaborative research center TRR 265, subproject B08, granted to GK, PK, and WS, and by the WIN-Kolleg 8 of the Heidelberg Academy of Sciences and Humanities granted to GK.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1077735/full#supplementary-material

References

Ahn, W.-Y., Gu, H., Shen, Y., Haines, N., Hahn, H. A., Teater, J. E., et al. (2020). Rapid, precise, and reliable measurement of delay discounting using a Bayesian learning algorithm. *Sci. Rep.* 10:12091. doi: 10.1038/s41598-020-68587-x

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychol. Bull.* 82, 463–509. doi: 10.1037/h0076860

Babor, T., De La Fuente, J., Saunders, J., and Grant, M. (1992). *The alcohol use disorders identification test: Guidelines for use in primary health care*. Geneva: World Health Organization.

Bailey, A. J., Romeu, R. J., and Finn, P. R. (2021). The problems with delay discounting: A critical review of current practices and clinical applications. *Psychol. Med.* 51, 1799–1806. doi: 10.1017/S0033291721002282

Batsikadze, G., Diekmann, N., Ernst, T. M., Klein, M., Maderwald, S., Deuschl, C., et al. (2022). The cerebellum contributes to context-effects during fear extinction learning: A 7T fMRI study. *NeuroImage* 253:119080.

Białaszek, W., Marcowski, P., and Cox, D. J. (2020). Comparison of multiplicative and additive hyperbolic and hyperboloid discounting models in delayed lotteries involving gains and losses. *PLoS One* 15:e0233337. doi: 10.1371/journal.pone.0233337

Bickel, W. K. (2015). Discounting of delayed rewards as an Endophenotype. *Biol. Psychiatry* 77, 846–847. doi: 10.1016/j.biopsych.2015.03.003

Boslaugh, S. (2012). *Statistics in a nutshell*, 2nd Edn. Sebastopol, CA: O'Reilly Media.

Cavagnaro, D. R., Aranovich, G. J., McClure, S. M., Pitt, M. A., and Myung, J. I. (2016). On the functional form of temporal discounting: An optimized adaptive test. *J. Risk Uncertain.* 52, 233–254. doi: 10.1007/s11166-016-9242-y

Chakroborty, P., Pinjari, A. R., Meena, J., and Gandhi, A. (2021). A psychophysical ordered response model of time perception and service quality: Application to level of service analysis at toll plazas. *Transp. Res. B Methodol.* 154, 44–64. doi: 10.1016/j.trb.2021.09.010

Cheng, Y. S., Ko, H. C., Sun, C. K., and Yeh, P. Y. (2021). The relationship between delay discounting and Internet addiction: A systematic review and meta-analysis. *Addict. Behav.* 114:106751. doi: 10.1016/j.addbeh.2020.106751

Collins, A. G., Albrecht, M. A., Waltz, J. A., Gold, J. M., and Frank, M. J. (2017). Interactions among working memory, reinforcement learning, and effort in value-based choice: A new paradigm and selective deficits in schizophrenia. *Biol. Psychiatry* 82, 431–439. doi: 10.1016/j.biopsych.2017.05.017

Cox, D. J., and Dallery, J. (2016). Effects of delay and probability combinations on discounting in humans. *Behav. Process.* 131, 15–23. doi: 10.1016/j.beproc.2016.08.002

da Matta, A., Gonçalves, F. L., and Bizarro, L. (2012). Delay discounting: Concepts and measures. *Psychol. Neurosci.* 5, 135–146. doi: 10.3922/j.psns.2012.2.03

Dagher, A., Owen, A. M., Boecker, H., and Brooks, D. J. (1999). Mapping the network for planning: A correlational PET activation study with the Tower of London task. *Brain* 122, 1973–1987. doi: 10.1093/brain/122.10.1973

Davison, M., and McCarthy, D. (1988). *The matching law: A research review*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Durstewitz, D., Koppe, G., and Toutounji, H. (2016). Computational models as statistical tools. *Curr. Opin. Behav. Sci.* 11, 93–99. doi: 10.1016/j.cobeha.2016.07.004

Ebert, J. E., and Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of the near and far future. *Manage. Sci.* 53, 1423–1438. doi: 10.1287/mnsc.1060.0671

Efron, B. (2021). Resampling plans and the estimation of prediction error (No. 4). *Stats* 4, 1091–1115. doi: 10.3390/stats4040063

Efron, B., and Hastie, T. (2021). *Computer age statistical inference: Algorithms, evidence, and data science*, Vol. 6. Cambridge: Cambridge University Press. doi: 10.1017/9781108914062

Estle, S. J., Green, L., Myerson, J., and Holt, D. D. (2006). Differential effects of amount on temporal and probability discounting of gains and losses. *Mem. Cogn.* 34, 914–928. doi: 10.3758/BF03193437

Fechner, G. T. (1858). Über ein wichtiges psychophysiches grundgesetz und dessen beziehung zur schäzung der sterngrössen. *Abk. K. Ges. Wissensch. Math. Phys. K1* 4, 455–532.

Frederick, S., Loewenstein, G., and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *J. Econ. Lit.* 40, 351–401. doi: 10.1257/jel.40.2.351

Goodhew, S. C., and Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Conscious. Cogn.* 69, 14–25. doi: 10.1016/j.concog.2019.01.008

Green, L., Fry, A. F., and Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychol. Sci.* 5, 33–36. doi: 10.1111/j.1467-9280.1994.tb00610.x

Grinband, J., Hirsch, J., and Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron* 49, 757–763. doi: 10.1016/j.neuron.2006.01.032

Grosskopf, C. M., Kroemer, N. B., Pooseh, S., Böhme, F., and Smolka, M. N. (2021). Temporal discounting and smoking cessation: Choice consistency predicts nicotine abstinence in treatment-seeking smokers. *Psychopharmacology* 238, 399–410. doi: 10.1007/s00213-020-05688-5

Hare, T. A., Camerer, C. F., and Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324, 646–648. doi: 10.1126/science.1168450

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*, Vol. 2. Berlin: Springer. doi: 10.1007/978-0-387-84858-7

Hedge, C., Powell, G., and Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* 50, 1166–1186. doi: 10.3758/s13428-017-0935-1

Hurvich, C. M., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297–307. doi: 10.1093/biomet/76.2.297

Huys, Q. J., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* 19, 404–413.

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., and Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biol. Mood Anxiety Disord.* 3, 1–16. doi: 10.1186/2045-5380-3-12

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 167, 748–751. doi: 10.1176/appi.ajp.2010.09091379

Jackson, S. L. (2011). *Research methods and statistics: A critical thinking approach*, 4th Edn. Belmont, CA: Cengage Learning.

Kanai, R., and Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Rev. Neurosci.* 12, 231–242. doi: 10.1038/nrn3000

Keidel, K., Rramani, Q., Weber, B., Murawski, C., and Ettinger, U. (2021). Individual differences in intertemporal choice. *Front. Psychol.* 12:643670. doi: 10.3389/fpsyg.2021.643670

Knorr, F. G., Neukam, P. T., Fröhner, J. H., Mohr, H., Smolka, M. N., and Marxen, M. (2020). A comparison of fMRI and behavioral models for predicting inter-temporal choices. *NeuroImage* 211:116634. doi: 10.1016/j.neuroimage.2020.116634

Koffarnus, M. N., Deshpande, H. U., Lisinski, J. M., Eklund, A., Bickel, W. K., and LaConte, S. M. (2017). An adaptive, individualized fMRI delay discounting procedure to increase flexibility and optimize scanner time. *NeuroImage* 161, 56–66. doi: 10.1016/j.neuroimage.2017.08.024

Kohler, R. J., Lichenstein, S. D., and Yip, S. W. (2022). Hyperbolic discounting rates and risk for problematic alcohol use in youth enrolled in the Adolescent Brain and Cognitive Development study. *Addict. Biol.* 27:2. doi: 10.1111/adb.13160

Koppe, G., Mallien, A. S., Berger, S., Bartsch, D., Gass, P., Vollmayr, B., et al. (2017). CACNA1C gene regulates behavioral strategies in operant rule learning. *PLoS Biol.* 15:e2000936. doi: 10.1371/journal.pbio.2000936

Koppe, G., Meyer-Lindenberg, A., and Durstewitz, D. (2021). Deep learning for small and big data in psychiatry. *Neuropsychopharmacology* 46, 176–190. doi: 10.1038/s41386-020-0767-z

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociol. Methods Res.* 33, 188–229. doi: 10.1177/0049124103262065

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Q. J. Econ.* 112, 443–478. doi: 10.1162/003355397555253

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Percept. Psychophys.* 63, 1279–1292. doi: 10.3758/BF03194543

Levitt, E., Sanchez-Roige, S., Palmer, A. A., and MacKillop, J. (2020). Steep discounting of future rewards as an impulsivity phenotype: A concise review. *Curr. Top. Behav. Neurosci.* 47, 113–138. doi: 10.1007/7854_2020_128

Loewenstein, G., and Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Q. J. Econ.* 107, 573–597. doi: 10.2307/2118482

Lonsdorf, T. B., and Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans-Biological, experiential, temperamental factors, and methodological pitfalls. *Neurosci. Biobehav. Rev.* 80, 703–728. doi: 10.1016/j.neubiorev.2017.07.007

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. *Quant. Anal. Behav.* 5, 55–73.

McKerchar, T. L., Pickford, S., and Robertson, S. E. (2013). Hyperboloid discounting of delayed outcomes: Magnitude effects and the gain-loss asymmetry. *Psychol. Rec.* 63, 441–451. doi: 10.11133/j.tpr.2013.63.3.003

Miedl, S. F. (2012). Altered neural reward representations in pathological gamblers revealed by delay and probability discounting. *Arch. Gen. Psychiatry* 69, 177–186. doi: 10.1001/archgenpsychiatry.2011.1552

Monterosso, J. R., Ainslie, G., Xu, J., Cordova, X., Domier, C. P., and London, E. D. (2007). Frontoparietal cortical activity of methamphetamine-dependent and comparison subjects performing a delay discounting task. *Hum. Brain Mapp.* 28, 383–393. doi: 10.1002/hbm.20281

Mumford, J. A. (2012). A power calculation guide for fMRI studies. *Soc. Cogn. Affect. Neurosci.* 7, 738–742. doi: 10.1093/scan/nss059

O'Donoghue, T., and Rabin, M. (2000). The economics of immediate gratification. *J. Behav. Decis. Mak.* 13, 233–250. doi: 10.1002/(SICI)1099-0771(200004/06)13:2<233::AID-BDM325>3.0.CO;2-U

Odum, A. L., Baumann, A. A. L., and Rimington, D. D. (2006). Discounting of delayed hypothetical money and food: Effects of amount. *Behav. Process.* 73, 278–284. doi: 10.1016/j.beproc.2006.06.008

Owen, L., Browder, J., Letham, B., Stocek, G., Tymms, C., and Shvartsman, M. (2021). Adaptive nonparametric psychophysics. *arXiv* [Preprint]. Available online at: http://arxiv.org/abs/2104.09549 (accessed April 19, 2021).

Peters, J., Miedl, S. F., and Büchel, C. (2012). Formal comparison of dual-parameter temporal discounting models in controls and pathological gamblers. *PLoS One* 7:e47225. doi: 10.1371/journal.pone.0047225

Phelps, E., and Pollak, R. (1968). On second-best national saving and game-equilibrium growth. *Rev. Econ. Stud.* 35, 185–199. doi: 10.2307/2296547

Pine, A., Seymour, B., Roiser, J. P., Bossaerts, P., Friston, K. J., Curran, H. V., et al. (2009). Encoding of marginal utility across time in the human brain. *J. Neurosci.* 29, 9575–9581. doi: 10.1523/JNEUROSCI.1126-09.2009

Pooseh, S., Bernhardt, N., Guevara, A., Huys, Q. J. M., and Smolka, M. N. (2018). Value-based decision-making battery: A Bayesian adaptive approach to assess impulsive and risky behavior. *Behav. Res. Methods* 50, 236–249. doi: 10.3758/s13428-017-0866-x

Prevost, C., Pessiglione, M., Metereau, E., Clery-Melin, M.-L., and Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *J. Neurosci.* 30, 14080–14090. doi: 10.1523/JNEUROSCI.2752-10.2010

Rabin, M., and O'Donoghue, T. (1999). Doing it now or later. *American* 89, 103–124. doi: 10.1257/aer.89.1.103

Rachlin, H. (2006). Notes on discounting. *J. Exp. Anal. Behav.* 85, 425–435. doi: 10.1901/jeab.2006.85-05

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R., and Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychol. Sci.* 9, 347–356. doi: 10.1111/1467-9280.00067

Rescorla, R. A., and Wagner, A. R. (1972). "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and Nonreinforcement," in *Classical conditioning II: Current research and theory*, eds A. H. Black and W. F. Prokasy (New York, NY: Appleton- Century-Crofts), 64–99.

Ripke, S., Hübner, T., Mennigen, E., Müller, K. U., Li, S.-C., and Smolka, M. N. (2014). Common neural correlates of intertemporal choices and intelligence in adolescents. *J. Cogn. Neurosci.* 27, 387–399. doi: 10.1162/jocn_a_00698

Ripke, S., Hübner, T., Mennigen, E., Müller, K. U., Rodehacke, S., Schmidt, D., et al. (2012). Reward processing and intertemporal decision making in adults and adolescents: The role of impulsivity and decision consistency. *Brain Res.* 1478, 36–47. doi: 10.1016/j.brainres.2012.08.034

Rodzon, K., Berry, M. S., and Odum, A. L. (2011). Within-subject comparison of degree of delay discounting using titrating and fixed sequence procedures. *Behav. Process.* 86, 164–167. doi: 10.1016/j.beproc.2010.09.007

Samuelson, P. (1937). A note on measurement of utility. *Rev. Econ. Stud.* 4, 155–161. doi: 10.2307/2967612

Schmidt-Atzert, L., Stefan, K., and Amelang, M. (2021). "Grundlagen diagnostischer verfahren," in *Psychologische diagnostik* (Berlin: Springer), 41–201.

Shamosh, N. A., and Gray, J. R. (2008). Delay discounting and intelligence: A meta-analysis. *Intelligence* 36, 289–305. doi: 10.1016/j.intell.2007.09.004

Shen, Y., and Richards, V. M. (2012). A maximum-likelihood procedure for estimating psychometric functions: Thresholds, slopes, and lapses of attention. *J. Acoust. Soc. Am.* 132, 957–967. doi: 10.1121/1.4733540

Spinella, M. (2007). Normative data and a short form of the Barratt Impulsiveness Scale. *Int. J. Neurosci.* 117, 359–368.

Stevens, S. (1957). On the psychophysical law. *Psychol. Rev.* 64, 153–181. doi: 10.1037/h0046162

Story, G. W., Vlaev, I., Seymour, B., Darzi, A., and Dolan, R. J. (2014). Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Front. Behav. Neurosci.* 8:76. doi: 10.3389/fnbeh.2014.00076

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Taubert, N., Christensen, A., Endres, D., and Giese, M. A. (2012). "Online simulation of emotional interactive behaviors with hierarchical Gaussian process dynamical models," in *Proceedings of the 2012 ACM symposium on applied perception*, Los Angeles, CA, 25–32. doi: 10.1145/2338676.2338682

Thomas, M. L., Brown, G. G., Patt, V. M., and Duffy, J. R. (2021). Latent variable modeling and adaptive testing for experimental cognitive psychopathology research. *Educ. Psychol. Meas.* 81, 155–181.

Thome, J., Pinger, M., Halli, P., Durstewitz, D., Sommer, W. H., Kirsch, P., et al. (2022). A model guided approach to evoke homogeneous behavior during temporal reward and loss discounting. *Front. Psychiatry* 13:846119. doi: 10.3389/fpsyt.2022.846119

van den Bos, W., and McClure, S. M. (2013). Towards a general model of temporal discounting. *J. Exp. Anal. Behav.* 99, 58–73. doi: 10.1002/jeab.6

Weber, E. H. (1835). *De pulsu, resorptione, auditu et tactu annotationes anatomica et physiologica*. (Charleston, SC: Nabu Press), 152.

Wichmann, F. A., and Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313. doi: 10.3758/BF03194544

Winer, B. J. (1971). *Statistical principles in experimental design*, 2nd Edn. New York, NY: McGraw-Hill.

Wood, G., Nuerk, H.-C., Sturm, D., and Willmes, K. (2008). Using parametric regressors to disentangle properties of multi-feature processes. *Behav. Brain Funct.* 4:38. doi: 10.1186/1744-9081-4-38

Yarkoni, T., and Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.* 12, 1100–1122. doi: 10.1177/1745691617693393

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fnins.2022.1060814

An objective model for diagnosing comorbid cognitive impairment in patients with epilepsy based on the clinical-EEG functional connectivity features

Zhe Ren[1†‡], Yibo Zhao[1†‡], Xiong Han[2*†], Mengyan Yue[3†], Bin Wang[2], Zongya Zhao[4], Bin Wen[5], Yang Hong[6], Qi Wang[1], Yingxing Hong[6], Ting Zhao[2], Na Wang[2] and Pan Zhao[2]

[1]Department of Neurology, Zhengzhou University People's Hospital, Zhengzhou, Henan, China, [2]Department of Neurology, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou, Henan, China, [3]Department of Rehabilitation, The First Hospital of Shanxi Medical University, Taiyuan, Shanxi, China, [4]School of Medical Engineering, Xinxiang Medical University, Xinxiang, Henan, China, [5]School of Life Sciences and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China, [6]Department of Neurology, People's Hospital of Henan University, Zhengzhou, Henan, China

OPEN ACCESS

EDITED BY
Reza Rastmanesh,
The Nutrition Society, United Kingdom

REVIEWED BY
Gian Marco Duma,
Eugenio Medea (IRCCS), Italy
Aleksandar Gavrilovic,
University of Kragujevac, Serbia

*CORRESPONDENCE
Xiong Han
✉ hanxiong@zzu.edu.cn

†These authors have contributed equally to this work

‡These authors share first authorship

SPECIALTY SECTION
This article was submitted to Translational Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 15 October 2022
ACCEPTED 28 December 2022
PUBLISHED 12 January 2023

CITATION
Ren Z, Zhao Y, Han X, Yue M, Wang B, Zhao Z, Wen B, Hong Y, Wang Q, Hong Y, Zhao T, Wang N and Zhao P (2023) An objective model for diagnosing comorbid cognitive impairment in patients with epilepsy based on the clinical-EEG functional connectivity features. *Front. Neurosci.* 16:1060814. doi: 10.3389/fnins.2022.1060814

Objective: Cognitive impairment (CI) is a common disorder in patients with epilepsy (PWEs). Objective assessment method for diagnosing CI in PWEs would be beneficial in reality. This study proposed to construct a diagnostic model for CI in PWEs using the clinical and the phase locking value (PLV) functional connectivity features of the electroencephalogram (EEG).

Methods: PWEs who met the inclusion and exclusion criteria were divided into a cognitively normal (CON) group (*n* = 55) and a CI group (*n* = 76). The 23 clinical features and 684 PLV_{EEG} features at the time of patient visit were screened and ranked using the Fisher score. Adaptive Boosting (AdaBoost) and Gradient Boosting Decision Tree (GBDT) were used as algorithms to construct diagnostic models of CI in PWEs either with pure clinical features, pure PLV_{EEG} features, or combined clinical and PLV_{EEG} features. The performance of these models was assessed using a five-fold cross-validation method.

Results: GBDT-built model with combined clinical and PLV_{EEG} features performed the best with accuracy, precision, recall, F1-score, and an area under the curve (AUC) of 90.11, 93.40, 89.50, 91.39, and 0.95%. The top 5 features found to influence the model performance based on the Fisher scores were the magnetic resonance imaging (MRI) findings of the head for abnormalities, educational attainment, PLV_{EEG} in the beta (β)-band C3-F4, seizure frequency, and PLV_{EEG} in theta (θ)-band Fp1-Fz. A total of 12 of the top 5% of features exhibited statistically different PLV_{EEG} features, while eight of which were PLV_{EEG} features in the θ band.

Conclusion: The model constructed from the combined clinical and PLV_{EEG} features could effectively identify CI in PWEs and possess the potential as a useful objective evaluation method. The PLV_{EEG} in the θ band could be a potential biomarker for the complementary diagnosis of CI comorbid with epilepsy.

KEYWORDS

epilepsy, cognitive impairment, EEG, phase locking value, GBDT, AdaBoost, diagnostic model, Fisher score

Introduction

Cognitive impairment (CI) is one of the very common comorbidities occurring in 70–80% of patients with epilepsy (PWEs) (Helmstaedter and Witt, 2017). Previous studies have revealed several factors that may induce CI in PWEs, including age at onset, duration of illness, surgical head trauma, perinatal injury, temporal lobe epilepsy, hippocampal abnormalities, seizures, status epilepticus, medications, and psychiatric factors (Black et al., 2010; Titiz et al., 2014; Vrinda et al., 2019; Jarcuskova et al., 2020; Wang et al., 2020; Novak et al., 2022a). Furthermore, interictal epileptiform discharges (IEDs) in electroencephalogram (EEG) recordings are an important indicator of CI in PWEs (Ung et al., 2017; Gavrilovic et al., 2019; Balcik et al., 2020), but the exact role of EEG in diagnosing CI in such patients has rarely been studied.

Cognitive scales serve as the primary method for diagnosing CI, with the Montreal Cognitive Assessment (MoCA) scale considered the most appropriate and more sensitive than the Mini-Mental State Examination (MMSE) scale for screening cognitive impairment in epileptic individuals (Montano-Lozada et al., 2021; Huang et al., 2022; Novak et al., 2022b). Notably, the MoCA-30 point scale is superior to the MoCA-20 scale for CI assessment in clinical practices (Bergeron et al., 2017; Del Brutto et al., 2019; Rodrigues et al., 2020; Melikyan et al., 2021). However, the scale has some shortcomings, most notably its susceptibility to subjective factors from both patients and physicians, which may lead to errors in the test. Although the MoCA scale is well suited to screening for CI in epileptic patients, however, it is a generic neurological screening tool for cognitive assessments. Therefore, there is an urgent need for developing an efficient objective assessment indicator for cognitive functions, specifically for individuals with epileptic symptoms.

Electroencephalogram plays a vital role in the diagnosis and management of epilepsy, as it provides an objective and accurate response to functional changes in the brain, thus avoiding the influence of subjective factors in the patient. A growing body of research has demonstrated a strong correlation between altered cognitive functions and the neural connectivity of different brain regions (He et al., 2018; Fadaie et al., 2021; Duma et al., 2022). Functional connectivity is a type of neural connectivity that mediates the temporal correlation between neurophysiological events at different brain regions and is primarily used to measure the degree of dependency and correlation between the signals. The phase locking value (PLV) is one of the quantitative indicators for functional connectivity (Elahian et al., 2017; Duma et al., 2021). Furthermore, EEG-based functional connectivity is employed to predict vagus nerve stimulation (VNS) responsiveness in children with refractory epilepsies (Ma et al., 2022), as well as to diagnose CI in patients comorbid with Parkinson's disease (PD) (Cai et al., 2021). However, this approach has not been applied to the diagnosis of cognitive dysfunctions in PWEs. The Adaptive Boosting (AdaBoost) and Gradient-Boosted Decision Trees (GBDT) are classic algorithms for ensemble learning (EL) and have been widely used in areas of neurologic disorders such as epilepsy, Alzheimer's disease (AD), PD, etc. (Peng et al., 2020; Wenbo et al., 2021; Zhang S. et al., 2021; Edeh et al., 2022). These follow the models constructed based on the clinical and PLV_{EEG} functional connectivity features of EL algorithms and have shown the potential of an efficient objective evaluation tool for diagnosing CI in PWEs.

Here, we used EL algorithms to construct three distinct models for the diagnosis of CI in PWEs, purely based on the clinical and PLV_{EEG} features. Additionally, we investigated to

Abbreviations: PWEs, patients with epilepsy; MoCA, Montreal Cognitive Assessment; CI, cognitive impairment; CON, cognitively normal; EEG, electroencephalogram; PLV, phase locking value; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; MRI, magnetic resonance imaging; AUC, area under the curve.

identify potential biomarkers for the diagnosis of cognitive functions in PWEs.

Materials and methods

Selection of the participants

A total of 131 PWEs from the outpatient clinic of the Epilepsy Center of Henan Provincial People's Hospital between June 2018 and May 2022 were retrospectively screened and enrolled in the study. The inclusion criteria were: (1) the patient must meet the criteria of the International League Against Epilepsy (ILAE) for the diagnosis of epilepsy, seizures, and other epileptic syndromes (Fisher et al., 2014); (2) the age range at the time of consultation must be 12–60 years; (3) the patient must had a MoCA test at the time of consultation and should not have any history of MoCA scale testing in the last year; (4) at least 20 min of outpatient scalp EEG at the time of consultation, along with the availability of retrospective EEG data; and (5) the patient must have a complete clinical history and previous cranial MRI findings. Subjects were excluded if: (1) the patient's age was less than 12 years or more than 60 years at the time of consultation; (2) the patient was diagnosed with psychogenic non-epileptic seizures, or epilepsy syndrome; (3) the patient was treated with drugs other than antiseizures medications that affect cognitive functions, such as benzodiazepines, anti-psychotics, and memory-enhancing drugs, at the time of consultation; and (4) the patient was missing the 20-min EEG recording data at the time of the enrollment.

Based on the patients' MoCA scores during their visits to the epilepsy clinic, 131 PWEs were recruited for the study and were subsequently divided into the control (CON) group (MoCA $\geq$ 26; n = 55) and the CI group (MoCA $<$ 26; n = 76) (**Figure 1** and **Table 1**). The study was approved by the Ethics Committee of Henan Provincial People's Hospital and all eligible subjects signed the written informed consent before their final recruitment to the study.

Clinical features

Based on the patients' medical history and clinical investigations at the time of the current visit to the epilepsy clinic, 23 clinical features were identified, in conjunction with previous studies: (1) age; (2) age at the first onset; (3) time from the first onset to current visit (Black et al., 2010); (4) gender; (5) family history of epilepsy (defined as whether a first or second degree relative had epilepsy); (6) history of previous head surgery or trauma; (7) history of previous the central nervous system (CNS) infections; (8) history of perinatal injuries due to premature birth, obstructed labor, hypoxia, and/or intracranial hematoma; (9) TLE; (10) MRI of the head for abnormalities; (11) hippocampal atrophy, or sclerosis (Titiz et al., 2014); (12) different types of seizures like generalized, focal, or both; (13) status epilepticus; (14) generalized tonic-clonic seizures (GTCS); (15) seizure frequency in the last year (Wang et al., 2020) (rare: $\leq$1 event; occasional: 2–3 events; frequent: $\geq$4 events); (16) class of antiseizures medications (Wang et al., 2020); (17) valproate (VPA) therapy in the last year; (18) phenytoin (PHT) therapy in the last year; (19) topiramate (TPM) therapy in the last year; (20) aura of epilepsy; (21) anxiety [according to the Hamilton Anxiety Inventory (HAI) scale rating: none, possible, definitely, or definitely obvious]; (22) depression [according to the Hamilton Depression Inventory (HDI) scale rating: none, possible, or definite]; and (23) educational attainment ($\leq$6 years, 7–9 years, 10–12 years, or $\geq$13 years) (**Table 2**).

EEG acquisition and preprocessing

All patients in both CON and CI groups had scalp EEG recordings monitored for at least 20 min during this visit. All tests were performed in the awake closed-eye state, while EEG recordings performed during the sleep and awake open-eye states were excluded. The EEG-1200°C machine (Nihon Kohden, Tokyo, Japan), with a sampling frequency of 256 Hz, an amplification multiplier of 1000×, a low-pass filter of 70 Hz, and a high-pass filter of 0.5 Hz, was used for this study. This system uniformly used the international 10–20 lead system for placing the scalp electrodes, including 19 recording leads, namely Fp1, Fp2, Fz, Cz, Pz, C3, C4, T3, T4, T5, T6, F3, F4, F7, F8, O1, O2, P3, and P4, and 2 reference leads A1 and A2.

Preprocessing of EEG data was performed using the EEGLAB toolbox in MATLAB software (Mathworks Inc., USA) (Delorme and Makeig, 2004). Briefly, the EEG recordings were first filtered to extract only the 0.5–30 Hz recordings. Afterward, the artifacts of eye movements in electromyogram (EMG) were removed using independent component analysis. Finally, the 20-min EEG recording of each patient was intercepted into 6 s segments, and PLV_{EEG} features were extracted.

Parameters setting for AdaBoost and GBDT

AdaBoost and GBDT are typical methods of boosting algorithm. In the AdaBoost model, the number and learning rate of base classifiers were also determined by grid search, ranging from 50 to 150 and 0 to 1, respectively and the algorithm of AdaBoost set to SAMME.R. The base classifier of AdaBoost was SVM, the kernel was RBF and the C and gamma of which were also determined by grid search, ranging from 2^{-10} to

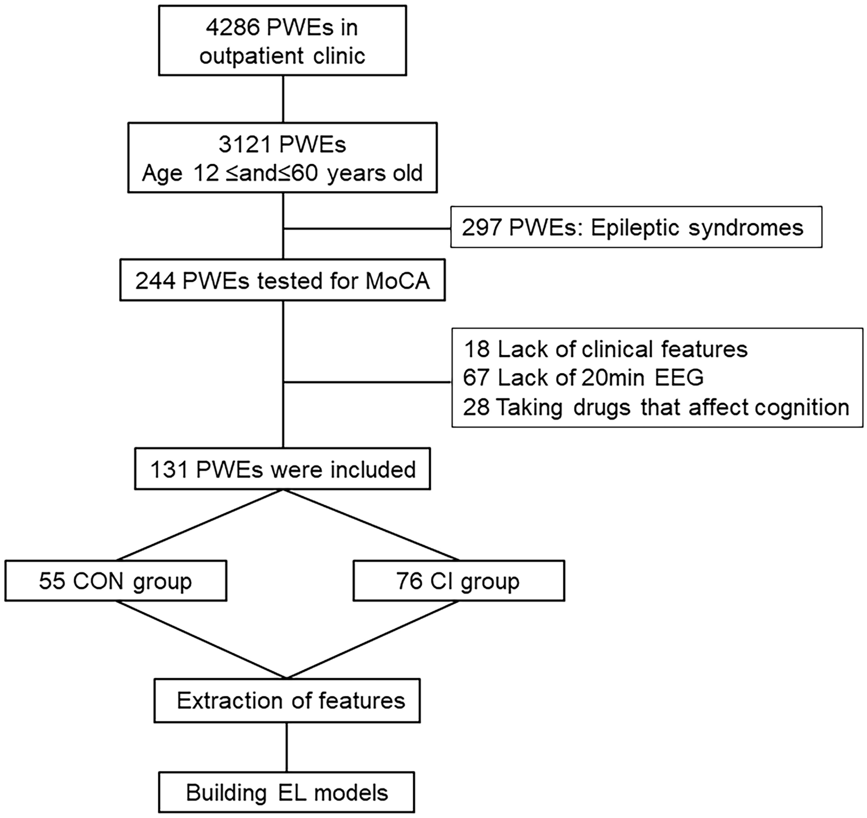

4286 PWEs in outpatient clinic
3121 PWEs Age 12 ≤and≤60 years old
297 PWEs: Epileptic syndromes
244 PWEs tested for MoCA
18 Lack of clinical features
67 Lack of 20min EEG
28 Taking drugs that affect cognition
131 PWEs were included
55 CON group
76 CI group
Extraction of features
Building EL models

FIGURE 1
Flow chart. PWEs, patients with epilepsy; MoCA, Montreal Cognitive Assessment; EEG, electroencephalogram; CI, cognitive impairment; CON, cognitively normal; EL, ensemble learning.

TABLE 1 Types of epilepsy in patients with epilepsy used in the study.

Epil. type	Unitemp.	Bitemp.	Par.	Occ.	Central	Front.	Undetermined
CON (*n* = 55)	25	2	0	9	8	2	9
CI (*n* = 76)	44	4	3	7	7	3	8

Epil. type, epilepsy type; Unitemp, unitemporal; Bitemp, bitemporal; Par, parietal; Occ, occipital; Front, frontal.

2^{10} and 0.0001 to 10, respectively. Other parameters were set to default values. In the GBDT model, the number, learning rate, and subsample of base classifiers were also determined by grid search, ranging from 50 to 150, 0 to 1 and 0.5 to 0.8, respectively. The base classifier of GBDT was CART, the max depth and the max leaf nodes of which were also determined by grid, search ranging from 10 to 15 and 10 to 30, respectively. Other parameters were set to default values. In order to reduce the contingency and improve the generalization ability, the five-fold cross-validation method was used to evaluate the performance of the model and select the best model. All of the above algorithms were programmed and realized by sklearn in PyCharm IDE using Python 3.7. The computer system is windows 10 professional, the CPU is Inter Core i7-10700K Processor @3.9 GHz, and the RAM is 32 GB. The final parameters of the model are shown in **Table 3**.

PLV-based functional connectivity features

Phase locking value is a type of connection characteristic, which quantifies the degree of phase synchronization between the two EEG signals (Aydore et al., 2013; Leguia et al., 2021). The Hilbert transform was first applied to the preprocessed EEG data to calculate the instantaneous amplitude and instantaneous phase for each lead site. The PLV indicator was then calculated using the following formula:

$$PLV_t = \frac{1}{N}\left|\sum_{n=1}^{N} exp\left(j\theta\left(t, n\right)\right)\right|$$

Where N denoted the number of EEG segments per subject, $\theta(t, n)$ presented the instantaneous phase difference

TABLE 2 Demographic information and clinical characteristics.

Clinical features	CON group (*n* = 55)	CI group (*n* = 76)	*P*-value
Age. y, mean ± SD	26.38 ± 10.49	31.34 ± 13.93	0.061
Age at first onset. y, mean ± SD	18.76 ± 11.02	20.71 ± 14.74	0.788
Time from first onset to current visit. y, mean ± SD	7.44 ± 7.79	10.63 ± 8.14	0.009*
Female	24	39	0.385
Family history of epilepsy. Y, *n*	2	5	0.730
History of previous head surgery or trauma. Y, *n*	6	17	0.089
History of previous CNS infections. Y, *n*	8	18	0.196
History of perinatal injury. Y, *n*	4	8	0.741
TLE. Y, *n*	27	48	0.108
MRI of the head for abnormalities. Y, *n*	28	51	0.061
Hippocampal atrophy, sclerosis. Y, *n*	14	37	0.004*
Seizure type, *n*			0.875
Generalized	13	21	
Focal	7	9	
Both	35	46	
Status epilepticus. Y, *n*	4	15	0.080
GTCS. Y, *n*	45	67	0.309
Seizure frequency, *n*			0.006*
Rare	17	12	
Occasionally	15	11	
Frequent	23	53	
Class of antiepileptic drugs ≥2. Y, *n*	18	41	0.016*
VPA. Y, *n*	17	40	0.013*
PTH. Y, *n*	1	2	1.000
TPM. Y, *n*	3	4	1.000
Aura of epilepsy. Y, *n*	22	24	0.319
Anxiety, *n*			0.444
None	14	12	
Possible	13	21	
Definitely	25	35	
Definitely obvious	3	8	
Depression, *n*			0.555
None	23	25	
Possible	31	48	
Definitely	1	3	
Educational attainment, *n*			<0.001*
≤6 y	1	23	
7–9 y	11	18	
10–12 y	15	19	
≥13 y	28	16	

y, year; Y, yes; CNS, central nervous system; TLE, temporal lobe epilepsy; MRI, magnetic resonance imaging; GTCS, generalized tonic-clonic seizures; VPA, valproate; PHT, phenytoin, TPM, topiramate.

$P < 0.05$ is considered as statistically significant. *The features that have statistically significance. For continuous variables, independent-samples *t*-test or Mann–Whitney *U*-test was carried out. For categorical variables, chi-square test or Fisher's exact test were carried out.

TABLE 3 The parameters of the models.

AdaBoost	Value	GBDT	Value
Clinical feature-based model			
Base_estimator	SVC	Base_estimator	CART
N_estimators	60	N_estimators	90
Learning_rate	0.2	Learning_rate	0.5
C	1024	Subsample	0.8
Gamma	0.0025	Max_depth	8
Kernel	RBF	Max_leaf_nodes	15
PLV$_{EEG}$ feature-based model			
base_estimator	SVC	Base_estimator	CART
N_estimators	100	N_estimators	90
Learning_rate	0.1	Learning_rate	0.2
C	256	Subsample	0.7
Gamma	0.25	Max_depth	10
Kernel	RBF	Max_leaf_nodes	13
Combined clinical-PLV$_{EEG}$ feature-based model			
Base_estimator	SVC	Base_estimator	CART
N_estimators	80	N_estimators	110
Learning_rate	0.3	Learning_rate	0.3
C	64	Subsample	0.7
Gamma	0.0125	Max_depth	12
Kernel	RBF	Max_leaf_nodes	15

between different leads of the same segment, $exp\left(j\theta\left(t, n\right)\right)$ represented the complex signal obtained with the help of Euler's formula using phase, and $\sum_{n\ 1}^{N} exp\left(j\theta\left(t, n\right)\right)$ represented the superimposed value of the complex signals of all segments of a patient, which was averaged to obtain the PLV feature value of a subject.

The PLV feature was then quantized into a value in the range [0,1]. When PLV = 1, the phase difference between the two signals was constant, i.e., perfectly synchronized. When PLV = 0, the phase difference was uniformly distributed over the complex plane unit circle according to time, indicating that there was no synchronization. Between 0 and 1, the signal difference exhibited an "overall convergence" nature, such that as PLV tended to 1, two close signals exhibited better synchronization.

Since it would be more accurate to calculate the instantaneous phase of narrowband signals using the Hilbert transform, the preprocessed EEG segments were divided into four narrow bands according to different frequency ranges, namely delta (δ) (1–4 Hz), θ (4–7), alpha (α) (8–13 Hz), and β (14–30 Hz) bands. The PLV$_{EEG}$ values of these four frequency bands were calculated separately for 200 windows (6 s) of each subject's 20-min EEG recording. Finally, 200 PLV$_{EEG}$ feature matrices of 19 × 19 in each of the four frequency bands were obtained for each subject and averaged into a single matrix for each frequeny band, so that each subject ended up with a total of four feature matrices for four frequency bands. These PLV$_{EEG}$ feature matrices would be further filtered and sorted characterized (**Figure 2**).

Feature extraction

As shown earlier, 23 clinical features were selected based on the previous studies and contents of available medical records. The EEG records of all subjects were divided into four different frequency bands. For each subject's 200 6 s segments in any of the frequency bands, 19 leads were paired as two by two, and a 19 × 19 PLV$_{EEG}$ functional connectivity matrix was calculated for each segment's EEG, excluding duplicate PLV$_{EEG}$ features that made comparisons with the leads themselves, to obtain a total of 171 PLV$_{EEG}$ features for the EEG recordings of a given subject. The PLV$_{EEG}$ features from 80 segments were then averaged. A total of 707 clinical-PLV$_{EEG}$ features, including 684 PLV$_{EEG}$ and 23 clinical features, were obtained in the four frequency bands for each subject. However, it was unknown which features were valid for a particular learning algorithm, and for this reason, we needed to filter all the features to select those that were beneficial to the learning algorithm. Filtering features not only optimized the algorithm to make the model more generalized but also reduced the running time of the algorithm resolving overfitting issues and the difficulty of the learning task, thereby improving the efficiency and the interpretability of the model.

Fisher score is a common feature filtering method (Zhang J. et al., 2021). Features with a strong discriminatory performance exhibit the smallest possible intra-class distance and the largest possible inter-class distance. The higher the inter-class variance and the lower the intra-class variance of PLV$_{EEG}$ features in the same frequency band from different patients, the higher the Fisher score value is. We ranked the features from the largest to the smallest, based on their Fisher score values, with the higher ranked features being theoretically more discriminative.

Modeling process

The classification models were trained using AdaBoost and GBDT platforms as classifiers. Models were constructed based on the pure clinical features, PLV$_{EEG}$ features, and combined clinical- PLV$_{EEG}$ features, as well. To improve the classification performance, generalization skills, and speed of each model, Fisher scores were used to filter the features. Five-fold cross-validation was used to construct the classification model, using 80% of the two sets of data each time, and the remaining 20% of the data was used for model validation.

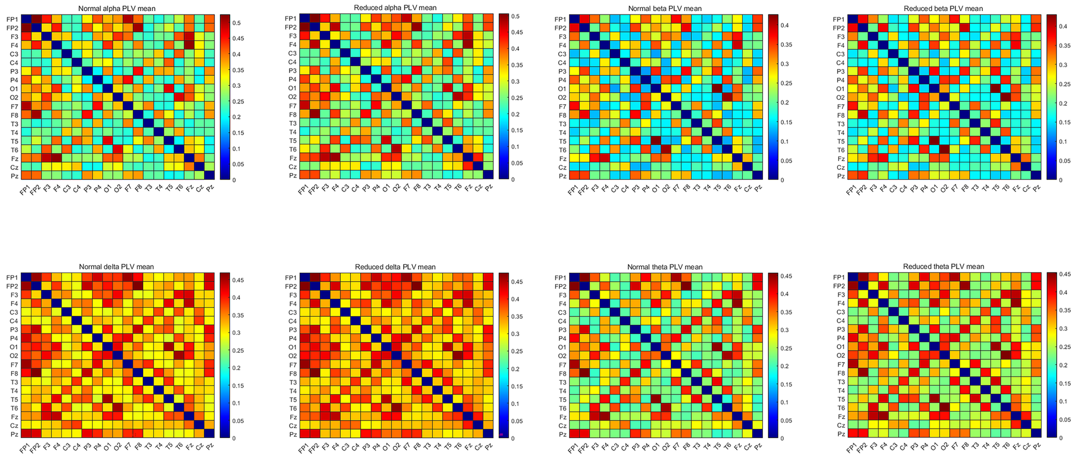


FIGURE 2
Mean PLV$_{EEG}$ features in the four frequency bands for the CON and the CI groups of PWEs. PWEs, patients with epilepsy; PLV, phase locking value; CI, cognitive impairment; CON, cognitively normal.

Statistical analysis

To compare the variability of clinical and normalized PLV$_{EEG}$ features between the CON and CI groups, the quantitative data were first tested for normality using the Shapiro–Wilk test, followed by a comparison of the data with a normal distribution expressed as mean ± standard deviation (SD) using the independent samples *t*-test, and the Mann–Whitney *U*-test was applied for data with an abnormal distribution expressed as median ± interquartile range (IQR). For qualitative information, the chi-squared (χ^2) test or Fisher's exact test was used to assess the variability between the two data sets. A *p*- or *p*'- value of < 0.05 was considered statistically significant, where *p*' referred to a *p*-value that was corrected by the false discovery rate (FDR) correction. We used SPSS v26.0 for all kinds of statistical analyses.

Results

Clinical feature-based model construction

Of the 23 clinical features, we used Fisher scores to filter the top 15 clinical features in terms of weightage to construct the diagnostic model (Table 4A). The selected features were educational attainment, seizure frequency, VPA, class of antiseizures medications, hippocampal atrophy and sclerosis, age, status epilepticus, MRI of the head for abnormalities, time from the first onset to the current visit, history of previous CNS infections, TLE, anxiety, age at the first onset, history of previous head surgeries or trauma, and gender. The features that showed significant statistical differences between the two groups were educational attainment, seizure frequency, VPA, class of antiseizures medications, hippocampal atrophy and sclerosis, and time from the first onset to the current visit. In the classification model, constructed based on the clinical features using AdaBoost, the model performances after a five-fold cross-validation for accuracy, precision, recall, F1-score, and AUC were 67.89, 66.69, 91.57, 76.71, and 0.75%, respectively. While, in case of the classification model built by GBDT, the final performances after cross-validation for accuracy, precision, recall, F1-score, and AUC were, respectively, 68.09, 70.80, 75.84, 72.62, and 0.76% (Figure 3 and Figure 4A). Therefore, these two algorithms were found to differ slightly in the construction of a model for identifying impaired consciousness in epilepsy patients using the clinical features only.

PLV$_{EEG}$ feature-based model construction

A total of 171 PLV$_{EEG}$ features were extracted for each of the 4 bands of the 20-min EEG recording for each patient, accounting for a total of 684 features (Table 4B). Then the model was constructed using those features with Fisher scores in the top 150 ranks. In the AdaBoost-based classification model, the model performance after a five-fold cross-validation for accuracy, precision, recall, F1-score, and AUC were 83.93, 84.76, 88.08, 86.30, and 0.91%, respectively. Likewise, for the GBDT-based classification model, the final performances after the cross-validation for accuracy, precision, recall, F1-score, and AUC were 88.58, 92.17, 88.17, 90.05, and 0.94%, respectively (Figure 3 and Figure 4B). Importantly, the GBDT was found to outperform AdaBoost in classification model construction using PLV$_{EEG}$ features, demonstrating that the GBDT-based model

TABLE 4 Ranking table of features affecting the model performance.

Rank	Clinic feature	FS-value	Rank	Clinic feature	FS-value
(A) Top 15 features affecting the pure clinical feature-based model.					
1	Educational attainment	0.2092	9	Time from first onset to current visit	0.0257
2	Seizure frequency	0.1037	10	History of previous CNS infections	0.0254
3	VPA	0.1033	11	TLE	0.0172
4	Class of antiepileptic drugs	0.0673	12	Anxiety	0.0134
5	Hippocampal atrophy, sclerosis	0.0558	13	Age at first onset	0.0128
6	Age	0.0453	14	History of previous head surgery or trauma	0.0118
7	Status epilepticus	0.038	15	Gender	0.0108
8	MRI of the head for abnormalities	0.0268			
Rank	**EEG feature**	**FS-value**	**Rank**	**EEG feature**	**FS-value**
(B) Top 20 features affecting pure PLV_{EEG}- based feature model.					
1	θ_T5-T6	0.1191	11	θ_F4-F7	0.0816
2	θ_Fp1-Pz	0.1082	12	θ_Fp2-T6	0.0815
3	δ_Fp1-Pz	0.1076	13	δ_F4-F7	0.0793
4	β_P3-F4	0.1003	14	α_Fp2-T4	0.079
5	β_C3-F4	0.0911	15	θ_P3-F8	0.079
6	α_Fp1-F8	0.0907	16	β_Fp1-F8	0.0787
7	β_F4-F7	0.0848	17	θ_P3-F4	0.0785
8	α_P3-T4	0.0829	18	θ_Fp1-F8	0.078
9	θ_P3-C4	0.0826	19	α_O2-C3	0.0764
10	α_Fp1-F7	0.082	20	β_Fp1-F3	0.0737
Rank	**Features**	**FS-value**	**Mean ± STD**	***P*-value**	***P'*-value**
(C) Features affecting the top 5% of the clinical-PLV_{EEG} feature-based model.					
1	MRI of the head for abnormalities	0.211	0.557 ± 0.497	0.061	<0.001*
2	Educational attainment	0.194	2.748 ± 1.108	<0.001	0.004*
3	β_C3-F4	0.077	0.155 ± 0.058	0.154	0.265
4	Seizure frequency	0.072	1.359 ± 0.820	0.006	0.052
5	θ_Fp1-Fz	0.072	0.205 ± 0.195	<0.001	<0.001*
6	Hippocampal atrophy, sclerosis	0.069	0.382 ± 0.486	0.004	0.019*
7	β_F3-F8	0.067	0.146 ± 0.048	0.216	0.411
8	β_C3-P4	0.059	0.205 ± 0.074	0.160	0.074
9	θ_C3-P4	0.057	0.237 ± 0.067	0.345	0.156
10	β_T5-T6	0.056	0.139 ± 0.049	0.028	0.012*
11	θ_P4-T5	0.054	0.220 ± 0.116	0.045	0.038*
12	θ_Fp2-T6	0.053	0.235 ± 0.099	0.003	0.008*
13	β_T5-F7	0.052	0.151 ± 0.074	0.028	0.019*
14	β_P3-P4	0.050	0.132 ± 0.050	0.830	0.655

(Continued)

TABLE 4 (Continued)

Rank	Features	FS-value	Mean ± STD	*P*-value	*P'*-value
(C) Features affecting the top 5% of the clinical-PLV$_{EEG}$ feature-based model.					
15	VPA	0.049	0.435 ± 0.496	0.013	0.369
16	β_F4-F7	0.048	0.159 ± 0.067	0.179	0.220
17	β_O1-T6	0.047	0.146 ± 0.064	0.282	0.106
18	Class of antiepileptic drugs	0.046	0.450 ± 0.498	0.016	0.125
19	θ_F3-F8	0.046	0.211 ± 0.057	0.467	0.213
20	θ_F4-F7	0.044	0.228 ± 0.067	<0.001	<0.001*
21	δ_P4-T5	0.043	0.293 ± 0.068	0.172	0.321
22	δ_F4-F7	0.042	0.305 ± 0.070	0.009	0.015*
23	β_Fp1-F8	0.042	0.282 ± 0.128	0.579	0.352
24	Time from first onset to current visit	0.040	9.290 ± 8.087	0.009	<0.001*
25	β_P3-F4	0.039	0.315 ± 0.146	0.006	<0.001*
26	Age	0.038	29.260 ± 12.746	0.061	0.075
27	θ_P3-F4	0.038	0.347 ± 0.169	0.013	0.049*
28	β_Fp2-T6	0.038	0.172 ± 0.065	0.130	0.063
29	θ_T5-F7	0.037	0.230 ± 0.082	0.012	0.025*
30	θ_Fp1-T6	0.036	0.290 ± 0.200	<0.001	<0.001*
31	δ_Fp2-T6	0.035	0.320 ± 0.085	0.450	0.157
32	β_Fp1-C3	0.035	0.144 ± 0.050	0.784	0.842
33	θ_O2-Pz	0.034	0.282 ± 0.095	0.211	0.082
34	α_C3-P4	0.034	0.260 ± 0.039	0.331	0.312
35	β_Fp2-F4	0.033	0.359 ± 0.096	0.093	0.165
36	θ_Fp1-F8	0.033	0.332 ± 0.121	0.046	0.025*

FS-value, Fisher score value; α, alpha; β, beta; δ, delta;θ, theta; For qualitative data, Chi-square tests were used; For normal data independent sample *t*-tests were used.
δ Fp1-Fz: δ band from Fp1-Fz and so on; *p* and *p'* < 0.05 is considered statistically significant, *p'* refers to *p*-value that is corrected by false discovery rate (FDR) correction. Although the selected features may not be statistically significant, they did have a classification value in the model.
*Is defined as features that have statistically significant between CI group and CON group.

could be more accurate in identifying epilepsy patients suffering from cognitive dysfunctions. It was also found that PLV$_{EEG}$ features in θ band T5-T6, θ band Fp1-Pz, δ band Fp1-Pz, β band P3-F4, and β band C3-F4 were the top 5 most important ones that might influence the model.

A combined clinical-PLV$_{EEG}$ feature-based model construction

The combined clinical-PLV$_{EEG}$ features were found the most appropriate for constructing the best performing classification models, using either AdaBoost or GBDT algorithm. A total of 707 features were screened using Fisher scores for 23 clinical features and 684 PLV$_{EEG}$ features. A total of 4 clinical features were selected within the top 10 weighted features, namely MRI of the head for abnormalities in the first rank, educational attainment in the second rank, seizure frequency in the fourth rank, and hippocampal atrophy or sclerosis in the sixth rank; all of which were significantly differed between the two groups. Between the two groups, the remaining PLV$_{EEG}$ features with significant differences were C3-F4 in the β-band, Fp1-Pz in the θ-band, F3-F8 in the β-band, C3-P4 in the β-band, C3-P4 in the θ-band, and T5-T6 in the β-band, with only Fp1-Pz in the θ-band, and T5-T6 in the β-band. Although many features were not statistically different between the two groups, they exhibited a very strong impact on the model after the Fisher score screening. Whereas a total of 12 PLV$_{EEG}$ features in the top 5% of features affecting the model performance were significantly different between the two groups, including eight features in the θ band and three PLV$_{EEG}$ features in the β band. We suspected that PLV$_{EEG}$ in the θ band might be the biomarker that could distinguish between these two groups (Table 4C and Figure 5).

For AdaBoost, the top 150 Fisher scores were selected to build the classification model, and the final performances

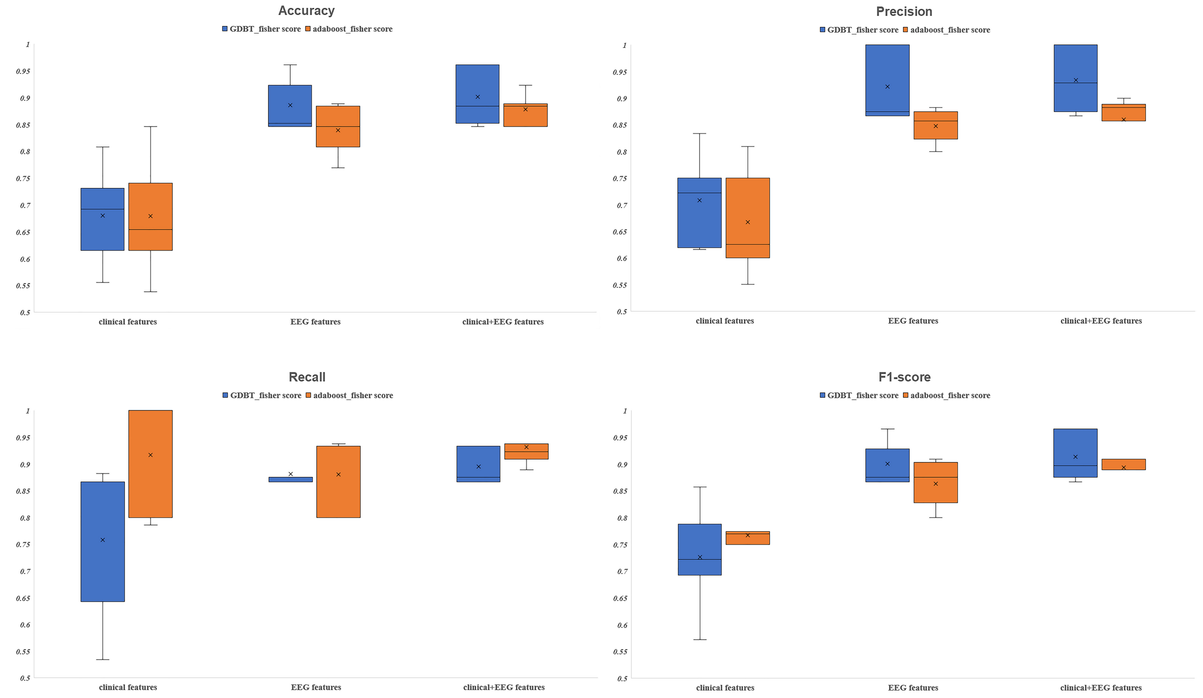


FIGURE 3
The evaluation indexes after five-fold cross-validation. GBDT, Gradient Boosting Decision Tree; AdaBoost, Adaptive Boosting.

after five-fold cross-validation were 87.78, 85.95, 93.17, 89.35, and 0.92% for accuracy, precision, recall, F1-score, and AUC, respectively. While for GBDT, the top 250 Fisher scores were selected to build the classification model, and the model performances after five-fold cross-validation were 90.11, 93.40, 89.50, 91.39, and 0.95% for accuracy, precision, recall, F1-score, and AUC, respectively (Figure 3 and Figure 4C). The recall performance of the AdaBoost model was found to be slightly higher than that of the GDBT, while GDBT outperformed AdaBoost in terms of other metrics.

Comparison between different models

Six models, based on the clinical features only, PLV_{EEG} features only, and combined clinical-PLV_{EEG} features, were constructed for 55 CON and 76 epilepsy patients suffering from cognitive dysfunctions, using the ensemble algorithms like AdaBoost and GBDT. We found that the models constructed with combined clinical-PLV_{EEG} features outperformed those developed with either pure clinical or pure PLV_{EEG} features for both the AdaBoost and GBDT algorithms. Notably, the models constructed solely with clinical features performed the worst. The cross-sectional comparisons also revealed that GBDT-built models outperformed the AdaBoost-based ones in both classification models constructed with PLV_{EEG} features. Furthermore, GBDT also outperformed AdaBoost in cases of both pure clinical features and combined clinical-PLV_{EEG} features, with an exception for recall performance (Table 5).

Not only that, but we could also identify potential biomarkers like EEG indicators using the combined clinical-PLV_{EEG} feature-based models that might be able to detect CI in epilepsy patients, which could be highly useful in the diagnosis of epilepsy in clinical settings. Additionally, many of the clinical features used have also been reported in previous studies suggesting their strong association with CI symptoms in epilepsy patients, but have not been ranked to the extent to which these clinical features might affect cognition. Therefore, we ranked these clinical features by their respective Fisher scores. Our findings suggest that EEG could be of great interest to subjects with cognitive deficits, especially those with epileptic symptoms. Previously, technical limitations were the main obstacle in improving the application of EEG for epilepsy diagnosis and treatment. By estimating the combined effects of clinical and PLV_{EEG} features, we could predict the current cognitive status in epilepsy patients, providing clinicians with more options for precise diagnosis and effective treatment plans.

Discussion

To the best of our knowledge, the present study is the first of its kind to use an integrated algorithm for the construction of a classification model for facilitating the diagnosis of

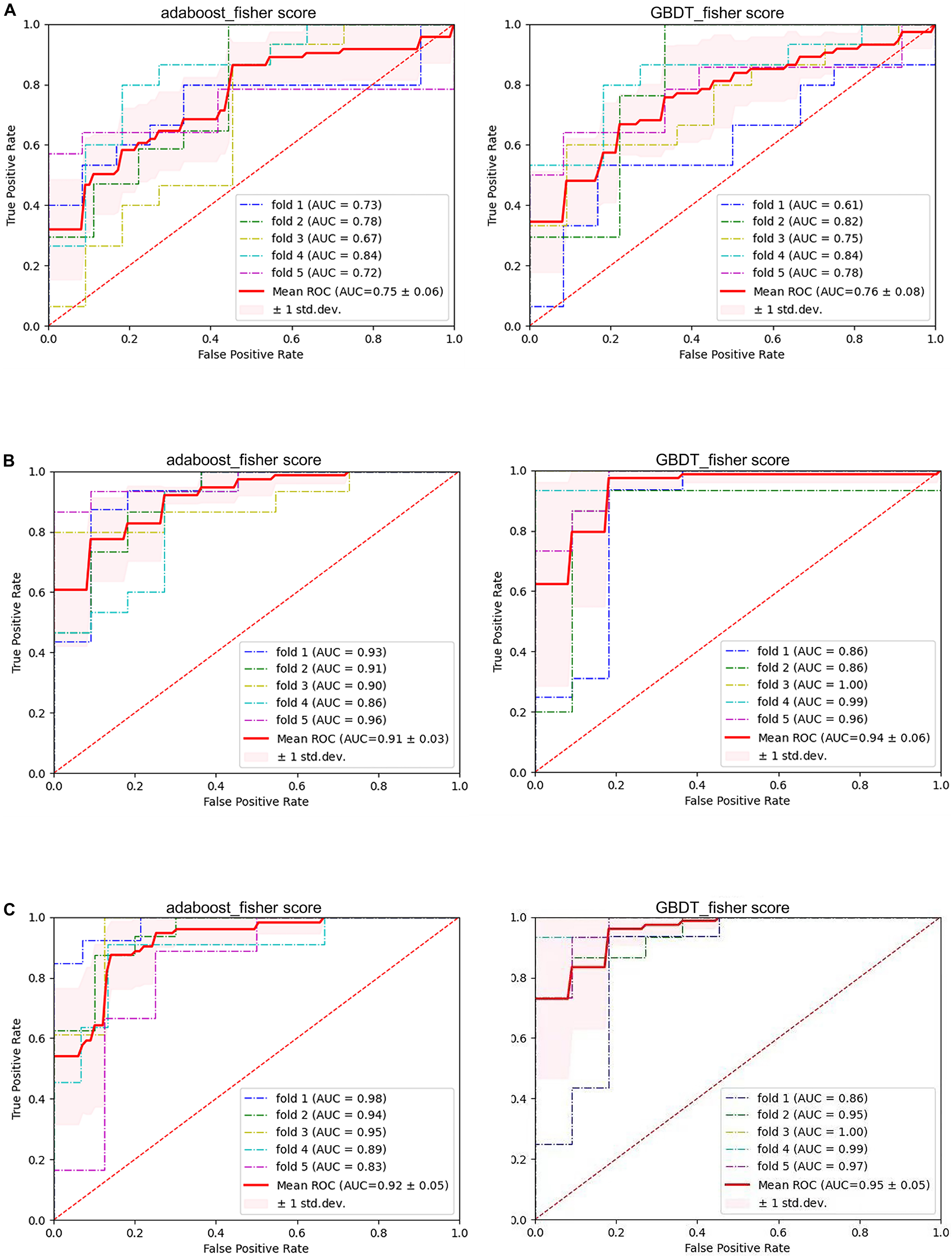


FIGURE 4

The performance of six models. **(A)** Pure clinical features. **(B)** Pure electroencephalogram (EEG) features. **(C)** Combined clinical and PLV_{EEG} features. GBDT, Gradient Boosting Decision Tree; AdaBoost, Adaptive Boosting; AUC, area under the curve; ROC, receiver operating-characteristic curve; std. dev, standard deviation.

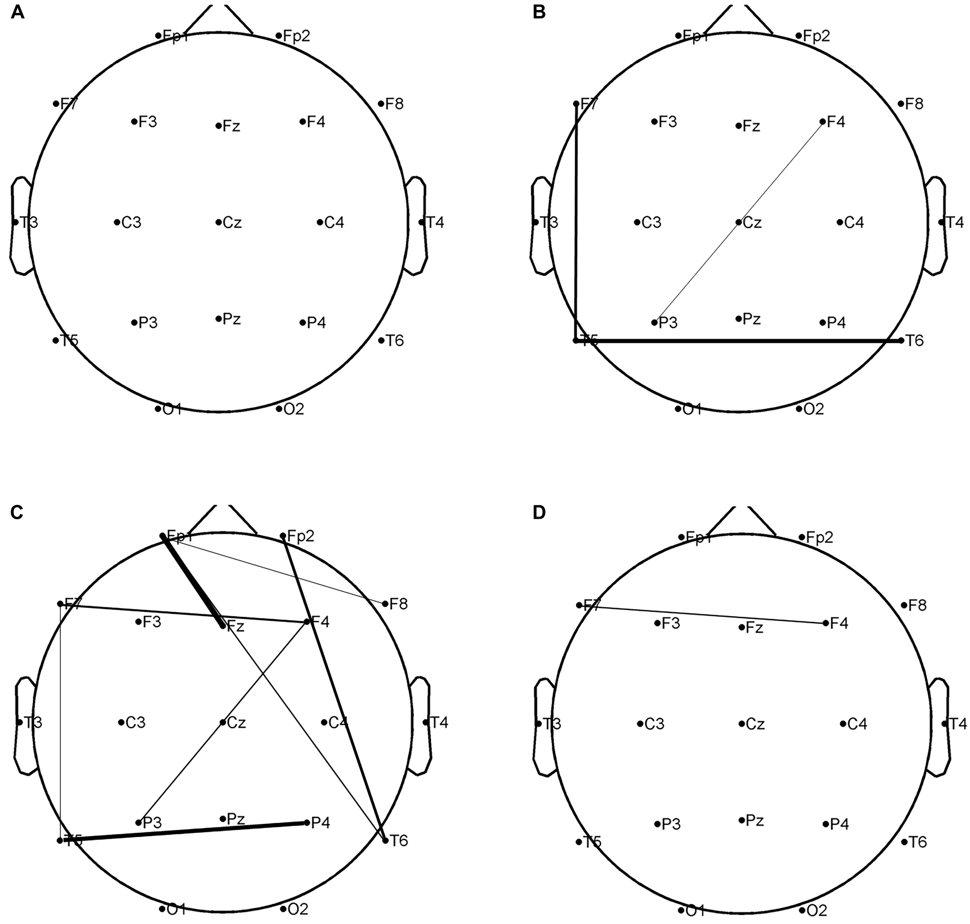


FIGURE 5
In the combined clinical-PLV$_{EEG}$ model, there were statistically significant differences in 12 PLV$_{EEG}$ features between the CON and CI groups of PWEs. The higher the fisher score, the tighter the connection between the leads. $P < 0.05$ is considered statistically significant. **(A)** Alpha band; **(B)** beta band; **(C)** theta band; **(D)** delta band. PWEs, patients with epilepsy; PLV, phase locking value; CI, cognitive impairment; CON, cognitively normal.

CI in PWE by combined clinical and PLV$_{EEG}$ functional connectivity features.

Advantages of combined clinical-PLV$_{EEG}$ features for classification model building

Although several risk factors affecting cognitive functions in epilepsy have been identified, however, only a few studies have used these clinical features to predict whether PWEs have a comorbid CI situation. Importantly, it's been difficult to determine the extent to which these clinical features might affect cognition with a background of epilepsy. A meta-analysis (Novak et al., 2022a) has found that duration of epilepsy, frequency of seizures, and use of antiseizures medications are important clinical features that can affect cognition. Moreover, some studies suggest that education, history of surgical head trauma, anxiety and depression, hippocampal abnormalities, TLE, and seizure types may influence cognitive functions in PWEs (Piazzini et al., 2006; Bell et al., 2011; Vrinda et al., 2019; Jarcuskova et al., 2020; Wang et al., 2020; Phuong et al., 2021; Elsherif and Esmael, 2022). A previous study (Lin et al., 2021) collected 12 clinical features from outpatients with epilepsy to construct a model for diagnosing CI with a performance accuracy, recall, precision, and AUC of 60, 51, 88, and 0.71%, respectively, and concluded that status epilepticus, history of previous surgical head trauma, and seizure frequency were the top three clinical features affecting cognition. However, the clinical features considered in this study were not comprehensive enough, for example, it did not take into account important factors affecting PWEs such as education level and the classes of antiseizures medications taken (Wang et al., 2020). It was previously thought that VPA, PHT, and TPM could cause cognitive dysfunctions in PWEs (Brunbech and Sabers, 2002; Dang et al., 2021; Lozano-Garcia et al., 2021), and for this reason,

TABLE 5 The performance of the six classifier models.

Features and algorithms	Performance	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean-value
Clinical features GBDT	Accuracy (%)	55.56	80.77	61.54	73.08	69.23	68.03
	Precision (%)	61.54	83.33	61.90	72.22	75.00	70.80
	Recall (%)	53.37	88.23	86.67	86.67	64.28	75.84
	F1-score (%)	57.17	85.71	72.22	78.79	69.23	72.62
	AUC	0.61	0.82	0.75	0.84	0.78	0.76
Clinical features AdaBoost	Accuracy (%)	74.07	84.62	65.38	61.54	53.85	67.89
	Precision (%)	75.00	80.95	62.50	60.00	55.00	66.69
	Recall (%)	80.00	100.00	100.00	100.00	78.57	91.57
	F1-score (%)	77.42	89.47	76.92	75.00	64.71	76.71
	AUC	0.73	0.78	0.67	0.84	0.72	0.75
EEG features GBDT	Accuracy (%)	85.19	84.62	92.31	96.15	84.62	88.58
	Precision (%)	87.50	86.67	100.00	100.00	86.67	92.17
	Recall (%)	87.50	86.67	86.67	93.33	86.67	88.17
	F1-score (%)	87.50	86.67	92.86	96.55	86.67	90.05
	AUC	0.86	0.86	1.00	0.99	0.96	0.94
EEG features AdaBoost	Accuracy (%)	88.89	84.62	80.77	76.92	88.46	83.93
	Precision (%)	88.24	82.35	85.71	80.00	87.50	84.76
	Recall (%)	93.75	93.33	80.00	80.00	93.33	88.08
	F1-score (%)	90.91	87.50	82.76	80.00	90.32	86.30
	AUC	0.93	0.91	0.90	0.86	0.96	0.91
Clinical+EEG features GBDT	Accuracy (%)	85.19	84.62	96.15	96.15	88.46	90.11
	Precision (%)	87.50	86.67	100.00	100.00	92.86	93.40
	Recall (%)	87.50	86.67	93.33	93.33	86.67	89.50
	F1-score (%)	87.50	86.67	96.55	96.55	89.66	91.39
	AUC	0.86	0.95	1.00	0.99	0.97	0.95
Clinical+EEG features AdaBoost	Accuracy (%)	88.89	88.46	92.31	84.62	84.62	87.78
	Precision (%)	85.71	88.24	90.00	76.92	88.89	85.95
	Recall (%)	92.31	93.75	100.00	90.91	88.89	93.17
	F1-score (%)	88.89	90.91	94.74	83.33	88.89	89.35
	AUC	0.98	0.94	0.95	0.89	0.83	0.92

GBDT, Gradient Boosting Decision Tree; AdaBoost, Adaptive Boosting; AUC, area under the curve.

the presence or absence of these three drugs was used as a clinical feature. The study showed that only VPA had significant weightage for this model, while PHT and TPM, probably due to insufficient data, were not statistically significant, and did not contribute to the construction of the model.

Of the models constructed using pure clinical features, the performance accuracy, recall, precision, and AUC for the AdaBoost/GBDT models were 67.89/68.03%, 91.57/75.84%, 66.69/70.80%, and 0.75/0.76%, respectively. Using Fisher scores, we selected 23 clinical features. Of these, education level, seizure frequency, and VPA therapy ranked the top three clinical characteristics affecting cognition in PWEs. Among the models constructed with combined clinical and PLV_{EEG} features, the accuracy, recall, precision, and AUC of the AdaBoost/GBDT models were 87.78/90.11%, 93.17/89.50%, 85.95/93.40%, and 0.92/0.95%, respectively. We applied the Fisher scoring method for the 23 clinical features and 684 PLV$_{EEG}$ features to jointly screen and rank. Among these features, MRI abnormalities, education level, and seizure frequency were the top 3 most influential clinical features. The performance of the models constructed using clinical features alone was better than that shown in previous studies for all metrics, except for the performance accuracy. While the performance of the models constructed using combined clinical and PLV$_{EEG}$ features was significantly improved than that reported previously. Thus, we concluded that combined clinical and PLV$_{EEG}$ features were more appropriate for PWEs and that a combination of different types of features would be an optimal choice for constructing diagnostic prediction models.

PLV$_{EEG}$ features are valid indicators for diagnosing CI in PWEs

PLV$_{EEG}$ is used to remotely examine the task-induced changes in neural activities, synchronized in EEG recordings, which is a classic metric for computing functional brain connectivity features. Jones et al. (2022) have used PLV$_{EEG}$ functional connectivity features as an evaluation metric for assessing the efficacy of transcranial alternating current stimulation (tACS) on age-associated cognitive decline. Li et al. (2022) have constructed a model combining the clinical and PLV$_{EEG}$ features to diagnose Alzheimer's disease (AD), which exhibits satisfactory performance and robustness. Another study (Lanzone et al., 2021) has found that PLV$_{EEG}$ in the α band of patients who were effective on treatment with perampanel as an add-on drug could be used as a biomarker to predict the responsiveness to perampanel drugs. Cho et al. (2017) have reported that PLV$_{EEG}$ in the γ band may be a potential biomarker for predicting seizures. In this study, the accuracy, recall, precision, and AUC of the AdaBoost/GBDT models were 83.93/88.58%, 84.76/92.17%, 88.08/88.17%, 86.30/90.05%, and 0.91/0.94%, respectively, when only the PLV$_{EEG}$ features were used for the model construction. The θ-band T5- T6, θ-band Fp1-Pz, and δ-band Fp1-Pz were the top three PLV$_{EEG}$ features affecting the model weightage, indicating that the PLV$_{EEG}$ functional connectivity features might be valid indicators for the diagnosis of cognitive dysfunctions comorbid with epilepsy.

PLV$_{EEG}$ features in the θ band may be a potential biomarker for diagnosing CI in PWEs

Here, we calculated the PLV$_{EEG}$ features of the four frequency bands (α, β, θ, δ), and found that the PLV$_{EEG}$ features, especially of the θ band, might be potential biomarkers to distinguish between epilepsy patients with or without comorbid CI. In our constructed model of the combined clinical and PLV$_{EEG}$ features, we employed Fisher scoring to rank individual features, which revealed 12 PLV features that ranked in the top 30 were significantly different between the CON and CI groups. Notably, eight of these features were related to the θ band and three to the β band.

The θ band has been found to have an important relationship with epilepsy and cognitive function in previous studies. One study (Douw et al., 2010) has demonstrated that functional connectivity features in the θ band could be used to aid in the diagnosis of epilepsy with a recall of 62% and a specificity of 72%. Other studies (Jun et al., 2020) have also suggested that stimulation of the hippocampus may increase the release of θ rhythms, thereby improving the associative memory function. These studies suggest that increasing the θ rhythm in the hippocampus may provide a theoretical basis for the neural mechanisms of memory enhancement. Moreover, Gupta et al. (2012) have identified that θ rhythms in the hippocampus of rats are associated with visuospatial abilities and executive abilities related to memory and cognition. Another study (Braithwaite et al., 2020) has revealed that increased power of the θ rhythm in children can be a valid biomarker for predicting non-verbal cognitive abilities. Furthermore, it (Ahmadlou et al., 2014) has been concluded that functional connectivity features in the θ band could be used to differentiate between patients with mild CI and healthy elderly populations. Briels et al. (2020) have found that functional connectivity indicators in the θ and β frequency bands in AD patients may help diagnose the disease severity. Other studies (Singh et al., 2018) have shown that a reduction in midfrontal θ wave frequency responds to the degree of effective control of cognitive functions in PD patients. The θ rhythms in the frontal lobe are highly correlated with cognitive function (Cavanagh and Frank, 2014), with Fp1-Fz being within the frontal lobe. Our results showed that the PLV$_{EEG}$ features of Fp1-Fz in the θ band were significantly different between the CON and CI groups of epilepsy patients, accounting for a high weightage in the diagnostic model. In this context, one study (Cao et al., 2022) has reported an important relationship between the θ rhythm and cognition in patients with schizophrenia, indicating that superior cognitive performance may be significantly associated with a smaller θ wave power, and altered θ rhythm and cognition are highly correlated mainly in the parieto-occipital lobe. The P4 and T5 were close to the occipital region in our investigation. The PLV$_{EEG}$ for P4-T5 were also significantly different between the two groups and accounted for a higher weightage in the model. Furthermore, it is shown (Usami et al., 2019) that β oscillations can enhance the responsiveness of the cerebral cortex to inputs from distant cortices, suggesting that β frequencies may have an important role in functional connectivity. Interestingly, α frequency is significantly increased in AD patients presenting with mild cognitive dysfunctions

(Moretti, 2015). The α frequency was found to be less influential in our study, in terms of statistical significance and the weightage of the model, possibly due to the exclusion of AD patients' data.

Previous studies have amply demonstrated the significance of functional connectivity features in the θ band in the diagnosis of epilepsy and cognitive dysfunctions. Therefore, our study demonstrated that PLV_{EEG} features in the θ band might be reliable biomarkers for diagnosing CI in PWEs, especially those with high Fisher scores.

Limitations and future directions

Despite these excellent results, there are still certain limitations to this study. First, this was a single-center retrospective study with data from only one institutional epilepsy center and a small sample population. Although the combined clinical and PLV_{EEG} features and advanced algorithms ensured the accuracy of our results, multi-center prospective studies are warranted for the generalization of our results. Here, we provided a theoretical basis and demonstrated the possibilities of further improving the diagnostic methods for PWEs comorbid with CI. Second, this study was based on the MoCA scale. However, we classified the features based on the total MoCA scores rather than the subtest scores. Although our model could address the issue of differentiating PWEs with or without cognitive deficits, the content of each subtest should be investigated more carefully in the future. Finally, the potential biomarkers that we extracted were mainly functional connectivity features of the EEG and a subset of clinical features. The future brain network features extracted from MRI examinations can be useful in improving the accuracy and superiority of the model. We propose to validate the performance of our models with larger datasets from multiple epilepsy centers in the future, as well as add new features to improve the accuracy of the model.

Conclusion

In this study, we constructed a diagnostic model for CI in PWEs based on the combined clinical and PLV_{EEG} features. Besides, we found that PLV_{EEG} functional connectivity features in the θ band might be potential biomarkers for the diagnosis of CI in PWEs.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by The Institutional Review Board of the Henan Provincial People's Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

ZR and MY designed the study. XH obtained funding. ZR, YZ, YH, YXH, and QW acquired the data. PZ, TZ, and NW analyzed EEG recordings. ZR, ZZ, and BWe worked on EEG preprocessing and machine learning process. ZR, ZZ, and BWa conducted the statistical analysis. ZR, ZZ, YH, YXH, and QW analyzed and interpreted the data. ZR and XH drafted and revised the manuscript. All authors revised this draft and read and approved the final manuscript.

Funding

This study was sponsored by Henan Province's Gong Jian Program (Authorization number: SB201901074) and 23456 Talent Engineering (Authorization number: ZC20200371).

Acknowledgments

We would like to thank the staff members at the Department of Neurology at the Henan Provincial People's Hospital for their assistance with this research study and we would like to express our most profound gratitude to our patients and their families.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmadlou, M., Adeli, A., Bajo, R., and Adeli, H. (2014). Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task. *Clin. Neurophysiol.* 125, 694–702. doi: 10.1016/j.clinph.2013.08.033

Aydore, S., Pantazis, D., and Leahy, R. M. A. (2013). Note on the Phase locking value and its properties. *Neuroimage* 74, 231–244. doi: 10.1016/j.neuroimage.2013.02.008

Balcik, Z. E., Senadim, S., Tekin, B., Ceyhan Dirican, A., Eren, F., Karahan, M. G., et al. (2020). Do interictal EEG findings reflect cognitive function in juvenile myoclonic epilepsy? *Epilepsy Behav.* 111:107281. doi: 10.1016/j.yebeh.2020.107281

Bell, B., Lin, J. J., Seidenberg, M., and Hermann, B. (2011). The neurobiology of cognitive disorders in temporal lobe epilepsy. *Nat. Rev. Neurol.* 7, 154–164. doi: 10.1038/nrneurol.2011.3

Bergeron, D., Flynn, K., Verret, L., Poulin, S., Bouchard, R. W., Bocti, C., et al. (2017). Multicenter validation of an MMSE-MoCA conversion table. *J. Am. Geriatr. Soc.* 65, 1067–1072. doi: 10.1111/jgs.14779

Black, L. C., Schefft, B. K., Howe, S. R., Szaflarski, J. P., Yeh, H. S., and Privitera, M. D. (2010). The effect of seizures on working memory and executive functioning performance. *Epilepsy Behav.* 17, 412–419. doi: 10.1016/j.yebeh.2010.01.006

Braithwaite, E. K., Jones, E. J. H., Johnson, M. H., and Holmboe, K. (2020). Dynamic modulation of frontal theta power predicts cognitive ability in infancy. *Dev. Cogn. Neurosci.* 45:100818. doi: 10.1016/j.dcn.2020.100818

Briels, C. T., Schoonhoven, D. N., Stam, C. J., de Waal, H., Scheltens, P., and Gouw, A. A. (2020). Reproducibility of EEG functional connectivity in Alzheimer's disease. *Alzheimers Res. Ther.* 12:68. doi: 10.1186/s13195-020-00632-3

Brunbech, L., and Sabers, A. (2002). Effect of antiepileptic drugs on cognitive function in individuals with epilepsy: A comparative review of newer versus older agents. *Drugs* 62, 593–604. doi: 10.2165/00003495-200262040-00004

Cai, M., Dang, G., Su, X., Zhu, L., Shi, X., Che, S., et al. (2021). Identifying mild cognitive impairment in Parkinson's disease with electroencephalogram functional connectivity. *Front. Aging Neurosci.* 13:701499. doi: 10.3389/fnagi.2021.701499

Cao, Y., Han, C., Peng, X., Su, Z., Liu, G., Xie, Y., et al. (2022). Correlation between resting theta power and cognitive performance in patients with schizophrenia. *Front. Hum. Neurosci.* 16:853994. doi: 10.3389/fnhum.2022.853994

Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/j.tics.2014.04.012

Cho, D., Min, B., Kim, J., and Lee, B. (2017). EEG-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1309–1318. doi: 10.1109/TNSRE.2016.2618937

Dang, Y. L., Foster, E., Lloyd, M., Rayner, G., Rychkova, M., Ali, R., et al. (2021). Adverse events related to antiepileptic drugs. *Epilepsy Behav.* 115:107657. doi: 10.1016/j.yebeh.2020.107657

Del Brutto, O. H., Mera, R. M., Del Brutto, V. J., Zambrano, M., Wright, C. B., and Rundek, T. (2019). Clinical and neuroimaging risk factors for cognitive decline in community-dwelling older adults living in rural ecuador. A population-based prospective cohort study. *Int. J. Geriatr. Psychiatry* 34, 447–452. doi: 10.1002/gps.5037

Delorme, A., and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Douw, L., de Groot, M., van Dellen, E., Heimans, J. J., Ronner, H. E., Stam, C. J., et al. (2010). 'Functional connectivity' is a sensitive predictor of epilepsy diagnosis after the first seizure. *PLoS One* 5:e10839. doi: 10.1371/journal.pone.0010839

Duma, G. M., Danieli, A., Mattar, M. G., Baggio, M., Vettorel, A., Bonanni, P., et al. (2022). Resting state network dynamic reconfiguration and neuropsychological functioning in temporal lobe epilepsy: An HD-EEG investigation. *Cortex* 157, 1–13. doi: 10.1016/j.cortex.2022.08.010

Duma, G. M., Danieli, A., Vettorel, A., Antoniazzi, L., Mento, G., and Bonanni, P. (2021). Investigation of dynamic functional connectivity of the source reconstructed epileptiform discharges in focal epilepsy: A graph theory approach. *Epilepsy Res.* 176:106745. doi: 10.1016/j.eplepsyres.2021.106745

Edeh, M. O., Dalal, S., Dhaou, I. B., Agubosim, C. C., Umoke, C. C., Richard-Nnabu, N. E., et al. (2022). Artificial intelligence-based ensemble learning model for prediction of hepatitis C disease. *Front. Public Health* 10:892371. doi: 10.3389/fpubh.2022.892371

Elahian, B., Yeasin, M., Mudigoudar, B., Wheless, J. W., and Babajani-Feremi, A. (2017). Identifying seizure onset zone from electrocorticographic recordings: A machine learning approach based on phase locking value. *Seizure* 51, 35–42. doi: 10.1016/j.seizure.2017.07.010

Elsherif, M., and Esmael, A. (2022). Hippocampal atrophy and quantitative EEG markers in mild cognitive impairment in temporal lobe epilepsy versus extra-temporal lobe epilepsy. *Neurol. Sci.* 43, 1975–1986. doi: 10.1007/s10072-021-05540-4

Fadaie, F., Lee, H. M., Caldairou, B., Gill, R. S., Sziklas, V., Crane, J., et al. (2021). Atypical functional connectome hierarchy impacts cognition in temporal lobe epilepsy. *Epilepsia* 62, 2589–2603. doi: 10.1111/epi.17032

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., et al. (2014). Ilae official report: A practical clinical definition of epilepsy. *Epilepsia* 55, 475–482. doi: 10.1111/epi.12550

Gavrilovic, A., Toncev, G., Boskovic Matic, T., Vesic, K., Ilic Zivojinovic, J., and Gavrilovic, J. (2019). Impact of epilepsy duration, seizure control and EEG abnormalities on cognitive impairment in drug-resistant epilepsy patients. *Acta Neurol. Belg.* 119, 403–410. doi: 10.1007/s13760-019-01090-x

Gupta, A. S., van der Meer, M. A., Touretzky, D. S., and Redish, A. D. (2012). Segmentation of spatial experience by hippocampal theta sequences. *Nat. Neurosci.* 15, 1032–1039. doi: 10.1038/nn.3138

He, X., Bassett, D. S., Chaitanya, G., Sperling, M. R., Kozlowski, L., and Tracy, J. I. (2018). Disrupted dynamic network reconfiguration of the language system in temporal lobe epilepsy. *Brain* 141, 1375–1389. doi: 10.1093/brain/awy042

Helmstaedter, C., and Witt, J. A. (2017). Epilepsy and cognition–a bidirectional relationship? *Seizure* 49, 83–89. doi: 10.1016/j.seizure.2017.02.017

Huang, H., Cui, G., Tang, H., Kong, L., Wang, X., Cui, C., et al. (2022). Relationships between plasma expression levels of microrna-146a and microrna-132 in epileptic patients and their cognitive, mental and psychological disorders. *Bioengineered* 13, 941–949. doi: 10.1080/21655979.2021.2015528

Jarcuskova, D., Palusna, M., Gazda, J., Feketeova, E., and Gdovinova, Z. (2020). Which clinical and neuropsychological factors are responsible for cognitive impairment in patients with epilepsy? *Int. J. Public Health* 65, 947–956. doi: 10.1007/s00038-020-01401-7

Jones, K. T., Johnson, E. L., Gazzaley, A., and Zanto, T. P. (2022). Structural and functional network mechanisms of rescuing cognitive control in aging. *Neuroimage* 262:119547. doi: 10.1016/j.neuroimage.2022.119547

Jun, S., Lee, S. A., Kim, J. S., Jeong, W., and Chung, C. K. (2020). Task-dependent effects of intracranial hippocampal stimulation on human memory and hippocampal theta power. *Brain Stimul.* 13, 603–613. doi: 10.1016/j.brs.2020.01.013

Lanzone, J., Ricci, L., Tombini, M., Boscarino, M., Mecarelli, O., Pulitano, P., et al. (2021). The effect of perampanel on EEG spectral power and connectivity in patients with focal epilepsy. *Clin. Neurophysiol.* 132, 2176–2183. doi: 10.1016/j.clinph.2021.05.026

Leguia, M. G., Andrzejak, R. G., Rummel, C., Fan, J. M., Mirro, E. A., Tcheng, T. K., et al. (2021). Seizure cycles in focal epilepsy. *JAMA Neurol.* 78, 454–463. doi: 10.1001/jamaneurol.2020.5370

Li, X., Zhou, T., and Qiu, S. (2022). Alzheimer's disease analysis algorithm based on no-threshold recurrence plot convolution network. *Front. Aging Neurosci.* 14:888577. doi: 10.3389/fnagi.2022.888577

Lin, F., Han, J., Xue, T., Lin, J., Chen, S., Zhu, C., et al. (2021). Predicting cognitive impairment in outpatients with epilepsy using machine learning techniques. *Sci. Rep.* 11:20002. doi: 10.1038/s41598-021-99506-3

Lozano-Garcia, A., Hampel, K. G., Villanueva, V., Gonzalez-Bono, E., and Cano-Lopez, I. (2021). The number of anti-seizure medications mediates the relationship between cognitive performance and quality of life in temporal lobe epilepsy. *Epilepsy Behav.* 115:107699. doi: 10.1016/j.yebeh.2020.107699

Ma, J., Wang, Z., Cheng, T., Hu, Y., Qin, X., Wang, W., et al. (2022). A prediction model integrating synchronization biomarkers and clinical features to identify responders to vagus nerve stimulation among pediatric patients with drug-resistant epilepsy. *CNS Neurosci. Ther.* 28, 1838–1848. doi: 10.1111/cns.13923

Melikyan, Z. A., Malek-Ahmadi, M., O'Connor, K., Atri, A., Kawas, C. H., and Corrada, M. M. (2021). Norms and equivalences for MoCA-30, MoCA-22, and MMSE in the oldest-old. *Aging Clin. Exp. Res.* 33, 3303–3311. doi: 10.1007/s40520-021-01886-z

Montano-Lozada, J. M., Lopez, N., Espejo-Zapata, L. M., Soto-Anari, M., Ramos-Henderson, M., Caldichoury-Obando, N., et al. (2021). Cognitive changes in patients with epilepsy identified through the moca test during neurology outpatient consultation. *Epilepsy Behav.* 122:108158. doi: 10.1016/j.yebeh.2021.108158

Moretti, D. V. (2015). Theta and alpha EEG frequency interplay in subjects with mild cognitive impairment: Evidence from EEG, MRI, and SPECT brain modifications. *Front. Aging Neurosci.* 7:31. doi: 10.3389/fnagi.2015.00031

Novak, A., Vizjak, K., and Rakusa, M. (2022a). Cognitive impairment in people with epilepsy. *J. Clin. Med.* 11:267. doi: 10.3390/jcm11010267

Novak, A., Vizjak, K., Gacnik, A., and Rakusa, M. (2022b). Cognitive impairment in people with epilepsy: Montreal cognitive assessment (MoCA) as a screening tool. *Acta. Neurol. Belg.* 8:4. doi: 10.1007/s13760-022-02046-4

Peng, T., Chen, X., Wan, M., Jin, L., Wang, X., Du, X., et al. (2020). The prediction of hepatitis E through ensemble learning. *Int. J. Environ. Res. Public Health* 18:159. doi: 10.3390/ijerph18010159

Phuong, T. H., Houot, M., Mere, M., Denos, M., Samson, S., and Dupont, S. (2021). Cognitive impairment in temporal lobe epilepsy: Contributions of lesion, localization and lateralization. *J. Neurol.* 268, 1443–1452. doi: 10.1007/s00415-020-10307-6

Piazzini, A., Canevini, M. P., Turner, K., Chifari, R., and Canger, R. (2006). Elderly people and epilepsy: Cognitive function. *Epilepsia* 47(Suppl. 5), 82–84. doi: 10.1111/j.1528-1167.2006.00884.x

Rodrigues, S. G., Gouveia, R. G., and Bentes, C. (2020). Moca as a cognitive assessment tool for absence status epilepticus. *Epileptic Disord.* 22, 229–232. doi: 10.1684/epd.2020.1149

Singh, A., Richardson, S. P., Narayanan, N., and Cavanagh, J. F. (2018). Mid-frontal theta activity is diminished during cognitive control in Parkinson's disease. *Neuropsychologia* 117, 113–122. doi: 10.1016/j.neuropsychologia.2018.05.020

Titiz, A. S., Mahoney, J. M., Testorf, M. E., Holmes, G. L., and Scott, R. C. (2014). Cognitive impairment in temporal lobe epilepsy: Role of online and offline processing of single cell information. *Hippocampus* 24, 1129–1145. doi: 10.1002/hipo.22297

Ung, H., Cazares, C., Nanivadekar, A., Kini, L., Wagenaar, J., Becker, D., et al. (2017). Interictal epileptiform activity outside the seizure onset zone impacts cognition. *Brain* 140, 2157–2168. doi: 10.1093/brain/awx143

Usami, K., Milsap, G. W., Korzeniewska, A., Collard, M. J., Wang, Y., Lesser, R. P., et al. (2019). Cortical responses to input from distant areas are modulated by local spontaneous alpha/beta oscillations. *Cereb. Cortex* 29, 777–787. doi: 10.1093/cercor/bhx361

Vrinda, M., Arun, S., Srikumar, B. N., Kutty, B. M., and Shankaranarayana Rao, B. S. (2019). Temporal lobe epilepsy-induced neurodegeneration and cognitive deficits: Implications for aging. *J. Chem. Neuroanat.* 95, 146–153. doi: 10.1016/j.jchemneu.2018.02.005

Wang, L., Chen, S., Liu, C., Lin, W., and Huang, H. (2020). Factors for cognitive impairment in adult epileptic patients. *Brain Behav.* 10:e01475. doi: 10.1002/brb3.1475

Wenbo, W., Yang, S., and Guici, C. (2021). Blood glucose concentration prediction based on VMD-KELM-adaboost. *Med. Biol. Eng. Comput.* 59, 2219–2235. doi: 10.1007/s11517-021-02430-x

Zhang, J., Xu, D., Hao, K., Zhang, Y., Chen, W., Liu, J., et al. (2021). FS-GBDT: Identification multicancer-risk module via a feature selection algorithm by integrating fisher score and Gbdt. *Brief Bioinform.* 22:bbaa189. doi: 10.1093/bib/bbaa189

Zhang, S., Wang, J., Li, X., and Liang, Y. (2021). M6A-GSMS: Computational identification of N(6)-methyladenosine sites with GBDT and stacking learning in multiple species. *J. Biomol. Struct. Dyn.* 40, 12380–12391. doi: 10.1080/07391102.2021.1970628

TYPE Original Research
PUBLISHED 15 February 2023
DOI 10.3389/fnins.2023.988394

Check for updates

OPEN ACCESS

EDITED BY
Reza Rastmanesh,
Nutrition Society, United Kingdom

REVIEWED BY
Max Ostrinsky Krucoff,
Medical College of Wisconsin and Marquette University, United States
Elise Rowe,
University of Canberra, Australia

*CORRESPONDENCE
Nathalie Kubis
✉ nathalie.kubis@aphp.fr
David Holcman
✉ david.holcman@ens.psl.eu

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Translational Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 07 July 2022
ACCEPTED 27 January 2023
PUBLISHED 15 February 2023

CITATION
Floyrac A, Doumergue A, Legriel S, Deye N, Megarbane B, Richard A, Meppiel E, Masmoudi S, Lozeron P, Vicaut E, Kubis N and Holcman D (2023) Predicting neurological outcome after cardiac arrest by combining computational parameters extracted from standard and deviant responses from auditory evoked potentials. *Front. Neurosci.* 17:988394. doi: 10.3389/fnins.2023.988394

Predicting neurological outcome after cardiac arrest by combining computational parameters extracted from standard and deviant responses from auditory evoked potentials

Aymeric Floyrac[1†], Adrien Doumergue[1†], Stéphane Legriel[2,3], Nicolas Deye[4,5], Bruno Megarbane[4,6], Alexandra Richard[7], Elodie Meppiel[7], Sana Masmoudi[7], Pierre Lozeron[7,8], Eric Vicaut[9], Nathalie Kubis[7,8]* and David Holcman[1]*

[1]Applied Mathematics and Computational Biology, Ecole Normale Supérieure-PSL, Paris, France, [2]Medical-Surgical Intensive Care Department, Centre Hospitalier de Versailles, Le Chesnay, France, [3]CESP, PsyDev Team, INSERM, UVSQ, University of Paris-Saclay, Villejuif, France, [4]Department of Medical and Toxicological Critical Care, APHP, Lariboisière Hospital, Paris, France, [5]INSERM U942, Paris, France, [6]INSERM UMRS 1144, Université Paris Cité, Paris, France, [7]Service de Physiologie Clinique-Explorations Fonctionnelles, APHP, Hôpital Lariboisière, Paris, France, [8]LVTS UMRS 1148, Hemostasis, Thrombo-Inflammation and Neuro-Vascular Repair, CHU Xavier Bichat Secteur Claude Bernard, Université Paris Cité, Paris, France, [9]Unité de Recherche Clinique Saint-Louis- Lariboisière, APHP, Hôpital Saint Louis, Paris, France

Background: Despite multimodal assessment (clinical examination, biology, brain MRI, electroencephalography, somatosensory evoked potentials, mismatch negativity at auditory evoked potentials), coma prognostic evaluation remains challenging.

Methods: We present here a method to predict the return to consciousness and good neurological outcome based on classification of auditory evoked potentials obtained during an oddball paradigm. Data from event-related potentials (ERPs) were recorded noninvasively using four surface electroencephalography (EEG) electrodes in a cohort of 29 post-cardiac arrest comatose patients (between day 3 and day 6 following admission). We extracted retrospectively several EEG features (standard deviation and similarity for standard auditory stimulations and number of extrema and oscillations for deviant auditory stimulations) from the time responses in a window of few hundreds of milliseconds. The responses to the standard and the deviant auditory stimulations were thus considered independently. By combining these features, based on machine learning, we built a two-dimensional map to evaluate possible group clustering.

Results: Analysis in two-dimensions of the present data revealed two separated clusters of patients with good versus bad neurological outcome. When favoring the highest specificity of our mathematical algorithms (0.91), we found a sensitivity of 0.83 and an accuracy of 0.90, maintained when calculation was performed using data from only one central electrode. Using Gaussian, K-neighborhood and SVM classifiers, we could predict the neurological outcome of post-anoxic comatose patients, the validity of the method being tested by a cross-validation procedure. Moreover, the same results were obtained with one single electrode (Cz).

Conclusion: statistics of standard and deviant responses considered separately provide complementary and confirmatory predictions of the outcome of anoxic comatose patients, better assessed when combining these features on a two-dimensional statistical map. The benefit of this method compared to classical EEG and ERP predictors should be tested in a large prospective cohort. If validated, this method could provide an alternative tool to intensivists, to better evaluate neurological outcome and improve patient management, without neurophysiologist assistance.

KEYWORDS

coma, electroencephalography, automatic classification algorithm, machine learning, neurological prognosis

Introduction

Sudden death by cardiac arrest (CA) is a major public health issue, affecting 55 patients out of 100,000 with nearly 40,000 cases per year in France (Sfar, 2007). Five to 30% of the patients resuscitated after CA are alive at 1 year (Pell, 2003; Carr et al., 2009; Chin et al., 2022). Despite the use of veno-arterial extracorporeal cardiopulmonary resuscitation (VA-ECPR), a contemporary resuscitation approach that increases patients' survival, prognosis remains grim (Miraglia et al., 2020). Favorable outcome after discharge relies mainly on the prognostic value of brain injury that outweighs the combined effects of all other terminal organ failures (Roberts et al., 2013; Rossetti et al., 2016).

Assessment of neurological damage is usually performed 48–72 h after CA and, optimally, after interruption of sedative drugs (Nolan et al., 2021). The evaluation is multimodal and combines, according to available local resources, clinical evaluation (Glasgow Coma Scale, photomotor and pupillary reflexes), biological markers of neural cell necrosis (NSE and S100bêta proteins), cerebral Magnetic Resonance Imaging and electrophysiological studies including electroencephalography (EEG), somatosensory evoked potentials (SSEP) and auditory evoked potentials (AEP). EEG analysis allows grading of post-anoxic encephalopathy (Synek, 1988), "highly malignant" EEG pattern (Westhall et al., 2016; André-Obadia et al., 2018), being associated with the least favorable prognosis. The absence of EEG reactivity can predict mortality and poor outcome. However, it is prone to large inter-rater variability when only determined using visual analysis. For this reason, quantitative methods developed to objectively measure EEG reactivity are promising (Duez et al., 2018; Admiraal et al., 2020; Bouchereau et al., 2022) and somatosensory and auditory evoked potentials can also be used to improve the accuracy of the patient outcome. The absence of cortical N20 response at SSEP after stimulation of median nerves has an almost 100% specificity for non-awakening prediction (Sandroni et al., 2014), while the presence of a "mismatch negativity" (MMN), an endogenous long latency negative potential at AEP (Rohaut et al., 2009) would rather indicate a good prognosis. The absence of cortical N20 response at SSEP after stimulation of median nerves has an almost 100% specificity for non-awakening prediction (Sandroni et al., 2014). The presence of a "mismatch negativity" (MMN), an endogenous long latency negative potential at AEP (Rohaut et al., 2009) would rather indicate a good prognosis.

Mismatch negativity consists in recording cortical potentials in response to auditory stimulation delivered by earphones, using electrodes placed on the scalp. The MMN (or N200), is a negative event-related potential (ERP) that occurs between 100 and 250 ms predominantly over the frontocentral scalp area and is obtained by the subtraction of oddball auditory stimuli (called deviant stimuli) randomly intermixed with repetitive frequent auditory stimuli also called standard or non-deviant stimuli. Thus, MMN reflects the ability to detect automatic auditory violations, but sensitivity to predict awakening is low (56%) with a high 93% specificity (Naccache et al., 2005). Because of lack of sensitivity in the ICU when interpreted only by visual analysis (present or absent) (Azabou et al., 2018), complementary statistical methods have been developed to analyze MMN more accurately, increasing thus the positive predictive value for awakening (Pfeiffer et al., 2017), at the cost of extension of the time of interpretation. Thus, multimodal approaches combining several prognostic factors of post-anoxic coma have been proposed (Bassetti et al., 1996; Fischer et al., 2006; Kim et al., 2012; Oddo and Rossetti, 2014) but the choice of these approaches has not yet succeeded to lead to automatic and predictive analyses.

Taking advantage of the considerable amount of information obtained at AEP, we conducted an explorative study in which we applied a machine learning classification approach based on EEG features arising from the distribution of the ERP fluctuations responses during the 20 min-recording, rather than to interpret the MMN as a binary response. We used data already acquired from a homogeneous cohort of patients admitted in the intensive care unit after CA and who all had EEG, SSEP and AEP recordings within 6 days after admission. We identified specific features from AEP, considering responses to standard and deviant auditory stimulations independently. Using a step-by-step data processing, we finally reported combined features in two-dimensional map where we observed that patients were clustered into two groups corresponding to a different outcome at discharge whether they were able to follow verbal command or not. We then estimated the probability for a patient to be classified into one of the two groups at the acute phase using several classifiers.

Patients and methods

Procedure

This study is a retrospective single-center study performed in 29 consecutive patients between January 2014 and March 2016, successfully resuscitated after CA, with persistent coma between the 3rd day and 6th day following admission in the Department of Medical and Toxicological Critical Care in Lariboisière Hospital (Paris), and who completed EEG, SSEP, and AEP recordings. From AEP recordings, we extracted individual features, and using a novel analysis method, we aimed to classify patients into two categories: communicating patients (assumed to have a good neurological outcome) and deceased or non communicating patients, according to their capacity to follow verbal command at discharge.

This study is an ancillary study of the PHRC CAPACITY AOR10109 and was approved by the ethics committee (Comité de Protection des Personnes, CPP Paris IV #2012/22). As this AEP processing was performed secondarily, physicians who were in charge of the patients could not have access to these data. Withdrawal of life-sustaining therapies was performed according to the usual guidelines (Société de réanimation de langue française., 2010).

Clinical data

Cardiac arrest characteristics, in-hospital management and outcome data were collected according to Utstein method by the intensivists in charge during hospitalization (Perkins et al., 2015). During ICU stay, the following data were collected: age, sex, past medical history; presumed etiology categorized into non-cardiac, cardiac and undetermined; shockable rhythm; time from collapse (CA) to return of spontaneous circulation (ROSC) dichotomized into $\leq$25 or >25 min (Oddo et al., 2008); interval from the time of collapse (presumed time of CA) to basic and/or advanced life support, defined as no-flow duration, and the interval from the beginning of life support until the return of spontaneous circulation or termination of resuscitative efforts, termed low-flow duration; hypothermia; Glasgow Coma Scale (GCS) on admission; SAPS II (Simplified Acute Physiology Score) (Le Gall et al., 2005); sedation.

Good neurological prognosis was defined by appropriate response to verbal command. Moreover, the Glasgow Outcome Scale Extended (GOS-E) was retrospectively collected at 3–6 months, when information was available.

Because of the retrospective design of our study, withdrawal of life-sustaining therapies decisions had been taken before our new analysis. They were multimodal and based upon European guidelines ERC-ESICM (2014).

Electrophysiological data

We used electrophysiological data acquired between day 3 and day 6 following admission, in order not to include patients with early predictable death. However, most of them had previous EEG recording in the first 48 h. All data were analyzed or double-checked by specialists in clinical neurophysiology with at least 10 years' experience.

EEG

Digital electroencephalography (EEG) recordings were performed for at least 20 min, with 21 scalp electrodes positioned according to the standard 10–20 system placement, reformatted to both bipolar and off-head referential montages, with filter settings at 0.3 and 70 Hz. Repetitive bilateral auditory and painful stimulations were systematically performed. These stimulations aimed to evaluate EEG reactivity and performed according to a standardized protocol for auditory (clapping noise, patient's name and patient's surname) and nociceptive stimulations (nail bed pressure plied to each upper limb) regularly applied in the same order. EEG was classified according Synek's classification (Synek, 1988), which defines precisely the five major EEG patterns based on the allocation of patients into five principal categories regarding their significance for survival (optimal, benign if persistent, uncertain, malignant if persistent and fatal).

Somatosensory evoked potentials

Median nerves were stimulated at the wrist to an intensity of 4–5 mA, greater than that needed to evoke a muscular response, and in the case of the use of neuromuscular blocking, the ERB potential amplitude was used to estimate the intensity of the stimulation. Pulse duration was 0.2 ms and stimulus rate 3 Hz. Active electrodes were placed at Erb's point and C3 and C4 points. At least two repetitions (averages of 300 responses) were performed to assess the reproducibility of the waveforms. N20 cortical response was dichotomized into absent or present.

Mismatch negativity

The auditory event related potentials were elicited using the classical odd-ball paradigm technique as already described (Fischer et al., 1999).

Event-related potentials were recorded with active electrodes (in an electrode cap) positioned at Fz, Cz, C3, C4 according to the International 10–20 system, reference electrode at the mastoid and ground reference at the forehead. Acoustic stimuli were delivered through earphones binaurally using a randomly intertwined sequence of standard and deviant stimuli in the proportion of 86 and 14%, respectively. Standard stimuli were delivered at a frequency of 800 Hz and lasting 75 ms each. Deviant stimuli were delivered at a frequency of 880 Hz and lasting 30 ms each to distinguish them from the standard stimuli (Fischer et al., 1999; Chausson et al., 2008; Comanducci et al., 2020). The interstimulus interval was 500 ms. EEG signals were band-pass filtered (0.5–75 Hz) using a time window of 500 ms. Each recording was performed during 20 min. Presence/absence of MMN defined as the negative peak obtained between the difference between deviant and standard response occurring in the 100–300 ms time interval following stimulation. In our experience, MMN is delayed in those critically ill and sedated patients, which explains this relatively wide time window.

Electrophysiological analysis

All data were analyzed by at least two different neurophysiologists, blind to the neurological outcome of the patients. When artifacts were too numerous leading to unreliable conclusion, data were not considered.

Statistical analyses for demographic and clinical data

In each group (good or bad neurological outcome), results of clinical and neurophysiological examinations were expressed as mean ± SD [min-max] and median [IQR 25–75], when appropriate. Statistical analyses were performed with Prism 5 software (Prism 5.03, GraphPad, San Diego, USA). Comparison of frequencies in each group was analyzed by the Fisher's exact test. A value of $p < 0.05$ was considered statistically significant.

Signal processing, features identification, and classification

This section is divided in three parts: 1-Signal processing, 2-Feature identification and 3-Classification using a two-dimensional map. Without *a priori* consideration, we considered specific features in a 1 s duration window, then in a shorter window of 500 ms, then at last in 320 ms, which contains the relevant features and gave similar results compared to the two other time windows. This time interval was chosen large to start (in order to take in consideration the maximum amount of information, then was restricted to the smallest time interval that still contained the whole information. The data corresponding to the responses obtained from standard and deviant auditory stimulations were considered independently, regardless of the MMN that was not considered here, and mathematical processing was applied as for any signal, independently of its potential significance. We chose specific independent features for the standard and deviant stimulations that allowed increasing the robustness of the results, and preventing a potential bias by choosing a single set of parameters. At last, we combined them into a two-dimensional map, and patients formed two clusters according to their outcome. All these steps were determined without *a priori* knowledge of the patient's prognosis.

Signal processing

Auditory evoked potential obtained with standard and deviant auditory stimulations were exported in the European Data Format (EDF), which is a simple and flexible format for storage of multichannel biological and physical signals, then anonymized through a specific software we designed. Analyses were performed on all four active electrodes then on one single Cz (central) electrode in order to see if we could obtain similar results with a simplified electrodes setting. *To quantify the auditory evoked responses recorded from post* CA patients in the intensive care unit, we studied separately standard and deviant responses (Figure 1), which is a novel and different paradigm compared to the classical MMN. We took into account the total 20 min extracted data, instead of the short interval response occurring in [100–300] ms following auditory stimulation. We filtered the signal in the [0.5–50] Hz band. Finally, all standard and deviant stimulations were averaged leading to a response in the time interval $[0-1000]$ ms, $[0--500$ ms$]$, and [0–320 ms], without difference in the analysis of the time intervals. To note, there was no difference either in the responses when they were computed in the interval [20–320 ms] that still contained the relevant information. Therefore, we converged to compute all statistics over a time window of [20–320] for all sounds, and results are presented in this interval.

We first focused on the ERP responses to standard periodic auditory stimuli, every 1s. We filtered the time series $X(t)$ using a Butterworth bandpass filter ($n = 4$) in the frequency range 0.5–50 Hz and obtained the output $X_f(t)$. Finally, we averaged the signal in the time interval $[0-1]s$, ensuring that auditory stimuli were produced at time $t = nT$ ($T = 1$s) leading to the response

$$X_p(s) = \frac{1}{N}\sum_{1}^{N} X_f(s+nT), s \in [0-1] \tag{1}$$

where N is the number of periods (typically of the order 10^3). This preliminary procedure therefore allowed obtaining an average response X_p that highlights any possible deterministic feature present in the response. We applied a similar averaging procedure for deviant stimuli (see below and Figure 1).

Analysis of responses to standard stimuli

For the analysis of standard stimulation, we divided the 20 min recording into two parts (two consecutive sequences of 10 min), to explore a possible adaptation between the first part of the acquisition and the last part. If patients' responses to auditory stimulations are able to fluctuate, this could indicate a better prognosis. This "reactivity" or ability to adapt is already used when interpreting the EEG in the ICU and indicates a better neurological outcome. We have introduced two parameters to that possible adaptation analysis: the variance of the signal computed over 10 min and the correlation between the two parts of the signal. The main parameters we extracted to study the response to standard stimulations were defined as follows:

We computed the standard deviation $\sigma_X$of the signal in the time interval $[20-320]\,ms$.

$$\sigma_X^2 = \frac{1}{t_2-t_1}\int_{t_1}^{t_2}\left(X(t) -< X(t) >\right)^2 dt, \tag{2}$$

where $t_2 = 320\,ms$ and $t_1 = 20\,ms$, and $X(t)$ is the average of the X variable over the time [t1, t2]. This time interval corresponds to time scales of the neural networks involved in cognitive tasks.

We then divided the acquisition time of 20 min into n equal parts. For $n = 2$, we got $[1-10]min$ and $[10-20]min$. We averaged the signals on each of these periods to obtain two responses $X(t)$ and $Y(t)$ in the interval $[0-1]s$. We computed the time correlation or similarity in $[20,\ 320]ms$ of these two signals:

$$r(X,Y) = \frac{<(X(t) -< X >)(Y(t) -< Y >)>}{\sigma_X\sigma_Y}, \tag{3}$$

where $< . >$ represents this time average.

We therefore used these two parameters to define the space state for the coordinates a patient: (1) the standard deviation computed over the entire sample of 20 min and (2) the similarity, computed in Eq. 3. These coordinates define a mathematical state space, which is not a specific of the medical state of the patient.

Analysis of responses to deviant stimuli

Deviant stimuli are random stimuli that account for 14% of the entire responses. The approach used for standard responses analysis is not well suited for deviant stimulations, as we did not expect any adaptation in time of such a random motif. We choose two parameters that are classically used for analysing oscillatory signals, the number of extrema and the total variation for the oscillation. We filtered the resulting signal X_d using a lowpass Butterworth filter ($n = 2$) with a cut off frequency at 10 Hz. Finally, we isolated responses in the different time windows described above and computed averaged responses

$$X_r(s) = \frac{1}{N}\sum_{1}^{N} X_d(s+nT). \tag{4}$$

The smooth signal is shown in Figure 1C. We computed two mathematical quantities on the signal:

(1) The number N_E of local extrema (minima and maxima) in the response attained at points e_i.

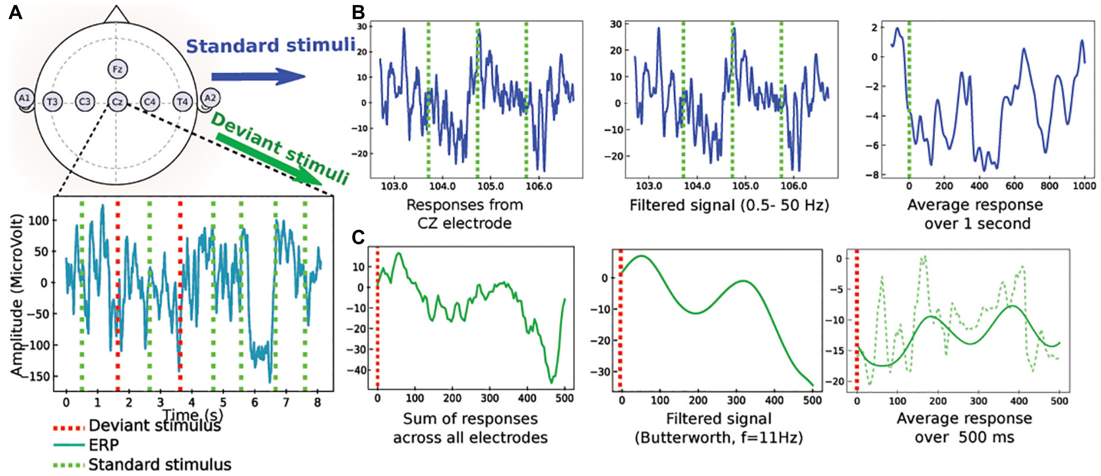


FIGURE 1
Pre-processing of the evoked auditory responses to standard and deviant stimuli (an example of data obtained from the CZ electrode is given for standard stimuli and an example from all electrodes is given for deviant stimuli). **(A)** Upper: standard position of the EEG electrodes. Lower: EEG traces during a protocol mixing standard (green) and deviant (red) stimulations. **(B)** Sample of standard stimuli (blue) the EEG signal from CZ-electrode is filtered 0.5–50 Hz. The output is an average filtered response over 1 s. **(C)** Pre-processing of deviant stimuli: (1) the signal is summed over electrodes, (2) a low-pass filter is applied (butterworth with $n = 2$, cutoff frequency at 10 Hz), (3) average filtered response (continuous green) in a window of 500 *ms* to a deviant stimulus, computed after synchronization to the stimulus. The non-filtered average response is also shown (dashed line).

(2) The total variation for the oscillation is measured by

$$|\Delta V| = \sum_i |V(e_i) - V(e_{i+1})|, \quad (5)$$

which is the sum of the absolute value of the difference between two consecutive extrema of the average evoked responses. This oscillation provides an information of the cumulative response amplitude; (ei) is the time point where the EEG signal is maximal or minimal.

Features identification associated to standard and deviant responses

For standard responses, we computed the standard deviation (formula 2) and the correlation function (formula 3) of the response computed between the response in the first and second time period (Figure 2A). To test the ability of these two parameters to separate the two categories of patients, we plotted the histogram of these two parameters for all patients in our data (Figure 2B), showing that each parameter individually could be potentially used for a classification.

For the deviant responses, as the signal showed different characteristics, we decided to use novel features, the number of extremum N_E present in the signal (Figure 3A) and the absolute value of the oscillation$|\Delta V|$, which represents the sum of the differences between the extrema (formula 5). The result of this classification is shown by histograms of the two parameters computed over the whole population of patients (Figure 3B).

Although two different types of parameters were studied for standard and deviant responses, each of them taken individually was not sufficient to clearly separate the two categories of patients.

Classification using a two-dimensional map

Based on the parameters we extracted in the previous subsection, we generated two-dimensional maps: for the map associated to standard stimuli, each patient has the $P = [\sigma_X, r(X, Y)]$ coordinates, while for deviant stimuli, we used the $P = (N_E, |\Delta V|)$ coordinates. In various plots, we normalized the coordinates in a population $(X_1, ..X_n)$ by:

$$\tilde{X}_i = \frac{X_i - < X_i >}{\sqrt{Var(X_1, ..X_n)}}, \quad (6)$$

Where X_i is the average over the points X_i and Var is the variance.

We mapped all points for all patients, where patients with bad *versus* good neurological outcome are shown in blue (vs red). Patients with good neurological outcome formed a cluster that will be the basis of the classification and prediction described below. The classification probability of a patient characterized by its coordinates was obtained by computing a score that measures the proximity to one of the two categories of patients.

To study the maps defined above as predictive tools, we used three independent statistical classifiers (SVM, Gaussian estimator, K-nearest neighbors). Because the present database did not contain many patients and to guarantee the robustness of our approach, we decided to use three classifiers (SVM, Gaussian mixture, and k-neighbors). As a small size database is also associated with overlearning, and to overcome this difficulty, we chose to use simple models for classification: Support Vector Machines (SVM) seems to be particularly suitable, as its classification is dependent only on a reduced number of patients. In fact, we wished to assign a good neurological outcome probability to any point that would be added on the map based on the ensemble of previous data points already classified. Using the assumption that statistics associated to patients (features) are independent from each other, we used a Bayesian classification.

SVM classification

To classify the data, we used the standard SVM algorithm (Cherkassky and Mulier, 1998), which determines the hyperplane that best separates the two classes. Briefly, the chosen hyperplane maximizes the distance between itself and the closest points of each class, while all points of a given class are located on one of the two

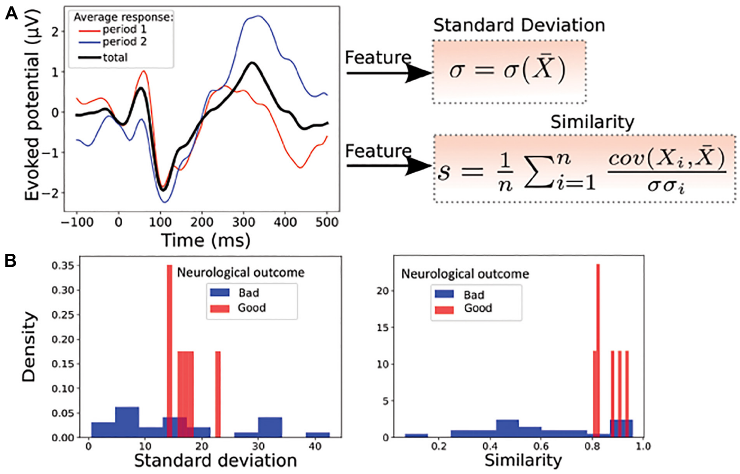


FIGURE 2

Statistical features associated to standard responses. **(A)** Left: average evoked responses computed over a time window of $\left[0-10\right]$ *min* (period 1, red), $\left[10-20\right]$ *min* (period 2, blue) and over the entire period ($\left[0-20\right]$ *min*, black). Right: the standard deviation σ and the average correlation function *s* (similarity), between the response over the entire period ($\left[0-20\right]$ *min*) and over one of the n periods ([0 – 20]*min* or $\left[10-20\right]$ *min*), here $n = 2$. **(B)** Example of features distribution of dataset from the Cz electrode: standard deviation (Left) and similarity (Right) computed over the entire period; red (good neurological outcome) and blue (bad neurological outcome). These two parameters taken separately are insufficient to properly discriminate the two groups of patients.

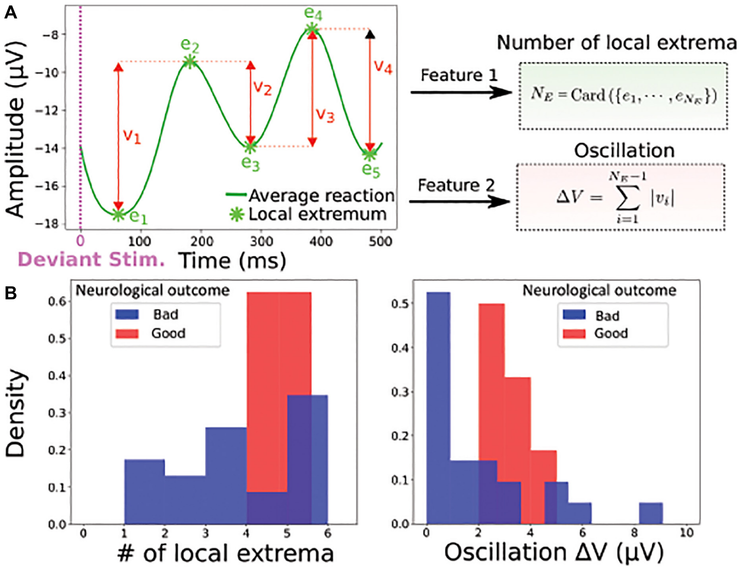


FIGURE 3

Statistical features associated to deviant responses. **(A)** Left: the average filtered evoked response (blue) to deviant stimuli computed over the entire time window contains N_E local extrema e_i (minimum or maximum), which is the first feature. The second feature is the oscillation $|\Delta V| = \sum_i |V(e_i) - V(e_{i+1})|$, which is the sum of the absolute value of the difference between two consecutive extrema of the average evoked response. **(B)** Example of feature distributions of dataset from all electrodes: local extrema (Left) and oscillation (Right) over the entire period; red (good neurological outcome) and blue (bad neurological outcome).

sides (Valiant, 1984). If no such hyperplane is found, which is the case here, the dimension of the space where the data are embedded is increased, a procedure known as kernelling (Aizerman, 1964). In a higher dimensional space, the classes are well separated by a higher dimensional hyperplane. If the two classes are still not well separated, a penalty is inflicted for every misclassified data point (Cortes and Vapnik, 1995). Here, the kernel is the Radial Basis Function $K(x, x^{'}) = \exp(-\gamma||x - x^{'}||^2)$, with $\gamma = 1$ and a penalty coefficient $C = 10$. Note that we obtained similar confusion matrix for all pairs (γ,C)∈[0.5,2.5] × [3,30] for SVM.

We implemented the SVM using the Scikit Learn module (Pedregosa et al., 2011; Buitinck et al., 2013). Data analyses and classification codes were performed using Python.

Gaussian estimator

In case of a Gaussian estimator, we estimated the mean and the covariance matrix for the 2 categories of patients. The probability of each class is computed empirically using the maximum likelihood estimator (**Supplementary methods**). We recall that for an ensemble of n data $\mathcal{S}_n = (x_1, ..x_n)$ that are separated into two classes, C_1 and

C_2, the probability that a patient X belongs to one class, conditioned on the ensemble $\mathcal{S}_n$:

$$p\left(X \in C_1 \mid X = x, \mathcal{S}_n\right) =$$

$$\frac{1}{1 + \frac{1-\Pi}{\Pi}\frac{|\Sigma_1|^{\frac{1}{2}}}{|\Sigma_1|^{\frac{1}{2}}} exp\ \left(-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right)} \tag{7}$$

where $(\mu_i, \Sigma_i)_{i=1,2}$ are the mean and variance computed from each class C_1 and C_2 from $\mathcal{S}_n$. We used the fraction $\pi = \frac{n_s}{n_s+n_d} =$, n_s for the number of patients with good neurological outcome at discharge and n_d for the other patients. Formula 7 is derived in the **Supplementary methods**.

K-nearest neighbors classifier and weighted K-nearest neighbors

To classify the standard stimuli, we used the K-nearest neighbors classifier. We computed the ratio for the probability of belonging to a class. For a given point X, the probability to belong to class C_1 ("good neurological outcome") given the distribution of point x is computed empirically as the number of neighbors out of a total of K.

$$p\left(X \in C_1 \mid x\right) = \frac{k_r}{K} \tag{8}$$

where k_r is the number of neighbors that belong to the class "good neurological outcome at discharge" among K closest points.

To classify deviant stimuli, we used a variant of the K-neighbors method by adding distance-relative weights to the points inside the dataset. The two classes labeled "bad neurological outcome" and "good neurological outcome" are defined as C_1 and $C_{2,}$ respectively. The ensemble of points $\mathcal{S}_n$ in dimension 2 are given by the coordinates $x_n = \left(N_{E,n}, \Delta V_n\right)$, extracted in subsection "Analysis of responses to deviant stimuli." To compute the classification probability, we defined K-nearest neighborhood $\mathcal{N}_K(x)$ for the point x as the K shortest points from x, computed from the Euclidean distance (between two points x_n, x_m),

$$d\left(x_n, x_m\right) = \sqrt{\left(N_{E,n} - N_{E,m}\right)^2 + \left(\Delta V_n - \Delta V_m\right)^2} \tag{9}$$

$$\mathcal{N}_K\left(x\right) = \left\{y_1, ..y_K \in \mathcal{S}_n, d\left(x, y_1\right) \le d\left(x, y_1\right) .. \le d\left(x, y_K\right)\right\}. \tag{10}$$

To obtain an accurate classifier, we used a different version of the K-neighbors classification, where the weight depends on the distance between the point to classify and the K-nearest neighbors (formula 11), defined by

$$p\left(x \in C_1 \mid x\right) = \frac{\sum_{i=1}^{N} \frac{1_{y_i \in C_1}}{d(y_i, x)}}{\sum_{i=1}^{N} \frac{1}{d(y_i, x)}}. \tag{11}$$

Cross-validation

We used a Leave-one-out cross-validation approach to validate the classification algorithm: we excluded a patient at a time and computed the probability of a good neurological outcome at discharge, based on the remaining elements in the data basis (Kohavi, 1995). In other words, we separated the patient database into a testing and a training group, with one patient out, 28 in the other group and ran this test 28 times so that each of the 29 patients was alternatively included in the 1 group patient. We then computed this probability using the three classifiers, SVM, Gaussian estimators and K-neighbors and compared the result to the true result. We followed the protocol: 1- a patient $P_i, i = 1..N$ is selected inside the data basis; 2- we trained the classification algorithm on the database of all patients $\{P_k, k = 1..N\} - P_i$. We evaluated the prediction of the model on the excluded patient, leading to a score s_i. We recall that $s_i = 1$ if the prediction is correct, otherwise, $s_i = 0$. We then replaced the patient P_i inside the database and reiterated the procedure until each patient has been exactly excluded once. This allowed us to reclassify with a given probability for each patient outcome based on the new map determined by the other patients. The final score of the model is computed as

$$s = \frac{1}{n}\sum_{1}^{n} s_i. \tag{12}$$

Finally, the confusion matrix defined as

$$C = \begin{pmatrix} T_p & F_N \\ F_p & T_N \end{pmatrix} \tag{13}$$

for the true positive T_p (number of patients who have a good neurological outcome at discharge and are classified correctly), true negative T_n (number of patients who have a bad neurological outcome and are classified correctly) and false positive F_p (number of patients who have a good neurological outcome and are classified incorrectly) and false negative F_n (number of patients who have a bad neurological outcome at discharge and are classified incorrectly). We calculated for each of the classifiers accuracy, sensitivity and specificity.

Combined probability for outcome decision

We proposed to use for the predictive decisional probability p_{dec} the minimum of the ones estimated for the standard (relation 8) and deviant (relation 11) classifications. For a patient of coordinate x in each map, survival probability is:

$$p_{dec}(x \in C_1|x) = \min\left(p_{dev}(x \in C_1|x)\right), p_{non-dev}(x \in C_1|x)). \tag{14}$$

Iteration and changing k-neighbors k

The approach developed here is iterative and any new additional case enriches the database and the classifications maps. For the K-neighbors approach, adding a point does not require any changes in the computation, although we expect that the number of neighbors that will enter progressively into the computation could diminish as the number of cases added in the map increases. For the Gaussian classification, the mean and the variance are recomputed following each new case.

Results

Overall patient characteristics

Data of twenty-nine consecutive patients were analyzed. Seven patients out of twenty-nine survived, but only 6 out of 7 were able to follow verbal command at hospital discharge. None of the patients was lost of follow-up. The last patient returned home but the degree of disability is unknown. At 3–6 months, GOS-E was scored at 3 for the patient who was unable to follow verbal command at discharge

TABLE 1 Comparison of clinical and electrophysiological characteristics between the two groups.

	Bad neurological outcome (*n* = 23 unless otherwise specified) (/29)	Good neurological outcome (*n* = 6) (/29)	*p*
Age (years), mean ± SD [min-max]	60 ± 16 [24–87]	47.5 ± 16 [26–64]	0.07
Median [IQR 25–75]	62 [54–68.5]	52.5 [34.75–59]	
Male, *n*	20	5	1
Shockable rhythm*, *n*	6	2	1
Etiology			0.57
Cardiac	10	4	
Non-cardiac	12	2	
Undetermined	1	0	
No-flow (minutes), mean ± SD	8.3 ± 8.4 (/21)	2.7 ± 4.3	0.07
Median [IQR 25–75]	7 [1–15]	0 [0–4.5]	
Low-flow (minutes), mean ± SD	24.9 ± 16.4	16.4 ± 11.7	0.22
Median [IQR 25–75]	20 [18.5–35]	15.5 [10.75–21]	
Time to ROSC (≤25 min), *n*	8 (/21)	4	0.20
GCS on admission /15), mean ± SD	3.1 ± 0.4	3.5 ± 1.2	0.21
Median [IQR 25–75]	3 [3–3]	3 [3–3]	
SAPS II score, mean ± SD	73 ± 15	62.5 ± 18	0.15
Median [IQR 25–75]	72 [63–85]	54 [49–76]	
Sedation, *n*	8	3	0.65
EEG Grade I: predominant alpha with some theta, *n*	0	1	
EEG Grade II: predominant theta with some alpha, *n*	0	0	
EEG Grade III: predominant theta, *n*	3	5	*0.0002*
EEG Grade IV: delta activity, *n*	7	0	
Generalized epileptiform periodic activity (GPEDs), *n*	6	0	
EEG Grade V electrocerebral silence, *n*	3	0	
Burst suppression patterns, *n*	4	0	
EEG reactivity, *n*	3	2	0.27
SSEP (N20 -), *n*	7 (/22)	1/5	0.64
AEP (MMN+), *n*	4	2	0.57
GOS-E (6 months) (*n*)	3 (1/23)	4–8 (5/6)	

*As the first documented rhythm; ROSC, return of spontaneous circulation; GCS, Glasgow Coma Scale; SAPS II score, simplified acute physiology score II; EEG, electroencephalography patterns according to the five major grades of severity scale for brain injury; SSEP, cortical somatosensory evoked potentials; AEP, auditory evoked potentials; MMN, mismatch negativity. No-flow data were missing in two patients and SSEP (N20 response) data in one patient (underlying Charcot Marie Tooth disease). GOS-E, Glasgow Outcome Scale-Extended.

and died 27 months later without neurological improvement. GOS-E was scored at 4 for one patient, at 5 for one patient, at 6 for one patient and at 8 for the last two patients. Age, sex, medical history, characteristics of CA and electrophysiological features are presented in Table 1. At the time of recording, all patients were still hypothermic (<35°C). Sedation was present in 11 out of the 29 patients (38%) at the moment of the electrophysiological recordings. For the non-surviving patients, 18 out of 22 died after withdrawal of life-sustaining therapies.

All six patients with good neurological outcome presented an EEG pattern graded between I to III for all, whereas 20 out of the 23 of the patients with final bad neurological outcome or death presented an EEG pattern graded IV or V ($p < 0.0002$), including the patient who survived 27 months with bad neurological outcome. EEG reactivity (2/6 versus 3/23) and presence of MMN (2/6 versus 4/23) were more frequent in the group with good neurological outcome, whereas N20 was less frequently absent (1/5 versus 7/22), but none of these last markers were statistically different between the two groups. Only 2 patients presented congruent favorable prognostic factors with a present N20 at SSEP, a positive MMN and EEG pattern graded I to III (areactive EEG for both), among whom one patient did not survive. By contrast, four patients presented congruent bad prognosis factors with absent N20, absent MMN and an EEG pattern graded IV or V and all of them died. ERP obtained at Cz location were the most reproducible and the only ones used for visual analysis. Artifacts prevented the interpretation of SSEP in one patient of each group.

Prognosis map constructed from bayesian statistical inference

Since each parameter taken individually for standard (standard deviation and similarity) and deviant (number of extremum N_E and oscillation$|\Delta V|$) responses were not sufficient to obtain a clear separation between the two patient categories, we decided to

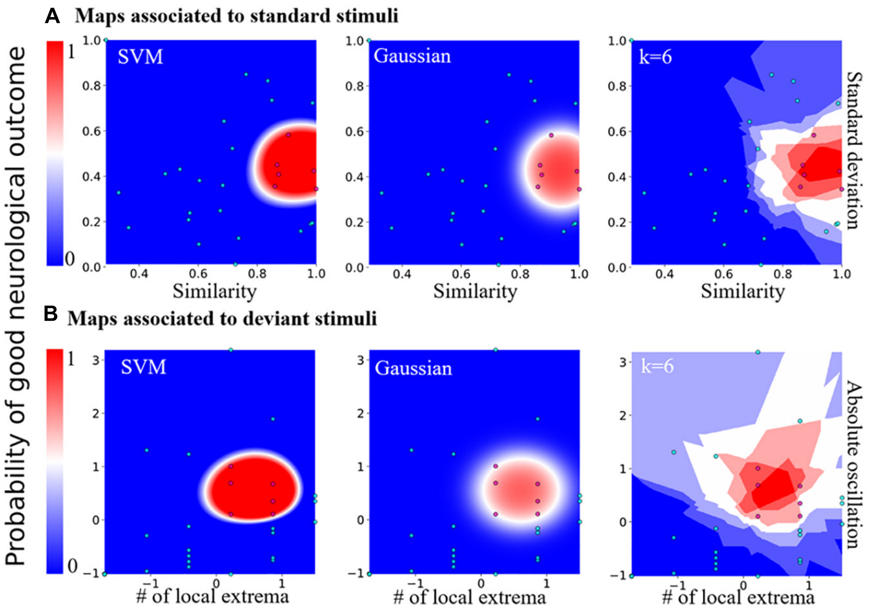


FIGURE 4
Predictive probability maps of good neurological outcome. **(A)** Probability maps computed from features of the standard stimuli responses. From left to right: maps computed from SVM, Gaussian, and the k-nearest neighbors classifier (k = 6, the worst case scenario). **(B)** Probability maps computed from features of the deviant stimuli features. From left to right: maps computed from SVM, Gaussian, and the k-nearest neighbors classifier with distance-related weights, k = 6 for example.

combine them into a two-dimensional map (Figure 4). Interestingly, we found that this map allowed a clear separation that we quantified using various *a priori* classifiers: SVM, Gaussian, and the K-neighbor classifiers (Hastie et al., 2001; Table 2). When mapping all the features first taken individually for standard and deviant responses, we found a cluster formed of patients with a "good" neurological outcome, bounded in red, well separated from the area in which were found the other patients (non-surviving or "bad" neurological outcome). This partition between two distinct areas was present in all classifiers: SVM, Gaussian, and k-neighbors, confirming that this partition was robust independently of the choice of the classification methods (Supplementary Figure 1 for other choices of k for the k-neighbor algorithm). Moreover, we found a similar partition into two categories of patients when classifying the standard or the deviant responses, which strengthens the robustness of our study (Figure 4). The present classification maps for both standard and deviant responses studied separately showed that the neurological outcome of post-anoxic comas can be predicted (Table 2). Combining the probability computed in each map, we proposed a decision probability with a high specificity, which does not misclassify patients with good neurological outcome in the category of patients with bad neurological outcome.

Classification efficiency of the two-dimensional maps

To test the predictive strength of the standard and deviant responses classification, we computed the confusion matrix (formula 13) as described in the methods. The confusion matrix computed for the Gaussian estimator showed a 89 % accuracy, and a 100% validation accuracy for the SVM classifier. The confusion matrix computed for the k-neighbors classifier showed that it was less performant than the SVM classifier. The sensitivity remained high and could be improved with the increasing number of classified patients (k = 4; similar results were obtained for other values of k). The distance-dependent weight showed this estimator introduces type I error, with an accuracy of 0.90, a sensitivity of 0.83 and a specificity of 0.91.

Finally, we also computed the confusion matrix obtained from a visual analysis of patient MMN, performed by a medical professional, and we obtained an accuracy of 0.72, a sensitivity of 0.33 and a specificity of 0.82. If MMN remained an interesting indicator, it showed a very weak sensitivity in these patients (Tables 2, 3).

Discussion

Our exploratory study was designed to identify mathematical parameters extracted from the AEP recording that could be more powerful than visual inspection of MMN in the routine ICU setting and used to predict neurological prognosis in these patients. The originality of the present strategy was to consider independently deviant from standard responses, not only in the time window of the mismatch negativity (that results from the difference between the two responses), but using the total amount of information that is generated during the procedure. We found that our new classification method, combining standard deviation and similarity (correlation) for standard auditory stimuli, and number of extrema and oscillations for deviant auditory stimuli, allowed clustering patients in two-dimensions, in one of the two categories of good or bad neurological prognosis. Importantly, we did not select these parameters *a priori* to obtain a best separation of patients as explained in the method's section.

TABLE 2 Accuracy, sensitivity, and specificity obtained by cross-validation for responses to standard, deviant stimuli; k is the number of neighbors used in the classification algorithm.

	k = 3	k = 4	k = 6	k = 8	SVM	Gaussian
Accuracy						
Standard responses	0.96	0.96	0.89	0.85	1.0	0.92
Deviant responses	0.89	0.89	0.89	0.89	1.0	0.89
Sensitivity						
Standard responses	0.83	0.83	0.83	0.66	1.0	0.66
Deviant responses	0.83	0.83	0.83	0.83	1.0	0.5
Specificity						
Standard responses	1.0	1.0	0.9	0.9	1.0	1.0
Deviant responses	0.91	0.91	0.91	0.91	1.0	1.0

TABLE 3 Classifications scores.

	Accuracy	Sensitivity	Specificity
SVM	1	1.0	1
Gaussian	0.89	0.5	1
k-neighbors	0.9	0.83	0.91
MMN	0.72	0.33	0.82

To evaluate the robustness of our method, we used three classifiers, showing similar maps classification results. Finally, using leave-one-out cross-validation, we computed a score for each classifier, demonstrating that any of the three classification methods was more robust than simply analyzing the MMN in a binary response, using logistic regression or single-trial topographic analysis (De Lucia and Tzovara, 2015). We showed that good neurological prognosis probability maps allow us to predict the neurological outcome of post-anoxic comatose patients with a very good accuracy of 0.90, sensitivity of 0.83 and specificity of 0.91 when considering the least efficient classifier (**Tables 2**, **3**).

We have used the standard deviation and the similarity index to analyze the standard responses, while we used the number of extrema and oscillations for the deviant in order to have two independent set of parameters and increase the robustness of the results, and preventing a potential bias by choosing a single set of parameters. We could have decided to use these two latest parameters in this study for all cases or use all four parameters that could have led to a more robust result, but also to a four-dimensional classification, that we wanted to avoid in order to obtain an easy-to-use tool. Moreover, the standard deviation and the similarity index would not really be appropriate to study the deviant sounds.

We can consider three other developments that could be built on this present investigation. The first one is to evaluate if repeating this procedure with this algorithm several days apart can present a potential additive value, as explained in Tzovara et al. (2013) who showed the additional prognostic value of repeating MMN. The second one is to test whether this procedure could be generalized to other auditory oddball paradigms. At last, it would be interesting to evaluate whether such a method could be applied to classical electroencephalography with more sparse auditory and nociceptive stimuli than the one developed here using auditory evoked potentials with frequent and regular auditory stimuli. Indeed, electroencephalography is a neurophysiological tool which is more widespread than auditory evoked potentials. Characterizing electrophysiological features to predict the outcome of post-anoxic coma remains a genuine challenge. There is currently no satisfactory, efficient and simple tool to predict comatose patient outcome accurately, especially at the acute phase, when patients are sedated and/or hypothermic. Standard electroencephalography is the most common method used to predict prognosis in those patients. If highly malignant pattern (suppressed background discharges without discharges or with continuous periodic discharges, or burst suppression background with or without discharges) is highly specific of poor outcome, as shown in our study, it has a sensitivity of only 50% (Westhall et al., 2016). The absence of cortical N20 response at SSEP after stimulation of median nerves has an almost 100% specificity for non-awakening prediction (Sandroni et al., 2014). By contrast, the predictive value of the visual analysis of MMN for post-CA comatose patients, limited to a binary response (presence/absence of a detectable peak of the MMN between the standard and deviant responses) is poorly sensitive, as shown in our study, even when choosing parameters that better discriminate standard and deviant sounds (Azabou et al., 2018). To overcome, the poor sensitivity of MMN at visual analysis, several statistical methods have been developed. Some are based on sample-by-sample paired t-test in the specific time window where MMN is ussually visualized. Others are based on wavelet transform, multivariate, cross-correlation and probabilistic methods (Fischer et al., 1999; Naccache et al., 2005, 2015; Daltrozzo et al., 2007; De Lucia and Tzovara, 2015, 2016; Gabriel et al., 2016; Juan et al., 2016). Tzovara et al. (2013) choose an alternative strategy: they showed that the progression of MMN auditory discrimination (and not one single analysis) over the first 2 days of coma was of good prognosis, suggesting that collecting repetitive data within days, or at an earlier phase, could reveal changes that could have a higher predictive value. Overall, this explains why a multimodal prognostication approach is still recommended in these patients, including clinical examination, serum biomarkers and brain imaging in addition to electrophysiological recordings (Sandroni et al., 2014; Nolan et al., 2021).

In that small series, none of the classical electrophysiological tools were sensitive or specific enough to give a reliable neurological prognosis. Only 2 patients presented congruent favorable prognostic factors with a present N20 at SSEP, a positive MMN and EEG pattern graded I to III (benign pattern according to the ACNS EEG terminology) (Westhall et al., 2015) and areactive for both, among whom one patient did not survive. By contrast, four patients presented congruent bad prognosis factors with absent N20, absent MMN and an EEG pattern graded IV or V (highly malignant pattern according to the ACNS EEG terminology) and all of them died, suggesting that congruent pejorative factors are strongest indicators of prognosis than congruent good prognosis factors, in accordance with literature. It is to note that one third of the patients with bad outcome and 50% of the patients with good outcome were under sedation at the time of recording, which is known to impede electrophysiology interpretation. Our study was not designed to compare our tool with classical electrophysiological examinations but sensitivity and specificity were higher in that small cohort that needs to be validated in a larger cohort. The total amount of data we collected for all epochs during the 20 min of auditory stimulations and not only during the time window used for MMN might explain our more sensitive results.

Our present study has several limitations. First, as a retrospective study, neurological prognosis was evaluated on the ability of the patient to follow verbal command at discharge, which remains a subjective assessment that may have led to patient's misclassification. However, in the 7 surviving patients, GOS-E was available for 6 of them at 3–6 months post-discharge and was found at 3 in the patient who was initially unable to follow verbal command and from 4 to 8 for the others, indicating that no patient was initially misclassified. Second, this cohort may not be representative of all post CA patients since electrophysiological assessment was performed relatively late, up to 6 days after admission, in patients still comatose at the time of the evaluation, and the relatively small sample size prevents generalization of our results that need to be replicated in a larger cohort. Third, our cohort between patients with good and bad prognosis was unbalanced, that we tried to offset using a leave-one out cross validation. Fourth, our new approach did not consider the order of the different sounds. For instance, a standard sound that would start a new sequence just after a deviant sound or ending a series of standard sounds just before a deviant sound, may not be processed the same. This point could deserve a specific attention in future studies, but as we averaged all our data, this probably does not bias our results.

To conclude, we developed a new promising classification method that could be self-sufficient, easily used by intensivists (only one electrode, with minimal cost and easy training), without the help of the neurophysiologists and in sedated and/or hypothermic patients, since these conditions represent actual limitations to electrophysiological data acquisition in the ICU. Moreover, electrophysiological recordings may be particularly difficult to acquire at the acute phase where patients combine aggressive care (extracorporeal membrane oxygenation (ECMO), haemodialysis, mechanical ventilation), and invasive methods of monitoring, generating artifacts. Finally, potential amplitudes are smaller under sedation and more difficult to extract from the background (Yppärilä et al., 2004). Our preliminary results suggest that all these issues could be addressed by this new method. The produced maps can be refined and upgraded by adding new cases and thus increase the performance of the probabilistic classifier. In the future, and according to the local human and logistical resources, the software could be implemented with other electrophysiological and clinical variables to provide an optimal estimated probability of the patient outcome, independently from neurophysiologists. Developing such algorithms, ready-to-use by the intensivits, would enable more aggressive management in patients with predicted good neurological outcome. Whether this approach could be secondarily applied to other predictive situations and generalized to other comas remains to be validated.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee (Comité de Protection des Personnes, CPP Paris IV #2012/22) (PHRC CAPACITY AOR10109). Informed consent was provided by next-of-kin for all participants as they were in a coma most of the time until their death, it was followed whenever possible, by informed consent from the patient.

Author contributions

AF, AD, and DH created the algorithm. SL, ND, and BM collected the clinical data. AR, EM, SM, PL, and NK performed the electrophysiological examinations. DH and NK wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study was funded by a grant from Programme Hospitalier de Recherche Clinique - PHRC CAPACITY AOR10109.

Conflict of interest

AF, AR, NK, and DH have a patent application for the prediction of coma outcome (French patent FR1852473, titled "Outil prédictif de la sortie du coma des patients après un arrêt cardio-respiratoire").

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.988394/full#supplementary-material

References

Admiraal, M. M., Horn, J., Hofmeijer, J., Hoedemaekers, C. W. E., van Kaam, C. R., Keijzer, H. M., et al. (2020). EEG reactivity testing for prediction of good outcome in patients after cardiac arrest. *Neurology* 95, e653–e661.

Aizerman, A. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automat. Remote Control.* 25, 821–837.

André-Obadia, N., Zyss, J., Gavaret, M., Lefaucheur, J.-P., Azabou, E., Boulogne, S., et al. (2018). Recommendations for the use of electroencephalography and evoked potentials in comatose patients. *Neurophysiol. Clin.* 48, 143–169.

Azabou, E., Rohaut, B., Porcher, R., Heming, N., Kandelman, S., Allary, J., et al. (2018). Mismatch negativity to predict subsequent awakening in deeply sedated critically ill patients. *Br. J. Anaesth.* 121, 1290–1297. doi: 10.1016/j.bja.2018.06.029

Bassetti, C., Bomio, F., Mathis, J., and Hess, C. W. (1996). Early prognosis in coma after cardiac arrest: a prospective clinical, electrophysiological, and biochemical study of 60 patients. *J. Neurol. Neurosurg. Psychiatry* 61, 610–615. doi: 10.1136/jnnp.61.6.610

Bouchereau, E., Marchi, A., Hermann, B., Pruvost-Robieux, E., Guinard, E., Legouy, C., et al. (2022). Quantitative analysis of early-stage EEG reactivity predicts awakening and recovery of consciousness in patients with severe brain injury. *Br. J. Anaesth.* 130, e225–e232. doi: 10.1016/j.bja.2022.09.005

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., et al. (2013). API design for machine learning software: experiences from the scikit-learn project. *arXiv [Preprint]*. Available online at: http://arxiv.org/abs/1309.0238 (accessed November 14, 2020).

Carr, B. G., Goyal, M., Band, R. A., Gaieski, D. F., Abella, B. S., Merchant, R. M., et al. (2009). A national analysis of the relationship between hospital factors and post-cardiac arrest mortality. *Intensive Care Med.* 35, 505–511. doi: 10.1007/s00134-008-1335-x

Chausson, N., Wassouf, A., Pegado, F., Willer, J.-C., and Naccache, L. (2008). [Electrophysiology: mismatch negativity and prognosis of coma]. *Rev. Neurol.* 164, F34–F35.

Cherkassky, V., and Mulier, F. (1998). *Learning from Data: Concepts, Theory, and Method*. New York, NY: Wiley.

Chin, Y., Yaow, C., Teoh, S., Foo, M., Luo, N., Graves, N., et al. (2022). Long-term outcomes after out-of-hospital cardiac arrest: a systematic review and meta-analysis. *Resuscitation* 171, 15–29. doi: 10.1016/j.resuscitation.2021.12.026

Comanducci, A., Boly, M., Claassen, J., De Lucia, M., Gibson, R. M., Juan, E., et al. (2020). Clinical and advanced neurophysiology in the prognostic and diagnostic evaluation of disorders of consciousness: review of an IFCN-endorsed expert group. *Clin. Neurophysiol.* 131, 2736–2765. doi: 10.1016/j.clinph.2020.07.015

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Daltrozzo, J., Wioland, N., Mutschler, V., and Kotchoubey, B. (2007). Predicting coma and other low responsive patients outcome using event-related brain potentials: a meta-analysis. *Clin. Neurophysiol.* 118, 606–614. doi: 10.1016/j.clinph.2006.11.019

De Lucia, M., and Tzovara, A. (2015). Decoding auditory EEG responses in healthy and clinical populations: a comparative study. *J. Neurosci. Methods* 250, 106–113. doi: 10.1016/j.jneumeth.2014.10.019

De Lucia, M., and Tzovara, A. (2016). Reply: replicability and impact of statistics in the detection of neural responses of consciousness. *Brain* 139:e32. doi: 10.1093/brain/aww063

Duez, C. H. V., Ebbesen, M. Q., Benedek, K., Fabricius, M., Atkins, M. D., Beniczky, S., et al. (2018). Large inter-rater variability on EEG-reactivity is improved by a novel quantitative method. *Clin. Neurophysiol.* 129, 724–730. doi: 10.1016/j.clinph.2018.01.054

Fischer, C., Luauté, J., Némoz, C., Morlet, D., Kirkorian, G., and Mauguière, F. (2006). Improved prediction of awakening or nonawakening from severe anoxic coma using tree-based classification analysis. *Crit. Care Med.* 34, 1520–1524.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., and Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clin. Neurophysiol.* 110, 1601–1610. doi: 10.1016/S1388-2457(99)00131-5

Gabriel, D., Muzard, E., Henriques, J., Mignot, C., Pazart, L., André-Obadia, N., et al. (2016). Replicability and impact of statistics in the detection of neural responses of consciousness. *Brain* 139:e30. doi: 10.1093/brain/aww065

Hastie, T., Firedman, J., and Tibshirani, R. (2001). *The Elements of Statistical Learning. Springer Series in Statistics*. New York, NY: Springer.

Juan, E., De Lucia, M., Tzovara, A., Beaud, V., Oddo, M., Clarke, S., et al. (2016). Prediction of cognitive outcome based on the progression of auditory discrimination during coma. *Resuscitation* 106, 89–95. doi: 10.1016/j.resuscitation.2016.06.032

Kim, J., Choi, B. S., Kim, K., Jung, C., Lee, J. H., Jo, Y. H., et al. (2012). Prognostic performance of diffusion-weighted MRI combined with NSE in comatose cardiac arrest survivors treated with mild hypothermia. *Neurocrit. Care* 17, 412–420. doi: 10.1007/s12028-012-9773-2

Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. Montreal, QC: ACM Digital Library, 1137–1145.

Le Gall, J., Neumann, A., Hemery, F., Bleriot, J., Fulgencio, J., Garrigues, B., et al. (2005). Mortality prediction using SAPS II: an update for French intensive care units. *Crit. Care* 9, R645–R652. doi: 10.1186/cc3821

Miraglia, D., Miguel, L., and Alonso, W. (2020). Long-term neurologically intact survival after extracorporeal cardiopulmonary resuscitation for in-hospital or out-of-hospital cardiac arrest: a systematic review and meta-analysis. *Resusc. Plus* 4:100045. doi: 10.1016/j.resplu.2020.100045

Naccache, L., King, J.-R., Sitt, J., Engemann, D., El Karoui, I., Rohaut, B., et al. (2015). Neural detection of complex sound sequences or of statistical regularities in the absence of consciousness? *Brain* 138:e395. doi: 10.1093/brain/awv190

Naccache, L., Puybasset, L., Gaillard, R., Serve, E., and Willer, J.-C. (2005). Auditory mismatch negativity is a good predictor of awakening in comatose patients: a fast and reliable procedure. *Clin. Neurophysiol.* 116, 988–989. doi: 10.1016/j.clinph.2004.10.009

Nolan, J. P., Sandroni, C., Böttiger, B. W., Cariou, A., Cronberg, T., Friberg, H., et al. (2021). European resuscitation council and European society of intensive care medicine guidelines 2021: post-resuscitation care. *Intensive Care Med.* 47, 369–421. doi: 10.1007/s00134-021-06368-4

Oddo, M., Ribordy, V., Feihl, F., Rossetti, A. O., Schaller, M.-D., Chioléro, R., et al. (2008). Early predictors of outcome in comatose survivors of ventricular fibrillation and non-ventricular fibrillation cardiac arrest treated with hypothermia: a prospective study. *Crit. Care Med.* 36, 2296–2301. doi: 10.1097/CCM.0b013e3181802599

Oddo, M., and Rossetti, A. O. (2014). Early multimodal outcome prediction after cardiac arrest in patients treated with hypothermia. *Crit. Care Med.* 42, 1340–1347. doi: 10.1097/CCM.0000000000000211

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Pell, J. P. (2003). Presentation, management, and outcome of out of hospital cardiopulmonary arrest: comparison by underlying aetiology. *Heart* 89, 839–842. doi: 10.1136/heart.89.8.839

Perkins, G. D., Jacobs, I. G., Nadkarni, V. M., Berg, R. A., Bhanji, F., Biarent, D., et al. (2015). Cardiac Arrest and cardiopulmonary resuscitation outcome reports: update of the utstein resuscitation registry templates for out-of-hospital cardiac arrest: a statement for healthcare professionals from a task force of the international liaison committee on resuscitation (American Heart Association, European Resuscitation Council, Australian and New Zealand Council on Resuscitation, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Council of Southern Africa, Resuscitation Council of Asia); and the American heart association emergency cardiovascular care committee and the council on cardiopulmonary, critical care, perioperative and resuscitation. *Circulation* 132, 1286–1300. doi: 10.1161/CIR.0000000000000144

Pfeiffer, C., Nguissi, N. A. N., Chytiris, M., Bidlingmeyer, P., Haenggi, M., Kurmann, R., et al. (2017). Auditory discrimination improvement predicts awakening of postanoxic comatose patients treated with targeted temperature management at 36 °C. *Resuscitation* 118, 89–95. doi: 10.1016/j.resuscitation.2017.07.012

Roberts, B. W., Kilgannon, J. H., Chansky, M. E., Mittal, N., Wooden, J., Parrillo, J. E., et al. (2013). Multiple organ dysfunction after return of spontaneous circulation in postcardiac arrest syndrome. *Crit. Care Med.* 41, 1492–1501. doi: 10.1097/CCM.0b013e31828a39e9

Rohaut, B., Faugeras, F., Bekinschtein, T.-A., Wassouf, A., Chausson, N., Dehaene, S., et al. (2009). Prédiction du réveil et détection de la conscience: intérêt des potentiels évoqués cognitifs. *Réanimation* 18, 659–663. doi: 10.1016/j.reaurg.2009.06.014

Rossetti, A. O., Rabinstein, A. A., and Oddo, M. (2016). Neurological prognostication of outcome in patients in coma after cardiac arrest. *Lancet Neurol.* 15, 597–609. doi: 10.1016/S1474-4422(16)00015-6

Sandroni, C., Cariou, A., Cavallaro, F., Cronberg, T., Friberg, H., Hoedemaekers, C., et al. (2014). Prognostication in comatose survivors of cardiac arrest: an advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. *Intensive Care Med.* 40, 1816–1831. doi: 10.1007/s00134-014-3470-x

Sfar, S. (2007). Prise en charge de l'arrêt cardiaque. *Ann. Fran. Anesth. Réanimat.* 26, 1008–1019. doi: 10.1016/j.annfar.2008.02.003

Société de réanimation de langue française. (2010). Limitation et arrêt des traitements en réanimation adulte. Actualisation des recommandations de la Société de réanimation de langue française. *Réanimation* 19, 679–698. doi: 10.1016/j.reaurg.2010.07.001

Synek, V. M. (1988). Prognostically important EEG coma patterns in diffuse anoxic and traumatic encephalopathies in adults. *J. Clin. Neurophysiol.* 5, 161–174. doi: 10.1097/00004691-198804000-00003

Tzovara, A., Rossetti, A. O., Spierer, L., Grivel, J., Murray, M. M., Oddo, M., et al. (2013). Progression of auditory discrimination based on neural decoding predicts awakening from coma. *Brain* 136, 81–89. doi: 10.1093/brain/aws264

Valiant, L. G. (1984). A theory of the learnable. *Commun. ACM* 27, 1134–1142. doi: 10.1145/1968.1972

Westhall, E., Rosén, I., Rossetti, A. O., van Rootselaar, A.-F., Wesenberg Kjaer, T., Friberg, H., et al. (2015). Interrater variability of EEG interpretation in comatose cardiac arrest patients. *Clin. Neurophysiol.* 126, 2397–2404. doi: 10.1016/j.clinph.2015.03.017

Westhall, E., Rossetti, A. O., van Rootselaar, A.-F., Wesenberg Kjaer, T., Horn, J., Ullén, S., et al. (2016). Standardized EEG interpretation accurately predicts prognosis after cardiac arrest. *Neurology* 86, 1482–1490. doi: 10.1212/WNL.0000000000002462

Yppärilä, H., Nunes, S., Korhonen, I., Partanen, J., and Ruokonen, E. (2004). The effect of interruption to propofol sedation on auditory event-related potentials and electroencephalogram in intensive care patients. *Crit. Care* 8, R483–R490. doi: 10.1186/cc2984

TYPE Original Research
PUBLISHED 02 March 2023
DOI 10.3389/fncom.2023.1108311

Check for updates

OPEN ACCESS

EDITED BY
Benjamin R. Pittman-Polletta,
Boston University, United States

REVIEWED BY
Stanley E. Lazic,
Prioris.ai Inc., Canada
James J. Foster,
University of Konstanz, Germany

*CORRESPONDENCE
Maura Mezzetti
✉ maura.mezzetti@uniroma2.it
Priscilla Balestrucci
✉ priscilla.balestrucci@uni-ulm.de

RECEIVED 25 November 2022
ACCEPTED 03 February 2023
PUBLISHED 02 March 2023

CITATION
Mezzetti M, Ryan CP, Balestrucci P, Lacquaniti F and Moscatelli A (2023) Bayesian hierarchical models and prior elicitation for fitting psychometric functions.
Front. Comput. Neurosci. 17:1108311.
doi: 10.3389/fncom.2023.1108311

Bayesian hierarchical models and prior elicitation for fitting psychometric functions

Maura Mezzetti[1]*, Colleen P. Ryan[2,3], Priscilla Balestrucci[4]*, Francesco Lacquaniti[2,3] and Alessandro Moscatelli[2,3]

[1]Department Economics and Finance, University of Rome "Tor Vergata", Rome, Italy, [2]Department of Systems Medicine and Centre of Space Bio-Medicine, University of Rome "Tor Vergata", Rome, Italy, [3]Department of Neuromotor Physiology, Istituto di Ricovero e Cura a Carattere Scientifico Santa Lucia Foundation, Rome, Italy, [4]Applied Cognitive Psychology, Ulm University, Ulm, Germany

Our previous articles demonstrated how to analyze psychophysical data from a group of participants using generalized linear mixed models (GLMM) and two-level methods. The aim of this article is to revisit hierarchical models in a Bayesian framework. Bayesian models have been previously discussed for the analysis of psychometric functions although this approach is still seldom applied. The main advantage of using Bayesian models is that if the prior is informative, the uncertainty of the parameters is reduced through the combination of prior knowledge and the experimental data. Here, we evaluate uncertainties between and within participants through posterior distributions. To demonstrate the Bayesian approach, we re-analyzed data from two of our previous studies on the tactile discrimination of speed. We considered different methods to include *a priori* knowledge in the prior distribution, not only from the literature but also from previous experiments. A special type of Bayesian model, the power prior distribution, allowed us to modulate the weight of the prior, constructed from a first set of data, and use it to fit a second one. Bayesian models estimated the probability distributions of the parameters of interest that convey information about the effects of the experimental variables, their uncertainty, and the reliability of individual participants. We implemented these models using the software Just Another Gibbs Sampler (JAGS) that we interfaced with R with the package rjags. The Bayesian hierarchical model will provide a promising and powerful method for the analysis of psychometric functions in psychophysical experiments.

KEYWORDS

psychophysics, PSE, generalized linear mixed models, Bayesian model, psychometric functions

1. Introduction

Psychophysical methods are largely used in behavioral neuroscience to investigate the functional basis of perception in humans and other animals (Pelli and Farell, 1995). Using a model called the psychometric function, it is possible to test the quantitative relation between a physical property of the stimulus and its perceptual representation provided by the senses. This model has a typical sigmoid shape and relates the actual stimulus intensity ("physics") on the abscissa to the probability of the response of the observer (i.e., perceptual response and "psychology") on the ordinate, as collected with a forced-choice experiment. It is possible to summarize the performance of an observer by the parameters that are computed by the psychometric function: the Point of Subjective Equality (PSE), the slope, and the Just Noticeable Difference (JND) (Knoblauch and Maloney, 2012). The PSE estimates the accuracy of the response and corresponds to the stimulus value associated with a probability of response at chance level ($p = 0.50$). In two-interval forced-choice experiments, a deviation of the PSE from the value of the reference stimulus may indicate a bias, for example, in perceptual illusions (Moscatelli et al., 2016, 2019). The JND measures the noise of the response; the higher the JND, the higher the perceptual noise (Prins, 2016). The JND is an inverse function of the slope parameter of the psychometric function that is a measurement of the precision of the response. It is possible to test the slope or the JND of the function to evaluate the precision (or the noise) of the response.

Typically, generalized linear models (GLMs) are applied to estimate the parameters of the psychometric functions for each individual participant (Knoblauch and Maloney, 2012). In our previous study, we showed the advantages of using generalized linear mixed models (GLMMs) to estimate the responses of multiple participants at the population level (Moscatelli et al., 2012). A fairly comprehensive literature on fitting GLM and GLMM exists in different programming languages, including R, Python, and Matlab (Linares and López-Moliner, 2016; Schütt et al., 2016; Moscatelli and Balestrucci, 2017; Prins and Kingdom, 2018; Balestrucci et al., 2022).

In GLMM, we distinguish between fixed- and random-effect parameters (Stroup, 2012). The former, akin to the parameters of the psychometric function, estimates the effects of the experimental variables. Typically, the random-effect parameters estimate the variability across individual participants. In more complex data-sets, it is possible to account for other sources of unobserved variation by means of random-effect parameters. Blocking or batch effects are common examples of other random-effects parameters. The addition of this random component is the distinguishing feature of mixed models. For GLMMs, we assume that the random-effect parameters are normally distributed variables. The goal is to estimate the variance of that distribution. The larger the variance, the larger the heterogeneity across participants for a given parameter. However, the mean (or other central tendencies) of that distribution can be treated as if fixed effects have been applied to standard models. The conditional modes of the model estimating the response of individual participants can be treated as the fixed effects in standard psychometric functions. For example, in Balestrucci et al. (2022), we used conditional modes to plot the model estimates for individual participants.

A natural reinterpretation of the mixed model is the Bayesian approach, where all parameters are naturally considered as random variables, each having its own probability distribution (Zhao et al., 2006; Fong et al., 2010). Bayesian models provide not only a point estimate but also a probability distribution of the population parameter. Therefore, a Bayesian approach allows a natural assessment of the uncertainty in the parameter estimation. The advantages of the hierarchical Bayesian framework have been established in different fields in experimental psychology (Gelman et al., 1995; Rouder et al., 2003) and item response (Fox and Glas, 2001; Wang et al., 2002). To the best of our knowledge, only a few studies evaluate the use of Bayesian inference for fitting psychometric functions (Alcalá-Quintana and García-Pérez, 2004; Kuss et al., 2005; Schütt et al., 2016; Houpt and Bittner, 2018). In addition to estimating the intercept and the slope of the model, the flexibility of a Bayesian approach allows the study of uncertainties of the PSE.

This article is organized as follows. In Section 2, the two-stage Bayesian hierarchical model is proposed and discussed. Section 2.1 focuses on the description of prior distributions and Section 2.2 is dedicated to the discussion of the computational aspects. In Section 3, the data from two published experiments are considered. In Section 3.1, a Bayesian hierarchical model is fitted and compared with the results of Dallmann et al. (2015), while in Section 3.2, a Bayesian hierarchical model is fitted and compared with the results of Picconi et al. (2022). In Section 4, the two studies considered in Section 3 are jointly analyzed. Two alternative approaches are proposed. The first one considers the combination of the two studies with the parameters from the first study used as a prior distribution. The second approach introduces a parameter, a_0, to quantify the uncertainty (or weight) of the first study that is considered as historical data—as detailed in Section 2.1. Finally, in Section 5, a discussion of the model is proposed and the results obtained are discussed.

2. Model

A typical data-set from a psychophysical experiment includes repeated responses from more than one participant. Fitting these types of data with ordinary generalized linear models (GLM) would produce invalid standard errors of the estimated parameters because they would treat the errors within the subject in the same manner as the errors between subjects. A viable approach to overcome this problem consists of applying a multilevel model (Morrone et al., 2005; Steele and Goldstein, 2006; Pariyadath and Eagleman, 2007; Johnston et al., 2008). First, the parameters of the psychometric function are estimated for each subject. Next, the individual estimates are pooled to perform the second-level analysis for statistical inference. Alternatively, it is possible to use the generalized linear mixed model (GLMM) that accounts separately for the experimental effects and the variability between participants using random- and fixed-effect parameters (Moscatelli et al., 2012).

Bayesian methods provide a viable solution for fitting models of the GLM and GLMM families (Gelman et al., 1995; Rouder and Lu, 2005). In particular, Kuss et al. (2005) have applied Bayesian methods for estimating a psychometric function, based on a binomial mixture model. A Bayesian hierarchical model is a statistical model written in multiple levels (hierarchical form) and estimates the parameters using Markov chain Monte Carlo (MCMC) sampling. Applying a Bayesian hierarchical model consists of the following processes: (i) model definition, including specification of parameters and prior distributions in different levels, (ii) update of the posterior distributions given the data, (iii) and Bayesian inference to analyze the parameters' posterior distributions (McElreath, 2020).

In the current study, we considered data from two-interval forced-choice discrimination tasks, as mentioned in the two example data-sets detailed in Sections 3.1 and 3.2. A two-stage Bayesian hierarchical model has been applied to these data-sets, with a probit model for each individual subject at the first stage. Let X denote the experimental variable (or variables), and let Y be the response variable that consists of binary responses. Thus $Y_{ij} = 1$ if subject i in trial j perceived a comparison stimulus with value x_{ij} as larger in magnitude (e.g., depending on the specific task, faster, stiffer, heavier, brighter, etc.) than a reference stimulus. As for the example data analyzed in this article (speed discrimination task), $Y_{ij} = 1$ if the subject perceived the comparison as "faster" than a reference one. The relationship between the response variable and the experimental variables is defined as:

$$Y_{ij} \sim Bern(\mu_{ij}) \tag{1}$$

$$\Phi^{-1}(\mu_{ij}) = \alpha_i + \beta_i x_{ij} \tag{2}$$

The model assumed that the forced-choice responses Y_{ij} are independent and identically distributed (i.i.d.) conditional on the individual parameters (α_i, β_i). In case of repeated measurement, for each subject and conditions, Equation (1) can easily be substituted by

$$Y_{ij} \sim Binom(\mu_{ij}, n_{ij}) \tag{3}$$

where Y_{ij} represents, the number of "faster" responses for subject i at condition x_{ij}.

The function Φ^{-1} in Equation (2) establishes a linear relationship between the response probability and the predictor that is fully described by two parameters α_i and β_i. The probit link function Φ^{-1} is the inverse of the cumulative distribution function of the standard normal distribution Z. That is:

$$\mu_{ij} = P\left(Z \leq \alpha_i + \beta_i x_{ij}\right) \qquad \forall i, j$$

For more details on probit link function refer to Agresti (2002) and Moscatelli et al. (2012). Other link functions like Logit and Weibull are also often used in psychophysics (Agresti, 2002; Foster and Zychaluk, 2009).

In the first stage, the model characterized the behavior of each individual participant i. The second level defines the model across all participants, similar to the GLMM described by Moscatelli et al. (2012). To this end, the second level estimates the overall effects across subjects by combining individual-specific effects. The parameters (a, b) describe the overall model and results from the combination of the subject-specific parameters, taking into account their uncertainties. Through a Bayesian hierarchical approach, the second level takes into account the uncertainties of the subject-specific parameters. It assumes the following distributions:

$$\alpha_i \sim Norm(a, \tau_\alpha) \tag{4}$$

$$\beta_i \sim Norm(b, \tau_\beta) \tag{5}$$

$$a \sim Norm(\mu_a, \sigma_a) \tag{6}$$

$$b \sim Norm(\mu_b, \sigma_b) \tag{7}$$

Appropriate hyperprior distributions for ($\tau_\alpha, \tau_\beta, \sigma_a$, and σ_b) need to be specified. The precision of the overall model and the between-subjects variability are gained by the posterior estimates of the parameters τ_α and τ_β, respectively. In the application in Section 3.1, we will discuss different prior distributions for τ_α and τ_β, which may be different for each subject or depend on other covariates. The proposed framework provides a reliable approach to account for the uncertainty of the fixed effects parameters.

The precision and the accuracy of the response are estimated by the parameters of the model. The slope parameters β_i link the inverse probit of the expected probability and the covariates x (i.e., the stimulus). Therefore, this parameter estimates the precision of the response, the higher is the estimated value of β_i, the more precise is the response. The interpretation of the location parameter of the psychometric function depends on the nature of the psychophysical task. In forced-choice discrimination tasks, as mentioned in the two examples detailed in Sections 3.1 and 3.2, the PSE estimates the accuracy of the response. The response is accurate if the PSE is equal to the value of the reference stimulus. The value of the PSE relative to observer i, pse_i is computed from intercept and slope in Equation (2) as follows:

$$pse_i = -\frac{\alpha_i}{\beta_i} \tag{8}$$

The PSE corresponds to the stimulus value yielding a response probability of 0.5, that is, the point at which participants are equally likely to choose the standard or the comparison stimulus in response to the task. In the examples mentioned later the PSE participants are equally likely to choose one stimulus or the other to the question "which stimulus moved faster?".

2.1. Prior distribution

According to the Bayesian paradigm, prior distributions and likelihood constitute a whole decision model. Ideally, a prior distribution describes the degree of belief about the true model parameters held by the scientists. If empirical data are available, then new information can coherently be incorporated via statistical models, through Bayesian learning. This process begins by documenting the available expert knowledge and uncertainty. A subjective prior describes the informed opinion of the value of a parameter before the collection of data.

Prior distributions as described in the previous paragraph are non-informative prior distributions. The flexibility of the Bayesian model allows to modify (Equations 4, 5) by considering, for example, partition or group of subjects between historical and current data. We assume that there is one relevant historical study available. However, the approaches proposed here can in principle be extended to multiple historical studies. Here, we recall the method based on the power prior proposed by Ibrahim and Chen (2000). This has emerged as a useful class of informative priors for a variety of situations in which historical data are available (Eggleston et al., 2017).

The power prior is defined as follows Ibrahim and Chen (2000). Suppose we have two data-sets from the current study and from a previous study that is similar to the current one, labeled as the current and the historical data, respectively. The historical data are indicated as $D_0 = (n_0, y_0, x_0)$, while the current data are indicated as $D = (n, y, x)$, n, and n_0 are the sample size, y and y_0 are the response vectors, respectively $n \times 1$ and $n_0 \times 1$ vectors. Finally, x and x_0 are (either $n \times p$ matrix or $n_0 \times p$ matrix) the covariates. Let indicate θ as the vector of parameters, $\pi_0(\theta)$ represents the initial prior distribution for θ before the historical data D_0 are observed. The parameter $L(\theta|D)$ indicates a general likelihood function for an arbitrary model, such as for linear models, generalized linear model, random-effects model, non-linear model, or a survival model with censored data. Given the parameter a_0, between 0 and 1, the power prior distribution of θ for the current study is defined as:

$$\pi(\theta|D_0, a_0) \propto L(\theta|D_0)^{a_0}\pi_0(\theta).$$

This way, a_0 represents the weights of the historical data relative to the likelihood of the current study. According to this definition, the parameter a_0 represents the impact of the historical data on $L(\theta|D)$.

Depending on the agreement between the historical and current data, the historical data may be down-weighted, reducing the value of a_0. The main question is what value of a_0 to use in the analysis, which means how to assess agreement between historical and current data and how to incorporate the historical data into the analysis of a new study. The easiest solution is to establish a hierarchical power prior by specifying a proper prior distribution for a_0. A uniform prior on a_0 might be a good choice, or a more informative prior would be to take a Beta distribution with moderate to large parameters. Although a prior for a_0 is attractive, it is much more computationally intensive than the a_0 fixed case. The a_0 random case has been extensively discussed (Ibrahim et al., 1999, 2015; Ibrahim and Chen, 2000; Chen and Ibrahim, 2006). Another approach, computationally more feasible, is to take a_0 as fixed and elicit a specific value for it and conduct several sensitivity analyzes about this value or to take a_0 as fixed and proceed, for example, with a model selection criterion.

2.2. Computational aspects

The large improvements in the availability of computational packages for implementing Bayesian analyzes have allowed the growth of applications of hierarchical Bayesian models. Many of the available packages permit the implementation of the Monte Carlo Markov Chain (MCMC) algorithm which saves time by avoiding technical coding. MCMC sampling is a simulation technique to generate samples from Markov chains that allow the reconstruction of the posterior distributions of the parameters. Once the posterior distributions are obtained, then the accurate and unbiased point estimates of model parameters are gained. Software for the application of Bayesian models is currently applied in several different fields (Palestro et al., 2018; Myers-Smith et al., 2019; Zhan et al., 2019; Dal'Bello and Izawa, 2021; Mezzetti et al., 2022). Gibbs sampling is an MCMC algorithm that can be implemented with the software Just Another Gibbs Sampler (JAGS), (Plummer, 2017). It is possible to interface JAGS with R using the CRAN package *rjags* developed by Plummer (2003). The reader may refer to the following tutorials for fitting hierarchical Bayesian models using JAGS (or STAN) and R (Plummer, 2003; Kruschke, 2014).

Once the model is defined in JAGS, it is possible to sample from the joint posterior distributions. The mean of samples from the posterior distribution of the parameters provides the posterior estimates of the parameters of interest. From the samples of the posterior distribution, it is also possible to extract the percentile and provide the corresponding 95% credible intervals.

As a diagnostic tool to assess whether the chains have converged to the posterior distribution, we use the statistic $\hat{R}$ (Gelman and Rubin, 1992). Each parameter has the $\hat{R}$ statistic associated with it (Gelman and Rubin, 1992), in the recent version (Vehtari et al., 2021); this is essentially the ratio of between-chain variance to within-chain variance (analogous to ANOVA). The $\hat{R}$ statistic should be approximately 1 ± 0.1 if the chain has converged.

To compare Bayesian models, different indicators can be adopted (Gelfand and Dey, 1994; Wasserman, 2000; Gelman et al., 2014). The sum of squared errors is a reasonable measure proposed. Although log-likelihood plays an important role in statistical model comparison, it also has some drawbacks, for example, the dependence on the number of parameters and on the sample size. A reasonable alternative is to evaluate a model through the log predictive density and its accuracy. Log pointwise predictive density (*lppd*) for a single value y_i is defined as Vehtari et al. (2017);

$$logp(y_i|y) = log \int p(y_i|\theta)p(\theta|y)d\theta$$

The log pointwise predictive density (*lppd*) is defined as the sum and can be computed using results from the posterior simulation

$$lppd = \sum_{i=1}^{n} logp(y_i|y) = \sum_{i=1}^{n} log \int p(y_i|\theta)p(\theta|y)d\theta$$

3. Fitting hierarchical bayesian models to the experimental data

Studies from our research group shed light on the interplay between slip motion and high-frequency vibrations (masking vibration) in the discrimination of velocity by touch (Dallmann et al., 2015; Picconi et al., 2022; Ryan et al., 2022). These and similar results are discussed in our recent review (Ryan et al., 2021). Using Bayesian hierarchical models, we combined two of these studies and evaluated the coherence of our findings across experiments. The two studies are summarized in Sections 3.1 and 3.2, respectively. Examples of the R and JAGS files for fitting our data are available in the following Github repository https://github.com/moskante/bayesian_models_psychophysics.

3.1. First data-set: The role of vibration in tactile speed perception

The data-set *touch-vibrations* was first published by Dallmann et al. (2015) and it is provided within the CRAN package *MixedPsy*. It consists of the forced-choice responses (i.e., the comparison stimulus is "faster" or "slower" than a reference) collected in a psychophysical study from nine human observers and the corresponding predictor variables. The task is as follows: In two separate intervals, participants were requested to compare the motion speed of a moving surface by touching it and reported whether it moved faster in the reference or the comparison stimulus. The speed of the comparison stimulus was chosen among seven values of speed ranging between 1.0 and 16.0 cm/s. In two separate blocks, participants performed the task either with masking vibrations (sinusoidal wave signal at 32 Hz) or without (control condition). Each speed and vibration combination was repeated 40 times in randomized order, resulting in a total of 560 trials for each participant.

According to Dallmann et al. (2015), GLMM with a probit link function was fitted to the data and the results presented in Supplementary Tables S1, S2 were obtained. Next, the data were fitted with a hierarchical Bayesian model in JAGS. Let Y_{ij}^h indicates the number of "faster" responses for subject i at speed x_j. Superscript h indicates the presence or absence of masking vibrations. That is, $h = 0$ masking vibrations were not present while $h = 1$ masking vibrations were present. n_{ij}^h is the total number of trials for subject i, speed x_j and vibration condition h. The model is the following:

$$Y_{ij}^h \sim Binom(\pi_{ij}^h, n_{i,j}^h) \tag{9}$$

$$\Phi^{-1}(\pi_{ij}^h) = \alpha_i^h + \beta_i^h x_j \qquad h = 0, 1 \tag{10}$$

The following set of priors are assumed:

$$\alpha_i^h \sim Norm(a^h, \tau_\alpha^h) \tag{11}$$

$$\beta_i^h \sim Norm(b^h, \tau_\beta^h) \tag{12}$$

$$\tau_\alpha^h \sim Gamma(1, 0.001) \tag{13}$$

$$\tau_\beta^h \sim Gamma(1, 0.001) \tag{14}$$

$$a^h \sim Norm(0, \sigma_a) \tag{15}$$

$$b^h \sim Norm(0, \sigma_b) \tag{16}$$

$$\sigma_a \sim Gamma(1, 0.01) \tag{17}$$

$$\sigma_b \sim Gamma(1, 0.01) \tag{18}$$

The model in Equation (10) can be parameterized as follows to allow focus on parameter PSE and the slope β_i^h:

$$Y_{ij}^h \sim Binom(\pi_{ij}^h, n_{i,j}^h) \qquad h = 0, 1 \tag{19}$$

$$\Phi^{-1}(\pi_{ij}^h) = -pse_i^h * \beta_i^h + \beta_i^h x_j \tag{20}$$

$$pse_i^h \sim Norm(PSE^h, \tau_{PSE}^h) \tag{21}$$

$$\beta_i^h \sim Norm(b^h, \tau_b^h) \tag{22}$$

$$\tau_{PSE}^h \sim Gamma(1, 0.001) \tag{23}$$

$$\tau_b^h \sim Gamma(1, 0.001) \tag{24}$$

$$PSE^h \sim Norm(0, \sigma_{PSE}) \tag{25}$$

$$b^h \sim Norm(0, \sigma_b) \tag{26}$$

$$\sigma_{PSE} \sim Gamma(1, 0.01) \tag{27}$$

$$\sigma_b \sim Gamma(1, 0.01) \tag{28}$$

We used the Greek letter β_i^h and the Latin letter b^h for the slope of subject i and the conditional value of slope common to all subjects, respectively. Similarly, we used the term pse_i^h and PSE^h for the estimate of the PSE in subject i and the conditional estimate.

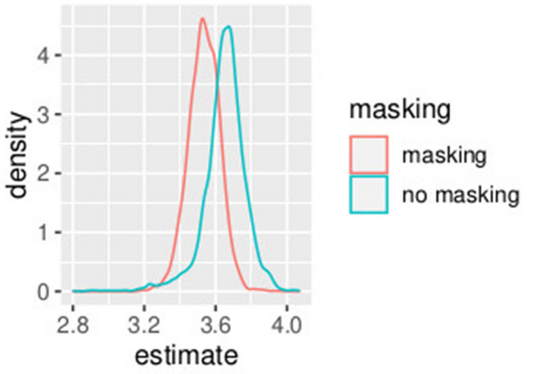


FIGURE 1
Posterior estimates of parameters b^h (slope). Experiment in Section 3.1.

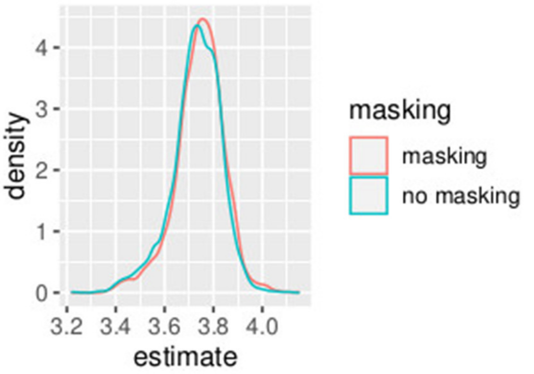


FIGURE 2
Posterior estimates of parameters PSE^h. Experiment in Section 3.1.

In this first example, non-informative prior distributions were adopted and the hierarchical Bayesian model confirmed the results obtained with the GLMM, as expected. Supplementary Table S3 presents the posterior estimates of a^h and b^h as defined in Equations (9)–(18), while Supplementary Table S4 presents posterior estimates of PSE^h as defined (Equations 19–28). Comparing Supplementary Table S2 (GLMM) and Supplementary Table S4 (Bayesian model), the PSE estimates result very close and the uncertainty is very similar with the two model approaches. Figures 1, 2 show the posterior distribution of the two parameters of the model b^h and PSE^h as defined in Equations (21), (22) that are common to all the subjects. The slope of the model is slightly higher without masking vibrations (b^0, in blue in the figure) as compared to masking vibrations (b^1, in red in the figure). The difference in PSE is negligible.

We considered the overlap between the posterior distributions as a measure of similarities and differences between parameters, where overlapping is defined as the area intersected by the two distributions. Overlapping was computed as the proportion of the

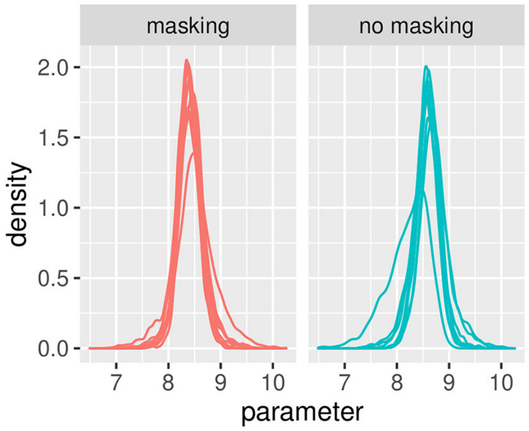


FIGURE 3
Posterior estimates of individual parameters of pse_i^h. The **(left)** figure illustrated with red lines represents conditions with masking vibrations, while the **(right)** figure illustrated with blue lines represents conditions without masking vibrations. Experiment in Section 3.1.

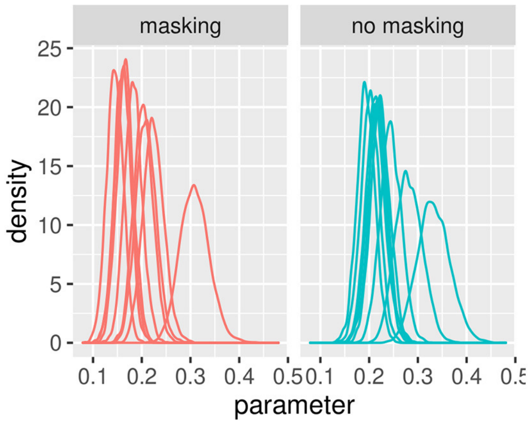


FIGURE 4
Posterior estimates of individual parameters of β_i^h. The **(left)** figure illustrated with red lines represents conditions with masking vibrations while the **(right)** figure illustrated with blue lines represents conditions without masking vibrations. Experiment in Section 3.1.

areas of the histograms belonging to the region shared by the two distributions. The idea of overlapping as a measure of similarity among data-sets or clusters is frequently used in different fields (Pastore and Calcagnì, 2019; Mezzetti et al., 2022).

An effect of vibration is present for the intercept. The overlap between the distribution of b^0 and b^1, the slope of the model, is 0.04. The overlap of the posterior distributions of PSE, in presence of vibration versus absence of vibration, is 0.58. This is consistent with our GLMM analysis where we found a small (yet significant) difference in slope but no differences in PSE.

Figures 3, 4 illustrate the posterior distributions of the parameters of the individual psychometric function, as specified in Equations (10), (21). It is interesting to notice that between-subject variability is present for the slope (parameter β_i^h), while subjects show similar behavior in posterior distribution respect to PSE (parameter pse_i^h). In fact in Figure 3, the between individual variability of PSE is quite negligible. Finally, Figures 5, 6 compare the predictions of the GLMM and of the hierarchical Bayesian model across the nine participants. The predictions of the two models are almost identical. To conclude, since we used a non-informative prior, the outcome of the Bayesian model does not differ substantially from the GLMM that was used in the original study.

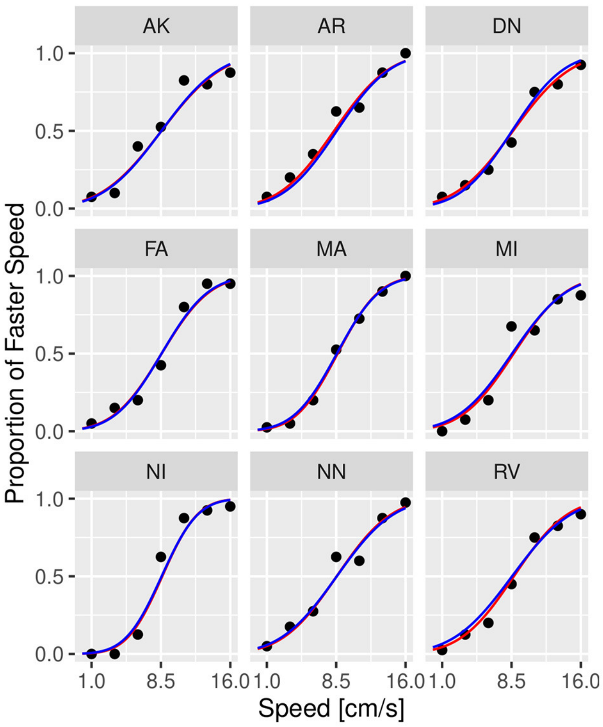


FIGURE 5
Psychometric functions of individual participants from Experiment 1 in conditions without masking vibrations. The scatter plot shows the observed (dots) versus predicted responses (solid lines) with data from individual participants illustrated in each panel. Blue lines correspond to the prediction by GLMM, while red lines correspond to predictions by the Bayesian model. Experiment in Section 3.1.

Different specifications of the prior distributions in Equations (23), (24) and in Equations (27), (28) were considered, based on the sum of squared errors and the uncertainties of parameters, measured with the length of credible intervals. In particular, alternative specification of Equations (21)–(24) was considered:

$$pse_i^h \sim Norm(PSE^h, \tau_{PSE}^i) \tag{29}$$

$$\beta_i^h \sim Norm(b^h, \tau_b^i) \tag{30}$$

$$\tau_{PSE}^i \sim Gamma(1, 0.001) \tag{31}$$

$$\tau_b^i \sim Gamma(1, 0.001) \tag{32}$$

Specifically, in the model earlier, each subject can have a different precision in the two parameters of PSE and slope—i.e., τ_{PSE}^i and τ_b^i may have different values depending on the participant. The previous choice of prior distributions assumed

TABLE 1 Comparison between the different models in data-set *touch-vibrations*.

Model	Effects	Log likelihood	LPPD	Sum errors	95% CI of PSE	Width CI
GLMM	Individual	-	-	-		
	Overall	−284.42	-	0.62	(0.52, 0.59)	0.07
Bayesian 1	Individual	−276.23	−14231.6 (1081.1)	0.42		
	Overall			0.61	(0.49, 0.55)	0.06
Bayesian 2	Individual	−278.03	−14323.0 (3.6)	0.41		
	Overall			0.63	(0.49, 0.50)	0.01
Bayesian 3	Individual	−276.29	14163.2 (2.0)	0.40		
	Overall			0.61	(0.57, 0.61)	0.04
Bayesian 4	Individual	−275.86	-14155.8 (2.3)	0.39		
	Overall			0.61	(0.58, 0.63)	0.05

For each model, we showed the log-likelihood and the LPPD of the model, and the sum or squared errors and length of the Credible Intervals of the PSE. The first two lines refer to the GLMM as described by Dallmann et al. (2015). The third and the fourth lines show the Bayesian model 1, as specified in Equations (19)–(28). In the fifth and the sixth lines, the Bayesian model 2 is shown, with a different specification of the prior distributions as in (29)–(32). In the Bayesian model 3, the distribution of $\tau^h_{\alpha,\beta} \sim Gamma(1, 0.1)$ was used, as in (21)–(24). The last two lines show the Bayesian model 4 with the distribution of τ is $\tau^i_{\alpha,\beta} \sim Gamma(1, 0.1)$. This means that the variability of the PSE and the slope was allowed to be different for each participant.

higher variability between subjects and evidenced a different outcome in the subject *NI* as compared to the others with respect to the intercept and the slope. The alternative specifications of prior distributions in Equations (29)–(32) provide similar values with respect to the sum of squared errors, and the length of credible intervals for the PSE was slightly lower than the model in Equations (27), (28). Table 1 shows the frequentist approach (GLMM) and the different specifications of the Bayesian model. Comparing the models with respect to the uncertainties in PSE estimation and model fitting, we justify the choice of the model proposed.

3.2. Second data-set: Tactile speed discrimination in people with type 1 diabetes

The second data-set, *touch-diabetes*, includes data from 60 human participants that were tested in a speed discrimination task similar to the one described in Section 3.1. The experimental procedure and the results are detailed by Picconi et al. (2022). Participants were divided into three groups, with 20 participants per group: healthy controls, participants with diabetes with mild tactile dysfunction, and participants with diabetes with moderate tactile dysfunction. The three groups were labeled as *controls, mild, and moderate*, respectively. As in *touch-vibration*, this experiment consisted of a force-choice, speed discrimination task. In each of the 120 trials, participants were requested to indicate whether a contact surface moved faster during a comparison or a reference stimulus interval. For this experiment, a smooth surface consisting of a glass plate was used. The motion speed of the comparison stimuli were as chosen pseudo-randomly from a set of five values ranging from 0.6 to 6.4 cm/s, with the speed of the reference stimulus equal to 3.4 cm/s. Participants performed the task with and without masking vibrations, with masking stimuli consisting of sinusoidal vibrations at 100 Hz.

As in the original study, we used the GLMM in Equations (33)–(35) to fit the data across groups and across masking vibration conditions:

$$Y^h_{ij} \sim Binom(\pi^h_{ij}, n^h_{i,j}) \qquad h = 0, 1 \tag{33}$$

$$\Phi^{-1}(\pi^h_{ij}) = \alpha^h_i + \beta^h_i x_j \tag{34}$$

The response variable Y^h_{ij} is the number of "faster" responses for subject i at speed x_j. The suprascript $h = 0$ represents conditions without masking vibrations and $h = 1$ represents conditions with masking vibration. The variable n^h_{ij} is the total number of trials. Considering two dummy variables for the two groups of participants with diabetes, *mild* (indicated with subscript 2) and *moderate* (indicated with subscript 3) patients with diabetes, the individual model with fixed effects is rewritten as:

$$\Phi^{-1}(\pi^h_{ij}) = \alpha^h + \alpha^h_2 + \alpha^h_3 + \beta^h x_j + \beta^h_2 x_j + \beta^h_3 x_j \tag{35}$$

We used the packages MixedPsy (Balestrucci et al., 2022) and lme4 (Bates et al., 2015) for model fitting. Supplementary Tables S5, S6 report results for the frequentist approach (GLMM). The slope of the model (referred to as *tactile sensitivity* in the study) was different across the three groups, with controls performing significantly better in the task than people with mild and moderate tactile dysfunctions. The difference between groups was larger without masking vibrations. As in the first data-set, masking vibrations reduced the values of the slope across all groups. We computed the values of PSE for all groups and conditions, see Supplementary Table S6. We expected no significant change in PSE, both between masking vibration conditions and between groups. This is because, in this task, the cues and the sensory noise are the same in the reference and comparison stimulus.

As in the previous example, we re-analyzed the data with a Bayesian hierarchical model. Let i indicates subject, j speed, h masking or no masking, and k indicates group.

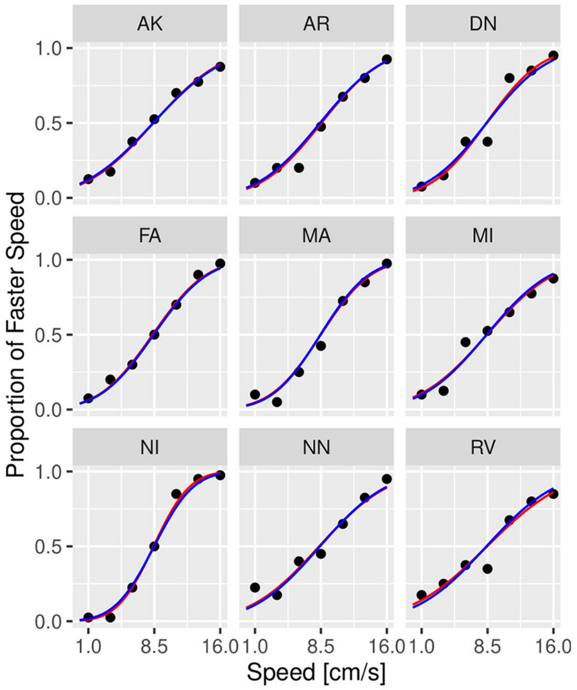


FIGURE 6
Psychometric functions of individual participants from Experiment 1 in conditions with 32 Hz masking vibrations. The scatter plot shows the observed (dots) versus predicted responses (solid lines) with data from individual participants illustrated in each panel. Blue lines correspond to the prediction by GLMM, while red lines correspond to predictions by the Bayesian model. Experiment in Section 3.1.

Similar to the analysis of the first data-set, the model was parameterized with respect to the PSE and the slope:

$$Y^h_{ij} \sim Binom(\pi^h_{ij}, n^h_{i,j}) \tag{36}$$

$$\Phi^{-1}(\pi^h_{ij}) = -pse^h_i * \beta^h_i + \beta^h_i x_j \qquad h = 0, 1 \tag{37}$$

The following prior and hyper-prior distributions are assumed:

$$pse^h_i \sim Norm(PSE^h_k, \tau^h_{PSE,k}) \quad h = 0, 1\, k = 1, 2, 3 \tag{38}$$

$$\beta^h_i \sim Norm(b^h_k, \tau^h_{\beta,k}) \tag{39}$$

$$\tau^h_{PSE,k} \sim Gamma(1, 0.001) \tag{40}$$

$$\tau^h_{\beta,k} \sim Gamma(1, 0.001) \tag{41}$$

$$PSE^h_k \sim Norm(0, \sigma_{PSE}) \tag{42}$$

$$b^h_k \sim Norm(0, \sigma_b) \tag{43}$$

$$\sigma_{PSE} \sim Gamma(1, 0.01) \tag{44}$$

$$\sigma_b \sim Gamma(1, 0.01) \tag{45}$$

The mean and the credible intervals of the parameters of the models b^h_k (slope) and PSE^h_k, as defined in Equations (33)–(45), are reported in Supplementary Table S7. The results confirmed the difference in slopes between the groups and between conditions. In conditions without masking vibrations, the slope was the highest in controls followed by the mild and moderate groups. The mean of the slope in controls is higher than the credible intervals of the mild group. Similarly, the mean of the slope of the mild group is higher than the credible intervals of the moderate group. The same effect can be observed in the masking vibration conditions, although the difference in slope is smaller between the control and mild groups. In Figure 7, the posterior distributions of the slope of the model are shown. We can observe the two effects of group (ordered from controls to moderate) and masking conditions. In particular, the group with moderate tactile dysfunction (illustrated in blue) is the one with the lowest values of slope.

In Figure 8, the posterior distributions of the PSE values, as specified in Equations (36)–(45) are shown. Uncertainties in the parameters PSE^h_k were comparable between the frequentist and the Bayesian models. This was expected because in this Bayesian model, we used a non-informative prior. Masking vibrations had a large effect on the slope and a much smaller effect on the PSE. Within the control group, the overlap between the posterior distributions of PSE with masking versus no masking is 0.04, and the overlap between the posterior distribution of the slope between masking and no masking is < 0.01. This supports our finding that masking vibration reduced tactile sensitivity. In Figures 9, 10, the posterior distributions of the individual parameters β_i and pse_i are shown. Again, it is interesting to notice that the posterior estimates of PSE have low subject variability. The individual posterior distributions show a higher overlapping, refer to Figure 10 for an almost perfect overlapping. Within groups, variability is lower for PSE compared to posterior distributions of the parameters representing the slopes.

4. Combined analysis of the two experiments

In this section, we propose two different approaches for the joint analysis of the two studies. In Section 4.1, the prior distributions of the parameters relative to the second study are defined from the data of the first study. In Section 4.2, a model approach based on the power prior distribution explained in Section 2.1 was applied to combine the two data-sets *touch-vibrations* and *touch-diabetes*.

The data-set *touch-vibrations* is considered historical data and indicated a $D_0 = (n_0, y_0, x_0)$, where n_0 is the sample size of the historical data, y_0 is the number of “faster” responses the $n_0 \times 1$ response vector, in this case number of, x_0 is a $n_0 \times 1$ vector of speed. The data-set *touch-diabetes* indicated the current study, we restrict the analysis only to the control group, we discarded the two diabetic groups because of their reduced tactile sensitivity. Data are denoted by $D = (n, y, x)$, where n denotes the sample size, y denotes the $n \times 1$ response vector, the number of “faster” responses, and x the $n \times 2$ matrix of covariates, indicator of cluster and speed.

4.1. Prior distribution defined on the first experiment

The two data-sets are jointly analyzed. Equations (33)–(45) are rewritten incorporating model (Equations 9, 10) in order to combine the two studies as follows:

$$Y^h_{ij} \sim Binom(\pi^h_{ij}, n^h_{i,j}) \tag{46}$$

$$\Phi^{-1}(\pi^h_{ij}) = \alpha^h_i + \beta^h_i x_j \tag{47}$$

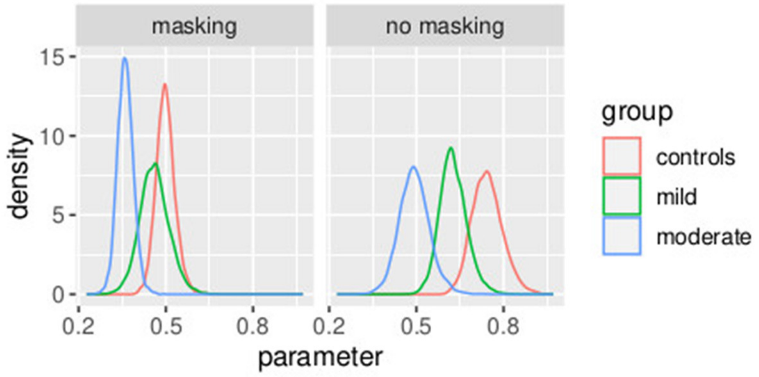


FIGURE 7
Posterior distributions of parameters b^h_k from the second stage of the hierarchical model. Experiment in Section 3.2.

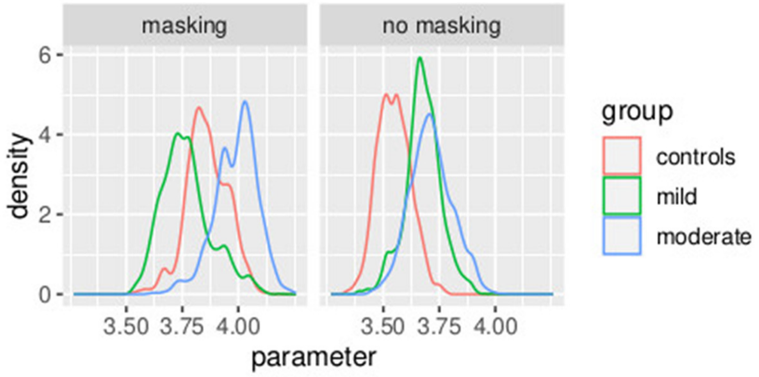


FIGURE 8
Posterior distributions of parameters of the second stage of the hierarchical model PSE^h_k. Experiment in Section 3.2.

$$Y^h_{0ij} \sim Binom(\pi^h_{0ij}, n^h_{0i,j}) \tag{48}$$

$$\Phi^{-1}(\pi^h_{0ij}) = \alpha^h_{0i} + \beta^h_{0i} x_{0j} \tag{49}$$

Because of Weber's Law, the sensitivity to speed and, therefore, the slope depends on the value of the stimulus. To address this issue, to combine the two experiments, we used the conversion factor in Equation (55).

$$\alpha^h_i \sim Norm(a^h, \tau^h_a) \tag{50}$$

$$\beta^h_i \sim Norm(b^h, \tau^h_b) \tag{51}$$

$$\alpha^h_{0i} \sim Norm(a^h_0, \tau^h_{a0}) \tag{52}$$

$$\beta^h_{0i} \sim Norm(b^h_0, \tau^h_{b0}) \tag{53}$$

$$a^h \sim Norm(a^h_0, \sigma^h_a) \tag{54}$$

$$b^h \sim Norm\left(b^h_0 \times \frac{\bar{x}}{\bar{x}_0}, \sigma^h_b\right) \tag{55}$$

$$a^h_0 \sim Norm(0, \sigma^0_a) \tag{56}$$

$$b^h_0 \sim Norm(0, \sigma^0_b) \tag{57}$$

$$\sigma^h_k \sim Gamma(1, 0.01) \qquad h = 0, 1, 2 \quad k = a, b \tag{58}$$

$$\tau^h_k \sim Gamma(1, 0.01) \qquad h = 1, 2 \quad k = a0, a, b0, b \tag{59}$$

From the posterior estimates of parameters σ^h_a and σ^h_b, we can gain information about whether the combination of two studies is appropriate for the same model. The posterior distributions of the precision parameters indicate a good agreement between the two studies and confirm the suitability of the choice for the prior distribution. High-posterior estimates of the precision of the prior distribution indicate good agreement between prior distribution and data.

4.2. Power prior model

Recalling Section 2.1, the prior distribution of parameters $\theta = (\alpha, \beta)$ is defined as follows:

$$\pi(\theta|D_0, a_0) \propto L(\theta|D_0)^{a_0} \pi_0(\theta). \tag{60}$$

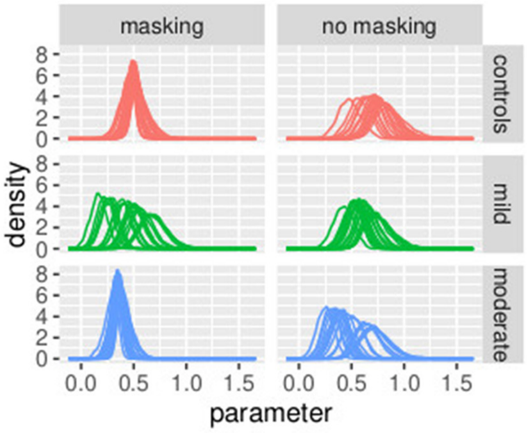


FIGURE 9
Posterior distributions of parameters of the first stage of the hierarchical model β_i^h, by group and masking condition. Experiment in Section 3.2.

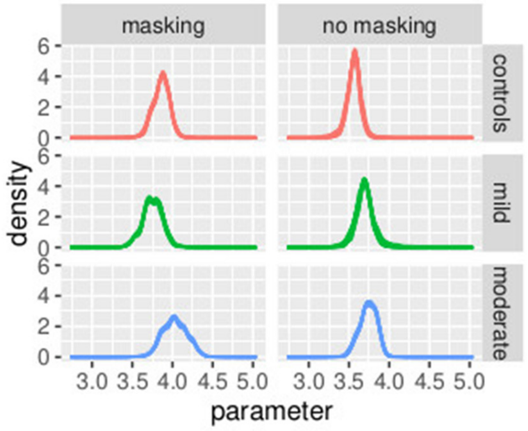


FIGURE 10
Posterior distributions of parameters of the first stage of the hierarchical model PSE_i^h, by group and masking condition. Experiment in Section 3.2.

The power parameter a_0 represents the weight of the historical data relative to the likelihood of the current study. The parameters represent how much data from the previous study is to be used in the current study. There are two special cases for a_0, the first case $a_0 = 0$ results in no incorporation of the data from the previous study relative to the current study. The second case $a_0 = 1$ results in full incorporation of the data from the previous study relative to the current study. Therefore, a_0 controls the influence of the data gathered from previous studies that is similar to the current study. This control is important when the sample size of the current data is quite different from the sample size of historical data or where there is heterogeneity between two studies (Ibrahim and Chen, 2000).

In Table 2, a comparison between all the models obtained by varying the parameter a_0 is shown. The choice of the value for a_0 is implemented by model comparison, taking into account the log-likelihood, the log point-wise predictive density, the sum of squared errors, of both the level of the model, that are the individual and overall model. Moreover, a comparison of the uncertainty in PSE estimation is computed. The uncertainty decreases as a_0 increases indicating that we are updating our informative knowledge for the correct model use. The likelihood increases as the value of a_0 increases. The measures of goodness of fit of the models are very similar increasing the value of a_0. We decide to favor the model that lowers the uncertainties in the estimation, that is the model with $a_0 = 0.7$.

In Table 3, three different prior distributions are compared. On one hand, an informative prior is assumed following Section 3.2; on the other hand, the first experiment is used to improve the understanding of experiment 2. A combination of the two studies [as in Equations (46)–(59)] illustrated in Section 4.1 is compared with power prior as in Section (4.2). In Figures 11, 12, a comparison of the posterior distributions of PSE and β, in the control group, obtained according to the three different prior distributions is shown. Again we favor the model that lowers the uncertainties of posterior estimates. Overall, combining the two studies with the power prior approach reduced the posterior estimate of the model parameters as can be clearly seen by comparing the three distributions in the figures.

5. Conclusion

In this study, we compared the outcome of a Bayesian approach to a frequentist mixed model (GLMM) approach. The comparison showed the importance of incorporating informative prior knowledge from previous studies for data analysis.

We re-analyzed data from two studies using GLMM and Bayesian models. First, we applied GLMM and four different Bayesian models to the data-set described by Dallmann et al. (2015). We compared the log-likelihood, LPPD, the sum of errors between the different models, and confidence interval of the two parameters of slope and PSE. The Bayesian approach allowed for more flexibility in the model fitting (see Table 1). Next, we applied Bayesian models to the second data-set for re-analysis of the results described by Picconi et al. (2022). With a non-informative prior, the Bayesian approach confirmed the estimation of the parameters of the frequentist model. Finally, we ran a joint analysis of the two data-sets using two different approaches, either by using the first data-set to choose the parameters of the prior or by using the power prior method. The informative prior in the power prior method reduced the credible intervals of the PSE and justified the choice of the model, as shown in Tables 2, 3.

The Bayesian approach provides useful features for the in-depth analysis of psychophysical data. Through a Bayesian approach, the random effects are estimated parameters, like the fixed effects, with the advantage of obtaining credible intervals for both the quantities. This allowed to estimate the effect of individual participants and the reliability of each of them. For example, in Figure 4, it is possible to identify a single participant with increased variability and higher slope as compared to the rest of the group. Potentially, this will simplify the identification of outliers or sources of unobserved variability. Another advantage of the hierarchical Bayesian approach is the possibility to incorporate information from past studies to reduce the uncertainty of the estimate. For example, compare the width of the three distributions in Figures 11, 12, with the non-informative prior having the larger width, i.e., the

TABLE 2 Comparison between the different models obtained by varying values of a_0 in Equation (60), as illustrated in subsection 4.2.

a_0	Effects	Log likelihood	LPPD	Sum errors	CI of PSE	Width CI
0	Individual	−291.76	−3169.67 (9.91)	2.24		
	Overall			2.68	(0.32, 0.37)	0.05
0.1	Individual	−292.85	−3183.66 (8.66)	2.24		
	Overall			2.64	(0.27, 0.26)	0.01
0.2	Individual	−293.47	−3202.41 (7.18)	2.21		
	Overall			2.64	(0.24, 0.27)	0.03
0.3	Individual	−294.07	−3221.91 (8.45)	2.21		
	Overall			2.64	(0.22, 0.27)	0.05
0.4	Individual	−295.08	−3243.93 (7.02)	2.20		
	Overall			2.65	(0.25, 0.24)	0.01
0.5	Individual	-295.48	-3242.41 (8.75)	2.20		
	Overall			2.66	(0.21, 0.25)	0.04
0.6	Individual	-296.00	−3253.47 (13.02)	2.20		
	Overall			2.66	(0.19, 0.27)	0.08
0.7	Individual	−296.02	−3256.63 (8.33)	2.23		
	Overall			2.68	(0.18, 0.22)	0.04
0.8	Individual	−296.84	−3276.64 (7.70)	2.1		
	Overall			2.68	(0.18, 0.25)	0.07
0.9	Individual	-296.77	−3268.68 (12.03)	2.23		
	Overall			2.68	(0.18, 0.2)	0.02
1	individual	−297.27	−3268.46 (10.7)	2.22		
	Overall			2.7	(0.17, 0.24)	0.07

For each model, we showed the log-likelihood, the LPPD of the model, the sum or squared errors, and length of the Credible Intervals of the PSE.

TABLE 3 Comparison between the different priors assumed for the data-set *touch-diabetes*.

Model	Effects	Log like	LPPD	Sum errors	CI of PSE	Width CI
Non Informative 1	Individual	−285.92	−3026.4 (3.4)	2.04		
	Overall			2.68	(0.3, 0.35)	0.05
Non Informative 2	Individual	−291.66	−3153.9 (6.2)	2.25		
	Overall			2.67	(0.26, 0.32)	0.06
Non Informative 3	Individual	-283.74	−3000.3 (2.8)	1.91		
	Overall			2.68	(0.31, 0.39)	0.08
Informative Prior Subsection 4.1	Individual	-292.14	−3156.2 (7.8)	2.2		
	Overall			2.65	(0.25, 0.36)	0.11
Informative Prior Subsection 4.2 with $a_0 = 0.7$	Individual	−296.02	−3256.63 (8.33)	2.23		
	Overall			2.68	(0.18, 0.22)	0.04

For each model, we showed the log-likelihood and the LPPD of the model, and the Sum or Squared Errors and the Credible Intervals of the PSE. The first three models refer to non-informative prior as illustrated in Equations (36)–(45). The first three models differ for hyperparameters in the Gamma distribution in Equations (40), (41). Non Informative 1 assumes $\tau_{\beta,PSE} \sim Gamma(1, 0.01)$. Non Informative 2 assumes $\tau_{\beta,PSE} \sim Gamma(1, 0.001)$ and Non Informative 3 assumes $\tau_{\beta,PSE} \sim Gamma(0.1, 0.01)$. The fourth and fifth models refers to the joint analyzes of the two data-sets. In particular, the fourth model refers to prior illustrated in Section 4.1. The fifth model refers to the prior illustrated in Section 4.2 with a_0 equal to 0.7.

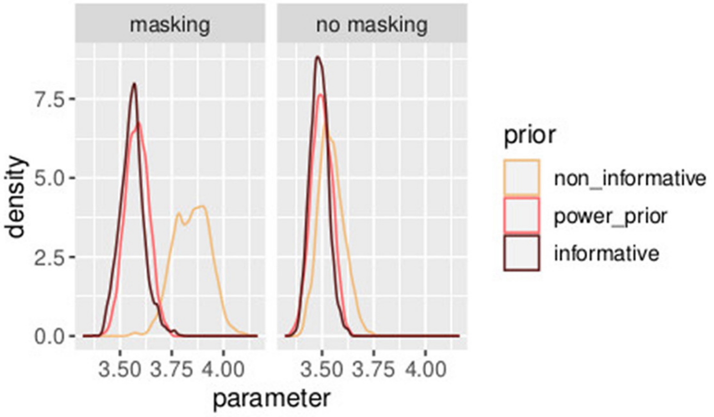


FIGURE 11
Posterior distributions of parameters PSE^h with different prior distributions for different values of a_0. The model with the informative prior ($a_0 = 1.0$) is illustrated in dark brown, the one with the power prior ($\alpha_0 = 0.7$) in orange, and the one with the non-informative prior ($\alpha_0 = 0.0$) in yellow. Experiment in Section 3.2.

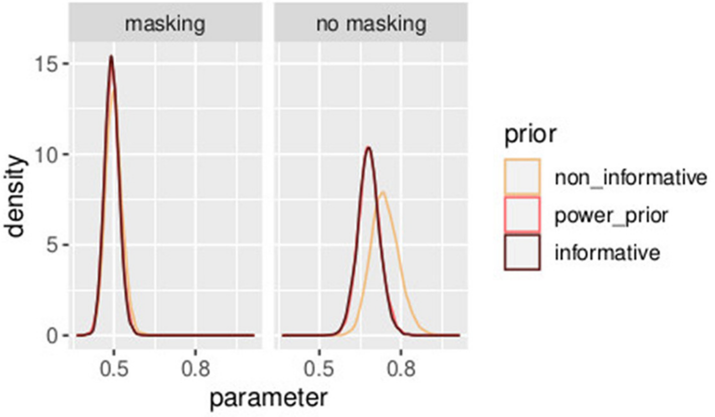


FIGURE 12
Posterior distributions of parameters b^h with different prior distributions for different values of a_0. The model with the informative prior ($a_0 = 1.0$) is illustrated in dark brown, the one with the power prior ($a_0 = 0.7$) in orange, and the one with the non-informative prior ($a_0 = 0.0$) in yellow. Experiment in Section 3.2.

higher variance. This will increase the power of the analysis. Finally, this approach allowed quantifying the coherence of multiple studies on a related topic through the parameter *a0*. The greater the value of *a0*, the higher the coherence across the studies.

Hierarchical modeling is a natural tool for combining several data-sets or incorporating prior information. In the current study, the method presented by Chen and Ibrahim (2006) has been used that provides a formal connection between the power prior and hierarchical models for the class of generalized linear models. Understanding the impact of priors on the current data and subsequently making decisions about these priors is fundamental for the interpretation of data (Koenig et al., 2021). One of the assumptions of the power prior approach is the existence of a common set of parameters for the old and current data and this assumption may not be met in practice. An alternative approach to incorporate historical data has been proposed by Neuenschwander et al. (2010) and van Rosmalen et al. (2018). This other method is based on meta-analytic techniques (MAP) and assumes exchangeability between old and current parameters.

Incorporating previous knowledge and insight into the estimation process is a promising tool (Van de Schoot et al., 2017) that is particularly relevant in studies with small sample sizes, as is often in psychophysical experiments. In our case, the sample size of the first data-set differed from the sample size of the second data-set. To take this into account, the power prior approach allowed us to assign a different weight to the historical data and the current data. It is possible to purposefully choose the hyperparameters of the prior, τ, to increase the precision of the posterior estimate. Zitzmann et al. (2015) suggested to specify a slightly informative prior to the group-level variance. As shown in Section 4, diffuse priors produce results that are aligned with the likelihood. On the other hand, using an informative prior that is relatively far from the likelihood, produces a shift in the posterior. It is possible to conduct a prior sensitivity analysis to fully understand its influence on posterior estimates (Van de Schoot et al., 2017).

Uncertainty quantification is an important issue in psychophysics. Hierarchical Bayesian models allow the researcher to estimate the uncertainty at a group level and the one specific to individual participants. This model approach will have an important impact on the evaluation of psychometric functions in psychophysical data.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/moskante/bayesian_models_psychophysics.

Author contributions

MM: conceptualization, methodology, visualization, software, formal analysis, writing—original draft, and writing—reviewing and editing. CR: conceptualization, data curation, visualization, software, writing—original draft preparation, and writing—review and editing. PB: conceptualization, software, and writing—reviewing and editing. FL: conceptualization, data curation, and writing—reviewing and editing. AM: conceptualization, data curation, visualization, software, formal analysis, writing—original draft, and writing—reviewing and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 251654672—TRR 161, by the Italian Ministry of Health (Ricerca Corrente, IRCCS Fondazione Santa Lucia), the Italian Ministry of University and Research (PRIN 2017SB48FP) and EU project HARIA—Human-Robot Sensorimotor Augmentation—Wearable Sensorimotor Interfaces and Supernumerary Robotic Limbs for Humans with Upper-limb Disabilities (grant agreement No. 101070292).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2023.1108311/full#supplementary-material

References

Agresti, A. (2002). *Categorical Data Analysis, Vol. 482.* Hoboken, NJ: John Wiley & Sons.

Alcalá-Quintana, R., and García-Pérez, M. A. (2004). The role of parametric assumptions in adaptive bayesian estimation. *Psychol. Methods* 9, 250. doi: 10.1037/1082-989X.9.2.250

Balestrucci, P., Ernst, M. O., and Moscatelli, A. (2022). Psychophysics with R: the R Package MixedPsy. *bioRxiv* 2022.06.20.496855. doi: 10.1101/2022.06.20.496855

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Chen, M.-H., and Ibrahim, J. G. (2006). The relationship between the power prior and hierarchical models. *Bayesian Anal.* 1, 551–574. doi: 10.1214/06-BA118

Dal'Bello, L. R., and Izawa, J. (2021). Task-relevant and task-irrelevant variability causally shape error-based motor learning. *Neural Netw.* 142, 583–596. doi: 10.1016/j.neunet.2021.07.015

Dallmann, C. J., Ernst, M. O., and Moscatelli, A. (2015). The role of vibration in tactile speed perception. *J. Neurophysiol.* 114, 3131–3139. doi: 10.1152/jn.00621.2015

Eggleston, B. S., Ibrahim, J. G., and Catellier, D. (2017). Bayesian clinical trial design using markov models with applications to autoimmune disease. *Contemp Clin. Trials* 63, 73–83. doi: 10.1016/j.cct.2017.02.004

Fong, Y., Rue, H., and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics* 11, 397–412. doi: 10.1093/biostatistics/kxp053

Foster, D. H., and Zychaluk, K. (2009). Model-free estimation of the psychometric function. *J. Vis.* 9, 30–30. doi: 10.1167/9.8.30

Fox, J.-P., and Glas, C. A. (2001). Bayesian estimation of a multilevel irt model using gibbs sampling. *Psychometrika* 66, 271–288. doi: 10.1007/BF02294839

Gelfand, A. E., and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *J. R. Stat. Soc. B* 56, 501–514. doi: 10.1111/j.2517-6161.1994.tb01996.x

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis.* New York, NY: Chapman and Hall; CRC.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Stat. Comput.* 24, 997–1016. doi: 10.1007/s11222-013-9416-2

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472. doi: 10.1214/ss/1177011136

Houpt, J. W., and Bittner, J. L. (2018). Analyzing thresholds and efficiency with hierarchical bayesian logistic regression. *Vision Res.* 148, 49–58. doi: 10.1016/j.visres.2018.04.004

Ibrahim, J. G., and Chen, M.-H. (2000). Power prior distributions for regression models. *Stat. Sci.* 15, 46–60. doi: 10.1214/ss/1009212673

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015). The power prior: theory and applications. *Stat. Med.* 34, 3724–3749. doi: 10.1002/sim.6728

Ibrahim, J. G., Chen, M.-H., and MacEachern, S. N. (1999). Bayesian variable selection for proportional hazards models. *Can. J. Stat.* 27, 701–717. doi: 10.2307/3316126

Johnston, A., Bruno, A., Watanabe, J., Quansah, B., Patel, N., Dakin, S., et al. (2008). Visually-based temporal distortion in dyslexia. *Vision Res.* 48, 1852–1858. doi: 10.1016/j.visres.2008.04.029

Knoblauch, K., and Maloney, L. T. (2012). *Modeling Psychophysical Data in R.* New York, NY: Springer New York.

Koenig, C., Depaoli, S., Liu, H., and Van De Schoot, R. (2021). Moving beyond non-informative prior distributions: achieving the full potential of bayesian methods for psychological research. *Front. Psychol.* 12, 809719. doi: 10.3389/fpsyg.2021.809719

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Cambridge, MA: Academic Press.

Kuss, M., Jäkel, F., and Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *J. Vis.* 5, 8–8. doi: 10.1167/5.5.8

Linares, D., and López-Moliner, J. (2016). quickpsy: an R package to fit psychometric functions for multiple groups. *R J.* 8, 122–131. doi: 10.32614/RJ-2016-008

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course With Examples in R and Stan*. New York, NY: Chapman and Hall; CRC.

Mezzetti, M., Borzelli, D., and d'Avella, A. (2022). A bayesian approach to model individual differences and to partition individuals: case studies in growth and learning curves. *Stat. Methods Appl.* 31, 1245–1271. doi: 10.1007/s10260-022-00625-6

Morrone, M. C., Ross, J., and Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nat. Neurosci.* 8, 950–954. doi: 10.1038/nn1488

Moscatelli, A., and Balestrucci, P. (2017). *Psychophysics with R: the R package MixedPsy*. R package version 1.0(0).

Moscatelli, A., Bianchi, M., Serio, A., Terekhov, A., Hayward, V., Ernst, M. O., et al. (2016). The change in fingertip contact area as a novel proprioceptive cue. *Curr. Biol.* 26, 1159–1163. doi: 10.1016/j.cub.2016.02.052

Moscatelli, A., Mezzetti, M., and Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: the generalized linear mixed model. *J. Vis.* 12, 26–26. doi: 10.1167/12.11.26

Moscatelli, A., Scotto, C. R., and Ernst, M. O. (2019). Illusory changes in the perceived speed of motion derived from proprioception and touch. *J. Neurophysiol.* 122, 1555–1565. doi: 10.1152/jn.00719.2018

Myers-Smith, I. H., Grabowski, M. M., Thomas, H. J., Angers-Blondin, S., Daskalova, G. N., Bjorkman, A. D., et al. (2019). Eighteen years of ecological monitoring reveals multiple lines of evidence for tundra vegetation change. *Ecol. Monogr.* 89, e01351. doi: 10.1002/ecm.1351

Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clin. Trials* 7, 5–18. doi: 10.1177/1740774509356002

Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., and Turner, B. M. (2018). A tutorial on joint models of neural and behavioral measures of cognition. *J. Math. Psychol.* 84, 20–48. doi: 10.1016/j.jmp.2018.03.003

Pariyadath, V., and Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS ONE* 2, e1264. doi: 10.1371/journal.pone.0001264

Pastore, M., and Calcagnì,, A. (2019). Measuring distribution similarities between samples: a distribution-free overlapping index. *Front. Psychol.* 10, 1089. doi: 10.3389/fpsyg.2019.01089

Pelli, D. G., and Farell, B. (1995). Psychophysical methods. *Handbook Optics* 1, 29–21.

Picconi, F., Ryan, C., Russo, B., Ciotti, S., Pepe, A., Menduni, M., et al. (2022). The evaluation of tactile dysfunction in the hand in type 1 diabetes: a novel method based on haptics. *Acta Diabetol.* 59, 1073–1082. doi: 10.1007/s00592-022-01903-1

Plummer, M. (2003). "Jags: a program for analysis of bayesian graphical models using gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vol. 124* (Vienna), 1–10.

Plummer, M. (2017). *Jags Version 4.3. 0 User Manual [computer software manual]*. Available online at: sourceforge.net/projects/mcmc-jags/files/Manuals/4.x2

Prins, N. (2016). *Psychophysics: A Practical Introduction*. Cambridge, MA: Academic Press.

Prins, N., and Kingdom, F. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the palamedes toolbox. *Front. Psychol.* 9, 1250. doi: 10.3389/fpsyg.2018.01250

Rouder, J. N., and Lu, J. (2005). An introduction to bayesian hierarchical models with an application in the theory of signal detection. *Psychon. Bull. Rev.* 12, 573–604. doi: 10.3758/BF03196750

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., and Zhou, D. (2003). A hierarchical bayesian statistical framework for response time distributions. *Psychometrika* 68, 589–606. doi: 10.1007/BF02295614

Ryan, C. P., Bettelani, G. C., Ciotti, S., Parise, C., Moscatelli, A., and Bianchi, M. (2021). The interaction between motion and texture in the sense of touch. *J. Neurophysiol.* 126, 1375–1390. doi: 10.1152/jn.00583.2020

Ryan, C. P., Ciotti, S., Cosentino, L., Ernst, M. O., Lacquaniti, F., and Moscatelli, A. (2022). Masking vibrations and contact force affect the discrimination of slip motion speed in touch. *IEEE Trans. Haptics* 15, 693–704. doi: 10.1109/TOH.2022.3209072

Schütt, H. H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and accurate bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Res.* 122, 105–123. doi: 10.1016/j.visres.2016.02.002

Steele, F., and Goldstein, H. (2006). "12 multilevel models in psychometrics," in *Psychometrics, Volume 26 of Handbook of Statistics*, eds C. Rao and S. Sinharay (Amsterdam: Elsevier), 401–420.

Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Boca Raton, FL: CRC Press.

Van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., and Depaoli, S. (2017). A systematic review of bayesian articles in psychology: the last 25 years. *Psychol. Methods* 22, 217. doi: 10.1037/met0000100

van Rosmalen, J., Dejardin, D., van Norden, Y., Löwenberg, B., and Lesaffre, E. (2018). Including historical data in the analysis of clinical trials: Is it worth the effort? *Stat. Methods Med. Res.* 27, 3167–3182. doi: 10.1177/0962280217694506

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Stat. Comput.* 27, 1413–1432. doi: 10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: an improved r for assessing convergence of mcmc (with discussion). *Bayesian Anal.* 16, 667–718. doi: 10.1214/20-BA1221

Wang, X., Bradlow, E. T., and Wainer, H. (2002). A general bayesian model for testlets: Theory and applications. *Appl. Psychol. Meas.* 26, 109–128. doi: 10.1177/0146621602026001007

Wasserman, L. (2000). Bayesian model selection and model averaging. *J. Math. Psychol.* 44, 92–107. doi: 10.1006/jmps.1999.1278

Zhan, P., Jiao, H., Man, K., and Wang, L. (2019). Using jags for bayesian cognitive diagnosis modeling: a tutorial. *J. Educ. Behav. Stat.* 44, 473–503. doi: 10.3102/1076998619826040

Zhao, Y., Staudenmayer, J., Coull, B. A., and Wand, M. P. (2006). General design bayesian generalized linear mixed models. *Statist. Sci.* 21, 35–51. doi: 10.1214/088342306000000015

Zitzmann, S., Lüdtke, O., and Robitzsch, A. (2015). A bayesian approach to more stable estimates of group-level effects in contextual studies. *Multivariate Behav. Res.* 50, 688–705. doi: 10.1080/00273171.2015.1090899

frontiers | Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 20 March 2023
DOI 10.3389/fncom.2023.1128694

Check for updates

OPEN ACCESS

EDITED BY
Jacob Raber,
Oregon Health and Science University,
United States

REVIEWED BY
Antonino Casile,
Italian Institute of Technology (IIT), Italy
Tim Verbelen,
Ghent University, Belgium

*CORRESPONDENCE
Ivilin Peev Stoianov
✉ ivilinpeev.stoianov@cnr.it

RECEIVED 23 December 2022
ACCEPTED 03 March 2023
PUBLISHED 20 March 2023

CITATION
Priorelli M and Stoianov IP (2023) Flexible intentions: An Active Inference theory. *Front. Comput. Neurosci.* 17:1128694.
doi: 10.3389/fncom.2023.1128694

Flexible intentions: An Active Inference theory

Matteo Priorelli and Ivilin Peev Stoianov*

Institute of Cognitive Sciences and Technologies (ISTC), National Research Council of Italy (CNR), Padua, Italy

We present a normative computational theory of how the brain may support visually-guided goal-directed actions in dynamically changing environments. It extends the Active Inference theory of cortical processing according to which the brain maintains beliefs over the environmental state, and motor control signals try to fulfill the corresponding sensory predictions. We propose that the neural circuitry in the Posterior Parietal Cortex (PPC) compute flexible intentions—or motor plans from a belief over targets—to dynamically generate goal-directed actions, and we develop a computational formalization of this process. A proof-of-concept agent embodying visual and proprioceptive sensors and an actuated upper limb was tested on target-reaching tasks. The agent behaved correctly under various conditions, including static and dynamic targets, different sensory feedbacks, sensory precisions, intention gains, and movement policies; limit conditions were individuated, too. Active Inference driven by dynamic and flexible intentions can thus support goal-directed behavior in constantly changing environments, and the PPC might putatively host its core intention mechanism. More broadly, the study provides a normative computational basis for research on goal-directed behavior in end-to-end settings and further advances mechanistic theories of active biological systems.

KEYWORDS

Active Inference, sensorimotor control, Posterior Parietal Cortex, intentions, Predictive Coding

1. Introduction

Traditionally, sensorimotor control in goal-directed actions like object-reaching is viewed as a sensory-response mapping involving several steps, starting with perception, movement planning in the body posture domain, translation of this plan in muscle commands, and finally movement execution (Erlhagen and Schöner, 2002). However, each of these steps is hindered by noise and delays, which make the approach unfeasible to operate in changing environments (Franklin and Wolpert, 2011). Instead, Predictive Coding or "Bayesian Brain" theories propose that prior knowledge and expectations over the environmental and bodily contexts provide crucial anticipatory information (Rao and Ballard, 1999). Under this perspective, motor control begins with target anticipation and motor planning even before obtaining sensory evidence. Here, we take on this view and extend an increasingly popular Predictive Coding based theory of action, Active Inference (Friston et al., 2010), with the formalization of flexible target-dependent motor plans. Moreover, based on extensive neural evidence for the role of the PPC in goal coding and motor planning (Snyder et al., 2000; Galletti et al., 2022), we propose that this cortical structure is the most likely neural correlate of the core intention manipulation process.

In primates, the dorsomedial visual stream provides critical support for continuously monitoring the body posture and the spatial location of objects to specify and guide actions, and for performing visuomotor transformations in the course of the evolving movement

(Cisek and Kalaska, 2010; Fattori et al., 2017; Galletti and Fattori, 2018). The PPC, located at the apex of the dorsal stream, is also bidirectionally connected to frontal areas, motor and somatosensory cortex, placing it in a privileged position to set goal-directed actions and continuously adjust motor plans by tracking moving targets and posture (Andersen, 1995; Gamberini et al., 2021) in a common reference frame (Cohen and Andersen, 2002). Undoubtedly, the PPC plays a crucial role in visually-guided motor control (Desmurget et al., 1999; Filippini et al., 2018; Gamberini et al., 2021)—with the specific subregion V6A involved in the control of reach-to-grasp actions (Galletti et al., 2022)—but its peculiar role is still disputed. The most consistent view is that the PPC estimates the states of both body and environment and optimizes their interactions (Medendorp and Heed, 2019). Others see the PPC as a task estimator (Haar and Donchin, 2020) or as being involved in endogenous attention and task setting (Corbetta and Shulman, 2002). Its underlying computational mechanism is not fully understood, especially as regards the definition of goals in motor planning and their integration within the control process (Shadmehr and Krakauer, 2008). For example, the prevailing Optimal Feedback Control theory defines motor goals through task-specific cost functions (Todorov, 2004). Neural-level details of motor goal coding are becoming increasingly important in light of the growing demand for neural interfaces that provide information about motor intents (Gallego et al., 2022) in support of intelligent assistive devices (Velliste et al., 2008; Srinivasan et al., 2021).

Intentions encode motor goals—or plans—set before the beginning of motor acts themselves and could be therefore viewed as memory holders of voluntary actions (Andersen, 1995; Snyder et al., 1997; Lau et al., 2004; Fogassi et al., 2005). Several cortical areas handle different aspects of this process: the Premotor cortex (PM) encodes structuring while the Supplementary Motor Area (SMA) controls phasing (Gallego et al., 2022). In turn, the PPC plays a role in building motor plans and their dynamic tuning, as different PPC neurons are sensitive to different intentions (Snyder et al., 2000). Notably, intention neurons respond not only when performing a given action but also during its observation, allowing observers to *predict the goal of the observed action and, thus, to "read" the intention of the acting individual* (Fogassi et al., 2005). Motor goals have also been observed down the motor hierarchy, which is an expression of Hierarchical Predictive Coding in the motor domain (Friston et al., 2011).

To investigate how neural circuitry in the PPC supports sensory-guided actions through motor intentions from a computational point of view, we adopted the Active Inference theory of cognitive and motor control, which provides fundamental insights of increasing appeal about the computational role and principles of the nervous system, especially about the perception-action loop (Friston and Kiebel, 2009; Friston et al., 2010; Bogacz, 2017; Parr et al., 2022). Indeed, Active Inference provides a formalization of these two cortical tasks, both of which are viewed as aiming to resolve the critical goal of all organisms: to survive in uncertain environments by operating within preferred states (e.g., maintaining a constant temperature). Accordingly, both tasks are implemented by dynamic minimization of a quantity called *free energy*, whose process generally corresponds to the minimization of high- and low-level prediction errors, that is, the satisfaction of prior and sensory expectations. There are two branches of Active Inference appropriate to tackle two different levels of control. The discrete framework can explain high-level cognitive control processes such as planning and decision-making, i.e., it evaluates expected outcomes to select actions in discrete entities (Pezzulo et al., 2018). In turn, dynamic adjustment of action plans in the PPC matches by functionality the Active Inference framework in continuous state space (Friston et al., 2011, 2012). In short, this theory departs from classical views of perception, motor planning (Erlhagen and Schöner, 2002), and motor control (Todorov, 2004), unifying and considering them as a dynamic probabilistic inference problem (Toussaint and Storkey, 2006; Kaplan and Friston, 2018; Levine, 2018; Millidge et al., 2020). The biologically implausible cost functions typical of Optimal Control theories are replaced by high-level priors defined in the extrinsic state space, allowing complex movements such as walking or handwriting (Friston, 2011; Adams et al., 2013).

In the following, we first outline the background computational framework and then elaborate on movement planning and intentionality in continuous Active Inference. Our most critical contributions regard the formalization of goal-directed behavior and the processes linking dynamic goals (e.g., moving visual targets) with motor plans through the definition of flexible intentions. We also investigate a more parsimonious approach to motor control based solely on proprioceptive predictions. We then provide implementation details and a practical demonstration of the theoretical contribution in terms of a simulated Active Inference agent, which we show is capable of detecting and reaching static visual goals and tracking moving targets. We also provide detailed performance statistics and investigate the effects of system parameters whose balance is critical to movement stability. Additionally, gradient analysis provides crucial insights into the causes of the movements performed. Finally, we discuss how intentions could be selected to perform a series of goal-directed steps, e.g., a multi-phase action, and illustrate conditions for neurological disorders.

2. Computational background

We first outline the computational principles of the underlying probabilistic and Predictive Coding approach and provide background on variational inference, free energy minimization, Active Inference, and variational autoencoders necessary to comprehend the following main contribution.

2.1. The Bayesian brain hypothesis

An interesting visual phenomenon, called *binocular rivalry*, happens when two different images are presented simultaneously to each eye: the perception does not conform to the visual input but alternates between the two images. How and why does this happen? It is well-known that priors play a fundamental role in driving the dynamics of perceptual experience, but dominant views of the brain as a feature detector that passively receives sensory signals and computes motor commands have so far failed to explain how such illusions could arise.

In recent years, there has been increasing attention to a radically new theory of the mind called the *Bayesian brain*, according to which our brain is a sophisticated machine that constantly makes use of Bayesian reasoning to capture causal relationships in the world and deliver optimal behavior in an uncertain environment (Doya, 2007; Hohwy, 2013; Pezzulo et al., 2017). At the core of the theory is the Bayes theorem, whose application here implies that posterior beliefs about the world are updated according to the product of prior beliefs and the likelihood of observing sensory input. In this view, perception is more than a simple bottom-up feedforward mechanism that detects features and objects from the current sensorium; rather, it comprises a predictive top-down generative model which continuously anticipates the sensory input to test hypotheses and explain away ambiguities.

According to the Bayesian brain hypothesis, this complex task is accomplished by *Predictive Coding*, implemented through message passing of top-down predictions and bottom-up prediction errors between adjacent cortical layers (Rao and Ballard, 1999). The former are generated from latent states maintained at the highest levels, representing beliefs about the causes of the environment, while the latter are computed by comparing sensory-level predictions with the actual observations. Each prediction will then act as a cause for the layer below, while the prediction error will convey information to the layer above. It is thanks to this hierarchical organization and through error minimization at every layer that the cortex is supposed to be able to mimic and capture the inherently hierarchical relationships that model the world. In this view, sensations are only needed in that they provide, through the computation of prediction errors, a measure of how good the model is and a cue to correct future predictions. Thus, ascending projections do not encode the features of a stimulus, but rather how much the brain is *surprised* about it, considering the strict correlation between surprise and model uncertainty.

2.2. Variational bayes

Organisms are supposed to implement model fit or error minimization by some form of variational inference, a broad family of techniques based on the *calculus of variations* and used to approximate intractable posteriors that would otherwise be infeasible to compute analytically or even with classical sampling methods like Monte Carlo (Bishop, 2006). Under the Bayesian brain hypothesis, we can assume that the nervous system maintains latent variables $\boldsymbol{z}$ about both the unknown state of the external world and the internal state of the organism. By exploiting a prior knowledge $p(\boldsymbol{z})$ and the partial evidence $p(\boldsymbol{s})$ of the environment provided by its sensors, it can apply Bayesian inference to improve its knowledge (Ma et al., 2006). To do so, given the observation $\boldsymbol{s}$, the nervous system needs to evaluate the posterior $p(\boldsymbol{z}|\boldsymbol{s})$:

$$p(\boldsymbol{z}|\boldsymbol{s}) = \frac{p(\boldsymbol{z},\boldsymbol{s})}{p(\boldsymbol{s})} \tag{1}$$

However, directly computing such quantity is infeasible due to the intractability of the marginal $p(\boldsymbol{s}) = \int p(\boldsymbol{z},\boldsymbol{s})d\boldsymbol{z}$, which involves integration over the joint density $p(\boldsymbol{z},\boldsymbol{s})$. What does the variational approach is approximating the posterior with a simpler to compute *recognition* distribution $q(\boldsymbol{z}) \approx p(\boldsymbol{z}|\boldsymbol{s})$ through minimization of the Kullback-Leibler (KL) divergence between them:

$$D_{KL}[q(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{s})] = \int_{\boldsymbol{z}} q(\boldsymbol{z}) \ln \frac{q(\boldsymbol{z})}{p(\boldsymbol{z}|\boldsymbol{s})} d\boldsymbol{z} \tag{2}$$

The KL divergence can be rewritten as the difference between log evidence $\ln p(\boldsymbol{s})$ and a quantity $\mathcal{L}(q)$ known as *evidence lower bound*, or ELBO (Bishop, 2006):

$$D_{KL}[q(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{s})] = \ln p(\boldsymbol{s}) - \int_{\boldsymbol{z}} q(\boldsymbol{z}) \ln \frac{p(\boldsymbol{z},\boldsymbol{s})}{q(\boldsymbol{z})} d\boldsymbol{z} = \ln p(\boldsymbol{s}) - \mathcal{L}(q) \tag{3}$$

Since the KL divergence is always nonnegative, the ELBO provides a lower bound on log evidence, i.e., $\mathcal{L}(q) \leq \ln p(\boldsymbol{s})$. Therefore, minimizing the KL divergence with respect to $q(\boldsymbol{z})$ is equivalent to maximizing $\mathcal{L}(q)$, which at its maximum corresponds to an approximate density that is closest the most to the real posterior, depending on the particular choice of the form of $q(\boldsymbol{z})$. In general, few assumptions are made about the form of this distribution—a multivariate Gaussian is a typical choice—with a trade-off between having a tractable optimization process and still leading to a good approximate posterior.

2.3. Free energy and prediction errors

How can Bayesian inference be implemented through a simple message passing of prediction errors? Friston (2002, 2005) proposed an elegant solution based on the so-called *free energy*, a concept borrowed from thermodynamics and defined as the negative ELBO. Accordingly, Equation (3) can be rewritten as:

$$\mathcal{F}(\boldsymbol{z},\boldsymbol{s}) = -\mathcal{L}(q) = D_{KL}[q(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{s})] - \ln p(\boldsymbol{s}) = \int_{\boldsymbol{z}} q(\boldsymbol{z}) \ln \frac{q(\boldsymbol{z})}{p(\boldsymbol{z},\boldsymbol{s})} d\boldsymbol{z} \tag{4}$$

Minimizing the free energy with respect to the latent states $\boldsymbol{z}$—a process called *perceptual inference*—is then equivalent to ELBO maximization and provides an upper bound on surprise:

$$\boldsymbol{z} = \arg\min_{\boldsymbol{z}} \mathcal{F}(\boldsymbol{z},\boldsymbol{s}) \tag{5}$$

In this way, the organism indirectly minimizes model uncertainty and is able to learn the causal relations between unknown states and sensory input, and to generate predictions based on its current representation of the environment. Free energy minimization is simpler than dealing with the KL divergence between the approximate and true posteriors as the former depends on quantities that the organism has access to, namely the approximate posterior and the *generative model*.

To this concern, it is necessary to distinguish between the latter and the real distribution producing sensory data, called *generative process*, which can be modeled with the following non-linear stochastic equations:

$$\begin{aligned} \boldsymbol{s} &= \boldsymbol{g}(\boldsymbol{z}) + \boldsymbol{w}_s \\ \dot{\boldsymbol{z}} &= \boldsymbol{f}(\boldsymbol{z}) + \boldsymbol{w}_z \end{aligned} \tag{6}$$

Where the function $\boldsymbol{g}$ maps latent states or causes $\boldsymbol{z}$ to observed states or sensations $\boldsymbol{s}$, the function $\boldsymbol{f}$ encodes the dynamics of the system, i.e., the evolution of $\boldsymbol{z}$ over time, while $\boldsymbol{w}_s$ and $\boldsymbol{w}_z$ are noise terms that describe system uncertainty.

Nervous systems are supposed to approximate the generative process by making a few assumptions: that (i) under the mean-field approximation the recognition density can be partitioned into independent distributions: $q(\boldsymbol{z}) = \prod_i q(\boldsymbol{z}_i)$, and that (ii) under the Laplace approximation each of these partitions is Gaussian: $q(\boldsymbol{z}_i) = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Pi}_i^{-1})$, where $\boldsymbol{\mu}_i$ represents the most plausible hypothesis—also called *belief* about the hidden state $\boldsymbol{z}_i$ - and $\boldsymbol{\Pi}_i$ is its precision matrix (Friston et al., 2007). In this way, the free energy does not depend on $\boldsymbol{z}$ and simplifies as follows:

$$\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{s}) = -\ln p(\boldsymbol{\mu}, \boldsymbol{s}) + C = -\ln p(\boldsymbol{s}|\boldsymbol{\mu}) - \ln p(\boldsymbol{\mu}) + C \quad (7)$$

Where C is a constant term. A more precise description of the unknown environmental dynamics can be achieved by considering not only the 1st order of Equation 6 but also higher temporal orders of the corresponding approximations: $\tilde{\boldsymbol{\mu}} = \{\boldsymbol{\mu}, \boldsymbol{\mu}', \boldsymbol{\mu}'', ...\}$—called *generalized coordinates* (Friston, 2008; Friston et al., 2008). This allows us to better represent the environment with the following generalized model:

$$\begin{aligned} \tilde{\boldsymbol{s}} &= \tilde{\boldsymbol{g}}(\tilde{\boldsymbol{\mu}}) + \boldsymbol{w}_s \\ \mathcal{D}\tilde{\boldsymbol{\mu}} &= \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\mu}}) + \boldsymbol{w}_\mu \end{aligned} \quad (8)$$

Where $\mathcal{D}$ is the differential (shift) operator matrix such that $\mathcal{D}\tilde{\boldsymbol{\mu}} = \{\boldsymbol{\mu}', \boldsymbol{\mu}'', ...\}$, $\tilde{\boldsymbol{s}}$ denotes the generalized sensors, while $\tilde{\boldsymbol{g}}$ and $\tilde{\boldsymbol{f}}$ denote the generalized model functions of all temporal orders. Note that in this system, the sensory data at a particular dynamical order $\boldsymbol{s}^{[d]}$—where $[d]$ is the order—engage only with the same order of belief $\boldsymbol{\mu}^{[d]}$, while the generalized equation of motion, or system dynamics, specifies the coupling between adjacent orders. Such equations are generated from the generalized likelihood and prior distributions, which can be expanded as follows:

$$\begin{aligned} p(\tilde{\boldsymbol{s}}|\tilde{\boldsymbol{\mu}}) &= \prod_d p(\boldsymbol{s}^{[d]}|\boldsymbol{\mu}^{[d]}) \\ p(\tilde{\boldsymbol{\mu}}) &= \prod_d p(\boldsymbol{\mu}^{[d+1]}|\boldsymbol{\mu}^{[d]}) \end{aligned} \quad (9)$$

As defined above, these variational probability distributions are assumed to be Gaussian:

$$\begin{aligned} p(\boldsymbol{s}^{[d]}|\boldsymbol{\mu}^{[d]}) &= \frac{\boldsymbol{\Pi}_s}{\sqrt{(2\pi)^L}} \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}_s^{[d]T}\boldsymbol{\Pi}_s\boldsymbol{\varepsilon}_s^{[d]}\right) \\ p(\boldsymbol{\mu}^{[d+1]}|\boldsymbol{\mu}^{[d]}) &= \frac{\boldsymbol{\Pi}_\mu}{\sqrt{(2\pi)^M}} \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}_\mu^{[d]T}\boldsymbol{\Pi}_\mu\boldsymbol{\varepsilon}_\mu^{[d]}\right) \end{aligned} \quad (10)$$

Where L and M are the dimensions of sensations and internal beliefs, respectively with precisions $\boldsymbol{\Pi}_s$ and $\boldsymbol{\Pi}_\mu$. Note that the probability distributions are expressed in terms of sensory and dynamics prediction errors:

$$\boldsymbol{\varepsilon}_s^{[d]} = \boldsymbol{s}^{[d]} - \boldsymbol{g}^{[d]}(\boldsymbol{\mu}^{[d]}) \quad (11)$$

$$\boldsymbol{\varepsilon}_\mu^{[d]} = \boldsymbol{\mu}^{[d+1]} - \boldsymbol{f}^{[d]}(\boldsymbol{\mu}^{[d]}) \quad (12)$$

The factorized probabilistic approximation of the dynamic model allows easy state estimation performed by iterative gradient descent over the generalized coordinates, that is, by changing the belief $\tilde{\boldsymbol{\mu}}$ over the hidden states at every temporal order:

$$\dot{\tilde{\boldsymbol{\mu}}} - \mathcal{D}\tilde{\boldsymbol{\mu}} = -\partial_{\tilde{\boldsymbol{\mu}}}\mathcal{F}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{s}}) \quad (13)$$

Gradient descent is tractable because the Gaussian variational functions are smooth and differentiable and the derivatives are easily computed in terms of generalized prediction errors, since the logarithm of Equation (7) vanishes the exponent of the Gaussian. The belief update thus turns to:

$$\dot{\tilde{\boldsymbol{\mu}}} = \mathcal{D}\tilde{\boldsymbol{\mu}} + \frac{\partial\tilde{\boldsymbol{g}}}{\partial\tilde{\boldsymbol{\mu}}}^T \tilde{\boldsymbol{\Pi}}_s\tilde{\boldsymbol{\varepsilon}}_s + \frac{\partial\tilde{\boldsymbol{f}}}{\partial\tilde{\boldsymbol{\mu}}}^T \tilde{\boldsymbol{\Pi}}_\mu\tilde{\boldsymbol{\varepsilon}}_\mu - \mathcal{D}^T\tilde{\boldsymbol{\Pi}}_\mu\tilde{\boldsymbol{\varepsilon}}_\mu \quad (14)$$

It is crucial to keep in mind the nature of the three components that compose this update equation: a likelihood error computed at the sensory level, a backward error arising from the next temporal order, and a forward error coming from the previous order. These terms represent the free energy gradients relative to the belief $\boldsymbol{\mu}^{[d]}$ of Equation (11) for the likelihood, and $\boldsymbol{\mu}^{[d+1]}$ and $\boldsymbol{\mu}^{[d]}$ of Equation (12) for the dynamics errors.

In short, by making a few plausible simplifying assumptions, the complexity of free energy minimization reduces to the generation of predictions, which are constantly compared with sensory observations to determine a prediction error signal. This error then flows back through the cortical hierarchy to adjust the distribution parameters accordingly and minimize sensory surprise—or maximize evidence—in the long run.

2.4. Active Inference

Describing the relationship between Predictive Coding and Bayesian inference still does not explain why has the cortex evolved in such a peculiar way. The answer comes from the so-called *free energy principle* (FEP), regard to which the Bayesian brain hypothesis is supposed to be a corollary. Indeed, learning the causal relationships of some observed data (e.g., what causes an increase in body temperature) is insufficient to keep organisms alive (e.g., maintaining the temperature in a vital range).

The FEP states that, for an organism to maintain a state of homeostasis and survive, it must constantly and actively restrict the set of latent states in which it lives to a narrow range of life-compatible possibilities, counteracting the natural tendency for disorder (Friston, 2012)—hence the relationship with thermodynamics. If these states are defined by the organism's phenotype, from the point of view of its internal model they are exactly the states that it expects to be less surprising. Thus, while perceptual inference tries to optimize the belief about hidden causes to explain away sensations, if on the other hand the assumptions defined by the phenotype are considered to be the true causes of the world, interacting with the external environment means that the agent will try to sample those sensations that make the assumptions true, fulfilling its needs and beliefs. *Active inference* becomes a self-fulfilling prophecy. In this view, there is no difference between a desire and a belief: we simply seek the states in which we expect to find ourselves (Friston et al., 2010; Buckley et al., 2017).

For achieving a goal-directed behavior, it is then sufficient to minimize the free energy also with respect to the action (see Equation 7):

$$\boldsymbol{a} = \arg\min_{\boldsymbol{a}} \mathcal{F}(\boldsymbol{\mu}, \boldsymbol{s}) \tag{15}$$

Given that motor control signals only depend on sensory information, we obtain:

$$\dot{\boldsymbol{a}} = -\partial_{\boldsymbol{a}}\mathcal{F}(\tilde{\boldsymbol{s}}, \tilde{\boldsymbol{\mu}}) = -\frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{s}}}\frac{\partial \tilde{\boldsymbol{s}}}{\partial \boldsymbol{a}} = -\frac{\partial \tilde{\boldsymbol{s}}}{\partial \boldsymbol{a}}^{T} \tilde{\boldsymbol{\Pi}}_s \tilde{\boldsymbol{\varepsilon}}_s \tag{16}$$

Minimizing the free energy of all sensory signals is certainly useful, as every likelihood contribution will drive the belief update; however, it requires the knowledge of an inverse mapping from exteroceptive sensations to actions (Baltieri and Buckley, 2019), which is considered a "hard problem" being in general highly non-linear and not univocal (Friston et al., 2010). In a more realistic scenario, only proprioception drives the minimization of free energy with respect to the motor signals; this process is easier to realize since the corresponding sensory prediction is already in the intrinsic domain. Control signals sent from the motor cortex are then not motor commands as in classical views of Optimal Control theories; rather, they consist of predictions that define the desired trajectory. Under this perspective, proprioceptive prediction errors computed locally at the spinal cord serve two purposes that only differ in how these signals are conveyed. They drive the current belief toward sensory observations—happening to realize perception—like for exteroceptive signals. But they also drive sensory observations toward the current belief by suppression in simple reflex arcs that activate the corresponding muscles—thus happening to realize movement (Adams et al., 2013; Parr and Friston, 2018; Versteeg et al., 2021).

In conclusion, perception and action can be seen as two sides of the same coin implementing the common vital goal of minimizing entropy or average surprise. In this view, what we perceive never tries to perfectly match the real state of affairs of the world, but is constantly biased toward our preferred states. This means that action only indirectly fulfills future goals; instead, it continuously tries to fill the gap between sensations and predictions generated from our already biased beliefs.

2.5. Variational autoencoders

Variational Autoencoders (VAEs) belong to the family of generative models, since they learn the joint distribution $p(\boldsymbol{z}, \boldsymbol{s})$ and can generate synthetic data similar to the input, given a prior distribution $p(\boldsymbol{z})$ over the latent space. VAEs use the variational Bayes approach to capture the posterior distribution $p(\boldsymbol{z}|\boldsymbol{s})$ of the latent representation of the inputs when the computation of the marginal is intractable (Goodfellow et al., 2016). A VAE is composed of two probability distributions, both of which are assumed to be Gaussian: a probabilistic *encoder* corresponding to the recognition distribution $q(\boldsymbol{z}|\boldsymbol{s})$, and a generative function $p(\boldsymbol{s}|\boldsymbol{z})$ called probabilistic *decoder* computing a distribution over the input space given a latent representation $\boldsymbol{z}$ (Figure 3C):

$$\begin{aligned} q(\boldsymbol{z}|\boldsymbol{s}) &= \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\phi) \\ p(\boldsymbol{s}|\boldsymbol{z}) &= \mathcal{N}(\boldsymbol{s}|\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) \end{aligned} \tag{17}$$

Although VAEs have many similarities with traditional autoencoders, they are actually a derivation of the AEVB algorithm when a neural network is used for the recognition distribution (Kingma and Welling, 2014). Unlike other variational techniques, the approximate posterior is generally not assumed to be factorial, but since the calculation of the ELBO gradient $\nabla_\phi \mathcal{L}_{\theta,\phi}(\boldsymbol{s})$ is biased, a method called *reparametrization trick* is used so that it is independent of the parameters $\boldsymbol{\phi}$. This method works by expressing the latent variable $\boldsymbol{z}$ by a function:

$$\boldsymbol{z} = r(\boldsymbol{\epsilon}, \boldsymbol{\phi}, \boldsymbol{s}) \tag{18}$$

Where $\boldsymbol{\epsilon}$ is an auxiliary variable independent of $\boldsymbol{\phi}$ and $\boldsymbol{s}$. The ELBO $\tilde{\mathcal{L}}_{\theta,\phi}(\boldsymbol{s})$ for a single data point can thus be expressed as:

$$\tilde{\mathcal{L}}_{\theta,\phi}(\boldsymbol{s}) = -D_{KL}[q(\boldsymbol{z}|\boldsymbol{s})||p(\boldsymbol{z})] + \frac{1}{M}\sum_{m}^{M} \log p(\boldsymbol{s}|\boldsymbol{z}_m) \tag{19}$$

Which can be minimized through backpropagation. Here, the KL divergence can be seen as a regularizer, while the second RHS term is an expected negative reconstruction term that depends on all the mth components of the latent variable $\boldsymbol{z}$.

3. A framework for flexible intentions

In what follows, we develop a computational theory of the circuitry controlling goal-directed actions in a dynamically changing environment through flexible intentions and discuss its putative neural basis in the PPC and related areas. We first elaborate on intentionality in Active Inference, then provide a proof-of-concept agent endowed with visual input. The theory is exemplified and assessed in the following sections through simulations of visually-guided behaviors. The theoretical work is motivated by basic research showing the critical role of the PPC in goal-directed sensorimotor control through intention coding (Andersen, 1995; Desmurget et al., 1999; Galletti and Fattori, 2018) and extends previous theoretical and applied research on Active Inference (Friston et al., 2009; Pio-Lopez et al., 2016; Lanillos and Cheng, 2018; Limanowski and Friston, 2020) and VAE-based vision support (Rood et al., 2020; Sancaktar et al., 2020). The simulations are inspired by a classical monkey reaching task (Breveglieri et al., 2014).

3.1. Flexible intentions

State-of-the-art implementations of continuous Active Inference have proven to successfully tackle a wide range of tasks, from oculomotion dynamics (Adams et al., 2015) to the well-known mountain car problem (Friston et al., 2009). Most simulations involve reaching movements in robotic experiments, where several strategies have been tried for designing goal states,

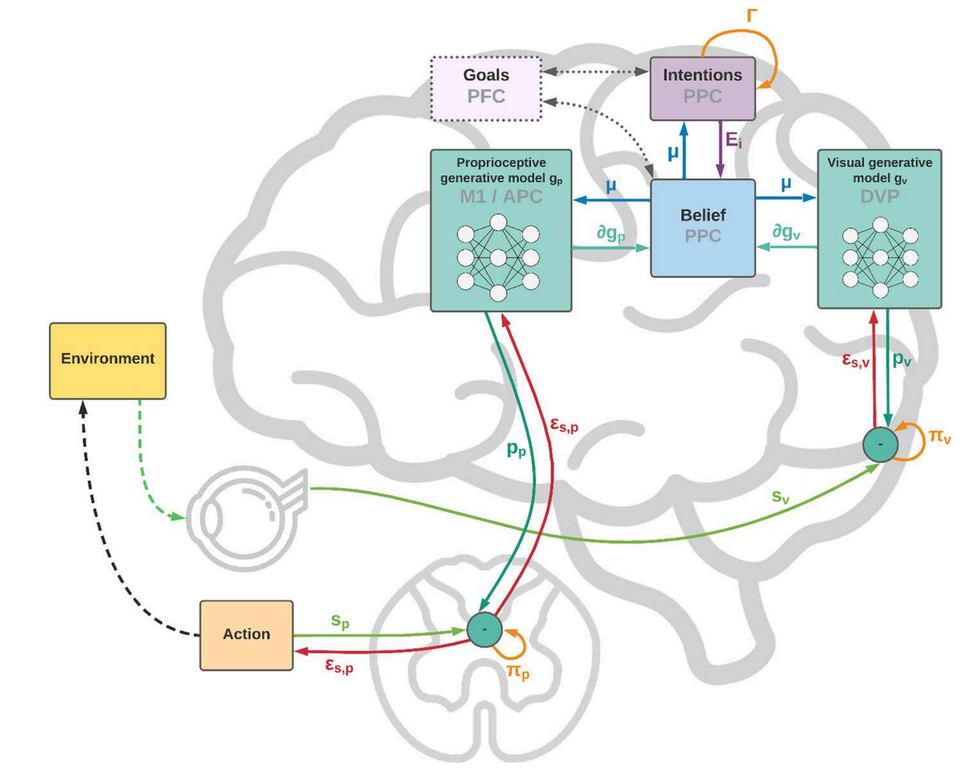


FIGURE 1
Functional architecture and cortical overlay. The process starts with the computation of future intentions $\boldsymbol{h}$ (not explicitly represented in the figure) in the PPC under the coordination of frontal and motor areas. In the middle of the sensorimotor hierarchy, the PPC maintains beliefs $\boldsymbol{\mu}$ over the latent causes of sensory observations $\boldsymbol{s}_p$ and $\boldsymbol{s}_v$, and computes proprioceptive and visual predictions through the somatosensory and dorsal visual pathways (for simplicity, we have omitted the somatomotor pathway and considered a single mechanism for both motor control and belief inference). The lower layers of the hierarchy compute sensory prediction errors $\boldsymbol{\varepsilon}_{s_p}$ and $\boldsymbol{\varepsilon}_{s_v}$, while the higher layers compute intention prediction errors $\boldsymbol{E}_i$; both are propagated back toward the PPC, which thus integrates information from multiple sensory modalities and intentions. Free energy is minimized throughout the cortical hierarchy by changing the belief about the causes of the current observation (perception) and by sending proprioceptive predictions from the motor cortex to the reflex arcs (action). An essential element of this process is the computation of gradients $\partial \boldsymbol{g}_p$ and $\partial \boldsymbol{g}_v$ by inverse mappings from the sensations toward the deepest latent states. In this process, intentions act as high-level attractors and the belief propagated down to compute sensorimotor predictions embeds a component directing the body state toward the goals.

which are expressed in terms of an attractor embedded in the system dynamics. However, there seem to be a few issues regarding biological plausibility. First, the goal state is usually static and the agent is not able to deal with continuously changing environments, expecting that the world will always evolve in the same way (Baioumy et al., 2020). For dynamic goals, one has to use low-level information of sensory signals (e.g., a visual input about a moving target) directly into the high-level dynamics function (Friston, 2011). Second, when goals are specified in an exteroceptive domain, one uses sensory predictions to obtain a belief update direction through backpropagation of the corresponding error (Oliver et al., 2019; Sancaktar et al., 2020). In this case, the same generative model that produces predictions and compares them with the actual observations, has to be duplicated into the system dynamics to further compare the belief with the desired cue. In other words, two specular mechanisms are used for the same model, with additional concerns when the latter can be changed by learning.

A common question seems to be behind these two similar issues: how does dynamic sensory information get available for generating high-level dynamic goals? The same inference process of environmental causes should be at work for the same signal flow, and a goal state should be computed locally without information passed inconsistently. How then to design a flexible exteroceptive attractor that avoids implausible scenarios?

Although the high-level latent state could be as simple as encoding body configurations only, an agent could also maintain a dynamically estimated belief over moving objects in the scene. An intention can then be computed by exploiting this new information to compute a future action goal in terms of body posture, so that the attractor—either defined in the belief domain or at the sensory level—is not fixed but depends on current perceptual and internal representations of the world (but also on past memorized experiences). This intention may also depend on priors generated from higher-level areas (Friston et al., 2011), so that the considered belief is located at an intermediate level between the generative models that produce sensory predictions, and the ones that define its evolution over time. In a non-trivial task, its dynamics may be generally composed of several contributions and not restricted to a single intention: we thus propose to decompose it into a set of functions, each one providing an independent expectation that the agent will find itself in a particular state. The belief is then constantly subject to several forces of two natures: one from lower hierarchical levels—proportional to *sensory prediction errors*—that pulls it toward what the agent is currently perceiving, and one from lateral or higher connections—which we call *intention prediction errors*—that pulls it toward what the agent expects to perceive in the future.

As shown in Figure 1, from a neural perspective the PPC is the ideal candidate for a cortical structure computing beliefs over bodily states and flexible intentions: on one hand, being at the apex of the Dorsal Visual Stream (DVS) and other sensory generative models, and on the other linked with motor and frontal areas that produce continuous trajectories and plans of discrete action chunks. The PPC is known to be an associative region that integrates information from multiple sensory modalities and encodes visuomotor transformations—e.g., area V6A is thought to encode object affordances during reaching and grasping tasks (Fattori et al., 2017; Filippini et al., 2017). Moreover, evidence suggests that the PPC encodes multiple goals in parallel during sequences of actions, even when there is a considerable delay between different goals (Baldauf et al., 2008).

In short, the agent constantly maintains plausible hypotheses over the causes of its percepts, either bodily states or objects in exteroceptive domains; by manipulating them, the agent dynamically constructs representations of future states, i.e., intentions, which in turn act as priors over the current belief. Thus, if the job of the sensory pathways is to compute sensory-level predictions, we hypothesize that higher levels of the sensorimotor control hierarchy integrate into the PPC previous states of belief with flexible intentions, each predicting the next plausible belief state.

3.2. Dynamic goal-directed behavior in Active Inference

For a more formal definition, we assume that the neural system perceives the environment and receives motor feedback through J noisy sensors $\boldsymbol{S}$ comprising multiple domains (most critically, proprioceptive and visual). Under the VB and Gaussian approximations of the recognition density, we also assume that the nervous system operates on beliefs $\boldsymbol{\mu} \in \mathbb{R}^{\mathrm{M}}$ that define an abstract internal representation of the world. Furthermore, we assume that the agent maintains generalized coordinates up to the 1st order resulting from free energy minimization in the generalized belief space $\tilde{\boldsymbol{\mu}} = \{\boldsymbol{\mu}, \boldsymbol{\mu}'\}$.

We then define *intentions* $\boldsymbol{h}_k$ as predictions of target goal states over the current belief $\boldsymbol{\mu}$ computed with the help of K functions $\boldsymbol{i}_k(\boldsymbol{\mu}) \in \mathbb{R}^{\mathrm{M}}$. Although both belief and intentions could be abstract representations of the world—comprising states in extrinsic and intrinsic coordinates—we assume a simpler scenario in which the intentions operate on beliefs in a common intrinsic motor-related domain, e.g., the joint angles space. As explained before, we assume that there are two conceptually different components in both the belief $\boldsymbol{\mu}$ and the output of the intention functions $\boldsymbol{i}_k$. The first component could represent the bodily states and serve to drive actions, while the second one could represent the state of other objects—mostly targets to interact with—which can be internally encoded in the joint angles space as well (the reason for this particular encoding will be clear later). These targets could be observed, but they could also be imagined or set by higher-level cognitive control frontal areas such as the PFC or PMd (Genovesio et al., 2012; Stoianov et al., 2016).

For the sake of notational simplicity, we group all intentions into a single matrix $\boldsymbol{H} \in \mathbb{R}^{\mathrm{MxK}}$:

$$\boldsymbol{H} = \boldsymbol{i}(\boldsymbol{\mu}) = \begin{bmatrix} \boldsymbol{i}_0(\boldsymbol{\mu}) & \dots & \boldsymbol{i}_K(\boldsymbol{\mu}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_0 & \dots & \boldsymbol{h}_K \end{bmatrix} \tag{20}$$

Intention prediction errors $\boldsymbol{e}_{i_k}$ are then defined as the difference between the current belief and every intention:

$$\boldsymbol{E}_i = \boldsymbol{H} - \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{e}_{i_0} & \dots & \boldsymbol{e}_{i_K} \end{bmatrix} \tag{21}$$

In turn, sensory predictions are produced by a set of generative models $\boldsymbol{g}_j$, one for each sensory modality. We group the predictions into a prediction matrix $\boldsymbol{P}$:

$$\boldsymbol{P} = \boldsymbol{g}(\boldsymbol{\mu}) = \begin{bmatrix} \boldsymbol{g}_0(\boldsymbol{\mu}) \\ \vdots \\ \boldsymbol{g}_J(\boldsymbol{\mu}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_0 \\ \vdots \\ \boldsymbol{p}_J \end{bmatrix} \tag{22}$$

Note that each term $\boldsymbol{p}_j$ is a multidimensional sensory-level representation that provides predictions for a particular sensory domain, with its own dimensionality, which we group into a single quantity for notational simplicity. Sensory prediction errors $\boldsymbol{\varepsilon}_{s_j}$ are then computed as the difference between sensations from each domain and the corresponding sensory-level predictions:

$$\mathcal{E}_s = \boldsymbol{S} - \boldsymbol{P} = \begin{bmatrix} \boldsymbol{\varepsilon}_{s_0} \\ \vdots \\ \boldsymbol{\varepsilon}_{s_J} \end{bmatrix} \tag{23}$$

Under the assumption of independence among intentions and sensations, we can factorize the joint probability of the generative model into a product of distributions for each sensory modality and intention, which expands as follows:

$$p(\tilde{\boldsymbol{\mu}}, \boldsymbol{s}) = p(\boldsymbol{\mu}) \prod_k^K p(\boldsymbol{\mu}'_k|\boldsymbol{\mu}) \cdot \prod_j^J p(\boldsymbol{s}_j|\boldsymbol{\mu}) \tag{24}$$

In the following, we will not consider the prior probability over the 0th order belief $p(\boldsymbol{\mu})$. The other probability distributions are assumed to be Gaussian:

$$\begin{aligned} p(\boldsymbol{\mu}'_k|\boldsymbol{\mu}) &= \mathcal{N}(\boldsymbol{\mu}'_k|\boldsymbol{f}_k(\boldsymbol{\mu}), \boldsymbol{\gamma}_k^{-1}) \\ p(\boldsymbol{s}_j|\boldsymbol{\mu}) &= \mathcal{N}(\boldsymbol{s}_j|\boldsymbol{g}_j(\boldsymbol{\mu}), \boldsymbol{\pi}_j^{-1}) \end{aligned} \tag{25}$$

Where $\boldsymbol{\gamma}_k$ and $\boldsymbol{\pi}_j$ are, respectively, the precisions of intention k and sensor j. Here, $\boldsymbol{\mu}'_k$ and $\boldsymbol{f}_k$ correspond to the kth component of the 1st order dynamics function:

$$\boldsymbol{f}_k(\boldsymbol{\mu}) = \lambda \boldsymbol{e}_{i_k} + \boldsymbol{w}_{\mu_k} \tag{26}$$

Where λ is the gain of intention prediction errors $\boldsymbol{E}_i$. Note that the goal states are embedded into these functions, acting as belief-level attractors for each intention, so that the agent expects to be pulled toward target states with a velocity proportional to the error. Although the generalized belief allows encoding information about the dynamics of the true generative process, in the simple case delineated the agent does not have any such prior. For example, the agent does not know the trajectory of a moving target in advance (whose prior, in a more realistic scenario, would be present and acquired through learning of past experiences) and will update the belief only relying on the incoming sensory information. Nevertheless, the agent maintains (false) expectations about target dynamics, and it is indeed the discrepancy between the evolution of the (real) generative process and that of the (internal and biased) generative model that makes it able to implement a goal-directed behavior.

The prediction errors of the dynamics functions can be grouped into a single matrix:

$$\mathcal{E}_\mu = \boldsymbol{\mu}' - \lambda \boldsymbol{E}_i \tag{27}$$

From Equation (14), we can now compute the free energy derivative with respect to the belief:

$$\dot{\tilde{\boldsymbol{\mu}}} = \begin{bmatrix} \dot{\boldsymbol{\mu}} \\ \dot{\boldsymbol{\mu}}' \end{bmatrix} = \mathcal{D}\tilde{\boldsymbol{\mu}} - \partial_{\tilde{\boldsymbol{\mu}}}\mathcal{F} = \begin{bmatrix} \boldsymbol{\mu}' + \boldsymbol{G}^T(\boldsymbol{\Pi} \odot \mathcal{E}_s) + (\boldsymbol{F} \odot \mathcal{E}_\mu)\boldsymbol{\Gamma}^T \\ -\mathcal{E}_\mu \boldsymbol{\Gamma}^T \end{bmatrix} \tag{28}$$

Here, $\odot$ is the element-wise product, $\boldsymbol{G}$ and $\boldsymbol{F}$ enclose the gradients of all sensory generative models and dynamics functions, while $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}$ comprise all sensory and intention precisions:

$$\boldsymbol{G} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\mu}} = \begin{bmatrix} \partial \boldsymbol{g}_0 \\ \vdots \\ \partial \boldsymbol{g}_J \end{bmatrix} \qquad \boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_0 \\ \vdots \\ \boldsymbol{\pi}_J \end{bmatrix}$$
$$\boldsymbol{F} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\mu}} = \begin{bmatrix} \partial \boldsymbol{f}_0 & \dots & \partial \boldsymbol{f}_K \end{bmatrix} \qquad \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\gamma}_0 & \dots & \boldsymbol{\gamma}_K \end{bmatrix} \tag{29}$$

In the following, we will neglect the backward error in the 0th order of Equation (28) since it has a much smaller impact on the overall dynamics, and treat as the actual attractor force the forward error at the 1st order:

$$\dot{\tilde{\boldsymbol{\mu}}} \approx \begin{bmatrix} \boldsymbol{\mu}' + \boldsymbol{G}^T(\boldsymbol{\Pi} \odot \mathcal{E}_s) \\ -\mathcal{E}_\mu \boldsymbol{\Gamma}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}' + \boldsymbol{\epsilon}_s \\ \boldsymbol{\epsilon}_i \end{bmatrix} \tag{30}$$

Where $\boldsymbol{\epsilon}_s$ and $\boldsymbol{\epsilon}_i$, respectively, stand for the contributions (in the belief domain) of precision-weighted sensory and intention prediction errors. Considering the 1st order forward error as attractive force instead of the 0th order backward error results in simpler computations since there is no gradient of the dynamics functions to be considered. Further studies are however needed to understand the relationships between these two forces in goal-directed behavior. We can interpret $\boldsymbol{\gamma}_k$ as a quantity that determines the relative attractor gain of intention k, so that intentions with greater strength have a more significant impact on the overall update direction; these gains could also be modulated by projections from higher-levels areas applying cognitive control. In turn, $\boldsymbol{\pi}_j$ corresponds to the confidence about each sensory modality j, so that the agent relies more on sensors with higher strength.

Similarly, we can compute control signals by minimizing the free energy with respect to the actions, expressing the mapping from sensations to actions by:

$$\begin{aligned} \frac{\partial \boldsymbol{s}}{\partial \boldsymbol{a}} &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{a}} \cdot \boldsymbol{G} \\ \dot{\boldsymbol{a}} &= -\partial_a \mathcal{F} = -\partial_a \boldsymbol{\mu}^T \boldsymbol{\epsilon}_s \end{aligned} \tag{31}$$

Where $\partial_a \boldsymbol{\mu}$ is an inverse model from belief to actions. If motor signals are defined in terms of joint velocities, we can decompose and approximate the inverse model as follows:

$$\partial_a \boldsymbol{\mu} = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{a}} \cdot \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{g}_p}{\partial \boldsymbol{a}} \cdot \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{g}_p} = \Delta_t \boldsymbol{G}_p^{-1} \tag{32}$$

Where $\boldsymbol{\theta}$ are the joint angles, the subscript p indicates the proprioceptive contribution and we approximated $\partial_a \boldsymbol{g}_p$ by a time constant Δ_t (Oliver et al., 2019). If we assume that the belief over hidden states is encoded in joint angles, the computation of the inverse model may be as simple as finding the pseudoinverse of a matrix. However, if the belief is specified in a more generic reference frame and the proprioceptive generative model is a non-linear function, it could be harder to compute the corresponding gradient, causing additional control problems like temporal delays on sensory signals (Friston, 2011). Alternatively, we can consider a motor control driven only by proprioceptive predictions, so that the control signal is already in the correct domain and may be achieved through simple reflex arc pathways (Adams et al., 2013; Versteeg et al., 2021). In this case, all that is needed is a mapping from proprioceptive predictions to actions:

$$\dot{\boldsymbol{a}} = -\partial_a \mathcal{F}_p = -\Delta_t \cdot \boldsymbol{\pi}_p \boldsymbol{\varepsilon}_p \tag{33}$$

Expressing in Equation (31) the mapping from sensations to actions by the product of the inverse model $\partial_a \boldsymbol{\mu}$ and the gradient of the generative models allows the control signals to be defined in terms of the weighted sensory contribution $\boldsymbol{\epsilon}_s$, already computed during the inference process. Such an approach may have some computational advantages (as will be explained later), but it is unlikely to be implemented in the nervous system as control signals are supposed to convey predictions and not prediction errors (Adams et al., 2013).

Algorithm 1 outlines a schematic description of the flow of dynamic computations. For simplicity, we used the term "intention" also when describing the dynamics functions and their precisions, but one has to keep in mind the difference between the intention prediction errors $\boldsymbol{E}_i$, which directly encode the direction toward target states, and the dynamics prediction errors $\mathcal{E}_\mu$, which arise from the derivation of the corresponding probability distributions.

3.3. Neural implementation

Figure 2 shows a schematic neuronal representation of the proposed agent, which further extends earlier perceptual inference schemes (Bogacz, 2017) to full-blown Active Inference. In this simple model, the intentions consist of a single layer with two neurons, and the goal states are implicitly defined in the dynamics functions; however, in a realistic setting the latter would be composed of networks of neurons where these states are explicitly encoded, and non-linear functions could also be used to achieve more advanced behaviors. Note also that intentions $\boldsymbol{h}_k$ and sensory generative models $\boldsymbol{g}_j$ are all part of the same architecture, the only difference being the location in the cortical hierarchy.

Low-level prediction errors for each sensory modality are represented by neurons whose dynamics depends on both observations and predictions of the sensory generative models:

$$\dot{\boldsymbol{\varepsilon}}_{s_j} = \boldsymbol{s}_j - \boldsymbol{g}_j(\boldsymbol{\mu}) - \frac{\boldsymbol{\varepsilon}_{s_j}}{\boldsymbol{\pi}_j} \tag{34}$$

Upon convergence of the neural activity, that is, $\dot{\boldsymbol{\varepsilon}}_{s_j} = 0$, we obtain the prediction error computation derived above. In turn, the internal activity of neurons corresponding to high-level prediction errors is obtained by subtracting the generated dynamics function from the 1st order belief:

$$\dot{\boldsymbol{\varepsilon}}_{\mu_k} = \boldsymbol{\mu}' - \boldsymbol{f}_k(\boldsymbol{\mu}) - \frac{\boldsymbol{\varepsilon}_{\mu_k}}{\boldsymbol{\gamma}_k} \tag{35}$$

```
Input: $\boldsymbol{S}, \boldsymbol{i}, \boldsymbol{g}, \partial_a\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Pi}, \lambda, \Delta_t$
$\boldsymbol{\mu}, \boldsymbol{\mu}', \boldsymbol{\mu}'', \leftarrow$ InitializeBelief()
while $t < T$ do
    $\boldsymbol{H} \leftarrow \boldsymbol{i}(\boldsymbol{\mu})$ ▷ Intentions and sensory predictions
    $\boldsymbol{P} \leftarrow \boldsymbol{g}(\boldsymbol{\mu})$
    $\mathcal{E}_\mu \leftarrow \boldsymbol{\mu}' - \lambda(\boldsymbol{H} - \boldsymbol{\mu})$ ▷ Prediction errors
    $\mathcal{E}_s \leftarrow \boldsymbol{S} - \boldsymbol{P}$
    $\boldsymbol{\epsilon}_i \leftarrow -\mathcal{E}_\mu \boldsymbol{\Gamma}^T$ ▷ Precision-weighted contributions
    $\boldsymbol{\epsilon}_s \leftarrow \boldsymbol{G}^T(\boldsymbol{\Pi} \odot \mathcal{E}_s)$
    $\dot{\boldsymbol{\mu}} \leftarrow \boldsymbol{\mu}' + \boldsymbol{\epsilon}_s$ ▷ Belief and action update
    $\dot{\boldsymbol{\mu}}' \leftarrow \boldsymbol{\mu}'' + \boldsymbol{\epsilon}_i$
    $\dot{\boldsymbol{a}} \leftarrow -\partial_a\boldsymbol{\mu} \cdot \boldsymbol{\epsilon}_s$
    $\tilde{\boldsymbol{\mu}} \leftarrow \tilde{\boldsymbol{\mu}} + \Delta_t\dot{\tilde{\boldsymbol{\mu}}}$ ▷ Gradient descent
    $\boldsymbol{a} \leftarrow \boldsymbol{a} + \Delta_t\dot{\boldsymbol{a}}$
end while
```

Algorithm 1. Active Inference agent with flexible intentions.

Having received information coming from the top and bottom of the hierarchy, the belief is updated by integrating every signal:

$$\dot{\boldsymbol{\mu}} = \sum_j^J \partial\boldsymbol{g}_j\boldsymbol{\varepsilon}_{s_j} + \sum_k^K \partial\boldsymbol{f}_k\boldsymbol{\varepsilon}_{\mu_k} \tag{36}$$

Which parallels the update formula derived above (Equation 28). Correspondingly, the 1st order component of the belief is updated as follows:

$$\dot{\boldsymbol{\mu}}' = -\sum_k^K \boldsymbol{\varepsilon}_{\mu_k} \tag{37}$$

The belief is thus constantly pushed toward a direction that matches sensations on one side and intentions on the other. We adopted the idea that the slow-varying precisions are encoded as synaptic strengths (Bogacz, 2017), but alternative views consider them as gains of superficial pyramidal neurons (Bastos et al., 2012). In any case, they could be dynamically optimized during inference in a direction that minimizes free energy—e.g., if a sensory modality does not help predict sensations, its weight will decrease. This is also true for the intention weights: by dynamically changing during the movement, they can act as modulatory signals that select the best intention to realize at every moment, which can be useful for solving simultaneous or sequential tasks. Nonetheless, the distinction is purely conceptual as the agent does not discriminate between modulating a future intention or increasing the confidence of a sensory signal. At the belief level, every element just follows the rules of free energy minimization.

4. Method

To demonstrate the feasibility of the approach and its capacity to successfully implement goal-directed behavior in dynamic environments, we simulated an agent consisting of an actuated upper limb with visual and proprioceptive sensors that allow it

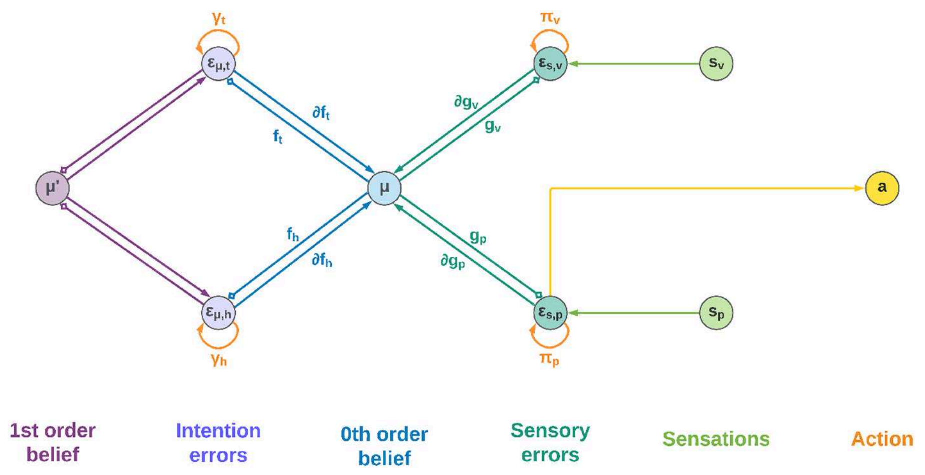


FIGURE 2
Neuronal representation with two intentions. Small squares stand for inhibitory connections.

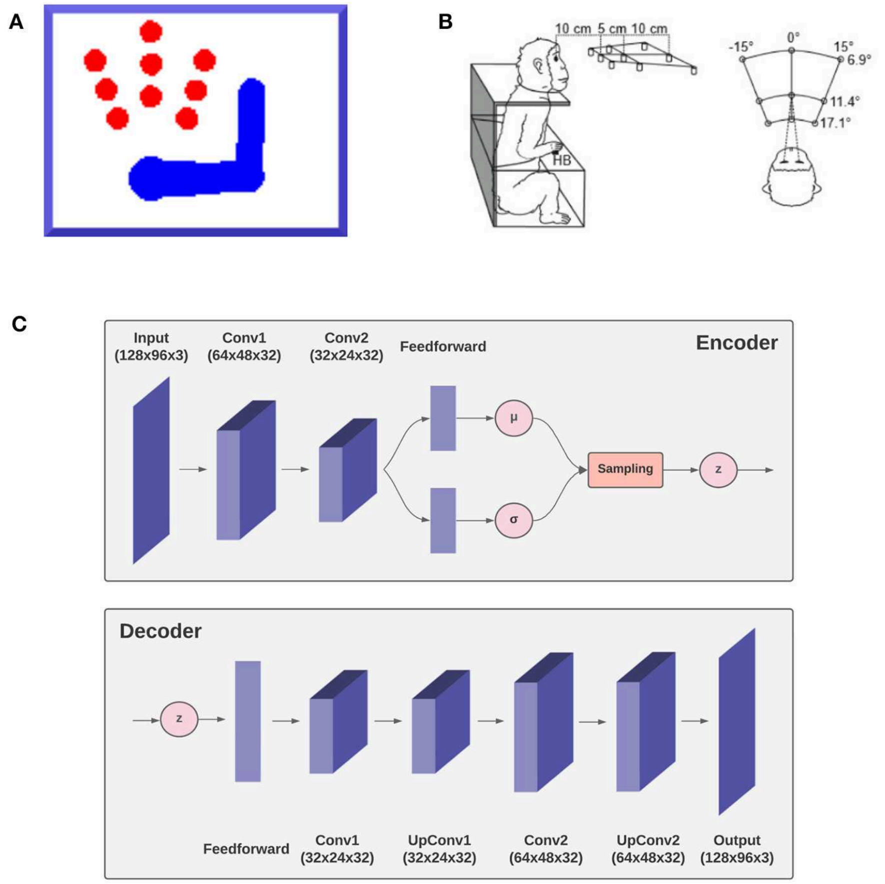


FIGURE 3
Simulation outline. The agent, a simulated 3-DoF actuated upper limb shown in **(A)** is set to reach one of the nine red circle targets as in the reference monkey experiment (Breveglieri et al., 2014) outlined in **(B)**. The agent is equipped with a fixed virtual camera providing peripersonal visual input and a visual model, the decoder g_v of a VAE shown in **(C)** simulating functions of the DVS.

to perceive and reach static and moving targets within its reach.[1] Figure 3A shows the size and position of the targets, as well as limb size and a sample posture. Since the focus here was on theoretical aspects, we simulated just a coarse 3-DoF limb model moving on a 2D plane. However, the approach easily generalizes to a more elaborated limb model and 3D movements. In the following, we describe the agent, the specific implementation, and the simulated task. Then, in the Results section we assess the agent's perceptual and motor control capabilities in static and dynamic conditions. The static condition simulated a typical monkey reaching task of peripersonal targets as in Figure 3 (Breveglieri et al., 2014). In turn, the dynamic condition involved a moving target that the agent had to track continuously.

4.1. Delayed reaching task

The primary testbed task is a simplified version of a delayed reaching monkey task in which a static target must be reached with a movement that can only start after a delay period (Breveglieri et al., 2014). Delayed actions are used to separately investigate neural processes related to action preparation (e.g., perception and planning) and execution in goal-directed behavior, and are thus useful to analyze the two main computational components of free energy minimization, namely, perceptual and active inference, which otherwise work in parallel. Delayed reaching could be implemented using various approaches: the update of the posture component of the belief dynamics could be blocked by setting the intention gain λ to zero during inference (implemented here): in this way, there are no active intentions and the belief only follows sensory information. Alternatively, action execution could be temporarily suspended by setting to zero the proprioceptive precision, so that the agent still produces proprioceptive predictions but does not trust their prediction errors: in this scenario, the belief dynamics includes a small component directed toward the intention, but the discrepancy produced is not minimized through movement.

Reach trials start with the hand placed on a home button (HB) located in front of the body center (i.e., the "neck"), and the belief is initialized with this configuration. Then one of the 9 possible targets of the reference experiment (Figure 3) is lit red. Follows a delay period of 100 time steps during which the agent is only allowed to perceive the visible target and the limb, and the inference process can only change the belief. After that, the limb is allowed

[1] Python code provided in https://github.com/priorelli/PACE.

to move and the joint angles are updated according to Equation (38). As in the reference task, upon target reaching the agent stops for a sufficiently long period, i.e., a total of 300 time steps per trial. After that, the agent reaches back the HB (not analyzed here). The simulation included 100 repetitions per target, i.e., 900 trials in total.

4.2. Body

The body consists of a simulated monkey upper limb composed of a moving torso attached to an anchored neck, an upper arm, and a lower arm, as shown in Figure 3. The three moving segments are schematized as rectangles, each with unit mass, while the joints (shoulder, elbow) and the tips (neck, hand) as circles. The proportions of the limb segment and the operating range of the joint angles were derived from monkey data *Macaca mulatta* (Kikuchi and Hamada, 2009). The state of the limb and its dynamics are described by the joint angles $\boldsymbol{\theta}$ and their first moment $\dot{\boldsymbol{\theta}}$. We assume noisy velocity-level motor control, whereby the motor efferents $\boldsymbol{a}$ noisily control the first moment of joint angles with zero-centered Gaussian noise:

$$\dot{\boldsymbol{\theta}} = \boldsymbol{a} + \boldsymbol{w}_a \tag{38}$$

4.3. Sensors

The agent receives information about its proprioceptive state and visual context. Simplified peripersonal visual input $\boldsymbol{s}_v$ was provided by a virtual camera that included three 2D color planes, each of them 128 x 96 pixels in size. The location and orientation of the camera were fixed so that the input provided full vision of peripersonal targets and the entire limb in any possible limb state within its operating range. The limb could occlude the target in some configurations.

As in the simulated limb, the motor control system also receives proprioceptive feedback through sensors $\boldsymbol{s}_p$, providing noisy information on the true state of the limb (Tuthill and Azim, 2018; Versteeg et al., 2021). We further assumed that $\boldsymbol{s}_p$ provides a noisy reading of the state of all joints only in terms of joint angles, ignoring other proprioceptive signals such as force and stretch (Srinivasan et al., 2021), which the Active Inference framework can natively incorporate.

4.4. Belief

We assume that both the orders of the generalized belief $\tilde{\boldsymbol{\mu}}$ comprise three components: (i) beliefs $\tilde{\boldsymbol{\mu}}_a$ over arm joint angles, or posture; (ii) beliefs $\tilde{\boldsymbol{\mu}}_t$ over the target location represented again in the joint angles space—i.e., the posture corresponding to the arm touching the target; and (iii) beliefs $\tilde{\boldsymbol{\mu}}_h$ over a memorized HB configuration. Thus, $\boldsymbol{\mu} = [\boldsymbol{\mu}_a, \boldsymbol{\mu}_t, \boldsymbol{\mu}_h]$. Note that the last two components can be interpreted as *affordances*, allowing the agent to implement interactions in terms of bodily configurations (Pezzulo and Cisek, 2016).

4.5. Sensory model

The sensory generative distribution has two components, one for each sensory modality: a simplified proprioceptive model $\boldsymbol{g}_p(\boldsymbol{\mu})$ and a full-blown visual model $\boldsymbol{g}_v(\boldsymbol{\mu})$:

$$\boldsymbol{g}(\boldsymbol{\mu}) = \begin{bmatrix} \boldsymbol{g}_p(\boldsymbol{\mu}) \\ \boldsymbol{g}_v(\boldsymbol{\mu}) \end{bmatrix} \tag{39}$$

Since the belief is already in the joint angles domain, we implemented a simple proprioceptive generative model $\boldsymbol{g}_p(\boldsymbol{\mu}) = \boldsymbol{G}_p\boldsymbol{\mu} = \boldsymbol{\mu}_a$, where $\boldsymbol{G}_p$ is a mapping that only extracts the first component of the belief:

$$\boldsymbol{G}_p = \begin{bmatrix} \mathbb{I} \,|\, \mathbf{0} \,|\, \mathbf{0} \end{bmatrix} \tag{40}$$

Where $\mathbf{0}$ and $\mathbb{I}$ are respectively 3 x 3 zero and identity matrices. Note that $\boldsymbol{g}_p(\boldsymbol{\mu})$ could be easily extended to a more complex proprioceptive mapping if the body and/or joint sensors have a more complex structure and the belief has a richer and abstract representation.

In turn, the visual generative model $\boldsymbol{g}_v$ is the decoder component of a VAE (see Figure 3C). It consists of one feedforward layer, two transposed convolutional layers, and two standard convolutional layers needed to smooth the output. Its latent space is composed of two elements, representing the joint angles of arm and target (example in Figure 13). The first component is used to generate, in the visual output, an arm with a specific joint configuration, while the second component is used to produce only the image of the target through direct kinematics of every joint angle. The VAE was trained in a supervised manner for 100 epochs on a dataset comprising 20.000 randomly drawn body-target configurations that uniformly spanned the entire operational space, and the corresponding visual images. The target size varied with a radius ranging from 5 to 12 pixels.

The proprioceptive gradient $\partial\boldsymbol{g}_p$ simplifies to the mapping $\boldsymbol{G}_p$ itself, while the visual component $\partial\boldsymbol{g}_v$ is the gradient of the decoder computed by backpropagation. Since the Cartesian position of the target is encoded in joint angles, this gradient implicitly performs a kinematic inversion. Therefore, predictions $\boldsymbol{P}$ and prediction errors $\mathcal{E}_s$ take the form:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{g}_v(\boldsymbol{\mu}) \end{bmatrix} \qquad \mathcal{E}_s = \begin{bmatrix} \boldsymbol{s}_p - \boldsymbol{\mu}_a \\ \boldsymbol{s}_v - \boldsymbol{g}_v(\boldsymbol{\mu}) \end{bmatrix} \tag{41}$$

Note that defining sensory predictions on both proprioceptive and visual sensory domains allows the agent to perform efficient goal-directed behavior also in conditions of visual uncertainty, e.g., due to low visibility. Indeed, since the belief is maintained over time, the agent remembers the last known target position and can thus accomplish reaching tasks also in case of temporarily occluded targets.

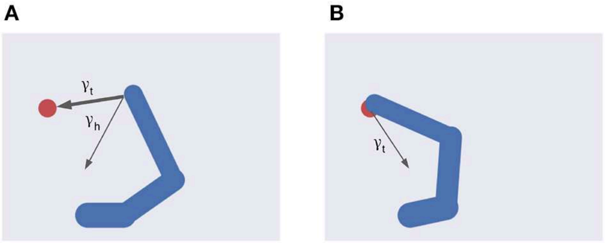


FIGURE 4
Multiple intentions. Intention prediction errors of variable strength (arrow width) controlled by intention precisions point at different target states. **(A)** A stronger target attraction (the red circle) implements the reaching action. **(B)** A stronger attraction of the invisible but previously memorized HB implements the return-to-home action.

4.6. Intentions

Stepping on the proposed formalization (Equation 20), we define two specific intentions (Figure 4) as follows:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{i}_t(\boldsymbol{\mu}) & \boldsymbol{i}_h(\boldsymbol{\mu}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_t\boldsymbol{\mu} & \boldsymbol{I}_h\boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_t & \boldsymbol{\mu}_h \\ \boldsymbol{\mu}_t & \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_h & \boldsymbol{\mu}_h \end{bmatrix} \tag{42}$$

Here $\boldsymbol{h}_t = \boldsymbol{i}_t(\boldsymbol{\mu})$ defines the agent's expectation that the arm belief is equal to the joint configuration corresponding to the target to be reached, and it is implemented as a simple mapping $\boldsymbol{I}_t$ that sets the first belief component equal to the second one. In turn, the intention $\boldsymbol{h}_h = \boldsymbol{i}_h(\boldsymbol{\mu})$ encodes the future belief of the agent that the arm will be at the HB position. The two intention mappings are defined by:

$$\boldsymbol{I}_t = \left[\begin{array}{c|c|c} \mathbf{0} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{I} \end{array}\right] \qquad \boldsymbol{I}_h = \left[\begin{array}{c|c|c} \mathbf{0} & \mathbf{0} & \mathbb{I} \\ \mathbf{0} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{I} \end{array}\right] \tag{43}$$

The corresponding intention prediction errors are then:

$$\boldsymbol{E}_i = \begin{bmatrix} \boldsymbol{e}_{i_t} & \boldsymbol{e}_{i_h} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_t - \boldsymbol{\mu} & \boldsymbol{h}_h - \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_t - \boldsymbol{\mu}_a & \boldsymbol{\mu}_h - \boldsymbol{\mu}_a \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{44}$$

These errors provide an update direction respectively toward the target and HB joint angles. As there is no intention to move the target or the HB, the second and third components of the prediction errors will be zero.

4.7. Precisions

Free energy minimization and Predictive Coding in general heavily depend on precisions modulation. To investigate their role, we parameterized the relative precisions of each intention and sensory domain with parameters α and β as follows:

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_p \\ \boldsymbol{\pi}_v \end{bmatrix} = \begin{bmatrix} 1-\alpha \\ \alpha \end{bmatrix} \qquad \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\gamma}_t & \boldsymbol{\gamma}_h \end{bmatrix} = \begin{bmatrix} 1-\beta & \beta \end{bmatrix} \tag{45}$$

The parameter α controls the relative strength of the error update due to proprioception and vision, while the parameter β controls the relative attraction by each intention. With these parameters, the sensory and intention weighted contributions are unpacked as follows:

$$\boldsymbol{\epsilon}_s = (1-\alpha)\cdot \boldsymbol{G}_p^T \boldsymbol{\varepsilon}_{s_p} + \alpha \cdot \partial \boldsymbol{g}_v^T \boldsymbol{\varepsilon}_{s_v} \tag{46}$$

$$\boldsymbol{\epsilon}_i = -\boldsymbol{\mu}' + \lambda[(1-\beta)\cdot \boldsymbol{e}_{i_t} + \beta \cdot \boldsymbol{e}_{i_h}] \tag{47}$$

Equation (46) shows the balance between visual and proprioceptive information. For example, if $\alpha = 0$ the agent will only use proprioceptive feedback, while for $\alpha = 1$ the belief will be updated only relying on visual feedback. Note that these are extreme conditions—e.g., the former may correspond to null visibility—and typical sensory systems provide balanced feedback. In turn, Equation (47) spells out the control of belief attraction. The agent will follow the first intention when $\beta = 0$, or the second one when $\beta = 1$ (Figure 4). Note that the introduction of a possible competing reach movement creates a conflict among intentions aiming to fulfill opposing goals (e.g., for intermediate values of β) while the agent can physically realize only one of them at a time (Figure 4). Thus, we assume that the control of intention selection is realized through mutual inhibition and higher-level bias. Finally, the parameter λ controls the overall attractor magnitude (see also Equation 26).

We can also use the precision parameter α to manipulate the strength of the free energy derivative with respect to the actions as follows:

$$\dot{\boldsymbol{a}} = -\Delta_t(1-\alpha)\cdot \boldsymbol{\varepsilon}_{s_p} \tag{48}$$

Note that by increasing α—i.e., more reliability on vision—the magnitude of the belief update remains constant, while action updates decrease because the agent becomes less confident about its proprioceptive information. Also, one could differentially investigate the effect of precision strength on belief and action by directly manipulating the precisions—e.g., visual precision $\boldsymbol{\pi}_v$ may include different components that follow the belief structure:

$$\boldsymbol{\pi}_v = [\alpha, \pi_{v_t}, \pi_{v_h}] \tag{49}$$

Where we used the parameter α only for the arm belief. For example, when $\alpha = 0$ and $\pi_{v_t} > 0$, the target belief is updated using visual input while the arm moves only using proprioception, a scenario that emulates movement in darkness with a lit target.

5. Results

In the following, we assess the capacities of the intention-driven Active Inference agent to perceive and perform goal-directed actions in reaching tasks with static and dynamic visual targets. The main testbed task was delayed reaching, but we simulated several other conditions.

Sensorimotor control that implements goal-directed behavior was investigated in various sensory feedback conditions, including pure proprioceptive or mixed visual and proprioceptive, in which the VAE decoder provided support for dynamic estimation of visual targets and bodily states. The latter is the typical condition of performing reaching actions and allows greater accuracy (Keele and Posner, 1968). In an additional *baseline* (BL) condition, the target was estimated by the decoder, but the movement was performed without visual feedback or proprioceptive noise, to allow comparisons with the typical approach in previous continuous Active Inference studies, e.g., Pio-Lopez et al. (2016). We also investigated the effects of sensory and intention precisions, motor control type, and movement onset policy. Finally, we analyzed the visual model and the nature of its gradients to provide critical information about the causes of the observed motor behavior.

Action performance was assessed with the help of several measures: (i) *reach accuracy*: success in approaching the target within 10 pixels of its center, i.e., the hand touching the target; (ii) *reach error*: L^2 hand-target distance at the end of the trial; (iii) *reach stability*: standard deviation of L^2 hand-target distance during the period from target reach to the end of the trial, in successful trials; (iv) *reach time*: number of time steps needed to reach the target in successful trials. We also assessed target perception through analog measures based on the L^2 distance between the target location and its estimate transformed from joint angles into visual position by applying the geometric (forward) model. Specifically, we defined the following measures: (v) *perception accuracy*: success in estimating the target location within 10 pixels; (vi) *perception error*: L^2 distance between the true and estimated target position at the end of the trial; (vii) *perception stability*: standard deviation of the L^2 distance between the target position and its estimation during the period starting from successful estimation until end of the trial; (viii) *perception time*: number of time steps needed to successfully estimate the target position.

Figure 5 illustrates key points of the delayed reaching task. During the delay period (Figures 5A–C), the posture does not change since the joint angles only follow the arm belief, which is kept fixed, while the target belief is attracted by the sensory evidence and gradually shifts toward it. When movement is allowed (Figures 5D–F) by setting $\lambda > 0$ and $\beta = 0.1$, the combined attractor produces a force that moves the arm belief toward the target, generating proprioceptive predictions—therefore motor commands—that let the real arm follow this trajectory. Reaching performance is summarized in Figure 6. Panels A-D show spatial statistics of the final hand location (with the corresponding belief) for each target, separately for reaching with proprioception only or proprioceptive and visual sensory feedback. Descriptive statistics revealed an important benefit of visual feedback (Figures 6E–H), in parallel with classical behavioral observations (Keele and Posner, 1968): reach accuracy was higher (with: 88.28%; without: 83.72%) and both reach stability and arm belief error were considerably better with visual feedback as well (stability: 1.35; error: 1.98px) compared to the condition with only proprioception (stability: 1.78; error: 2.87px).

5.1. Precision balance

The effects of sensory feedback led to a further systematic assessment of the effects of sensory and intention precisions α, π_{v_t} and λ (see Equations 45, 49). The assessment was carried out following the structure of the delayed reaching task. We varied the above precisions one at a time, using levels shown on the abscissas in Figure 7, while keeping the non-varied precisions at their default values. Note that $\alpha = 0$ corresponds to reaching without visual feedback, while the conditions $\alpha > 0$ may be interpreted as reaching with different levels of arm visibility. We recall that the baseline condition (BL) performs reaching movements without visual feedback and proprioceptive noise, i.e., $\alpha = 0$ and $w_p = 0$. To obtain a systematic evaluation, each condition was run on a rich set of 1,000 randomly selected targets that covered the entire operational space. Finally, we only considered the target-reaching intention, i.e., $\beta = 0$; everything else was the same as in the main task.

The results are shown in Figure 7. The panels in the left column show the effect of α compared to the BL agent with noiseless proprioception. Active Inference with only proprioception (i.e., $\alpha = 0$) has a lower reach accuracy and higher error, while the best performance is obtained with balanced proprioceptive and visual input, in corroboration with the observations of the basic delayed reaching task. In the latter case, the motor control circuitry continuously integrates all available sensory sources to implement visually-guided behavior (Saunders and Knill, 2003). However, accuracy and stability rapidly decrease for excessively high values of α, due to the discrepancy in update directions between the belief—which makes use of all available sensory information, including the more precise visual feedback—and action—which in this case relies on excessively noisy proprioception. Furthermore, as in the main experiment, the effects of visual precision are evident in the stability of the arm belief, which gradually improves with increasing values (Figure 8): In addition to the reliability of the visual input, this effect is also a consequence of the smaller action updates due to the reduced proprioceptive precision.

In turn, the panels in the middle column of Figure 7 reveal the effects of the attractor gain λ; to remind the reader, the greater the gain, the greater the contribution of intention prediction errors in the belief updates. The results show that as the intention gain λ increases reach accuracy generally improves, and the number of time steps needed to reach the target decreases. However, beyond a certain level, the accuracy tends to decrease since the trajectory dynamics becomes unstable; thus, excessively strong action drag is counterproductive to the implementation of smooth movements. Finally, the panels in the right column of Figure 7 show the effects of the target precision π_{v_t}, which directly affects the quality of target perception. Note that better performances are generally obtained in terms of accuracy, error, and perception time for values of π_{v_t} higher than the arm visual precision, which corresponds to a classical effect of contrast on perception, but also means here that the arm and target beliefs follow different dynamics.

5.2. Motor control

We described earlier two different ways of implementing motor control in Active Inference: making use of all sensory information, or proprioception only. The first method requires significantly more computations since the agent needs to know the inverse mapping from every sensory domain to compute the control signals. However, given the assumptions we made,

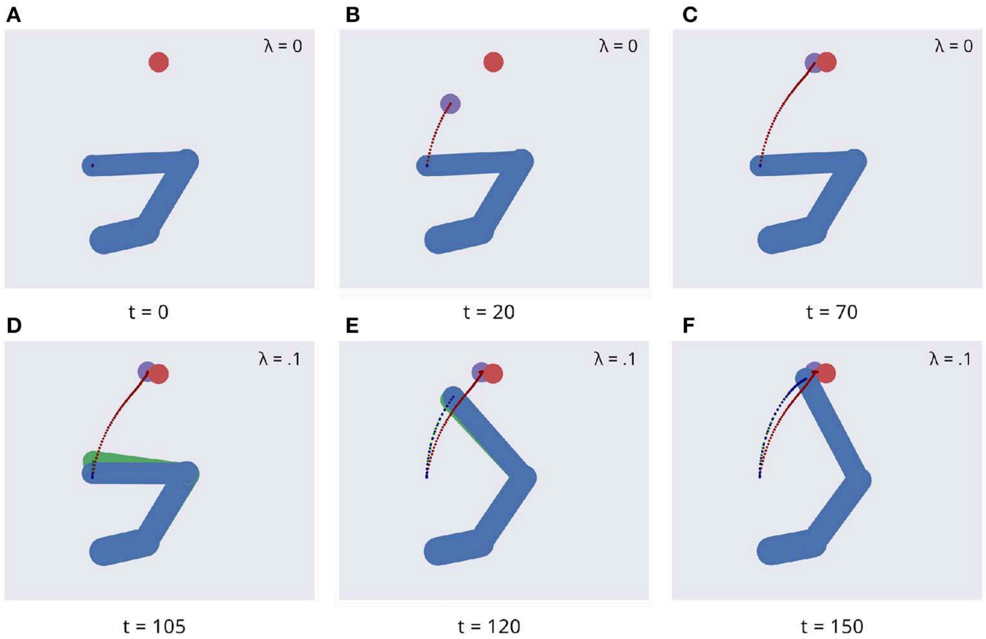


FIGURE 5
Dynamics of the delayed reaching task. At trial onset **(A)** a visual target (red circle) appears, the arm (in blue) is located on the HB position, and the arm belief (in green) is set at the true arm state. During the delay period, the perceptual inference process gradually drives the target belief (purple circle) toward the real position **(B, C)**. During this phase, the intention gain λ is set to 0, so that movement is inhibited and the arm belief does not change given the unchanged proprioceptive evidence. After movement onset, the arm freely follows its belief **(D, E)** until they both arrive at the goal state **(F)**.

this approach is potentially more stable because it updates both belief and action with the same information. On the other hand, a pure proprioception-based control mechanism could produce potentially incorrect movements because the motor control commands result from comparing proprioceptive predictions with noisy observations. Greater cost-effectiveness of the second method thus might come at the cost of worsened performances, which we investigate here.

Figure 9 shows a comparison of the two control methods and the BL agent, evaluated under the same conditions we used to investigate precision balance, including 1,000 random targets. Performance was measured in terms of belief and reach stability and reach accuracy. The results reveal, first, that the expected decreased belief stability of the full model with respect to the BL agent (Figure 9C) does not affect hand stability (Figure 9B), although the proprioceptive noise apparently contributed to decreased reach accuracy (Figure 9A). More importantly, the results confirm our expectations that pure proprioception control has considerably lower reach stability caused by incorrect update directions of the motor control signals, resulting in a greater decreased reach accuracy relative to the full model.

5.3. Movement onset policy

We also investigated the effects of movement onset using several policies, which differ by the duration of the period of pure perception preceding full Active Inference. One such policy we investigate here is characteristic of actions performed under time pressure, in which movement starts along with perception, i.e., action is *immediate*. Another policy that could be considered typical for acting under normal conditions has movement beginning with the satisfaction of a certain perception criterion. This policy *dynamically* deliberates the onset of movement. Various perception criteria could be used: here, the action starts when the norm of the target belief $\dot{\boldsymbol{\mu}}_t$ remained below a given threshold (i.e., 0.01) for a certain period (i.e., 5 time steps). These parameters were arbitrarily chosen in consideration of exploratory delayed reaching simulations. Finally, we include the previously used delayed action policy in which movement onset is delayed by a *fixed* period (here, 100 time steps, sufficient to obtain a precise target estimation). To obtain systematic observations, each policy was again run on 1,000 randomly selected targets. Measurements included reach and perception accuracy, motor control stability after reach, target perception stability, as well as reach time since the beginning of the trial or after movement onset.

Figure 10 shows the results with the three different policies. Although the reach error is approximately the same in all tested conditions, the agent controlled by the immediate policy reached the target within the lowest total number of time steps: target perception and intention setting were dynamically computed along with movement onset. However, if we consider the total task time, the number of time steps is the highest in this condition, since the arm belief and the arm itself move along with the slow visual target estimation. In turn, if the agent starts the movement when the uncertainty about the target position is already minimized (either in the *dynamic* or *fixed* condition), the movement time decreases, although if added on top of the perception time results in slower actions relative to the immediate movement condition. Finally, we note that target perceptual stability somewhat decreases for dynamic and fixed policies; this somewhat unexpected result is

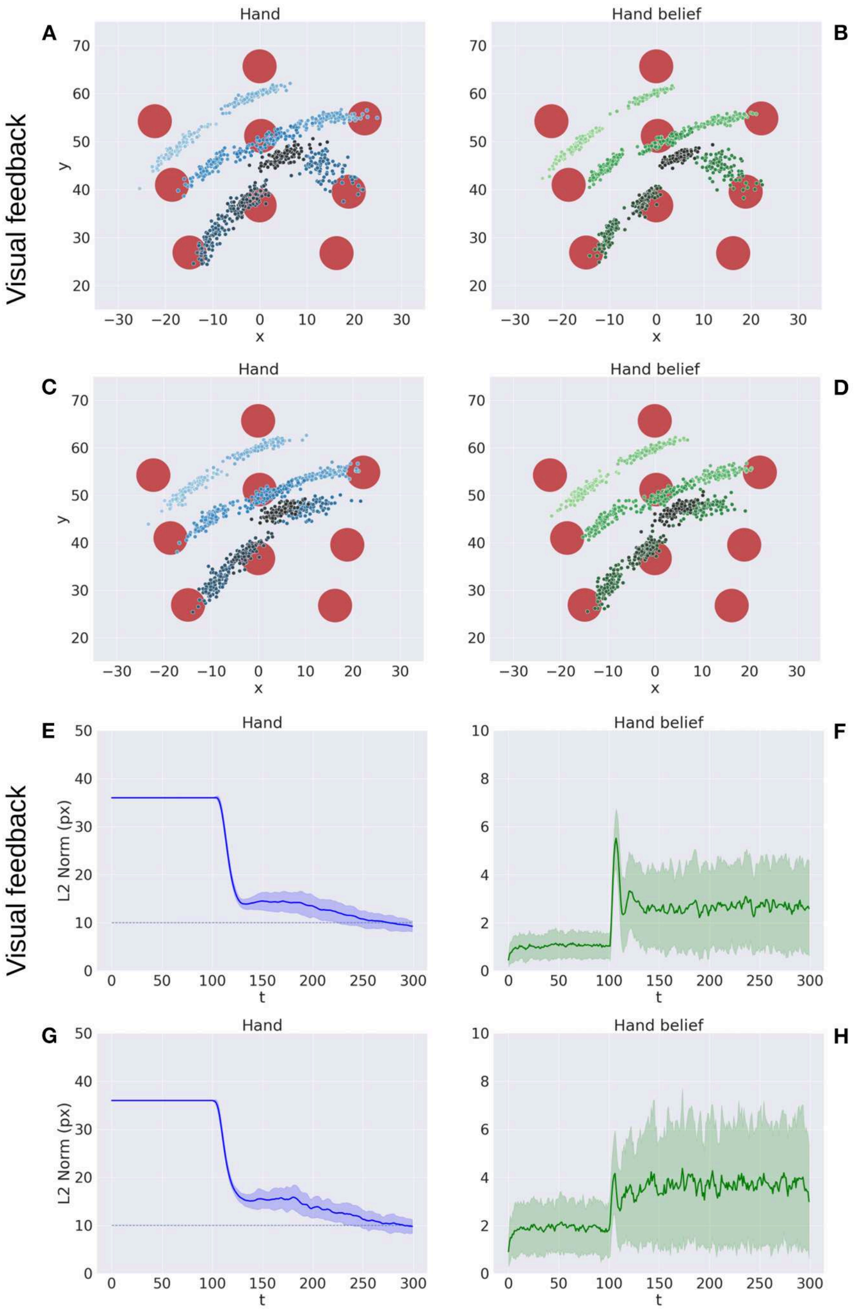


FIGURE 6
Performance of the delayed reaching task. **(A–D)** Spatial distribution of hand positions **(A, C)** and corresponding beliefs **(B, D)** per target at the end of the reach movements, with **(A, B)** and without **(C, D)** visual feedback. Each point represents a trial (100 trials per target). Reach error **(E, G)** and belief error **(F, H)** over time, with **(E, F)** and without **(G, H)** visual feedback (bands represent C.I.). The reach criterion of the hand-target distance is visualized as a dotted line. L^2 norm for the hand belief is computed by the difference between real and estimated hand positions. Reaching with visual feedback resulted in a more stable hand belief.

encouraging for dynamic target tracking tasks in which immediate movement onset is mandatory.

5.4. Tracking dynamic targets

In a second testbed task, the agent was required to track a smooth-moving target whose initial location was randomly chosen from the entire operational space. In each trial, the targets received an initial velocity of 0.1px per step in a direction uniformly spanning the 0–360° range. When the target reached a border, its movement was reflected. As in the previous simulations, the belief was initialized at the HB configuration and the trial time limit was 300 time steps. However, for the agent to correctly follow the targets, both the belief and action were dynamically

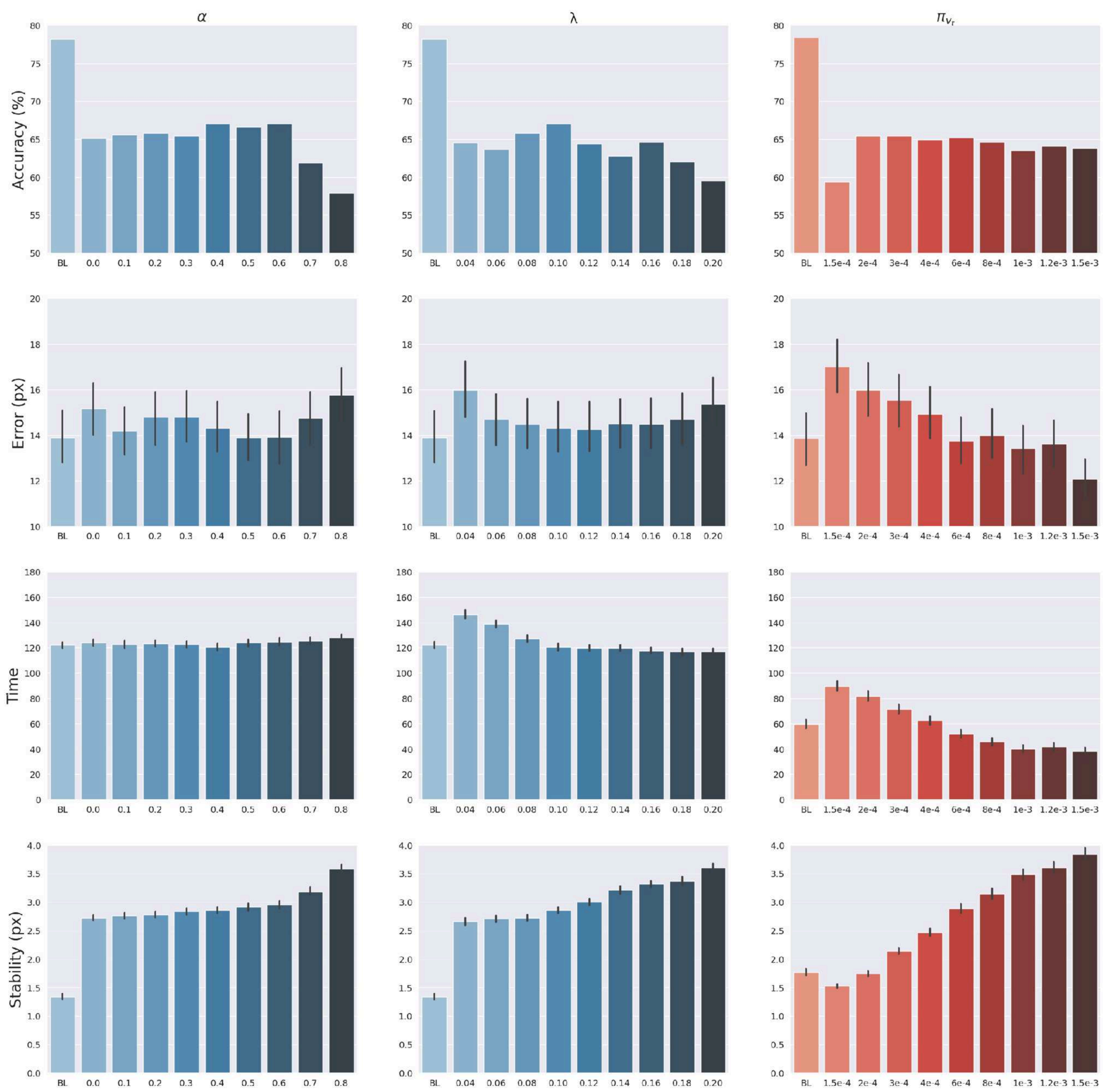


FIGURE 7
Effects of precision balance. Reach accuracy (1st row), reach error, i.e., L^2 hand-target distance (2nd row), reach time (3rd row), and reach stability (4th row). Reach performance is shown as a function of arm sensory precision α **(left)** and attractor strength λ **(middle)**. In turn, perception performance is shown as a function of target sensory precision π_{v_t} **(right)**. Vertical bars represent C.I.

and continuously inferred in parallel, i.e., without a pure perceptual period.

Figure 11 shows the reach trajectory in dynamic target tracking for 10 random trials. The left panel shows the evolution over time of L^2 hand-target distance, while the right panel represents the error between estimated and true target positions. The results suggest that the agent is generally able to correctly and dynamically estimate the beliefs over both target and arm for almost every trial, also in the case of moving targets. In some cases however, mainly when the target is out of reach, it is temporarily or permanently "lost" in terms of its belief, which has also the consequences of losing the target in terms of reach. Further analysis with a more realistic bodily configuration and visual sensory system—as well as comparisons with actual kinematic data—should provide further insights into the capabilities of Active Inference to perform dynamic reaching.

5.5. Free energy minimization

Here we illustrate the dynamics of free energy minimization in delayed reaching, which is at the heart of continuous Active Inference. To that aim, we run 10 new reaching trials with static

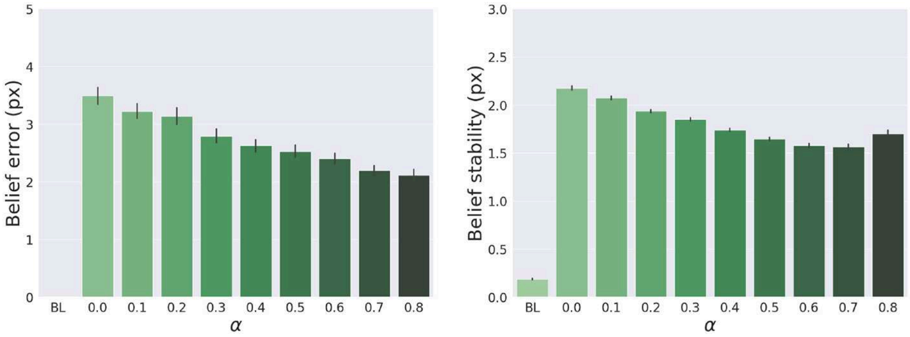


FIGURE 8
Belief error and stability—representing the difference between real and estimated hand positions—for different values of α. Vertical bars represent C.I.

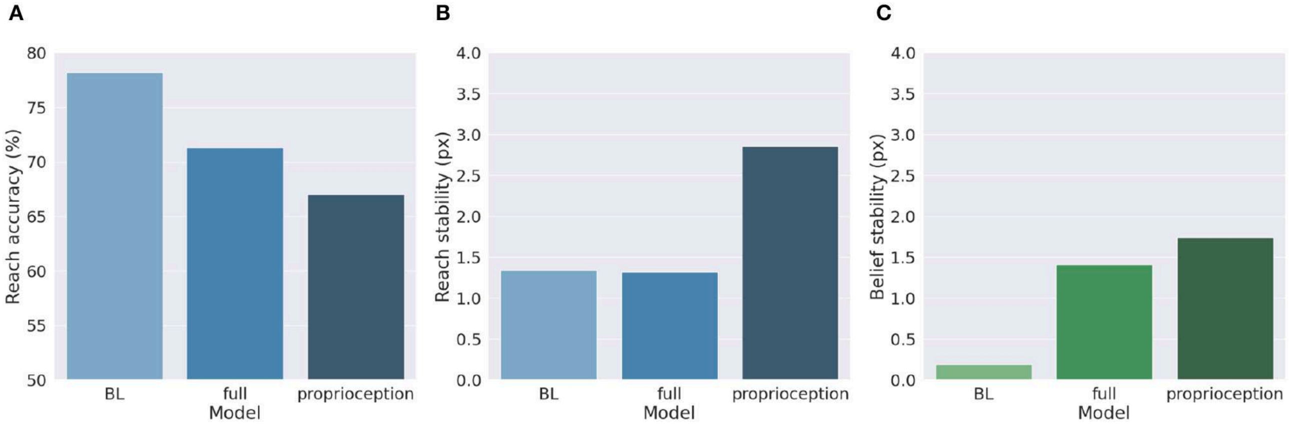


FIGURE 9
Motor control methods. Reach accuracy **(A)**, reach stability **(B)**, and belief stability **(C)** for a BL agent and the two different implementations of motor control, based either on all sensory information (full control) or on proprioception only.

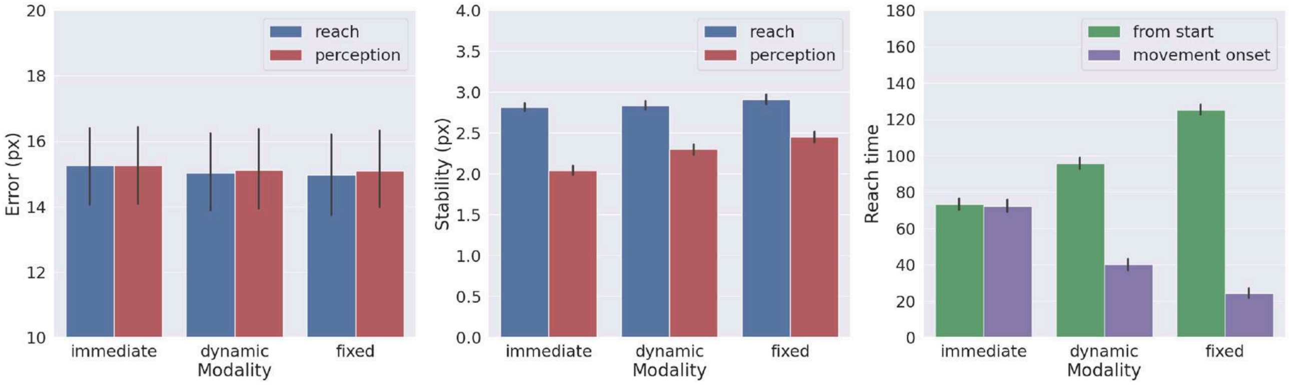


FIGURE 10
Effects of movement onset policy. Reach error **(left)**, stability **(middle)**, and time **(right)** across several policies (immediate, dynamic, and fixed delay). Vertical bars represent C.I.

and dynamic targets and recorded the free energy derivatives with respect to generalized belief and action.

Figures 12A–F shows the trajectory of the free energy derivatives with respect to the arm and target components during delayed reaching of a static target; the two columns show the trends for the last two joints, i.e., the arm and forearm segments, that most strongly articulate the reaching action. Note that the gradients of the free energy with respect to the target belief are

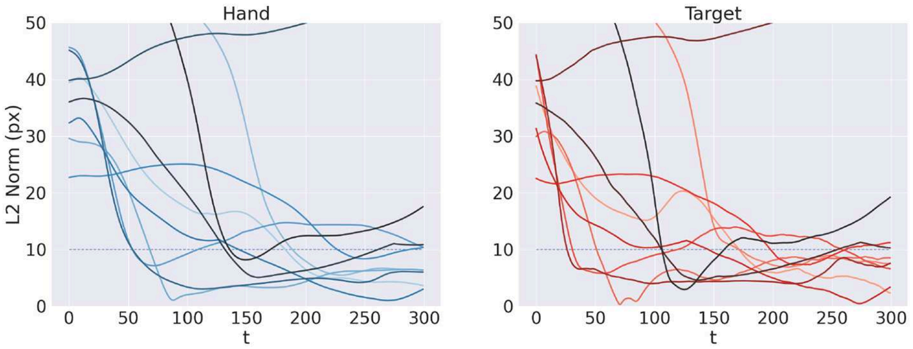


FIGURE 11
Tracking dynamic targets. Reach **(left)** and perception **(right)** error over time for 10 random trials.

rapidly minimized during the initial perceptual phase while the arm gradients remain still. Upon action execution (indicated by a vertical line), the arm gradients rapidly change as well, resulting in updated proprioceptive predictions that drive arm movements. However, arm movements cause changes in the visuals scene, resulting in a secondary effect over the just minimized free energy on target belief. Figures 12G–J goes even deeper, showing a direct comparison between $\dot{\boldsymbol{\mu}}$, $\boldsymbol{\mu}'$, and the difference $\dot{\boldsymbol{\mu}} - \boldsymbol{\mu}'$, on sample static (G-H) and dynamic (I-J) targets. We recall that free energy minimization implies that the two reference frames (the path of the mode $\dot{\boldsymbol{\mu}}$ and the mode of the path $\boldsymbol{\mu}'$) should overlap at some point in time, when the agent has inferred the correct trajectory of the generalized hidden states. This is crucial especially in dynamic reaching, in which the aim is to capture the instantaneous trajectory of every object in the scene. The decreasing free energy gradients (blue lines) show that this aim is indeed successfully achieved in both static and dynamic tasks.

5.6. Visual model analysis

Here, we provide an assessment of the visual model whose performance is critical for accurate visually-guided motor control. To recall, the visual model is implemented with a VAE trained offline to reconstruct images of arm-target configurations such as the one in Figure 13A. A critical VAE parameter is the variance of the recognition (encoder) density Σ_ϕ (see Equation 17). We therefore evaluated its effect on perception and action by training several VAEs with different variance levels. VAE performance was assessed on other 10.000 randomly selected configurations that uniformly sampled the space, with a target size of 5 pixels (the default condition for the Active Inference tests).

Most critical was the VAE capacity to generate adequate images of joint arm-target configurations, which we measured with the help of the L^2 norm between visual observations, and VAE-generated images. To provide more insights on the two VAE processes, decoding and encoding, we proceeded as follows: first, *decoding* was assessed by generating images for given body-target states such as that in Figure 13B. The decoded images were compared with the ground truth images produced by applying the geometric model for the same state of the body target (Figure 13A). Second, full VAE performance was assessed by computing the average L^2 norm between observed images and their full VAE *reconstruction*, i.e., first encoding and then decoding them (as in Figure 13C). Third, we directly assessed the specific effects of the recognition density variance on Active Inference using the BL condition of the delayed reaching task as a measure.

Figure 13D represents the results of the perceptual assessment tests, showing the L^2 norm between the original and generated images as a function of recognition density variance. As expected, lower variances generally resulted in lower errors with respect to both pure decoding and full encoding-decoding. Surprisingly, however, the accuracy of Active Inference in the reaching task behaved somewhat differently: the best accuracy was obtained not for predictions with low variance, but for intermediate variance levels (Figure 13E). This could be explained by the fact that low-variance images imply highly non-linear gradients that prevent correct gradient descent on free energy. On the other hand, as the variance increases the reconstructed image becomes somewhat blurred, which helps obtain a smoother gradient that correctly drives free energy minimization and therefore improves movement accuracy (more on this in the next section). However, as the variance continues to increase, the reconstructed images become too blurry, degrading both belief inference and motor control.

5.7. Visual gradient analysis

To further investigate the cause of the unexpected low variance issue, we analyzed the consistency of the visual gradient $\partial \boldsymbol{g}_v$ of the decoder for several encoder variance values. To this aim, we computed the gradients for different reference states over the entire operational space.

Figures 14A–C reveals that a decoder with intermediate variance values (green line) causes smaller but smoother gradients, while a too-low variance (orange line) causes sharp peaks near the reference point and even incorrect gradient directions in some cases. Therefore, too low encoder variances seem to make the decoder prone to overfitting, while higher variance values help extract a smoother relationship between irregular multidimensional sensory domains and regular low-dimension causes. Figures 14D–G further illustrates the arm and target

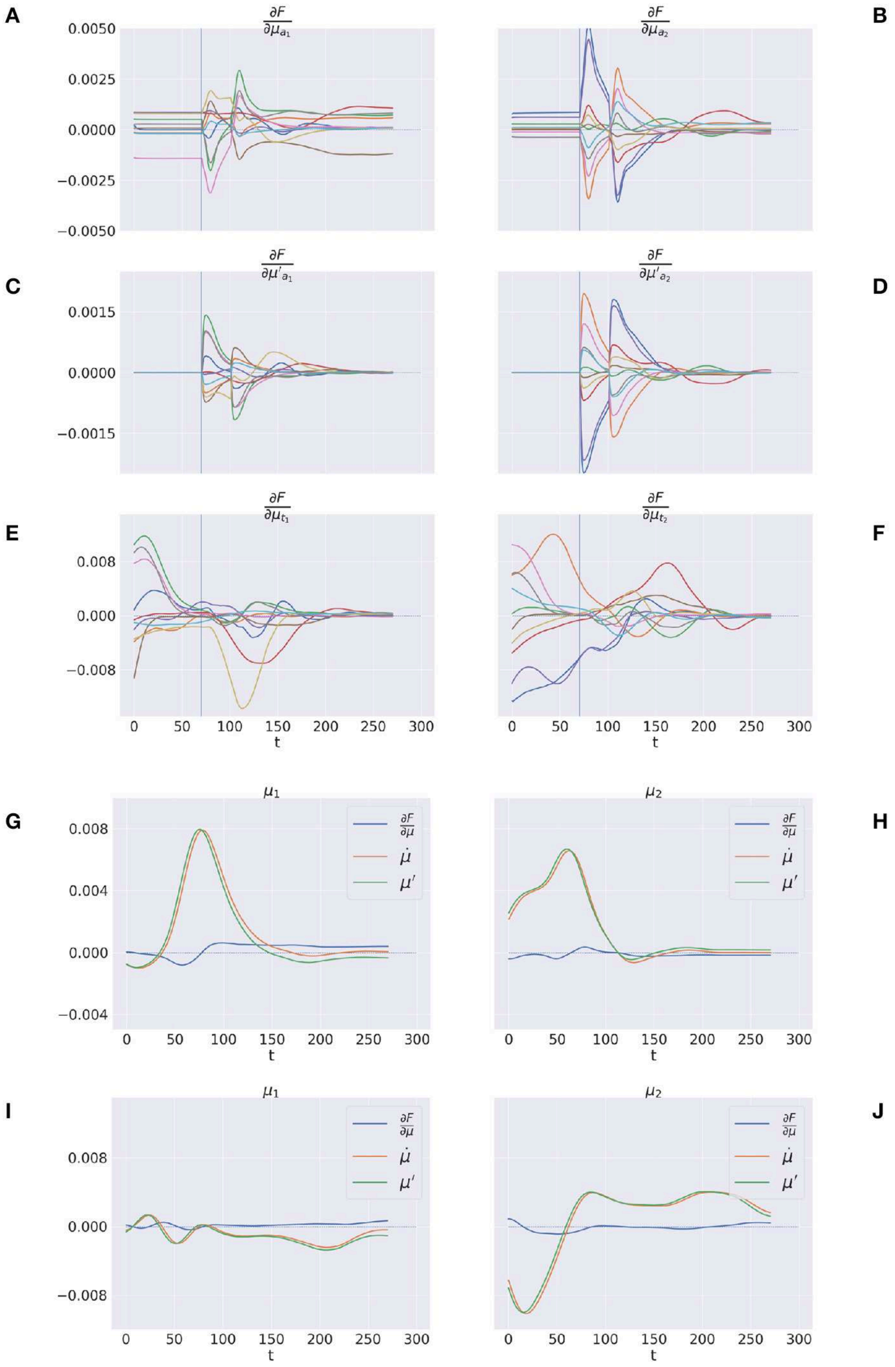


FIGURE 12
Free energy minimization. **(A–F)** Free energy derivative with respect to the 0th and 1st order belief for arm **(A–D)** and target **(E, F)**. **(G–J)** Comparison between the reference frames of the belief—the path of the mode $\dot{\mu}$ and the mode of the path μ'—for sample static **(G, H)** and dynamic **(I, J)** trials. The left/right columns refer to the arm/forearm segments. Trials data are smoothed with a 30 time-step moving average.

gradients relative to a sample reference posture and target location (the result is similar for other configurations) in both Cartesian and polar coordinates; the polar plot shows the two joints most relevant to the reaching action.

The plots reveal greater arm gradients (upper panels) in the vicinity of the target location; in that subspace, the decoder has less uncertainty about which direction to choose to minimize the error. Notably, the gradients tend to compose curved directions, a characteristic of biological motions. The polar plot provides critical insights into the causes of the circular pattern: the gradients are mostly parallel to the horizontal axis, which corresponds to a movement consisting essentially of pure shoulder rotation. Thus, they provide a strong driving force on the shoulder almost throughout the operational space, while the area in which the elbow is controlled is limited to the vicinity of the target location. These gradients result in a two-phase reaching of static targets in

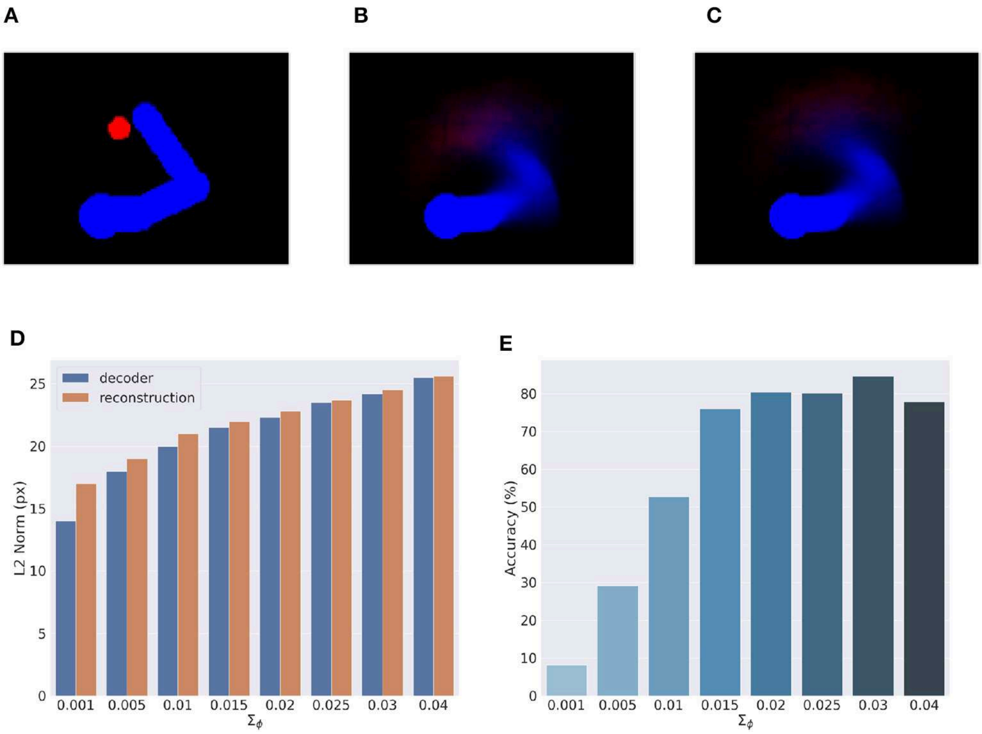


FIGURE 13
Visual model analysis. **(A–C)** Sample visual observation **(A)** and its decoding from joint angles **(B)** and through a complete encoding-decoding process **(C)**. **(D, E)** Visual model performance. Quality of perception is measured as the L^2 norm between observed and reconstructed images **(D)** and accuracy of Active Inference **(E)**, as a function of the recognition density variance Σ_ϕ.

which the agent first rotates the shoulder— resulting in a horizontal positioning—and then starts to rotate the elbow as soon as the latter enters its attraction area. The same gradients can explain the linear motion pattern of an arm tracking dynamically moving targets when the arm is close to the target: in that case, all gradients provide motion force as explained above. On the other hand, the gradients of the target belief (bottom panels) behave somewhat differently: since this belief is unconstrained and can freely move in the environment, update directions more directly approach the target in all angular coordinates (see the polar plot to the right). Yet, linear belief updates in the polar space still translate to curve directions in the Cartesian space.

6. Discussion

We presented a normative computational theory based on Active Inference of how the neural circuitry in the PPC and DVS may support visually-guided actions in a dynamically changing environment. Our focus was on the computational basis of encoding dynamic action goals in the PPC through flexible motor intentions and its putative neural basis in the PPC. The theory is based on Predictive Coding (Doya, 2007; Hohwy, 2013), Active Inference (Friston, 2010), and evidence suggesting that the PPC performs visuomotor transformations (Cisek and Kalaska, 2010; Fattori et al., 2017; Galletti and Fattori, 2018) and encodes motor plans (Andersen, 1995; Snyder et al., 1997). Accordingly, the PPC is proposed to maintain dynamic expectations of both current and desired latent states over the environment and use them to generate proprioceptive predictions that ultimately generate movements through reflex arcs (Adams et al., 2013; Versteeg et al., 2021). In turn, the DVS encodes a generative model that translates latent state expectations into visual predictions. Discrepancies between sensory-level predictions and actual sensations produce prediction errors sent back through the cortical hierarchy to improve the internal representation. The theory unifies research on intention coding (Snyder et al., 1997) and current views that the PPC estimates the body and environmental states (Medendorp and Heed, 2019), providing specific computational hypotheses regarding the involvement in goal-directed behavior. It also extends some perception-bound Predictive Coding interpretations of the PPC dynamics (FitzGerald et al., 2015) and provides a more comprehensive account of movement planning (Erlhagen and Schöner, 2002), tightly integrated into the overall sensorimotor control process.

The core novelty with respect to state-of-the-art implementations in continuous time Active Inference is that we first considered an internal belief over not only bodily states but also every object in the scene, where the latter are encoded in the joint angles space as well, simulating a visuomotor reference frame that the PPC is supposed to encode. Then, we decomposed the belief dynamics into a set of independent intentions each depending on the current belief and predicting the next plausible state. Such formalization has several advantages. First, since

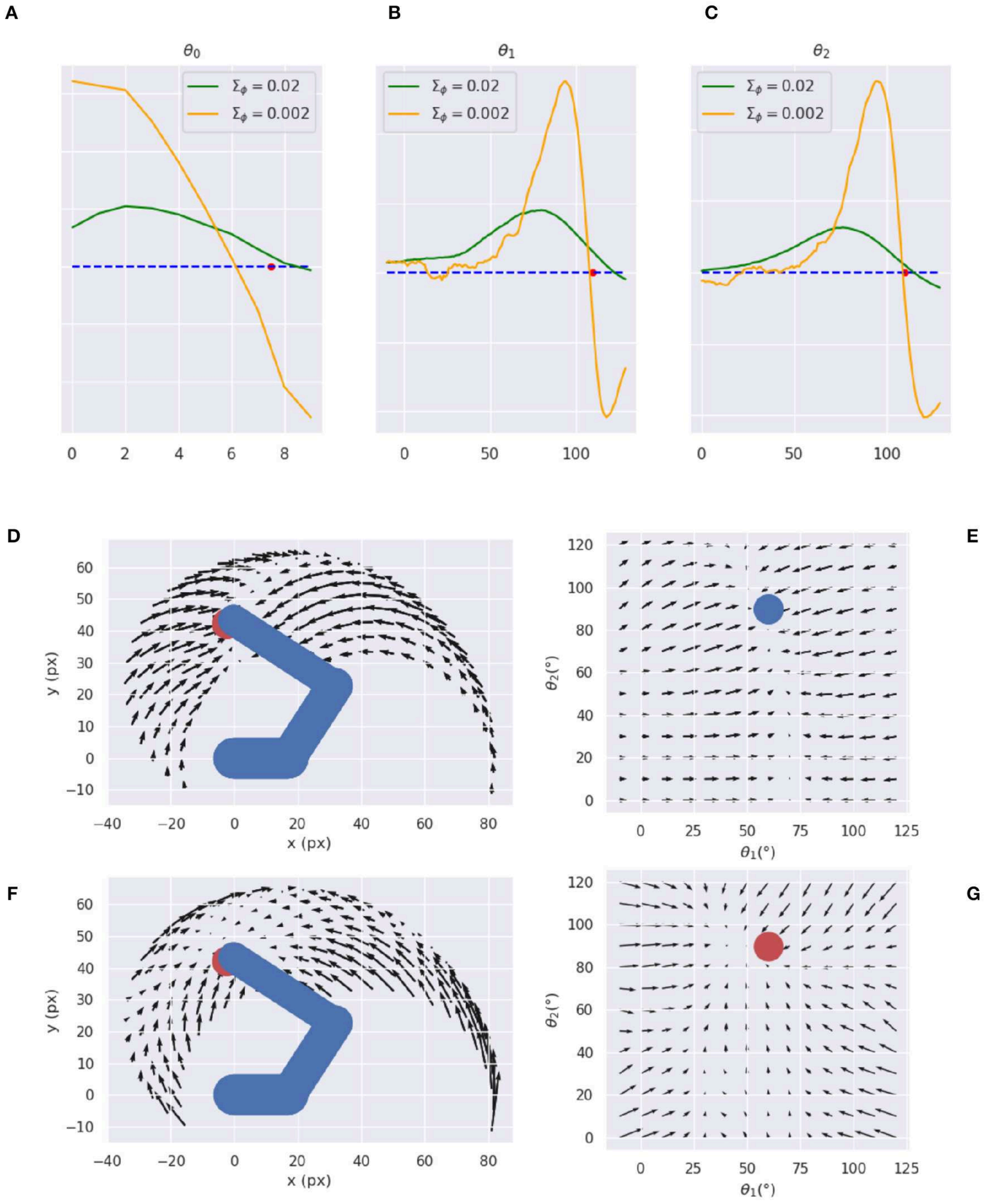


FIGURE 14
Visual gradient analysis. Marginal gradients for each joint, i.e., neck **(A)**, shoulder **(B)**, and elbow **(C)**, and for two values of the recognition density variance Σ_ϕ (green/orange line) computed by backpropagating the error between images with different arm configurations (abscissa: joint angle) and a reference image (whose angle is represented by the red dot on the abscissa). **(D–G)** Gradients for arm **(D, E)** and target **(F, G)** in both Cartesian **(D, F)** and joint **(E, G)** space.

attractors are dynamically generated at each time step, the agent can also follow moving targets and interact with a constantly changing environment, in contrast to static reaching tasks where a desired fixed state is specified in the belief dynamics (Baioumy et al., 2020). Second, expressing the target position in terms of a possible joint configuration—either imposed by higher levels for realizing specific affordances or freely inferred by the exteroceptive models—results in simple intentions, without the need to directly use sensory information or duplicating lower-level generative models, which leads to implausible scenarios (Lanillos et al., 2020; Sancaktar et al., 2020). It should be however noted that, although an intrinsic-only attractor is faster and more parsimonious, continuous activation of visual low-level attractors may provide more precise motor control. Indeed, it seems that motor areas are able, at a certain "neural energy" cost, to interact with and generate predictions in multiple sensory domains. The key difference is that in the former a single prediction—which is already biased toward a future state—is compared with the sensory input; in the latter, a prediction of the current state is compared with the desired exteroceptive goal, biasing the belief through the backpropagated gradient. Further studies are thus needed to implement low-level attractors in a biologically plausible way—e.g., intentions could generate through parallel pathways their own future sensory predictions that are compared with the observations in the usual way, with a particular intention that can be viewed as trying to continuously predict the current sensory input—and analyze the differences between the two modalities. Last, maintaining different belief components also allows easy encoding of previously memorized states which can be especially useful when implementing a sequence of actions, since only the intention precisions have to be adjusted. Indeed, it seems that the PPC explicitly encodes and maintains such goals during the whole unfolding of sequential actions (Baldauf et al., 2008). A specific goal is selected among other competitive intentions possibly under the control of the PFC and PMd (Stoianov et al., 2016) and fulfilled by setting it as a predominant belief trajectory as an attractor with a strong gain (see Equations 44, 26 and Figure 4). For example, in a typical reaching task, the goal of reaching a specific visual target corresponds to the future expectation that the agent's arm will be over that target; thus, if the agent maintains a belief over the latter, the corresponding intention links the expected belief over the future body posture with the inferred target, expressed in joint angles, encoding a specific interaction to realize.

We tested the computational feasibility of the theory on a delayed reaching task—a classical experiment in electrophysiology—in which a monkey is required to reach with its hand a visual-spatial target, starting the movement from an HB (Breveglieri et al., 2014). To do this, we simulated an agent consisting of a coarse 3-DoF limb model and noisy visual and proprioceptive sensors (Figure 3A). Simplified proprioceptive sensors provided a noisy reading of the state of the limb in joint angles, while visual input was provided by a fixed camera and consisted of an image of the target and limb. Predictive visual sensory processing simulating the DVS was implemented with a VAE trained to infer body state and target location, both in the joint angles domain (Figure 3C). The limbs were animated at the velocity level with motor control signals computed by the visually-guided Active Inference controller. The computational analysis showed, first and most importantly, that the controller could correctly infer the position of the visual targets (Figure 5, t = 70), use it to compute and set motor goals in terms of prior beliefs on the future body state through intention functions (Figure 5, t = 105), and perform adequate and smooth reach movements (Figure 5 t = 105–150), with and without visual feedback (Figure 6). The greater accuracy obtained with visual feedback parallels classical results in a similar classical behavioral comparison of reaching (Keele and Posner, 1968).

We then systematically investigated the effects of noise on various functional components (Figure 7), starting with the balance of the precision between proprioceptive and visual sensory models: a noiseless Active Inference agent (BL condition) resulted in the best performance, with a stable final approach and accuracy only limited by the quality of the visual target estimation. Among the noisy conditions, pure proprioceptive control resulted in the lowest performance, as expected. Motor control driven by both proprioceptive and visual feedback with balanced precision between the two domains resulted in improved reach accuracy and greatly improved arm belief stability (Figure 8). The effect on accuracy was mainly due to the inclusion of visual information in the inference process, but also to slower updates of the motor control signals due to decreased confidence about proprioceptive input. The increased stability of the arm belief did not improve movement stability as increasing confidence about visual input also increased the discrepancy between belief and action updates, the latter only relying on noisy proprioceptive observations. In fact, we showed that if we remove the plausibility constraint that motor control is driven only by proprioceptive predictions and thus let actions minimize prediction errors from all sensory domains, the reach performance greatly increases (Figure 9). Nonetheless, any combination of visual and proprioceptive feedback improved performance relative to a control driven by feedback from a single sensory domain. The instability due to the difference in update directions between belief and action could be balanced by other mechanisms that we have not considered here. For example, we assumed that the same pathway is used for both control and belief inference, but it seems that the motor cortex generates different predictions depending on the brain areas which it interacts with: purely proprioceptive predictions for motor control, whose prediction error is suppressed at the lowest level of the hierarchy, and rich somatosensory predictions for latent state inference, which integrates somatic sensations at different hierarchical levels (Adams et al., 2013). Intention precisions or attractor gains affected performance as well. First, they affected reach time: as expected, the greater the gain, the faster the movement. However, fast movements come at a cost: increased gains generally resulted in less precise movements and decreased stability during the final reach period. Finally, higher visual target precisions decreased perception time and improved perception accuracy but decreased perception stability.

We also investigated the effects of movement onset policies: response delay allows investigating perceptual and motor preparatory processes separately from the motor control and action execution. We found that delayed response decreased movement time with respect to a policy that requires an immediate response (Figure 10), which fits the behavioral pattern (Shenoy et al., 2013). Apparently, this is due to the need to estimate, in the latter condition, the target position "on the fly," and constantly adapt the intention according to the updated target estimate. The advantage of allowing some preparatory time becomes clear in an anecdotal fly-catching task, which results in faster movement and increased chances of success. This comes with the critical contribution of PPC neurons that systematically modulate their activity during the preparatory period (Shenoy et al., 2013), which here provided specific predictions for the computations performed in the PPC. Notably, the immediate-response policy allowed the Active Inference controller to perform actions under dynamic environmental conditions, such as tracking moving objects. Free energy minimization resulted in rapid target detection also in this case, and maintained subthreshold perception error on moving targets (Figure 11) which allowed precise tracking after an initial reaching period.

Intention-driven Active Inference in continuous time largely compares to classical neural-level hypotheses of motor planning such as the Dynamic Field Theory (Erlhagen and Schöner, 2002), with the advantage of stepping on an established Predictive Coding framework, dynamic approximate probabilistic inference, and end-to-end sensorimotor control. The Dynamic Field Theory estimates the parameters of the desired movement—such as movement direction and target velocity—from sensory and task features encoding environmental descriptors, which closely compares to motor goal coding through flexible intentions in our model. The two theories have in common a dynamic activation of the internal representations in continuous time, governed by a dynamic system, but they differ in the nature of the signals and their coding. Movement descriptors in Dynamic Field Theory are represented by a dynamically activated multidimensional space, each encoded on a population of competing neurons, while Active Inference approximates movement properties with their central moments (belief) and dispersion (precisions). While population coding allows a complete description of a probabilistic distribution, it could be overshot when used to code single magnitudes (although it is essential when encoding discrete categories). Yet, such representation allows coding multiple competing targets on the same population of neurons, while in our scheme each target should be encoded by a dedicated unit. Notably, the brain encodes scalar variables using a variety of number coding schemes, including monotonic and distributed (Stoianov and Zorzi, 2012). The latter, known as a "mental number line" (Stoianov et al., 2008), could be an interesting hypothesis to explore also in the context of feature coding in continuous Active Inference. Currently, distributed coding is used only in discrete Active Inference and other probabilistic models to investigate computationally high-level cognitive functions such as planning, navigation, and control (Stoianov et al., 2016, 2022; Pezzulo et al., 2018). The two theories also differ in the nature of the input to their dynamic systems. In Active Inference, system input encodes generalized prediction errors, which are integrated into higher-level moments. Instead, input in Dynamic Field Theory directly encodes state values. Coding based on prediction errors has the advantage of minimizing the quantity of transmitted information—hence, energy. Finally, the theories also differ in scope: Active Inference provides a full account of the entire sensorimotor control process, while Dynamic Field Theory describes only movement planning.

6.1. Precision balance and conditions for disorders

Based on our computational analysis, it becomes clear that some motor and behavioral disorders could be due to the lack of proper sensory and intention precisions (Adams et al., 2021). Here, we illustrate the normal condition and two types of potentially improper precision balance that could become a causal condition for neurological disorders. Figure 15A illustrates the condition for normal functioning, which is such that the contribution of a single intention to the belief update (which, as a reminder to the reader, is proportional to the gain of intentions λ) is sufficiently small with respect to the sensory contribution. In this case, during free energy minimization, the system dynamics smoothly moves the belief toward the strongest goal, along with precise tracking of the true latent state and sensory signal of the limbs, allowing thus to compute correct motor control errors and perform smooth action execution. A critical abnormal condition arises when the intention gain λ is too strong, as illustrated in Figure 15B. In this case, the belief moves too rapidly toward the goal without being able to match the proprioceptive observations, which results in computing incorrect motor control signals. Another abnormal condition is caused by too close precisions $\boldsymbol{\gamma}_k$ of competitive intentions, which is likely to result in opposing belief updates and thus prevent the fulfillment of any of the competing goals (as in Figure 15C). This situation might manifest in terms of motor onset failure or oscillatory behavior.

6.2. Neural-level predictions

One peculiarity of Active Inference based theories of motor control is that proprioceptive predictions are sent through efferents down to the spinal cord and that specific muscle control signals are computed at that level by reflex arcs, so that action attempts to suppress proprioceptive prediction errors (Adams et al., 2013). This prediction critically differs from competing modern theories such as the Optimal Control (Todorov and Jordan, 2002), according to which the efferents convey muscle control signals computed at the cortical level. A general aspect of Active Inference regards the dynamic inferential process, which predicts with increasing precision the internal representation of the sensorium—including estimation of targets and body posture—starting from noisy priors that gradually converge to ideal states. This kind of precision trend should be observed in an experiment with multiple repetitions of the same action and target, with variability of cell activity encoding the target and body that gradually decreases in time within trials. While this prediction is generally shared with Predictive Coding based theories, classical stimulus-response theories would predict invariant variability of cell activity across time. Another general aspect regards coding of prediction errors. In fact, body-environment transitions involving a change of states and tasks result in transient bursts of activity in error-conveying cells until the error is minimized. Prediction errors conveying upstream information are supposed to be encoded by pyramidal cortical cells in superficial layers while downstream predictions are encoded by deep pyramidal neurons (Parr et al., 2022).

In light of the considerations so far, we predict several different types of correlates that should be found in the PPC related to coding environment, task, and bodily states. The former two include correlates of potential spatial targets and selected motor goals, which indeed have been consistently found in the PPC (Andersen, 1995; Snyder et al., 1997; Filippini et al., 2018). The latter includes correlates of intention-biased bodily state estimates, which thus are not precise representations of the true states. To this concern,

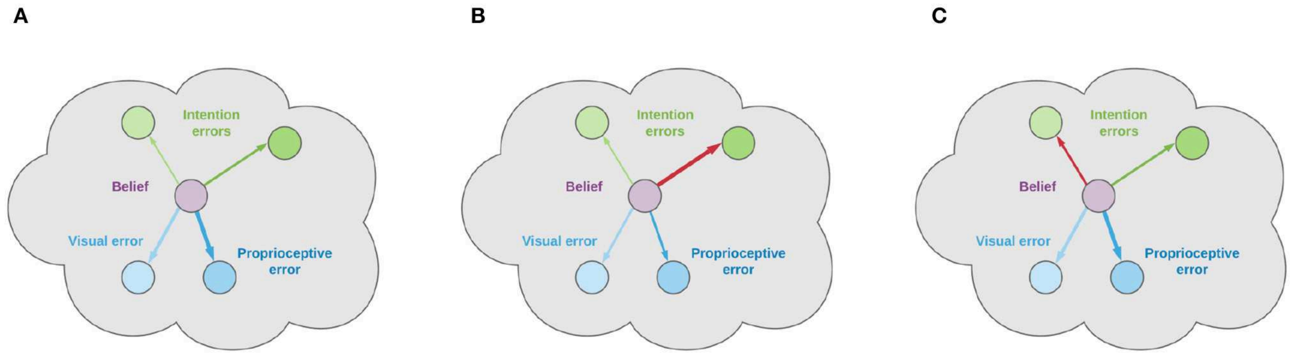


FIGURE 15
Normal and abnormal intention gains and precisions. Arrow widths represent strengths of prediction errors. In normal conditions, the sensory contribution to free energy minimization is bigger than that of the intentions and there is a clear unambiguous intention to fulfill **(A)**. An abnormally strong intention gain λ drives the belief away from the true joint states **(B)**. Too close precision levels γ also hinder implementing competing intentions **(C)**.

a key expected neural correlate of the proposed mechanism in the PPC includes signals encoding intention prediction errors between the current belief and future states corresponding to targets to interact with, both encoded in a visuomotor reference frame. To investigate this, one can manipulate high-level priors, e.g., by inducing an abrupt change of the intention, which should then be observed as a fast decaying change of the corresponding prediction error. A related hypothesis is that in tasks comprising several targets—like the classical monkey experiment analyzed here—each goal generates its own intention and prediction error. In normal conditions only one intention is selected at a time, and this behavior should be observed in the relative dynamics between all the intention prediction errors encoded simultaneously. Finally, the use of generalized beliefs in Active Inference predicts that the PPC encodes not only static states but also a detailed estimate of body dynamics, up to a few temporal orders. Indeed, a body of literature report motion-sensitive, or Vision-for-Action activity in the DVS and PPC (Galletti and Fattori, 2018). The validation of all these correlates will be the subject of further studies with real monkey experiments similar to the one described in Figure 3B.

6.3. Limitations and future directions

Our focus here was on intention coding in the PPC, which directly deals with motor plans and motor control. Further elaborations will extend the theory with higher-level aspects of cognitive control, including intention structuring (dealt by PM), phasing (SMA) (Gallego et al., 2022), planning, and goal selection (HC, PFC) (Stoianov et al., 2016, 2018; Pezzulo et al., 2019). Motor control operated here in an inner belief space belonging to the joint angles domain, which is generally suboptimal in the external Cartesian space. Although a kinematic transformation was implicitly performed by the VAE, we assumed that neural activity in the PPC encodes generalized beliefs over targets and body only in a motor-related domain; however, neural data suggest that neurons in the motor cortex encode motor trajectories also in extrinsic coordinates (Cohen and Andersen, 2002; Adams et al., 2013), and a more realistic model should include representations encoding states in both intrinsic and extrinsic reference frames. A functional correlate of the motor cortex should represent future states—which were defined here implicitly in the intention prediction errors and dynamics functions—and transform desired trajectories from Cartesian coordinates to proprioceptive predictions in the intrinsic state-space. This transformation is different from Optimal Control planning since the optimization of a classical inverse model reduces to a more manageable inference problem.

Since our focus was on the theoretical introduction of intentionality in Active Inference, every analysis was only partially characterized by a simple reaching task. However, fundamental properties of the physical model, including geometry, mass, and friction, strongly influence the resulting motion dynamics—hence the entire inferential process. This implementation does not adopt other important neural and biomechanical specificities such as signal delay and joint friction (Wolpert and Flanagan, 2016), and just partially covers the three main domains of sensorimotor learning through a predictive forward control; for example, it does not fully include reactive, stimuli-driven control such as obstacle avoidance, although we showed that it can successfully perform static and dynamic tasks. However, it could be easily extended to accommodate additional sensory modalities—e.g., tactile sensations—with rich generative models such as the VAE implemented here. Further planned computational analyses will use a richer belief space, a more realistic physical arm model, and additional actuators, and expand the complexity of the intention functions to investigate the capacity of the theory to explain in-depth neural levels, cognitive, and kinematic phenomena related to motor learning, motion perception, motor planning, and so on. Planned future studies with a more articulated agent will also challenge the theory at the behavioral and neural level against other empirical findings regarding movement preparation and motor control, in either delayed or direct response settings. For example, we will test the model for stimulus-stimulus congruency and stimulus-response compatibility effects (Kornblum et al., 1990). As for the former, it is intuitive that a greater sensory dimension overlap predicts faster target-belief convergence—thus faster intention setting. Less intuitive is that stimulus-response compatibility effects should emerge due to differences in the dynamic transition from one belief state to another in the proprioceptive domain. For example, a belief over the effector state should change more, requiring more time to converge when reaching a contralateral position than an ipsilateral one.

Although we considered an Active Inference model with just a single layer of intentions, the structure represented in Figure 2 could be scaled hierarchically and intermediate goals could be considered between high-level intentions and low-level sensory generative models, e.g., by combining discrete and continuous Active Inference for planning and movement execution (Friston et al., 2017a,b; Parr et al., 2020; Sajid et al., 2021). According to the free energy principle, the agent will then choose goals and subgoals and rely on specific sensory modalities such that free energy is minimized at every hierarchical level based on prediction errors coming from the level below. This formalization will provide an explicit basis for motor planning, including tasks like object manipulation. Indeed, although the current implementation performs well on spatial tasks like reaching in a dynamically changing environment, it cannot implement composite goals which the brain needs to handle. On the other hand, an agent that can encode higher-level goals in a discrete domain and infer policies based on the *expected free energy* will be able to dynamically modify its behavior and react to environmental changes. An extended implementation of this kind—showing the interplay between discrete goals and continuous intentions—will be the subject of future work.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/priorelli/PACE.

Author contributions

MP developed the computational method, wrote the code, run the simulations, analyzed the results, and wrote the draft. IS developed theoretical and methodological ideas and wrote the draft. All authors contributed to the article and approved the submitted version.

Funding

This research was received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under H2020-EIC-FETPROACT-2019 Grant Agreement 951910 (MAIA) to IS, Grant Agreement No 945539 (Human Brain Project SGA3), the European Research Council under Grant Agreement No. 820213 (ThinkAhead), and from the Italian Ministry for Research MIUR under Grant Agreement PRIN 2017KZNZLN (PACE) to IS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adams, R. A., Aponte, E., Marshall, L., and Friston, K. J. (2015). Active inference and oculomotor pursuit: the dynamic causal modelling of eye movements. *J. Neurosci. Methods* 242, 1–14. doi: 10.1016/j.jneumeth.2015.01.003

Adams, R. A., Shipp, S., and Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* 218, 611–643. doi: 10.1007/s00429-012-0475-5

Adams, R. A., Vincent, P., Benrimoh, D., Friston, K. J., and Parr, T. (2021). Everything is connected: Inference and attractors in delusions. *Schizophrenia Res.* 245, 5–22. doi: 10.1016/j.schres.2021.07.032

Andersen, R. A. (1995). Encoding of intention and spatial location in the posterior parietal cortex. *Cereb. Cortex* 5, 457–469. doi: 10.1093/cercor/5.5.457

Baioumy, M., Duckworth, P., Lacerda, B., and Hawes, N. (2020). Active inference for integrated state-estimation, control, and learning. *arXiv.* doi: 10.1109/ICRA48506.2021.9562009

Baldauf, D., Cui, H., and Andersen, R. A. (2008). The posterior parietal cortex encodes in parallel both goals for double-reach sequences. *J. Neurosci.* 28, 10081–10089. doi: 10.1523/JNEUROSCI.3423-08.2008

Baltieri, M., and Buckley, C. L. (2019). PID control as a process of active inference with linear generative models. *Entropy* 21, 257. doi: 10.3390/e21030257

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *J. Math. Psychol.* 76, 198–211. doi: 10.1016/j.jmp.2015.11.003

Breveglieri, R., Galletti, C., Dal Bò, G., Hadjidimitrakis, K., and Fattori, P. (2014). Multiple aspects of neural activity during reaching preparation in the medial posterior parietal area V6A. *J. Cogn. Neurosci.* 26, 879–895. doi: 10.1162/jocn_a_00510

Buckley, C. L., Kim, C. S., McGregor, S., and Seth, A. K. (2017). The free energy principle for action and perception: a mathematical review. *J. Math. Psychol.* 81, 55–79. doi: 10.1016/j.jmp.2017.09.004

Cisek, P., and Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* 33, 269–298. doi: 10.1146/annurev.neuro.051508.135409

Cohen, Y. E., and Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nat. Rev. Neurosci.* 3, 553–562. doi: 10.1038/nrn873

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., and Grafton, S. T. (1999). PPC and visually directing reaching to targets. *Nature Ne* 2, 563–567. doi: 10.1038/9219

Doya, K. (2007). Bayesian *Brain: Probabilistic Approaches to Neural Coding*. Cambridge, MA: The MIT Press.

Erlhagen, W., and Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychol. Rev*. 109, 545–572. doi: 10.1037/0033-295X.109.3.545

Fattori, P., Breveglieri, R., Bosco, A., Gamberini, M., and Galletti, C. (2017). Vision for prehension in the medial parietal cortex. *Cereb. Cortex* 27, 1149–1163. doi: 10.1093/cercor/bhv302

Filippini, M., Breveglieri, R., Ali Akhras, M., Bosco, A., Chinellato, E., and Fattori, P. (2017). Decoding information for grasping from the macaque dorsomedial visual stream. *J. Neurosci*. 37, 4311–4322. doi: 10.1523/JNEUROSCI.3077-16.2017

Filippini, M., Breveglieri, R., Hadjidimitrakis, K., Bosco, A., and Fattori, P. (2018). Prediction of reach goals in depth and direction from the parietal cortex. *Cell Rep*. 23, 725–732. doi: 10.1016/j.celrep.2018.03.090

FitzGerald, T. H., Moran, R. J., Friston, K. J., and Dolan, R. J. (2015). Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation. *Neuroimage* 107, 219–228. doi: 10.1016/j.neuroimage.2014.12.015

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolotti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667. doi: 10.1126/science.1106138

Franklin, D. W., and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron* 72, 425–442. doi: 10.1016/j.neuron.2011.10.006

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol*. 4, e1000211. doi: 10.1371/journal.pcbi.1000211

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci*. 11, 127–138. doi: 10.1038/nrn2787

Friston, K. (2011). What is optimal about motor control? *Neuron* 72, 488–498. doi: 10.1016/j.neuron.2011.10.018

Friston, K. (2012). The history of the future of the Bayesian brain. *Neuroimage* 62, 1230–1233. doi: 10.1016/j.neuroimage.2011.10.004

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. B Biol. Sci*. 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234. doi: 10.1016/j.neuroimage.2006.08.035

Friston, K. J. (2002). Functional integration and inference in the brain. *Progr. Neurobiol*. 68, 113–143. doi: 10.1016/S0301-0082(02)00076-X

Friston, K. J. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond B Biol. Sci*. 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J., Daunizeau, J., and Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE* 4, e6421. doi: 10.1371/journal.pone.0006421

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern*. 102, 227–260. doi: 10.1007/s00422-010-0364-z

Friston, K. J., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern*. 104, 137–160. doi: 10.1007/s00422-011-0424-z

Friston, K. J., Parr, T., and de Vries, B. (2017a). The graphical brain: belief propagation and active inference. *Netw. Neurosci*. 1, 381–414. doi: 10.1162/NETN_a_00018

Friston, K. J., Rosch, R., Parr, T., Price, C., and Bowman, H. (2017b). Deep temporal models and active inference. *Neurosci. Biobehav. Rev*. 77, 388–402. doi: 10.1016/j.neubiorev.2017.04.009

Friston, K. J., Samothrakis, S., and Montague, R. (2012). Active inference and agency: optimal control without cost functions. *Biol. Cybern*. 106, 523–541. doi: 10.1007/s00422-012-0512-8

Friston, K. J., Trujillo-Barreto, N., and Daunizeau, J. (2008). DEM: A variational treatment of dynamic systems. *Neuroimage* 41, 849–885. doi: 10.1016/j.neuroimage.2008.02.054

Gallego, J. A., Makin, T. R., and McDougle, S. D. (2022). Going beyond primary motor cortex to improve brain-computer interfaces. *Trends Neurosci*. 45, 176–183. doi: 10.1016/j.tins.2021.12.006

Galletti, C., and Fattori, P. (2018). The dorsal visual stream revisited: Stable circuits or dynamic pathways? *Cortex* 98, 203–217. doi: 10.1016/j.cortex.2017.01.009

Galletti, C., Gamberini, M., and Fattori, P. (2022). The posterior parietal area V6A: an attentionally-modulated visuomotor region involved in the control of reach-to-grasp action. *Neurosci. Biobehav. Rev*. 141, 104823. doi: 10.1016/j.neubiorev.2022.104823

Gamberini, M., Passarelli, L., Filippini, M., Fattori, P., and Galletti, C. (2021). Vision for action: thalamic and cortical inputs to the macaque superior parietal lobule. *Brain Struct. Funct*. 226, 2951–2966. doi: 10.1007/s00429-021-02377-7

Genovesio, A., Tsujimoto, S., and Wise, S. P. (2012). Encoding goals but not abstract magnitude in the primate prefrontal cortex. *Neuron* 74, 656–662. doi: 10.1016/j.neuron.2012.02.023

Goodfellow, I. J., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.

Haar, S., and Donchin, O. (2020). A revised computational neuroanatomy for motor control. *J. Cogn. Neurosci*. 32, 1823–1836. doi: 10.1162/jocn_a_01602

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press UK. doi: 10.1093/acprof:oso/9780199682737.001.0001

Kaplan, R., and Friston, K. J. (2018). Planning and navigation as active inference. *Biol. Cybern*. 112, 323–343. doi: 10.1007/s00422-018-0753-2

Keele, S. W., and Posner, M. I. (1968). Processing of visual feedback in rapid movements. *J. Exp. Psychol*. 77, 155–158. doi: 10.1037/h0025754

Kikuchi, Y., and Hamada, Y. (2009). Geometric characters of the radius and tibia in Macaca mulatta and Macaca fascicularis. *Primates* 50, 169–183. doi: 10.1007/s10329-008-0120-3

Kingma, D. P., and Welling, M. (2014). "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014-Conference Track Proceedings* (Banff), 1–14. doi: 10.48550/arXiv.1312.6114

Kornblum, S., Hasbroucq, T., and Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility-a model and taxonomy. *Psychol. Rev*. 97, 253–270. doi: 10.1037/0033-295X.97.2.253

Lanillos, P., and Cheng, G. (2018). "Adaptive robot body learning and estimation through predictive coding," in *IEEE International Conference on Intelligent Robots and Systems* (Madrid: IEEE), 4083–4090.

Lanillos, P., Pages, J., and Cheng, G. (2020). "Robot self/other distinction: active inference meets neural networks learning in a mirror," in *ECAI 2020* (Santiago de Compostela). doi: 10.48550/arXiv.2004.05473

Lau, H. C., Rogers, R. D., Haggard, P., and Passingham, R. E. (2004). Attention to Intention. *Sicence* 303, 1208–1210. doi: 10.1126/science.1090973

Levine, S. (2018). Reinforcement learning and control as probabilistic inference: tutorial and review. *ArXiv [Preprint]*. doi: 10.48550/arXiv.1805.00909

Limanowski, J., and Friston, K. (2020). Active inference under visuo-proprioceptive conflict: simulation and empirical results. *Sci. Rep*. 10, 1–14. doi: 10.1038/s41598-020-61097-w

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci*. 9, 1432–1438. doi: 10.1038/nn1790

Medendorp, W. P., and Heed, T. (2019). State estimation in posterior parietal cortex: distinct poles of environmental and bodily states. *Progr. Neurobiol*. 183, 101691. doi: 10.1016/j.pneurobio.2019.101691

Millidge, B., Tschantz, A., Seth, A. K., and Buckley, C. L. (2020). On the relationship between active inference and control as inference. *Commun. Comput. Inf. Sci*. 1326, 3–11. doi: 10.1007/978-3-030-64919-7_1

Oliver, G., Lanillos, P., and Cheng, G. (2019). Active inference body perception and action for humanoid robots. *ArXiv [Preprint]*. doi: 10.48550/arXiv.1906.03022

Parr, T., and Friston, K. J. (2018). The anatomy of inference: Generative models and brain structure. *Front. Comput. Neurosci*. 12, 90. doi: 10.3389/fncom.2018.00090

Parr, T., Pezzulo, G., and Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/12441.001.0001

Parr, T., Rikhye, R. V., Halassa, M. M., and Friston, K. J. (2020). Prefrontal computation as active inference. *Cereb. Cortex* 30, 682–695. doi: 10.1093/cercor/bhz118

Pezzulo, G., and Cisek, P. (2016). Navigating the affordance landscape: feedback control as a process model of behavior and cognition. *Trends Cogn. Sci*. 20, 414–424. doi: 10.1016/j.tics.2016.03.013

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., and Castelfranchi, C. (2019). The body talks: sensorimotor communication and its brain and kinematic signatures. *Phys. Life Rev*. 28, 1–21. doi: 10.1016/j.plrev.2018.06.014

Pezzulo, G., Donnarumma, F., Iodice, P., Maisto, D., and Stoianov, I. (2017). Model-based approaches to active perception and control. *Entropy* 19, 266. doi: 10.3390/e19060266

Pezzulo, G., Rigoli, F., and Friston, K. J. (2018). Hierarchical active inference: a theory of motivated control. *Trends Cogn. Sci*. 22, 294–306. doi: 10.1016/j.tics.2018.01.009

Pio-Lopez, L., Nizard, A., Friston, K., and Pezzulo, G. (2016). Active inference and robot control: a case study. *J. R. Soc. Interface* 13, 122. doi: 10.1098/rsif.2016.0616

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci*. 2, 79–87. doi: 10.1038/4580

Rood, T., van Gerven, M., and Lanillos, P. (2020). "A deep active inference model of the rubber-hand illusion," in *Active Inference. IWAI 2020. Communications in Computer and Information Science, Vol. 1326*, eds T. Verbelen, P. Lanillos, C. L. Buckley and C. De Boom (Cham: Springer).

Sajid, N., Ball, P. J., Parr, T., and Friston, K. J. (2021). Active inference: demystified and compared. *Neural Comput*. 33, 674–712. doi: 10.1162/neco_a_01357

Sancaktar, C., van Gerven, M. A. J., and Lanillos, P. (2020). "End-to-end pixel-based deep active inference for body perception and action," in *2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (Valparaiso: IEEE), 1–8.

Saunders, J. A., and Knill, D. C. (2003). Humans use continuous visual feedback from the hand to control fast reaching movements. *Exp. Brain Res*. 152, 341–352. doi: 10.1007/s00221-003-1525-2

Shadmehr, R., and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Exp. Brain Res*. 185, 359–381. doi: 10.1007/s00221-008-1280-5

Shenoy, K. V., Sahani, M., and Churchland, M. M. (2013). Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci*. 36, 337–359. doi: 10.1146/annurev-neuro-062111-150509

Snyder, L. H., Batista, A. P., and Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature* 386, 167–170. doi: 10.1038/386167a0

Snyder, L. H., Batista, A. P., and Andersen, R. A. (2000). Intention-related activity in the posterior parietal cortex: a review. *Vision Res*. 40, 1433–1441. doi: 10.1016/S0042-6989(00)00052-3

Srinivasan, S. S., Gutierrez-Arango, S., Teng, A. C. E., Israel, E., Song, H., Bailey, Z. K., et al. (2021). Neural interfacing architecture enables enhanced motor control and residual limb functionality postamputation. *Proc. Natl. Acad. Sci. U.S.A*. 118, e2019555118. doi: 10.1073/pnas.2019555118

Stoianov, I., Genovesio, A., and Pezzulo, G. (2016). Prefrontal goal codes emerge as latent states in probabilistic value learning. *J. Cogn. Neurosci*. 28, 140–157. doi: 10.1162/jocn_a_00886

Stoianov, I., Kramer, P., Umiltà, C., and Zorzi, M. (2008). Visuospatial priming of the mental number line. *Cognition*. 106, 770–779. doi: 10.1016/j.cognition.2007.04.013

Stoianov, I., Maisto, D., and Pezzulo, G. (2022). The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning. *Progr. Neurobiol*. 217, 1–20. doi: 10.1016/j.pneurobio.2022.102329

Stoianov, I., Pennartz, C., Lansink, C., and Pezzulo, G. (2018). Model-based spatial navigation in the hippocampus-ventral striatum circuit: a computational analysis. *PLoS Comput. Biol*. 14, 1–28. doi: 10.1371/journal.pcbi.1006316

Stoianov, I., and Zorzi, M. (2012). Emergence of a 'visual number sense' in hierarchical generative models. *Nat. Neurosci*. 15, 194–196. doi: 10.1038/nn.2996

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nat. Neurosci*. 7, 907–915. doi: 10.1038/nn1309

Todorov, E., and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nat. Neurosci*. 5, 1226–1235. doi: 10.1038/nn963

Toussaint, M., and Storkey, A. (2006). Probabilistic inference for solving discrete and continuous state Markov Decision Processes. *ACM Int. Conf. Proceed. Ser*. 148, 945–952. doi: 10.1145/1143844.1143963

Tuthill, J. C., and Azim, E. (2018). Proprioception. *Curr. Biol*. 28, R194-R203. doi: 10.1016/j.cub.2018.01.064

Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature* 453, 1098–1101. doi: 10.1038/nature06996

Versteeg, C., Rosenow, J. M., Bensmaia, S. J., and Miller, L. E. (2021). Encoding of limb state by single neurons in the cuneate nucleus of awake monkeys. *J. Neurophysiol*. 126, 693–706. doi: 10.1152/jn.00568.2020

Wolpert, D. M., and Flanagan, J. R. (2016). Computations underlying sensorimotor learning. *Curr. Opin. Neurobiol*. 37, 7–11. doi: 10.1016/j.conb.2015.12.003

frontiers | Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 27 July 2023
DOI 10.3389/fncom.2023.1132160

Check for updates

OPEN ACCESS

EDITED BY
Benjamin R. Pittman-Polletta,
Boston University, United States

REVIEWED BY
Kelong Lu,
Wenzhou Medical University, China
Zhipeng He,
Sun Yat-sen University, China

*CORRESPONDENCE
Lilia Costa
✉ liliacosta@ufba.br

RECEIVED 26 December 2022
ACCEPTED 13 June 2023
PUBLISHED 27 July 2023

CITATION
do Nascimento DC, Santos da Silva JR, Ara A, Sato JR and Costa L (2023) Hyperscanning fNIRS data analysis using multiregression dynamic models: an illustration in a violin duo. *Front. Comput. Neurosci.* 17:1132160. doi: 10.3389/fncom.2023.1132160

Hyperscanning fNIRS data analysis using multiregression dynamic models: an illustration in a violin duo

Diego Carvalho do Nascimento[1], José Roberto Santos da Silva[2,3], Anderson Ara[4], João Ricardo Sato[5] and Lilia Costa[2]*

[1]Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile, [2]Department of Statistics, Federal University of Bahia, Salvador, Brazil, [3]EcMetrics Pesquisa de Mercado, Salvador, Brazil, [4]Departamento de Estatística, Universidade Federal do Parana, Curitiba, Brazil, [5]Center of Mathematics, Computing and Cognition, Universidade Federal do ABC, São Bernardo do Campo, Brazil

Introduction: Interpersonal neural synchronization (INS) demands a greater understanding of a brain's influence on others. Therefore, brain synchronization is an even more complex system than intrasubject brain connectivity and must be investigated. There is a need to develop novel methods for statistical inference in this context.

Methods: In this study, motivated by the analysis of fNIRS hyperscanning data, which measure the activity of multiple brains simultaneously, we propose a two-step network estimation: Tabu search local method and global maximization in the selected subgroup [partial conditional directed acyclic graph (DAG) + multiregression dynamic model]. We illustrate this approach in a dataset of two individuals who are playing the violin together.

Results: This study contributes new tools to the social neuroscience field, which may provide new perspectives about intersubject interactions. Our proposed approach estimates the best probabilistic network representation, in addition to providing access to the time-varying parameters, which may be helpful in understanding the brain-to-brain association of these two players.

Discussion: The illustration of the violin duo highlights the time-evolving changes in the brain activation of an individual influencing the other one through a data-driven analysis. We confirmed that one player was leading the other given the ROI causal relation toward the other player.

KEYWORDS

dynamic network, state-space models, causal inference, dual brain, interactive social neuroscience

1. Introduction

The brain is formed by a network in which different regions share information Horwitz (2003). This brain network can be studied through *functional connectivity*, which represents the patterns of statistical dependence on the activity of distinct brain regions, or through *effective connectivity*, which means the causal influences of the activity of one region over another. The variance-covariance matrix and the Bayesian network (BN) are examples of methods used to estimate functional connectivity. Other methods can be used to study effective connectivity, such as dynamic causal modeling (DCM) and the multiregression dynamic model (MDM). For a given directed network structure, the MDM models the data at each node as a linear combination of the parent nodes with time-varying connectivity parameters. According to Queen and Smith (1993), the MDM can distinguish between directed graphs corresponding to the same statistical dependence structure (which map onto the same undirected graphs), allowing for the accurate estimation of the directions of edges (a simple example of this is also discussed here). Moreover, the MDM can be observed as more than a static network (similar to BN). Alternative examples of these dynamic methods can be found in Burger et al. (2009), who used the dynamic Bayesian network (DBN) and hidden Markov models (HMM) used in human-robot interaction (DBN and HMM are a particular case of MDM).

In any case, the problem of finding a common pattern of brain connectivity for a given individual profile (healthy or with a specific disease, e.g., Alzheimer's disease) is not trivial owing to the presence of noise and the high-dimensionality of the data (Nascimento et al., 2020; Pinto-Orellana et al., 2020). For the MDM, Costa et al. (2015) presented a score-based learning network approach using a linear programming problem that finds the most likely network structure while considering the subset comparison through their maximum posterior probability (MAP) estimation. The authors demonstrated the usefulness of their method on functional Magnetic Ressonance Imaging (fMRI) data as it becomes unfeasible as the number of nodes (i.e., brain regions) increases.

In addition to this challenge, in the field of social neuroscience, understanding how the activity of the brain might influence the activity of another brain, which is known as brain-to-brain activity correlation, is also desirable. As examples, we considered a classroom where the teacher and the students interact or an orchestra where the musicians and the conductor interact. Konvalinka and Roepstorff (2012) describes how mutually interacting brains can be useful in social interaction. Balconi et al. (2017) studied the effects of strategic cooperation on intra- and inter-brain connectivity by functional near-infrared spectroscopy (fNIRS). Jiang et al. (2019) developed a study entitled "BrainNet: a multi-person brain-brain interface for direct collaboration between brains," among others.

Hyperscanning studies—measuring the activity of multiple brains simultaneously—is a promising (flexible) paradigm regarding the measurement of brain activity from two or more people simultaneously while they are interacting. This could reveal interpersonal brain mechanisms underlying interaction-mediated brain-to-brain coupling Scholkmann et al. (2013). One experiment that could be conducted to this end, focusing on two brains' observations, is the study of violin duos playing together. The fNIRS could be used to overcome functional magnetic resonance imaging constraints, but few dynamic data-driven models have been proposed. Thus, we aimed to apply a dynamic graphical model to show dynamic changes in intersubject brain activity dependence over time.

1.1. Interaction-mediated brain-to-brain activity correlation

In recent decades, part of the neuroscience field has focused on demonstrating the nervous system and its function through individuals' behavior (and inter-relations) (Liu and Pelowski, 2014). For instance, some studies have discussed the brain connectivity structure by gender (Wang et al., 2009; Baker et al., 2016; Pan et al., 2017), age (Gong et al., 2009), or using other characteristics such as intelligence (Song et al., 2008; Van Den Heuvel M. et al., 2008; van den Heuvel M. P. et al., 2008), psychoactive ingestion (Palhano-Fontes et al., 2019), and meditative states (Brefczynski-Lewis et al., 2007; Brewer et al., 2011; Hasenkamp and Barsalou, 2012). Nevertheless, all of them have targeted different methodologies related to neuroanatomy. These methodologies also understand the brain connection patterns in human actions, such as opening and closing eyes (or moving any other body part), reading, writing, playing sports, learning, sleeping, creating memories, and recalling these memories (Hahn et al., 2018).

However interpersonal neural synchronization (INS) demands a greater understanding of the influence that a brain may carry on others rather than observing only a single brain response per time (for further details, see Babiloni and Astolfi, 2014). Hyperscanning studies are based on the simultaneous acquisition of brain dynamics during a cooperative task, as a joint action or decision-making (Liu et al., 2016, 2017).

Li et al. (2020) studied the cooperative behavior among basketball players, in which significant INS was observed due to the performed joint-drawing task but not the control task. Nguyen et al. (2020) investigated the neural processes related to transferring information across brains during naturalistic teaching and learning, underlying the effective communication of complex information across brains in classroom settings.

With more than only linking actions across subjects, studies have revealed that inter-individuals' neural representation can even build memories, thereby promoting brain integration at some influential level. Zadbood et al. (2017) uncovered the intimate correspondences between memory encoding and event construction and highlighted the essential role that our common language plays in the process of transmitting one's memories to other brains. Chen et al. (2017) elucidated that the neural patterns during perception are systematically altered across people into shared memory representations for real-life events.

Most methods used in hyperscanning fMRI and fNIRS studies are static or temporal correlation (Cui et al., 2012; Reindl et al., 2022; Balconi and Angioletti, 2023; Morgan et al., 2023; Wei et al., 2023) and Granger-based causality (Zhang et al., 2017; Chen et al., 2020, 2023; Pan et al., 2021; Zhao et al., 2022). Examples of the former method are the partial correlation coefficient and wavelet transform coherence (WTC). These methods are used to estimate functional connectivity and, therefore, do not distinguish the causal relationships between nodes. Nonetheless, according to neuroimaging literature, the latter is used to estimate directed functional connectivity (Bilek et al., 2022), and according to some studies, Granger causality theory cannot be suitable for hemodynamic data (Smith et al., 2011; Babiloni and Astolfi, 2014). Therefore, these approaches do not study putative causal synchrony between brains (Bilek et al., 2022). Thus, Bilek et al. (2022) used dynamic causal modeling (DCM) in the study of social interaction to estimate the causal effect one brain might have on another. However, DCM is a method for testing hypotheses, and initially specifying some candidate network structures is necessary.

This study uses the MDM with the Bayes factor (MDM-BF) that considers the contemporaneous relationship between regions, i.e., the nodes are related at the same time, in contrast, for example, to a DBN in which the past of the parents is connected with the present of the child. Moreover, the Kalman filtering method is used to estimate the effective connectivity in a simple way. However, in contrast to DCM, it can capture the dynamic nature of social interaction. A similar objective can be observed in Li et al. (2021) and Wang et al. (2022), in which the researchers used a data-driven approach based on sliding windows and k-mean clustering to capture the dynamic modulation of inter-brain synchrony patterns. However, it is based on temporal correlation and does not estimate effective connectivity.

The MDM appears to accommodate fMRI data well (see e.g., Costa et al., 2015); therefore, it has been used in this study for the first time with fNIRS data. Furthermore, this study proposes a new method that can be used to learn the directed acyclic graph (DAG) structure using the MDM faster than the method already available in the literature (MDM-IPA) because this method does not create the need to check all possible parents for each node. This can be especially useful in social neuroscience—which involves estimating both inter-brain and intra-brain connections and thus studying brain function on the subject and dyadic levels.

This novel method consists of two steps: in the first one, the tabu search algorithm would be applied to find a partial conditional directed acyclic graph (partial conditional DAG). The tabu search is a combinatorial optimization algorithm used to find an optimal network structure by local searches, as explained in the next section. Then, the Markov equivalence class would be found, that is, DAGs that encode the same statistical properties, and the DAG with the highest log predictive likelihood (LPL) score from the MDM would be chosen. This search method can also be used to estimate individual brain networks.

Based on such evidence, which highlights the possibility of studying the brain-to-brain activity correlation, in the next subsection, we have discussed an extension class of DBNs that can be used to represent these brain dynamic and causal structures (from now on, whenever we refer to causality, it is associated with effective connectivity via MDM-BF, unless indicated differently).

This study is divided into four parts. In Section 2, we have described the fNIRS data analyzed and the methods used to estimate brain connectivity as a graph-based model. Section 3 describes the evaluation, through synthetic data, of the robustness of the dynamic graphical model. Then, Section 4 presents the empirical results, and finally, Section 5 presents the discussion of the proposed method and the findings.

2. Materials and methods

We present a proof-of-concept based on a hyperscanning experiment in which the human interaction is investigated from brain-to-brain activity dependence. The methodological approach adopted in this study was divided into four main steps, aiming to estimate the brain's dynamics and interactions. The developed R script in this study is available at https://github.com/ProfNascimento/MDM-BF (accessed on April 20th, 2023).

2.1. The data

This study dataset was first presented as a case study experiment (Balardin et al., 2017) that considered two individuals who played in a violin duo. In the current study, we have investigated the brain-to-brain coupling (and the direction) and explored which brain regions of a violinist are linked to the other.

The fNIRS signals acquired are demonstrated in Figure 1 (for further experiment details, see Balardin et al., 2017). Hemodynamic changes were obtained from the optical changes collected using the continuous wave functional near-infrared spectroscopy system (NIRScout 16x16, NIRx Medical Technologies, Glen Head, NY) with 16 LED light sources (760 and 850 nm) and 16 detectors per musician, at a sampling rate of 7.81 Hz. Channel aggregation was conducted by considering the EEG 10-10 system in which the optodes were placed.

The participants were at a professional level, right-handed, and men aged 41 and 50 years old. They were instructed to play a 32-s stretch of Allegro, by Antonio Vivaldi, from Concerto No 1 in E major, op. 8, RV 269, "Spring." Hyperscanning was performed considering 23 channels of the right motor hemisphere and the temporoparietal junction of the two violinists (Balardin et al., 2017). The first 36 s of acquisition refers to the duo playing and the remaining refers to a resting-state condition.

2.2. Dataflow

Figure 2 demonstrates a data processing flow chart. Given the computational cost of searching for the likely topology of the graph, at first, the tabu search algorithm was applied using Bayesian networks to reduce the sub-graph structure to be sought. Then, the result was transformed into a partial conditional directed acyclic graph (DAG), enabling it to proceed under the causal inference paradigm (Pearl, 2009; Oates et al., 2015). After that, the most likely undirected network structure found was implied in Markov equivalent graphs, and then, the MDM was applied to unravel directionality through the maximization of the LPL, that is, Bayes factor. By adopting a particle filter supposing Gaussian noises (often known as the Kalman filter), we compared the MAP graphs to obtain the most likely DAG. Once the DAG is defined, the MDM-BF can present the dynamic strength of these estimated links. This adopted methodology enables the estimation of complex brain structures whenever the number of vertices (nodes) is >11 with a sample size $>$ 100 points (for further details, see Costa et al., 2015, Table 01, p. 456), without computational constraints.

2.3. The multiregression dynamic model

The MDM models multivariate time series, studying putative causal relations among its variables over time (Queen and Smith, 1993; Queen and Albers, 2009). This class of models is extremely powerful, given that it can discriminate complex multivariate relations up to a finite r-th time series, with length t, set as $(Y_t(1), Y_t(2), ..., Y_t(r))$. Moreover, the joint distribution $(P(Y_t(1), Y_t(2), ..., Y_t(r)))$ is estimated regardless of the presence

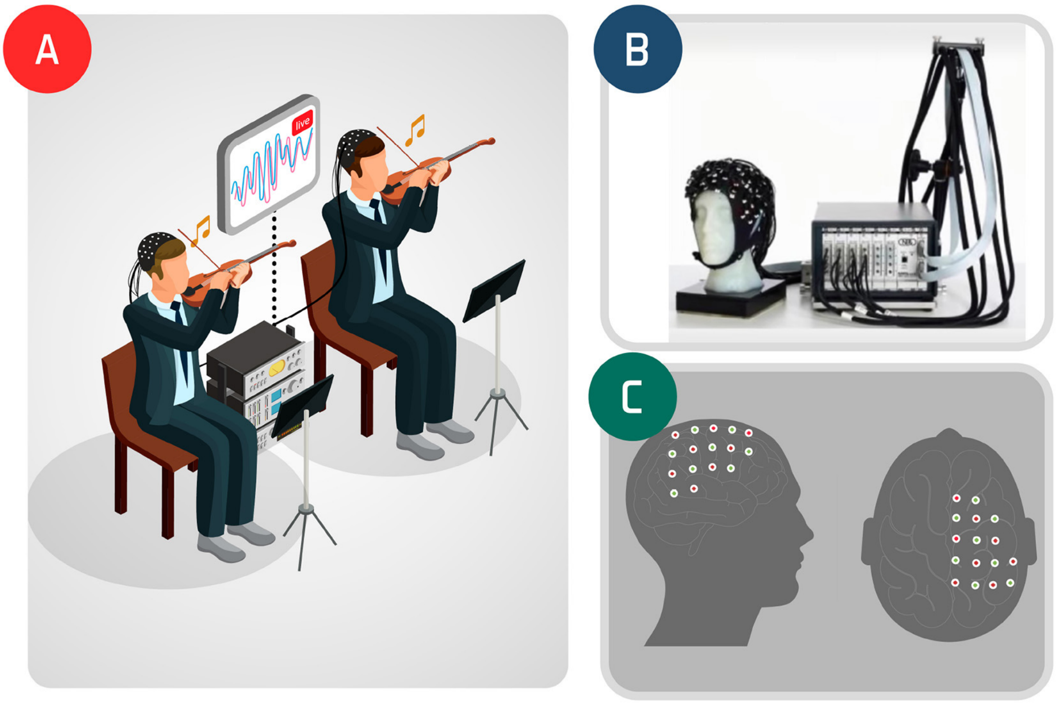


FIGURE 1
Violin duo experiment: inter-subjects' experiment icon **(A)**; fNIRS **(B)** and the observed brain region **(C)**.

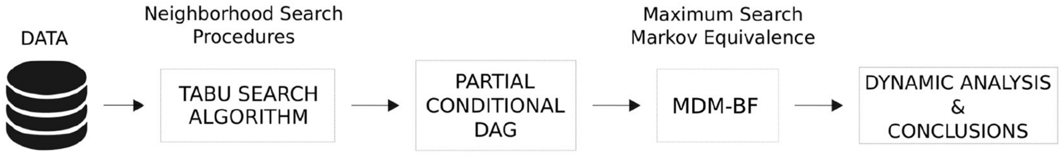

DATA
Neighborhood Search Procedures
TABU SEARCH ALGORITHM
PARTIAL CONDITIONAL DAG
Maximum Search Markov Equivalence
MDM-BF
DYNAMIC ANALYSIS & CONCLUSIONS

FIGURE 2
Visual summary of the methodological framework. Tabu search algorithm reduced the sub-graph structure to be sought. Then, the result was transformed into a partial conditional DAG, and the outcome was compared using the MDM-BF.

of a Gaussianity (for further details, please see Queen and Smith, 1993, who have retained the proof of the consistency of this method of non-Gaussian processes). The MDM is formed by using univariate regression dynamic linear models (DLMs), in which the observation $Y_t(r)$ is regressed onto its parents, with Gaussian residuals, such as in Equation (1).

$$\begin{aligned} Y_t(r) &\sim \mathcal{N}(\mathbf{F}_t(r)'\boldsymbol{\theta}_t(r), V_t(r)) \\ \boldsymbol{\theta}_t &\sim \mathcal{N}(\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t), \end{aligned} \tag{1}$$

where $Y_t(r)$ is an observable variable at time t and brain region r, $r = 1, \ldots, n$ regions, $t = 1, \ldots, T$ time points, $\mathcal{N}$ denotes the Gaussian distribution, $\boldsymbol{\theta}'_t = (\boldsymbol{\theta}_t(1)', \ldots, \boldsymbol{\theta}_t(n)')$, $\boldsymbol{\theta}_t(r)'$ is the p_r-dimensional parameter vector for $Y_t(r)$, and, when it is not intercepted, it represents the effective connectivity between node r and its descendent (also called parents). $\mathbf{F}_t(r)$ is the set of the parents, and for nodes that do not have parents, $F_t(r) = 1$. $\mathbf{G}_t$ increments the state equation in the form, giving extra variance.

In addition, $\mathbf{W}_t(r)$ are p_r square matrices that form $\mathbf{W}_t = \text{blockdiag}\{\mathbf{W}_t(1), \ldots, \mathbf{W}_t(n)\}$. Note that, when $\mathbf{W}_t(r)$ is a matrix with all elements equal to zero, the MDM becomes the BN. The parameters can be estimated using well-known Kalman filter recurrences over time (see, for example, West and Harrison, 2006). By so doing, the DLM is described by the set $\{\mathbf{F}_t(r), Vt, \mathbf{G}_t, \mathbf{W}_t\}$, although, in practice, establishing the $\mathbf{W}_t$ is challenging; therefore, a strategy called "discounting" (stochastic shifting) is adopted.

$$\mathbf{W}_t = \frac{1-\delta}{\delta} \times \boldsymbol{C}_{t-1}, \tag{2}$$

where $\mathbf{W}_t$ is specified directly through a discount factor $\delta \in (0, 1]$, and $\boldsymbol{C}_{t-1}$ is the posterior variance of $\boldsymbol{\theta}_t$.

Before proceeding, three terminologies are important for distinguishing estimation processes: (i) Filtering is a procedure that aims to update the current estimates as new data are observed, i.e., $\mathbb{P}(\theta_t \mid Y_{1:t})$; ($ii$) smoothing is a retrospective analysis that has all the observations and calculates the conditional distribution θ given the heading from the complete data, $\mathbb{P}(\theta_t \mid Y_{1:T})$; and (*iii*) prediction is a forecast procedure that estimates the next observation based on the distribution, $\mathbb{P}(\theta_{t+1} \mid Y_{1:t})$.

2.4. The proposed learning network

The learning network process used in this study is 2-fold: (i) an estimation process of a Bayesian network structure, the tabu search algorithm, and (ii) choosing a structure via the MDM in Markov equivalent networks, that is, partial conditional DAG → MDM-BF. This methodological combination is an alternative to reduce the np-hard (dimensional) complexity search problem of the network estimation.

First, the initial estimation process is related to traditional methods in Bayesian networks (time-invariant structure). This approach was performed using a score-based method via standard tabu search with Bayesian information criterion (BIC). In general, this method searches for a Bayesian network structure that maximizes BIC. A Tabu search (Glover, 1986) may be viewed as a meta-heuristic algorithm to perform a greedy search and to avoid local minima. Thus, the procedure records information about changes recently made in BN structures, using one or more tabu lists. The tabu lists are managed by recording moves in a sequential order. Each time a new link is added to the end of a list, the oldest arc on the list is dropped from the beginning. Thus, each structure generated by adding or removing links is appraised by the BIC scoring (Nagarajan et al., 2013). The tabu algorithm adopted here can be found in the bnlearn package from the R software (R Core Team, 2022). Furthermore, every statistical analysis used in this study adopted the software R.

As it is well known that the BN search approaches have trouble distinguishing Markov equivalent structures, the next step was to find the graphs that are Markov equivalent to one resulting from the tabu search. Afterward, the network structure with the largest score of MDM among these Markov equivalent graphs was chosen. The pcalg package was used to obtain the partial conditional DAG.

Once the partial DAG structure was established, only a few subsets of possibilities remain to be sought. At this point, the maximum likelihood approach was adopted to determine the best options for the subgroup. The assumption from the MDM is that the standardized conditional one-step forecast errors have an approximate Gaussian distribution, although not based on stationary time series, and are serially independent with constant variance. Under these assumptions, the joint log predictive likelihood (LPL) has the closed form of a noncentral t distribution and is easily found in the Kalman filter (Costa et al., 2015). Remembering that $\mathbb{Y} = \{Y_t(1), \cdots, Y_t(r)\}$ if time-invariant $\mathbb{Y} = \{Y(1), \cdots, Y(r)\}$, considering a multivariate non-central t distribution function

$$f(\mathbb{Y}|\mu, \sigma^2\Sigma, \nu) = \frac{\Gamma[(\nu+r)/2]}{(\pi\nu)^{r/2}|\sigma^2\Sigma|^{1/2}\Gamma[\nu/2]} \left(1 + \frac{(\mathbb{Y}-\mu)'\Sigma^{-1}(\mathbb{Y}-\mu)}{\sigma^2\nu}\right)^{-(\nu+r)/2} \tag{3}$$

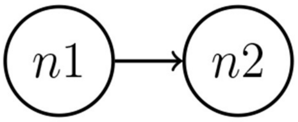


FIGURE 3
Data were simulated considering two nodes. The MDM method through Kalman filter estimation performs an estimate of direction and its time-varying strength.

in which μ is the vector of the means and Σ is the variance-covariance matrix under the Bayesian framework

$$\mathbb{P}(\tau, \mathbb{Y}|\mu, \sigma^2\Sigma, \nu) \propto \mathbb{P}(\mathbb{Y}|\tau, \mu, \sigma^2\Sigma)\mathbb{P}(\tau|\nu) \tag{4}$$

$$\mathbb{Y}|\tau, \mu, \sigma^2\Sigma \sim N(\mu, (\sigma^2\Sigma/\tau)) \tag{5}$$

$$\tau|\nu \sim Ga(\nu/2, \nu/2) \tag{6}$$

and then assuming that the conditional distribution of each $Y_t(r)$ is given by the previous information set $\mathcal{F}_{t-1}$, one can simply consider a regression structure for the conditional mean $\mu_t = \mathbf{F}_t(r)'\boldsymbol{\theta}_t(r)$ and $\Sigma_t = V_t$

$$\begin{aligned} &log(f(Y_t(1), \cdots, Y_t(r)|\mathcal{F}_{t-1})) = log(L(\mathbf{F}_t(r)'\boldsymbol{\theta}_t(r), V_t|\mathcal{F}_{t-1})) \\ &= LPL(\mathbf{F}_t(r)'\boldsymbol{\theta}_t(r), V_t|\mathcal{F}_{t-1}) \end{aligned} \tag{7}$$

Therefore, the LPL is the score of the MDM used in the learning network process, and in the following section, Section 3, a simple example of the ability of this score to distinguish two Markov equivalent graphs is given. It must be mentioned that local Gaussian models do not imply, necessarily, a posterior symmetrical multivariate distribution (for further details, see Queen and Smith, 1993).

3. Simulating the MDM

This simulation study aimed to present the performance of the MDM in estimating the network structure and relationship strength (parameter θ) between the two nodes over time. Each time series contains 300 observations, that is, $t = \{1, \ldots, 300\}$. Figure 3 represents the theoretical (known) network, in which node 1 (n1) is the parent of node 2 (n2). The data simulation was performed by using R software.

For instance, let us suppose that two signals, $n1(t)$ and $n2(t)$, are related, and they can be written as a first-order linear-Gaussian state-space model (these models presenting Gaussian noises are often called the Kalman filter, which is a special case of a particle filter for contemporaneous influence), as demonstrated in the following equations:

$$n1_t = \theta_t^{(1)} + v_t^{(1)}, v_t^{(1)} \sim \mathcal{N}(0, 0.1^2) \tag{8}$$

$$n2_t = \theta_t^{(2)} + \theta_t^{(3)} n1_t + v_t^{(2)}, v_t^{(2)} \sim \mathcal{N}(0, 0.1^2), \tag{9}$$

$$\theta_t^{(k)} = \theta_{t-1}^{(k)} + w_t^{(k)}, w_t^{(k)} \sim \mathcal{N}(0, 0.1^2) \tag{10}$$

in which $k = \{1, 2, 3\}$, and $v^{(1)}$, $v^{(2)}$, and $w'(k)$ are independent. There are structural equations containing time-varying parameters

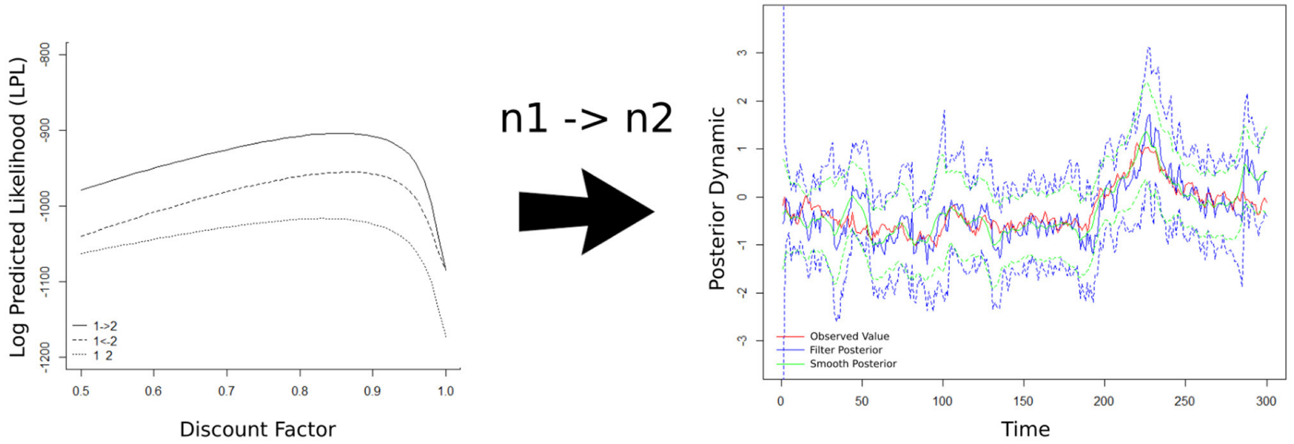


FIGURE 4
The result of the data-driven model MDM-BF is represented in two parts, *structure selection* and *dynamic estimation*. The left-hand panel describes the score (LPL) of each possible network arrangement (as a time-varying score). Three possible network outcomes were adjusted, n1 impacting n2 (solid line), n2 impacting n1 (dash line), and both nodes being independent (dot line), and at all times, n1→ n2 presents a higher score (then, the most likely structure). The right-hand panel shows the dynamic of the link strength, given the selected network (n1→ n2), that is, the dynamic of $\theta_t^{(3)}$. The red line represents the true value, the blue solid line is the smooth posterior mean (and the dashed lines are the credible intervals containing 95%), and the green line is the filter posterior mean.

$\theta_t^{(1)}$, $\theta_t^{(2)}$, and $\theta_t^{(3)}$. The parameters $\theta_t^{(1)}$ and $\theta_t^{(2)}$ are the drift that translates the strength for each node i at time t. The parameter $\theta_t^{(3)}$ is assumed to represent the form of the exchangeable sample information, in which (n1) impacts into (n2), and then, later, this is observed as a causal strength (in neuroscience, the effective neuronal connectivity).

After generating and processing the synthetic network, the left-hand panel of Figure 4 shows the estimated LPL for each possible network set (that is, n1→ n2, n2→ n1, and both independent nodes) by a discount factor (DF). In the inference process of the MDM-BF, $\mathbf{W}_t^{(r)}$ can be written in the function of a DF that represents the loss of information in the change of parameter $\boldsymbol{\theta}$ between times $t-1$ and t. The DF varies between zero and one, in a way that the closer the DF is to one, the more stable the system is. When DF assumes the value one, $\mathbf{W}_t^{(r)}$ is the matrix of zeros, and the MDM becomes a BN (Costa et al., 2015). After selecting the most likely model, the strength dynamism of the connection is calculated through a time-varying parameter (θ_t) approach. The right-hand panel of Figure 4 shows the true value and the MDM dynamic estimation regarding the causal effect between the nodes.

The steps are summarized in Algorithm 1 summarized as follows:

```
Read the DATA

Apply the TABU search using the Bayesian Network to
estimate the invariant structure

if $n1 \nrightarrow n2$ or $n2 \nrightarrow n1$ then
    $n1, n2$ are independent
else
    $n1 \rightarrow n2$ or $n2 \rightarrow n1$
    Then, n1 is connected to n2 but partial
conditional, that is, the direction will be ignored
at this point. For a greater number of nodes, every
combination will be tested.
end if

Calculate LPL from the TABU search subgroup, the
partial conditional DAG ($n1 \rightarrow n2$ or $n2 \rightarrow n1$).

Then, the choice will be the DAG with the maximum
LPL (that is, the directions that are established).

Once the DAG is set, the dynamic linear model
is adjusted on the DAG regression structure
(time-varying parameters estimation step, that
is MDM-BF).
```

Algorithm 1. Causal inference MDM-BF schematic (based on Figure 2).

In this case, the network with the highest LPL values was network n1 → n2, which generated the data. Markov equivalent networks have the same dependency relations between the nodes and have equivalent/equal LPL. Therefore, when the discount factor is 1, as we mentioned, there is no variation in the state parameters over time, and the MDM simply becomes a BN. Then, unsurprisingly, the direction n1→ n2 or n2→ n1 does not matter (see the left-hand panel of Figure 4). Thus, this study presents an indication that the MDM-BF is efficient in distinguishing structures that can be Markov equivalent. Here, we described the simplest case of a network structure (with only two nodes) for the sake of simplicity and visualization; nevertheless, the results are expandable to higher complexities (see e.g., in Costa et al., 2015). The next section discusses the results obtained in neuroscience application tasks.

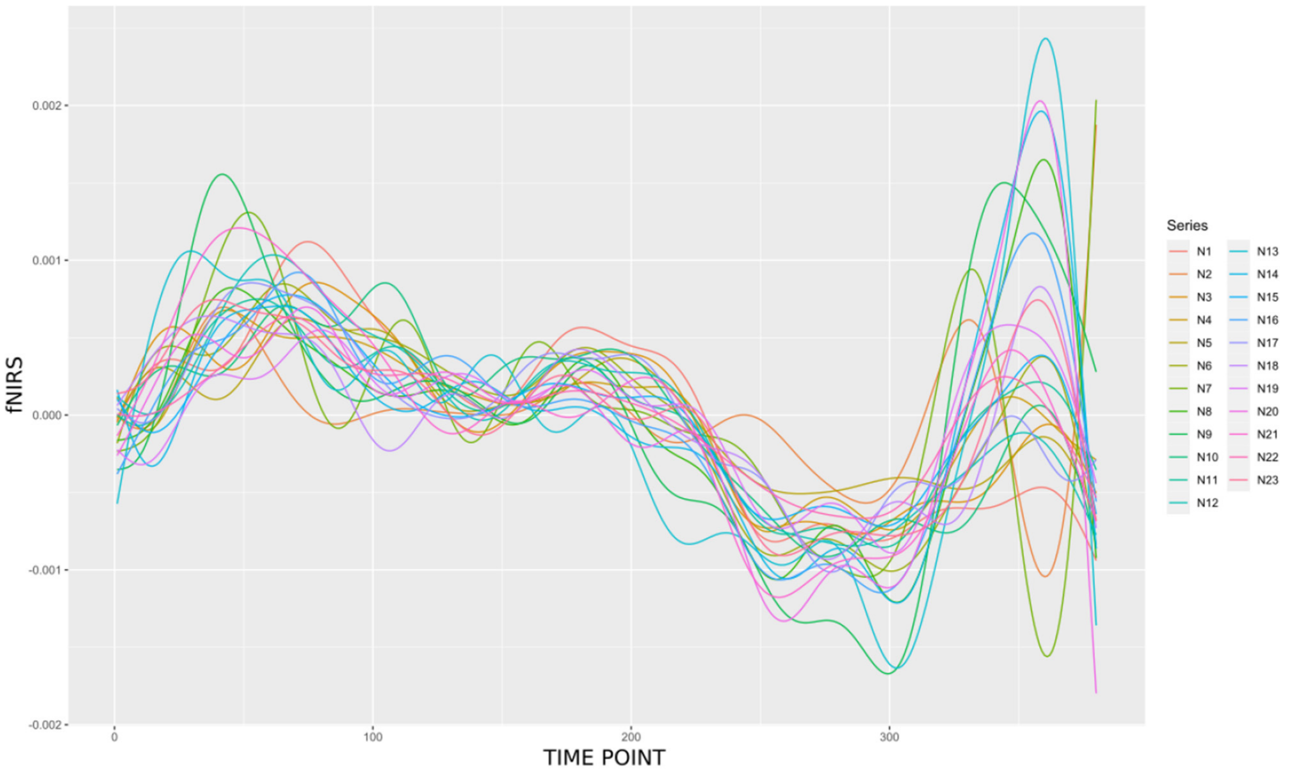


FIGURE 5
A butterfly plot illustrating the 23 oxyhemoglobin (HbO) signals from violinist A.

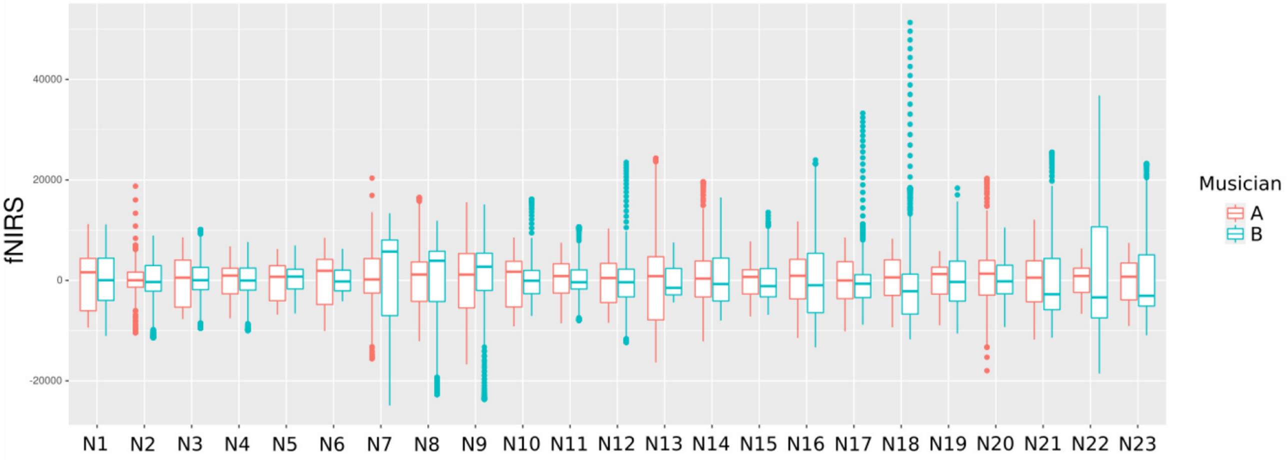


FIGURE 6
A boxplot comparing the 23 HbO signals among both musicians. This graph depicts the dispersion of the HbO of each violinist.

4. Experimental results

fNIRS enables simultaneous recording, making it possible to study the influence of brain-to-brain coupling through social interaction experiments. Figure 5 shows the fNIRS data during the music duration (218 time points) from violinist A in the 23 channels.

4.1. Dynamic brain-to-brain evolution

The study of the network involved the brains of the two violinists and considered 46 nodes, the first 23 ones corresponding to the first subject, and the remaining ones to the second subject (Figure 6). The learning of the network structure was carried out by comparing the Markov equivalent networks to the graph

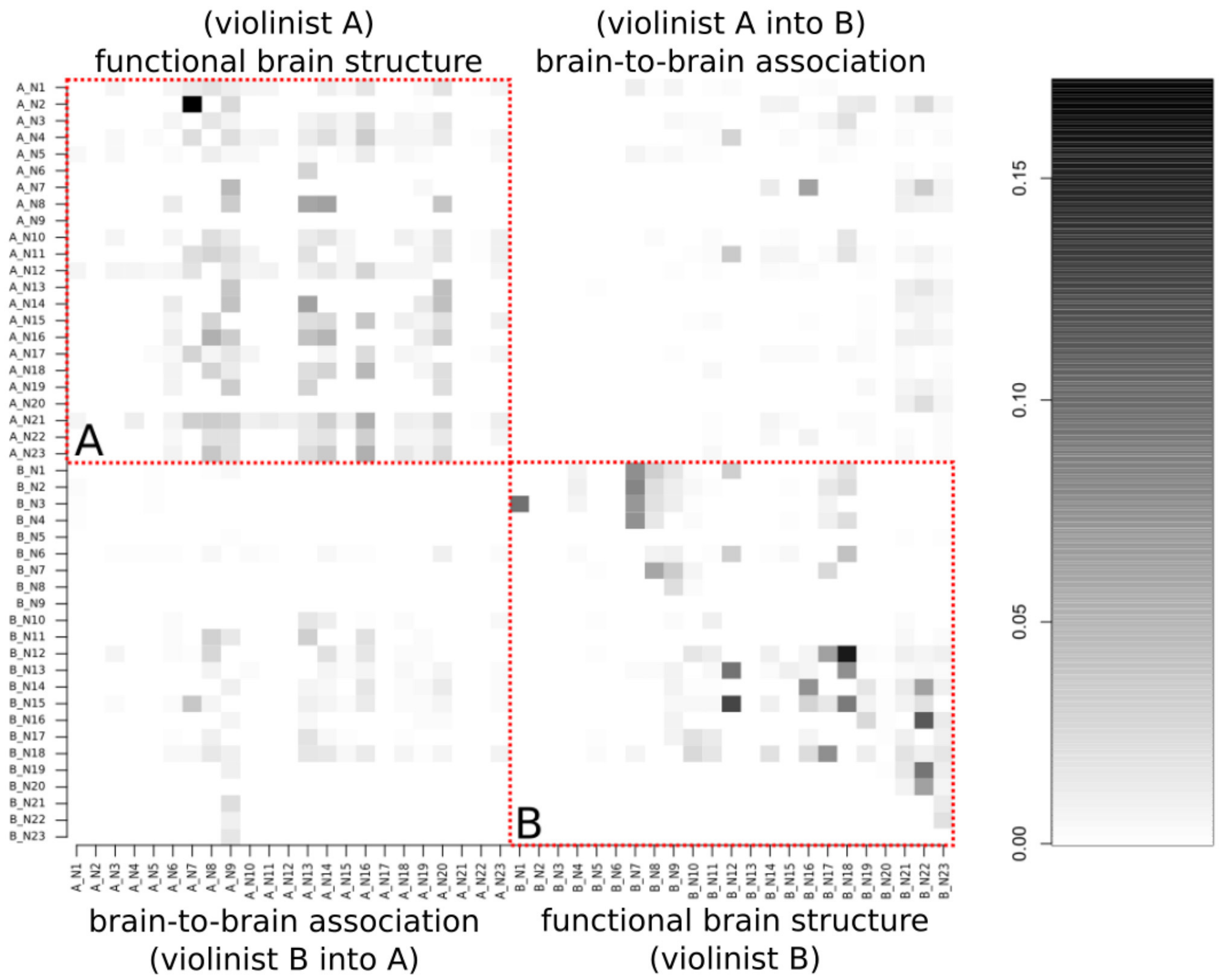


FIGURE 7
The average of the posterior mean of the strength of connectivity *i* → *j* over time, in which *i* indexes rows and *j* columns (representing each observed HbO). This matrix is divided into four blocks, in which the diagonal block is related to the musicians' activation of their own brain (in the red dashed line), and the anti-diagonal block to the brain-to-brain activity correlations. **(A)** (top-left quadrant) is from the functional brain structure of violinist A, and **(B)** (bottom-right quadrant) is from violinist B.

estimated by the tabu method and using the LPL of the MDM. The combination of the tabu search algorithm, partial conditional DAG, and the MDM-BF helped enhance the computation efficiency, bringing back the best network structure chosen for each subject.

The brain activation dynamic was analyzed by using the state-space model, obtained from the MDM-BF through its posterior mean smoothing process. It is worth mentioning that only positive connections (ignoring the few small negative estimates, as physiological interpretations are difficult to make) were presented, which enabled us to take into account their neurological interpretability. Moreover, these connections represent the neural activation resulting from one region's influences over another.

Figure 7 presents the results of the graph-based MDM, as a matrix in which each element is the average of the posterior mean of the strength of connectivity $i \rightarrow j$ over time, in which i (parents) indexes rows and j (children) indexes columns (the matrix causal relation direction is described from the row to the column). Moreover, this matrix is divided into four blocks, in which the diagonal block is intrasubject connectivity, for violinist A, at the top left square and for violinist B, at the lower right square. In contrast, the antidiagonal block shows intersubject connectivity, in which the influence of the brain regions of violinist A to B is at the top right, whereas the influence of violinist B to A is at the lower left.

Stronger connections are represented in the matrix by the darker color, while lighter regions represent weak or absent connections. As expected, the strongest connections are in the primarily diagonal block, which represents the intrasubject brain connections. The anti-diagonal block reveals that the intersubject connectivities are less prevalent and less strong. Thus, based on a standard 10-10 EEG montage, we aggregated the channel numbers 1, 2, 3, and 4 as dorsal frontal Regions of Interest (ROIs) 6, 7, 8, 10, and 11 as sensorimotor ROIs, and 12, 13, 15, 16, and 21 as temporoparietal junction (TPJ) and calculated the mean of the influence of each region. A summary of the intra-individual

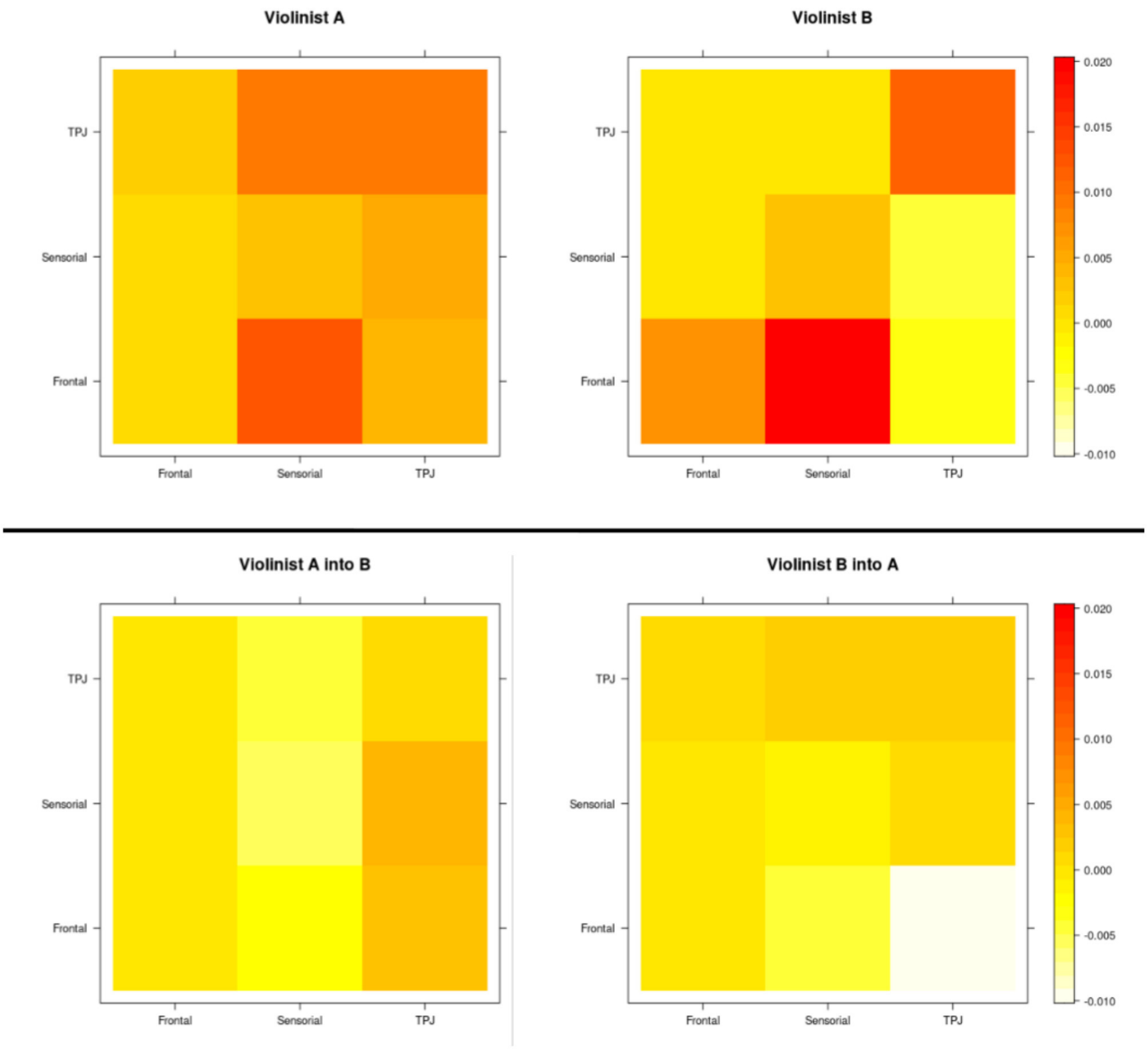


FIGURE 8
ROIs' mean representation, according to the 10-10 EEG montage, of the intra-individuals connectivity as well as the inter-individual based on the MDM-BF.

connectivity (summed through the ROIs' mean according to the 10-10 EEG montage) vs. inter-individual connectivity is represented in Figure 8. The causal direction is from the row to the column, that is, the highest ROI activity from violinist A was from the frontal into the sensorimotor, whereas for violinist B, the causal relation from TPJ into sensorimotor was not strong. Moreover, the strongest observed values across inter-brains were from all three ROIs from violinist A to the TPJ from violinist B (left-bottom picture in the third column).

Broadly speaking, these causal relationships are estimated based on the best-adjusted joint probability distribution between the NIRScout (16 LED light sources leading to 23 time series from each violinist) represented as a network. In the best model, first, the partial DAGs obtained can be said to present the intra- and inter-individual connections, and then, the conditional independence of the time series is incorporated according to the assumptions of the model. First, the best network structure for each participant is estimated independently, and then, the hyperscanning network structure is also estimated independently from the others. Nonetheless, the three network dynamics cannot be regarded as totally independent because only thetas can show that (if they are zeros). Moreover, a "partializing relationship" can be observed across structures conditioned to the inter-individual vs. intra-individual as the obtained DAGs.

Figure 9 shows the visual representation of the summation of this antidiagonal block as a graph. For instance, the most influenced regions were sensorimotor and TPJ, as results of the INS, and the results demonstrated that violinist B was influenced by violinist A, as the highest positive value goes from the sensorimotor (violinist A) → TPJ (musician B), and the highest negative value goes from

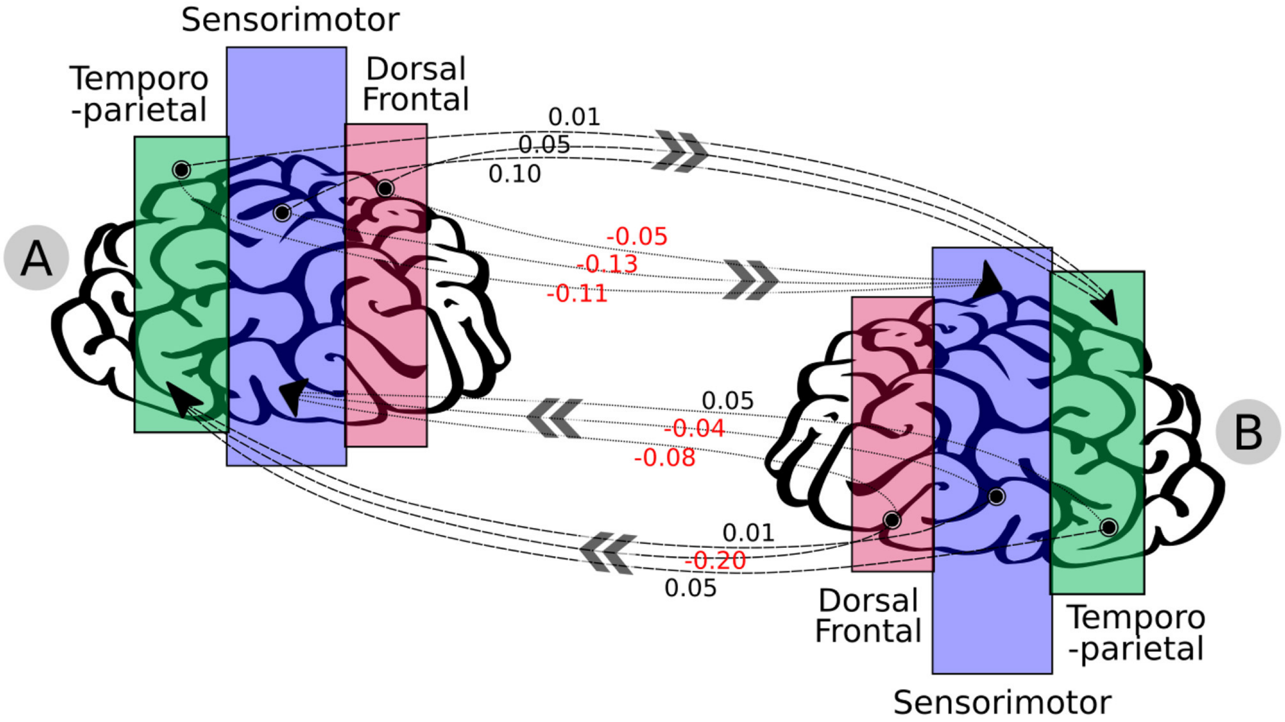


FIGURE 9
Graph-based representation resulting from the musicians' brain-to-brain synchrony through the ROIs: dorsal frontal, sensorimotor, and temporoparietal. Moreover, this graph is the summation of channels that represent the mutual interactions across the musicians' brains (estimated from the MDM's antidiagonal blocks demonstrated in Figure 7). High activity was seen to influence the sensorimotor and temporoparietal for both musicians; nonetheless, no effective correlation of brain-to-brain activity was observed directly toward the dorsal frontal. The numbers in black are positive values, and those in red are negative values. The results highlight that musician B followed musician A, as the highest positive value goes from sensorimotor (musician A) → TPJ (musician B), and the highest negative value goes from dorsal frontal (musician B) → TPJ (musician A), suggesting a reverse causal direction.

dorsal frontal (violinist B) → TPJ (musician A), suggesting a reverse causal direction.

The uncertainty can be associated with confidence intervals (CI) toward this ROI causal connectivity, which was obtained through the non-parametric bootstrap algorithm (Carpenter and Bithell, 2000), using the MDM average of each posterior. We used the nptest package while considering the mean statistic method, a confidence level of 0.95, and the number of replicates of 50,000 (Table 1). A statistical significance was observed from the dorsal frontal channels' behavior (musician A) → TPJ (musician B), sensorimotor channels' behavior (musician A) → sensorimotor (musician B). In the other direction, it was observed from the dorsal frontal channels' behavior (musician B) → sensorimotor (musician A), dorsal frontal (musician B) → TPJ (musician A), and TPJ (musician B) → TPJ (musician A). The other relations were not statistically significant.

Additionally, by using the MDM class, one can make inferences regarding the time-varying strength of the network's links. For instance, the dynamic change among some channels was noticeable, especially during the resting-stage period (delimited by after the red line), as shown in Figure 10. It is clear that the estimated dynamic of the network links was captured by the MDM.

It is clear that intra-subject effective connectivity is stronger than the brain-to-brain coupling strength. Tasks involving music were reported previously and appear to induce brain activation (Li et al., 2015). Berkowitz and Ansari (2010) discussed the importance of the observed brain region (right TPJ, also called rTPJ) in musicians. Luo et al. (2014) showed neuroimaging toward long-term musical training, which shows an impact on emotional and cognitive function, suggesting the presence of neuroplasticity in the rTPJ.

The sensorimotor and TPJ ROIs presented a greater activation influence from the INS; furthermore, the MDM could capture that musician B was following musician A, which also provides some evidence toward the brain synchronization theory. The hypotheses for the ROIs' inter-individual connections relate to a distinct activation, for instance, highlighted in the literature as resulting from the assessment of different body movements (Kimura, 1977)

TABLE 1 Non-parametric bootstrap of the ROIs' mean.

Regions influence	CI 95%	
Frontal_A → TPJ_B	0.00001	0.00581
Sensor_A → Sensor_B	−0.01053	−0.00107
Frontal_B → Sensor_A	−0.00790	−0.00161
Frontal_B → TPJ_A	−0.02136	−0.00311
TPJ_B → TPJ_A	0.00080	0.00422

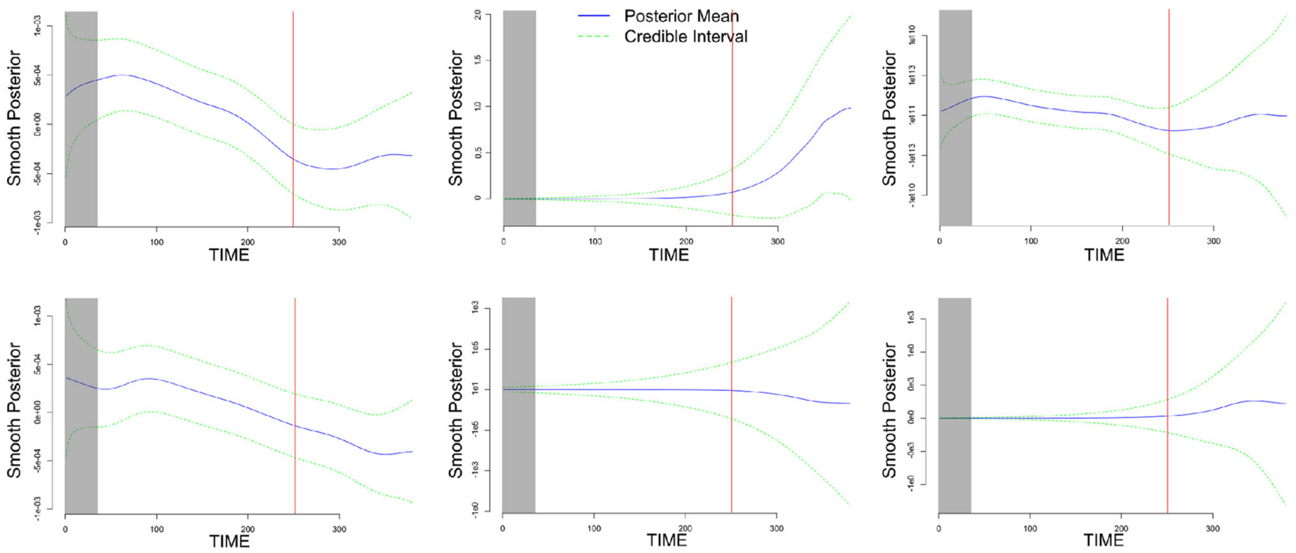


FIGURE 10
An illustration of the dynamics of the six parameters related to the INS inter-activated fNIRS, through their smoothing posterior mean time-varying parameters (solid blue line), with 95% credible interval (dashed green line). Thus, the resting-stage moment is represented by the rectangular-shaded areas and the red-solid line represents the time-point of the end of the music. Based on the two higher causal ROI relations (in module), the three top charts are associated with some sensorimotor channels' behavior (musician A) → TPJ (musician B), and the three bottom charts are associated with some dorsal frontal channels' behavior (musician B) → TPJ (musician A).

or even emotions felt through visual stimuli due to the execution of the activity (Zaitchik et al., 2010).

5. Final remarks

The current study proposes the MDM-BF for fNIRS data obtained in hyperscanning experiments, i.e., simultaneous acquisition, while two or more subjects are interacting. The illustration in a violin duo confirmed the existence of influences of one brain over the other. In the individual brain network analysis for each violinist, it was observed that, although the brain regions work together, some areas play different roles. In other words, some regions connect to others with greater strength. Moreover, this data-driven analysis demonstrated, through their INS estimation, that the influence between violinists is not symmetric and also time-evolving. Therefore, the MDM-BF appears to be a competitive model that is better for hyperscanning studies (due to estimating the effective connectivity) than other methods based on correlation or the consideration of static connections (which only estimate functional connectivity), corroborating similar results that have already been presented in other fields of neuroscience (Costa et al., 2015). In addition, the MDM-BF estimated the inter-brain network using the contemporaneous relationship between regions, without needing to consider the Granger causality.

In the INS analysis [also known as Thinking Through Other Minds (TTOM)], the regions activated on the violinists are represented by the ROI activation and, through the data-driven model, corresponded to the expected results observed in the experimentation (in which musician A was the leader in the duo); that is, the quantification obtained from the MDM-BF brain region connections are highlighted, as shown in Figure 9. However, as this study considered only a pair of violinists, further studies targeting the brain mapping should be conducted to associate the pattern with more in-depth details regarding those connections. In general, the connections estimated by the MDM-BF for the joint matrix of connections represent the brain ROIs' activity correlations and their dynamic over time, in which all regions present positive meaning and strong connections.

Different for DCM, in this study, social brain network structures could be better explored. The study also analyzed the synchronized dynamical system = globally, as well as the communication of specific parts of the brain. Moreover, the novel procedure for the learning structure network that is presented in this study, or others that are used with the MDM (as the MDM-BF or the MDM-DGM) can be easily applied in other scenarios, such as communication and computer-mediated cooperation games. Furthermore, this approach can be suitable for other neuroscience studies that aim to estimate brain networks and have a large number of nodes. A natural next step will be to incorporate informative priors, in which targets transform the researchers' prior knowledge into hyper-parameters. In addition, parametric space shrinkage should be investigated as an alternative to score-based structure selection. In other words, as a complement to the MDM-BF method, the number of time-varying parameter estimations can be reduced based on some a priori information or some specific criteria.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

The authors are grateful for the funding partially provided by the Brazilian agencies CNPq, FAPESP, and CAPES. JRSat is grateful to FAPESP (grants 2018/04654-9 and 2018/21934-5).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The opinions, hypotheses, conclusions, and recommendations of this study are those of the authors and do not necessarily represent the opinions of the funding agencies.

References

Babiloni, F., and Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neurosci. Biobehav. Rev.* 44, 76–93. doi: 10.1016/j.neubiorev.2012.07.006

Baker, J. M., Liu, N., Cui, X., Vrticka, P., Saggar, M., Hosseini, S. H., et al. (2016). Sex differences in neural and behavioral signatures of cooperation revealed by fNIRS hyperscanning. *Sci. Rep.* 6, 26492. doi: 10.1038/srep26492

Balardin, J. B., Zimeo Morais, G. A., Furucho, R. A., Trambaiolli, L., Vanzella, P., Biazoli, C. Jr., et al. (2017). Imaging brain function with functional near-infrared spectroscopy in unconstrained environments. *Front. Hum. Neurosci.* 11, 258. doi: 10.3389/fnhum.2017.00258

Balconi, M., and Angioletti, L. (2023). Inter-brain hemodynamic coherence applied to interoceptive attentiveness in hyperscanning: why social framing matters. *Information* 14, 58. doi: 10.3390/info14020058

Balconi, M., Pezard, L., Nandrino, J.-L., and Vanutelli, M. E. (2017). Two is better than one: the effects of strategic cooperation on intra-and inter-brain connectivity by fNIRS. *PLoS ONE* 12, e0187652. doi: 10.1371/journal.pone.0187652

Berkowitz, A. L., and Ansari, D. (2010). Expertise-related deactivation of the right temporoparietal junction during musical improvisation. *Neuroimage* 49, 712–719. doi: 10.1016/j.neuroimage.2009.08.042

Bilek, E., Zeidman, P., Kirsch, P., Tost, H., Meyer-Lindenberg, A., and Friston, K. (2022). Directed coupling in multi-brain networks underlies generalized synchrony during social exchange. *Neuroimage* 252, 119038. doi: 10.1016/j.neuroimage.2022.119038

Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., and Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *Proc. Nat. Acad. Sci. U. S. A.* 104, 11483–11488. doi: 10.1073/pnas.0606552104

Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y.-Y., Weber, J., and Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proc. Nat. Acad. Sci. U. S. A.* 108, 20254–20259. doi: 10.1073/pnas.1112029108

Burger, B., Infantes, G., Ferrané, I., and Lerasle, F. (2009). "DBN versus hmm for gesture recognition in human-robot interaction," in *International Workshop on Electronics, Control, Modelling, Measurement and Signals (ECMS'09)* (Spain: University of Mondragon), 59–65.

Carpenter, J., and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Stat. Med.* 19, 1141–1164. doi: 10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* 20, 115–125. doi: 10.1038/nn.4450

Chen, L., Qu, Y., Cao, J., Liu, T., Gong, Y., Tian, Z., et al. (2023). The increased inter-brain neural synchronization in prefrontal cortex between simulated patient and acupuncturist during acupuncture stimulation: evidence from functional near-infrared spectroscopy hyperscanning. *Hum. Brain Mapp.* 44, 980–988. doi: 10.1002/hbm.26120

Chen, M., Zhang, T., Zhang, R., Wang, N., Yin, Q., Li, Y., et al. (2020). Neural alignment during face-to-face spontaneous deception: does gender make a difference? *Hum. Brain Mapp.* 41, 4964–4981. doi: 10.1002/hbm.25173

Costa, L., Smith, J., Nichols, T., Cussens, J., Duff, E. P., Makin, T. R., et al. (2015). Searching multiregression dynamic models of resting-state fMRI networks using integer programming. *Bayesian Anal.* 10, 441–478. doi: 10.1214/14-BA913

Cui, X., Bryant, D. M., and Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage* 59, 2430–2437. doi: 10.1016/j.neuroimage.2011.09.003

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Comp. Operat. Res.* 13, 533–549. doi: 10.1016/0305-0548(86)90048-1

Gong, G., Rosa-Neto, P., Carbonell, F., Chen, Z. J., He, Y., and Evans, A. C. (2009). Age-and gender-related differences in the cortical anatomical network. *J. Neurosci.* 29, 15684–15693. doi: 10.1523/JNEUROSCI.2308-09.2009

Hahn, A., Gryglewski, G., Nics, L., Rischka, L., Ganger, S., Sigurdardottir, H., et al. (2018). Task-relevant brain networks identified with simultaneous pet/mr imaging of metabolism and connectivity. *Brain Struct. Funct.* 223, 1369–1378. doi: 10.1007/s00429-017-1558-0

Hasenkamp, W., and Barsalou, L. W. (2012). Effects of meditation experience on functional connectivity of distributed brain networks. *Front. Hum. Neurosci.* 6, 38. doi: 10.3389/fnhum.2012.00038

Horwitz, B. (2003). The elusive concept of brain connectivity. *Neuroimage* 19, 466–470. doi: 10.1016/S1053-8119(03)00112-5

Jiang, L., Stocco, A., Losey, D. M., Abernethy, J. A., Prat, C. S., and Rao, R. P. (2019). Brainnet: a multi-person brain-to-brain interface for direct collaboration between brains. *Sci. Rep.* 9, 1–11. doi: 10.1038/s41598-019-41895-7

Kimura, D. (1977). Acquisition of a motor skill after left-hemisphere damage. *Brain* 100, 527. doi: 10.1093/brain/100.3.527

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6, 215. doi: 10.3389/fnhum.2012.00215

Li, C.-W., Chen, J.-H., and Tsai, C.-G. (2015). Listening to music in a risk-reward context: the roles of the temporoparietal junction and the orbitofrontal/insular cortices in reward-anticipation, reward-gain, and reward-loss. *Brain Res.* 1629, 160–170. doi: 10.1016/j.brainres.2015.10.024

Li, L., Wang, H., Luo, H., Zhang, X., Zhang, R., and Li, X. (2020). Interpersonal neural synchronization during cooperative behavior of basketball players: a fNIRS-based hyperscanning study. *Front. Hum. Neurosci.* 14, 169. doi: 10.3389/fnhum.2020.00169

Li, R., Mayseless, N., Balters, S., and Reiss, A. L. (2021). Dynamic inter-brain synchrony in real-life inter-personal cooperation: a functional near-infrared spectroscopy hyperscanning study. *Neuroimage* 238, 118263. doi: 10.1016/j.neuroimage.2021.118263

Liu, N., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E., and Reiss, A. L. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative jenga game with face-to-face communication. *Front. Hum. Neurosci.* 10, 82. doi: 10.3389/fnhum.2016.00082

Liu, T., and Pelowski, M. (2014). A new research trend in social neuroscience: towards an interactive-brain neuroscience. *PsyCh J.* 3, 177–188. doi: 10.1002/pchj.56

Liu, T., Saito, G., Lin, C., and Saito, H. (2017). Inter-brain network underlying turn-based cooperation and competition: a hyperscanning study using near-infrared spectroscopy. *Sci. Rep.* 7, 1–12. doi: 10.1038/s41598-017-09226-w

Luo, C., Tu, S., Peng, Y., Gao, S., Li, J., Dong, L., et al. (2014). Long-term effects of musical training and functional plasticity in salience system. *Neural Plast.* 2014:13. doi: 10.1155/2014/180138

Morgan, J. K., Santosa, H., Conner, K., Fridley, R., Forbes, E. E., Iyengar, S., et al. (2023). Mother-child neural synchronization is time linked to mother-child positive affective state matching. *Soc. Cogn. Affect. Neurosci.* 18:nsad001. doi: 10.1093/scan/nsad001

Nagarajan, R., Scutari, M., and Lèbre, S. (2013). *Bayesian Networks in R*. New York, NY: Springer, 125–127.

Nascimento, D. C., Pinto-Orellana, M. A., Leite, J. P., Edwards, D. J., Louzada, F., and Santos, T. E. (2020). Brainwave nets: are sparse dynamic models susceptible to brain manipulation experimentation? *Front. Syst. Neurosci.* 14, 527757. doi: 10.3389/fnsys.2020.527757

Nguyen, M., Chang, A., Micciche, E., Meshulam, M., Nastase, S. A., and Hasson, U. (2020). Teacher-student neural coupling during teaching and learning. *bioRxiv*. 40. [preprint] doi: 10.1101/2020.05.07.082958

Oates, C. J., Costa, L., and Nichols, T. E. (2015). Toward a multisubject analysis of neural connectivity. *Neural Comput.* 27, 151–170. doi: 10.1162/NECO_a_00690

Palhano-Fontes, F., Barreto, D., Onias, H., Andrade, K. C., Novaes, M. M., Pessoa, J. A., et al. (2019). Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. *Psychol. Med.* 49, 655–663. doi: 10.1017/S0033291718001356

Pan, Y., Cheng, X., Zhang, Z., Li, X., and Hu, Y. (2017). Cooperation in lovers: an fNIRS-based hyperscanning study. *Hum. Brain Mapp.* 38, 831–841. doi: 10.1002/hbm.23421

Pan, Y., Guyon, C., Borragán, G., Hu, Y., and Peigneux, P. (2021). Interpersonal brain synchronization with instructor compensates for learner's sleep deprivation in interactive learning. *Biochem. Pharmacol.* 191, 114111. doi: 10.1016/j.bcp.2020.114111

Pearl, J. (2009). *Causality: Models, Reasoning & Inference*. Cambridge: Cambridge University Press.

Pinto-Orellana, M. A., Nascimento, D. C., Mirtaheri, P., Jonassen, R., Yazidi, A., and Hammer, H. L. (2020). 1*A Hemodynamic Decomposition Model for Detecting Cognitive Load Using Functional Near-Infrared Spectroscopy. arXiv*. [preprint]. doi: 10.48550/arXiv.2001.08579

Queen, C. M., and Albers, C. J. (2009). Intervention and causality: forecasting traffic flows using a dynamic bayesian network. *J. Am. Stat. Assoc.* 104, 669–681. doi: 10.1198/jasa.2009.0042

Queen, C. M., and Smith, J. Q. (1993). Multiregression dynamic models. *J. R. Stat. Soc. Ser. B* 55, 849–870. doi: 10.1111/j.2517-6161.1993.tb01945.x

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reindl, V., Wass, S., Leong, V., Scharke, W., Wistuba, S., Wirth, C. L., et al. (2022). Multimodal hyperscanning reveals that synchrony of body and mind are distinct in mother-child dyads. *Neuroimage* 251, 118982. doi: 10.1016/j.neuroimage.2022.118982

Scholkmann, F., Holper, L., Wolf, U., and Wolf, M. (2013). A new methodical approach in neuroscience: assessing inter-personal brain coupling using functional near-infrared imaging (fNIRI) hyperscanning. *Front. Hum. Neurosci.* 7, 813. doi: 10.3389/fnhum.2013.00813

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for fMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., et al. (2008). Brain spontaneous functional connectivity and intelligence. *Neuroimage* 41, 1168–1176. doi: 10.1016/j.neuroimage.2008.02.036

Van Den Heuvel, M., Mandl, R., and Pol, H. H. (2008). Normalized cut group clustering of resting-state fMRI data. *PLoS ONE* 3, e2001. doi: 10.1371/journal.pone.0002001

van den Heuvel, M. P., Stam, C. J., Boersma, M., and Pol, H. H. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* 43, 528–539. doi: 10.1016/j.neuroimage.2008.08.010

Wang, J., Wang, L., Zang, Y., Yang, H., Tang, H., Gong, Q., et al. (2009). Parcellation-dependent small-world brain functional networks: a resting-state fMRI study. *Hum. Brain Mapp.* 30, 1511–1523. doi: 10.1002/hbm.20623

Wang, X., Zhang, Y., He, Y., Lu, K., and Hao, N. (2022). Dynamic inter-brain networks correspond with specific communication behaviors: using functional near-infrared spectroscopy hyperscanning during creative and non-creative communication. *Front. Hum. Neurosci.* 16, 907332. doi: 10.3389/fnhum.2022.907332

Wei, Y., Liu, J., Zhang, T., Su, W., Tang, X., Tang, Y., et al. (2023). Reduced interpersonal neural synchronization in right inferior frontal gyrus during social interaction in participants with clinical high risk of psychosis: an fNIRS-based hyperscanning study. *Progr. Neuropsychopharmacol. Biol. Psychiatry* 120, 110634. doi: 10.1016/j.pnpbp.2022.110634

West, M., and Harrison, J. (2006). *Bayesian Forecasting and Dynamic Models*. New York, NY: Springer Science & Business Media.

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., and Hasson, U. (2017). How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb. Cortex* 27, 4988–5000. doi: 10.1093/cercor/bhx202

Zaitchik, D., Walker, C., Miller, S., LaViolette, P., Feczko, E., and Dickerson, B. C. (2010). Mental state attribution and the temporoparietal junction: an fMRI study comparing belief, emotion, and perception. *Neuropsychologia* 48, 2528–2536. doi: 10.1016/j.neuropsychologia.2010.04.031

Zhang, M., Liu, T., Pelowski, M., Jia, H., and Yu, D. (2017). Social risky decision-making reveals gender differences in the TPJ: a hyperscanning study using functional near-infrared spectroscopy. *Brain Cogn.* 119, 54–63. doi: 10.1016/j.bandc.2017.08.008

Zhao, H., Li, Y., Wang, X., Kan, Y., Xu, S., and Duan, H. (2022). Inter-brain neural mechanism underlying turn-based interaction under acute stress in women: a hyperscanning study using functional near-infrared spectroscopy. *Soc. Cogn. Affect. Neurosci.* 17, 850–863. doi: 10.1093/scan/nsac005

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 24 October 2023
DOI 10.3389/fnins.2023.1222749

OPEN ACCESS

EDITED BY
Jacob Raber,
Oregon Health and Science University,
United States

REVIEWED BY
Johan Nakuci,
Georgia Institute of Technology, United States
Jonathan Touryan,
United States Army Research Laboratory,
United States

*CORRESPONDENCE
Ken-ichi Morishige
✉ kmorishi@pu-toyama.ac.jp

†These authors have contributed equally to this work and share first authorship

RECEIVED 15 May 2023
ACCEPTED 03 October 2023
PUBLISHED 24 October 2023

CITATION
Yoshiiwa S, Takano H, Ido K, Kawato M and Morishige K-I (2023) Group analysis and classification of working memory task conditions using electroencephalogram cortical currents during an n-back task.
Front. Neurosci. 17:1222749.
doi: 10.3389/fnins.2023.1222749

Group analysis and classification of working memory task conditions using electroencephalogram cortical currents during an n-back task

Shinnosuke Yoshiiwa[1†], Hironobu Takano[2], Keisuke Ido[3], Mitsuo Kawato[2,4] and Ken-ichi Morishige[2,5*†]

[1]Graduate School of Engineering, Toyama Prefectural University, Imizu, Japan, [2]Department of Intelligent Robotics, Toyama Prefectural University, Imizu, Japan, [3]Center of Liberal Arts and Science, Toyama Prefectural University, Imizu, Japan, [4]Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International, Kyoto, Japan, [5]Neural Information Analysis Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan

Electroencephalographic studies of working memory have demonstrated cortical activity and oscillatory representations without clarifying how the stored information is retained in the brain. To address this gap, we measured scalp electroencephalography data, while participants performed a modified n-back working memory task. We calculated the current intensities from the estimated cortical currents by introducing a statistical map generated using Neurosynth as prior information. Group analysis of the cortical current level revealed that the current amplitudes and power spectra were significantly different between the modified n-back and delayed match-to-sample conditions. Additionally, we classified information on the working memory task conditions using the amplitudes and power spectra of the currents during the encoding and retention periods. Our results indicate that the representation of executive control over memory retention may be mediated through both persistent neural activity and oscillatory representations in the beta and gamma bands over multiple cortical regions that contribute to visual working memory functions.

KEYWORDS

working memory, EEG, hierarchical Bayesian estimation, sparse logistic regression, artifact

1. Introduction

Although the human brain can temporarily store information, such as numbers or strings, it remains unclear how the stored information is retained in the brain (Postle, 2006; D'Esposito and Postle, 2015; Constantinidis and Klingberg, 2016; Chai et al., 2018).

Baddeley's model of working memory consists of one central executive and three subsystems: the phonological loop, the visuospatial sketchpad, and the episodic buffer. The phonological loop stores verbal information and revives auditory memory. A visuospatial sketch pad is a storage system that holds and processes non-verbal information. An episodic buffer is a temporary storage system that integrates visual, spatial, and verbal information with time sequencing. The central executive acts as a supervisory system and controls the flow of information from and to its subsystems, thus focusing on and dividing attention and switching and activating long-term memory to support goal-oriented behavior (Baddeley and Hitch, 1974; Baddeley, 2010).

Human functional magnetic resonance imaging (fMRI) studies have shown that the prefrontal and anterior cingulate cortices play major roles in implementing the concept of working memory (Osaka et al., 2003; Postle, 2015). Electroencephalography (EEG) and magnetoencephalography (MEG) studies have demonstrated that oscillatory activity is related to working memory content and load (Sarnthein et al., 1998; Miltner et al., 1999). Miller et al. proposed a model in which executive control acts via the interplay between gamma network oscillations in superficial cortical layers and alpha and beta oscillations in deep cortical layers (Lundqvist et al., 2016; Miller et al., 2018). However, how Baddeley's psychological model (particularly the representation of the executive control involved in memory retention) is implemented in the nervous system remains an open question.

fMRI has been widely used in working memory studies in humans. This method, which has the advantage of high spatial resolution, can be used to identify brain regions related to working memory and investigate their functional connectivity. However, fMRI cannot acquire high-resolution temporal data due to its measurement principles.

However, EEG and MEG are candidates for recording high-resolution temporal data used for brain activity. EEG/MEG studies on working memory have demonstrated cortical activity and oscillatory representations. However, it is difficult to use the EEG method to acquire high-resolution spatial data because of volume conduction effects and large interelectrode distances. MEG has a significant advantage over EEG because magnetic fields pass through the head without distortion; however, a higher spatial resolution is required. Moreover, a visual stimulus may cause task-related eye movements that induce eye artifacts in the EEG/MEG data. These eye artifacts have some correlation with brain activity, and separating the components of brain activity and artifact components is difficult using conventional statistical methods such as principal component analysis (PCA) or independent component analysis (ICA).

We simultaneously estimated both cortical currents and multiple extra-brain source currents from contaminated EEG/MEG data. Although the measured EEG/MEG data were contaminated by eye artifacts, the proposed method separated the effects of artifacts and estimated the cortical currents of the entire brain using the extra-dipole method (Morishige et al., 2014, 2021). The sparse logistic regression (SLR) method can automatically select, in a data-driven manner, truly important features of working memory calculated from the estimated cortical currents in multiple cortical regions (Yamashita et al., 2008). Furthermore, it can predict the task conditions of the working memory from selected current sources. In this study, by combining the extra-dipole method and SLR, we predicted working memory task conditions from brain regions and investigated the types of information represented in these cortical regions.

Two hypotheses have been proposed to explain the brain mechanisms used for memory retention in working memory, based on the following question: Is it a simple persistent spiking pattern or a periodic pattern of theta, alpha, beta, and gamma bandwidths? If memory retention is achieved by sustained firing patterns of neurons, some differences should exist in the intensity of the estimated current at each dipole. If the function is implemented in periodic patterns, the spectral features of the estimated currents will differ. We examined differences in the magnitudes of the estimated currents in response to different memory loads and found significant differences in the encoding and retention periods. Furthermore, the spectral features of beta and gamma waves were significantly different in several cortical regions.

2. Materials and methods

2.1. Participants

Fourteen adults [11 men and 3 women; aged 21–51 years, mean age = 31.6 ± 12.2 (standard deviation) years] took part in this study. All participants had normal or corrected-to-normal visual acuity. All participants participated in the EEG experiments. Five other participants also participated in the fMRI experiment; however, these data were not included in the study. All experiments were approved by the Ethics Committee of Toyama Prefectural University, the Safety Committee of the Advanced Telecommunications Research Institute International (ATR), and the Ethics Committee of the Hokuriku Health Service Association. All experiments were performed in accordance with approved guidelines and regulations. Written informed consent was obtained from each participant before the experiment.

2.2. EEG data collections

We continuously recorded EEG data using a 64-channel ActiveTwo EEG system (BioSemi, Amsterdam, Netherlands) with electrodes attached to a nylon cap based on the extended 10–20 international system. The participants sat on a comfortable chair 50 cm away from a 24-inch LCD monitor (60-Hz refresh rate) in an electromagnetically shielded room. We recorded an electrooculogram (EOG) from four electrodes located at the left and right temples and above and below the left eye. We recorded a neck electromyogram (EMG) using two electrodes placed on the left sternocleidomastoid muscle. We also recorded finger electromyograms (EMGs) by using two electrodes placed in tandem on the extensor digitorum muscles of the right arm. To verify the timing of the visual stimulus, we measured its onset on the screen using a photodiode. We used either the 2-Button Response Pad (Current Designs, Inc., Philadelphia, PA) or the BSGP815GY GamePad (Buffalo, Inc., Aichi, Japan) as a response box to obtain participants' feedback and measure the response time. However, due to a malfunction of the response box, the response time could not be measured for the two participants.

2.3. Magnetic resonance imaging data collection

T1-weighted structural images were obtained using either a 3 T Siemens Magnetom Prisma Fit scanner (Siemens AG, Erlangen, Germany) or a Vantage Orian 1.5 T Magnetic resonance imagining (MRI) system (Canon Medical Systems, Ohtawara, Japan), with a magnetization-prepared rapid gradient-echo (MPRAGE) sequence. The scanning parameters of the Siemens Magnetom Prisma Fit were as follows: repetition time (TR), 2,300 ms; echo time (TE), 2.98 ms; flip angle, 9°; voxel size, 1 mm; number of slices, 208; matrix size, 256 × 256; and field of view, 256 × 256 mm. Those of the Vantage Orian were as follows: TR, 20 ms; TE, 4.00 ms; flip angle, 15°; voxel size, 0.5 mm; number of slices, 400; matrix size, 512 × 512; and field of view, 256 × 256 mm.

2.4. Task design and procedure

In the original version of the n-back task, figures were presented sequentially on the screen, and participants had to remember these sequences (Kirchner, 1958; WU-Minn HCP Consortium, 2015). This protocol is widely used; however, it poses difficulties for EEG data analysis in isolating brain activity during the encoding and retention periods. In this study, we modified the n-back working memory task. This task comprised three periods (Figure 1A). (a) During the encoding period, the modified 2-back task and delayed match-to-sample (DMTS) task were randomly presented. In the modified 2-back task, seven stimuli chosen from four types of arrows (left, right, up, and down) were presented and replaced sequentially on a monitor. One stimulus was randomly presented as a red arrow. Participants were instructed to memorize the direction of the arrow that appeared two steps before the red arrow. In the DMTS task, the serial presentation of a stimulus was the same as that in the modified 2-back task, except that a single-arrow stimulus chosen from the four types of arrows was used. The same arrow stimulus was presented seven times on a monitor. (b) Information is maintained for 3 s. A random pattern was presented to avoid visual aftereffects [Figure 1A (3)]. (c) During the retrieval period, the participants judged whether the probe arrow direction matched the retained direction by pressing one of the two buttons with their right index or middle finger [Figure 1A (4)]. The participants received visual feedback regarding the correctness of their responses [Figure 1A (5)].

The process comprised a single trial. Each session consisted of 20 trial repetitions, and each task consisted of eight sessions. Each participant performed 160 trials (20 trials × eight sessions). The order of the modified 2-back and DMTS tasks was counterbalanced across participants (left/right/up/down:36 trials; DMTS:16 trials). EEG and fMRI experiments were conducted on different days. The participants followed identical experimental protocols for the EEG and fMRI experiments.

2.5. EEG data analysis

We preprocessed the raw EEG data in the following steps using EEGLAB version 14.1.2 (Delorme and Makeig, 2004) running in MATLAB 2014b. The data were band-pass filtered in the range of 0.4–512 Hz (FIR filter of order 16,897; 0.2 Hz and 512.2 Hz cutoff frequencies (−6 dB); zero-phase) to remove the low-frequency drift components and the high-frequency noise components. Then, we applied a notch filter of 59–61 Hz to remove powerline noise (FIR filter of order 6761; 59.5 and 60.5 Hz cutoff frequencies; zero-phase).

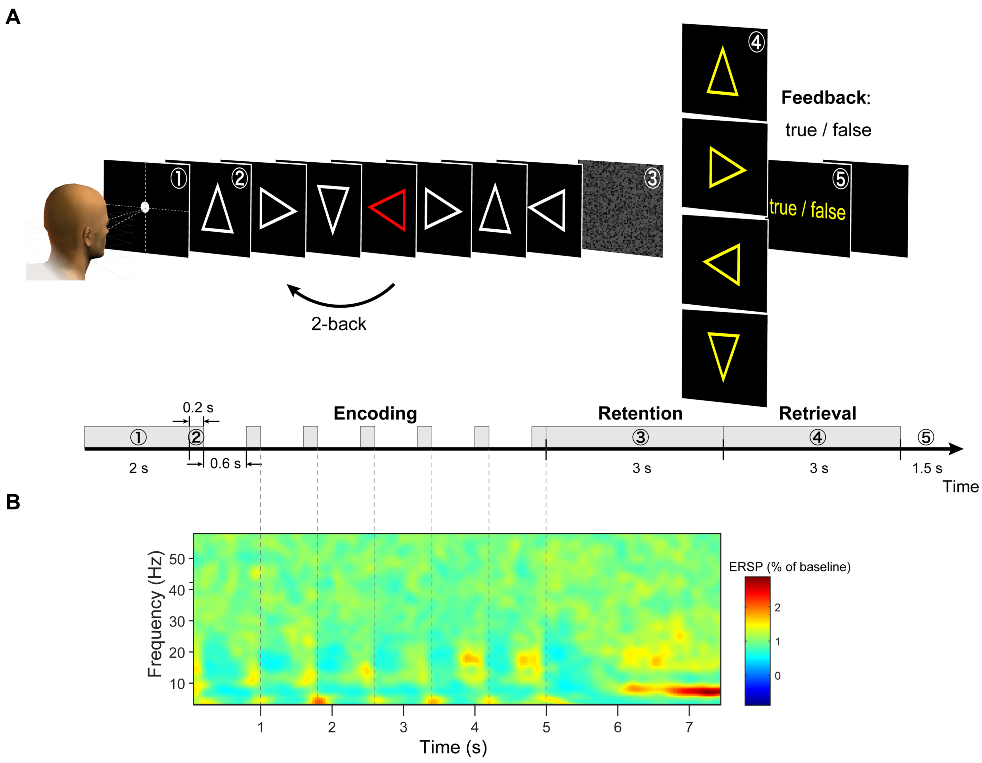


FIGURE 1
(A) Illustration of task design. **(B)** We extracted each trial from −0.5 to 8.0 s and calculated a grand average of the ERSP (event-related spectral perturbation) spectrogram of EEG signals across all channels (using wavelet analysis). During the retention period, the largest periodic change was observed between 6 and 7.5 s.

Next, the EEG data were downsampled from 2,048 to 512 Hz. We extracted single-trial EEG data epochs from −0.5 to 8.0 s with respect to the encoding onset (Figure 1B). After the extraction, we corrected the baselines to the pre-stimulus period (−0.5 to 0 s). During all sessions, noisy channels due to poor electrode contact and broken electrodes were identified by visual inspection and excluded. The data were re-referenced using the average reference (the reference signal was the average of all the electrodes). Signal deviations in the vertical EOG channel of more than 350 µV within the retention period were identified as eyeblinks. Signal deviations in all EEG channels of more than 200 µV within a retention period were identified as large artifacts. Trial data contaminated with eyeblinks and large artifacts were excluded from the analysis. Trials with incorrect responses during the retrieval period were excluded from the analysis. The remaining trials accounted for 79.6% of the total trials (Supplementary Table S1) and were used for the data analysis.

2.6. Meta-analysis fMRI prior

We generated a meta-analysis statistical map synthesized by Neurosynth (Yarkoni et al., 2011)[1] by selecting the term "working memory" to express functional activities during the n-back task. After generation, the statistical map was co-registered to the participant's structural image using the FSL tools FLIRT and FNIRT (Smith et al., 2004; Figure 2A). As the synthesized meta-analysis maps were defined on voxels, they were transformed into cortical surfaces using an inverse-distance weighted interpolation method. An imported map was used to calculate the parameters in the probability distribution of the prior current variances for hierarchical Bayesian estimation according to a previously established method (Suzuki and Yamashita, 2021).

2.7. Head and source models

We constructed a polygon cortical surface model for all participants using the FreeSurfer software (version 6.0.0; http://surfer.nmr.mgh.harvard.edu/; Dale et al., 1999) with a T1-structural image for each participant. The number of cortical surface dipoles in the participants was 10,004. The cortical current sources were located at the vertex points of the cortical surface model, and current sources were oriented perpendicular to the cortical surface. A positive current was defined as the one directed toward the interior of the cortex. The main noise source for the left and right eye movements was assumed to be the center of each eyeball. The position of each eyeball was obtained from the T1-structural images by visual inspection. Each extra-brain source was modeled using the resultant three-dimensional dipole current in the x–y–z direction. Six dipoles (two extra-brain sources × three directions) were located as described in our previous study (Morishige et al., 2014).

We used the three-shell boundary element method (BEM) derived from the MRI dataset (Mosher et al., 1999). The conductivities of the brain, skull, and skin were assumed to be 0.62, 0.03, and 0.62 S/m, respectively.

2.8. Cortical and extra-brain source current estimation

We calculated the cortical and extra-brain source currents using an extra-dipole method (Morishige et al., 2014) based on a hierarchical Bayesian method (Sato et al., 2004; Yoshioka et al., 2008) and simultaneously estimated the cortical and extra-brain source currents by placing the dipoles on both the cortical and extra-brain sources. This method can be used to estimate the cortical currents from EEG data contaminated with extra-brain sources (Figure 2B).

2.9. Group analysis for estimated cortical currents and oscillatory activities

Takeda et al. proposed a group analysis method for the time series of the estimated source currents (Takeda et al., 2019). We applied this method to examine the differences in the amplitudes and power spectra of the source currents estimated from EEG data.

We calculated the time series of trial-averaged source currents and scaled their amplitude, so they had a mean of 0 and a standard deviation of 1 in a baseline period (−0.5 to 0 s). The time series of the trial-averaged source currents was calculated from the normalized source currents for a single retention period. Then, we split the encoding and retention periods into 12 subperiods (0.2–1.0 s, 1.0–1.8 s, 1.8–2.6 s, 2.6–3.4 s, 3.4–4.2 s, 4.2–5.0 s, 5.0–5.5 s, 5.5–6.0 s, 6.0–6.5 s, 6.5–7.0 s, 7.0–7.5 s, and 7.5–8.0 s), and then, we compared all participants' current amplitude in an encoding/retention subperiod between modified n-back and DMTS conditions with a paired *t*-test at each current source. To examine the differences in the spectral features of the two conditions, we estimated the power spectral density using Welch's method for each source current in each trial in a baseline period and an encoding/retention subperiod from the estimated source currents and calculated the sum of power spectral densities in each frequency band of interest: theta (4–8 Hz), alpha (8–13 Hz), low beta (13–20 Hz), high beta (20–30 Hz), and gamma waves (30–50 Hz). We normalized the mean power spectral density of each frequency band using the baseline period values and converted them to a decibel scale using a log base (Cohen, 2014). We compared the normalized mean power spectral densities between the modified n-back and DMTS conditions using a paired t-test at each sampling time. The *p*-values for the paired t-test were corrected for multiple comparisons using Benjamini and Hochberg's false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995). The FDRs were controlled at 0.05.

2.10. Classification

To investigate the representation of working memory in cortical regions, we classified information on the task conditions of working memory using current amplitudes and power spectral densities during the encoding and retention periods. We selected 100 cortical dipole currents in the order of t-values generated by the Neurosynth meta-analysis statistical map and used them for classification. We computed

[1] https://neurosynth.org

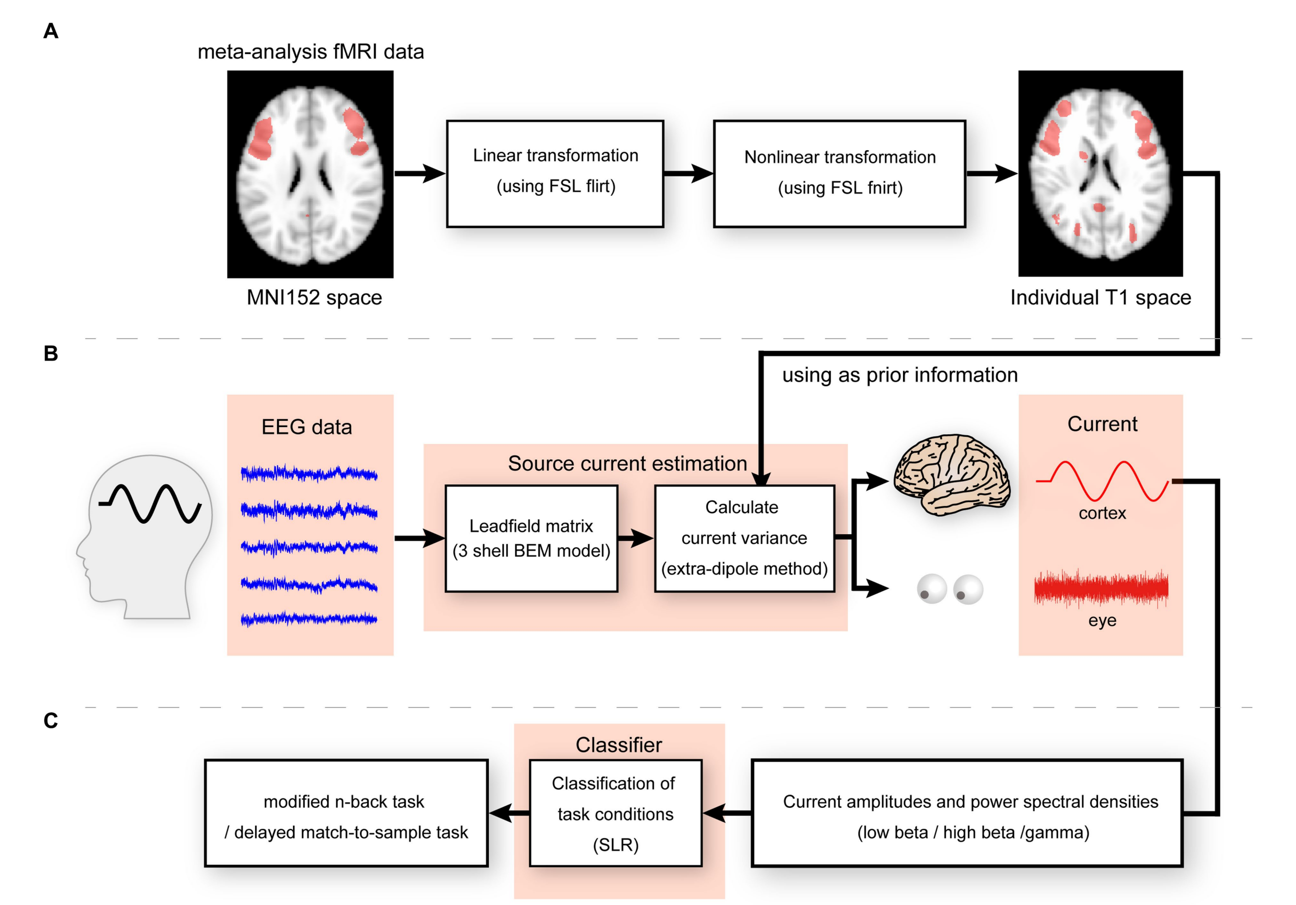


FIGURE 2
(A) Linear and non-linear transformation for meta-analysis fMRI data from MNI152 to individual T1 spaces. **(B)** Illustration of the extra-dipole method. **(C)** Illustration of classification of modified n-back and DMTS task conditions.

the sum of the absolute current amplitudes and power spectral densities of the low beta, high beta, and gamma waves using Welch's method. Sparse logistic regression was used to reduce the input dimensions of the current amplitudes and power spectral densities, which were then divided into two classes (modified n-back or DMTS tasks; Yamashita et al., 2008; Figure 2C) and evaluated using leave-one-out cross-validation. A permutation test was performed by randomizing the labels 100 times to determine whether the performance of the classifiers was statistically meaningful. The one-sided *p*-values of the test were calculated as the proportion of sampled permutations where the differences in means were greater than the test statistic. The accuracy, precision, recall, F-measure, and balanced accuracy were calculated and used for the evaluations. The p-values for the permutation test were corrected for multiple comparisons using Benjamini and Hochberg's false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995). The FDRs were controlled at 0.05.

The ratio of the trial numbers for the modified n-back and DMTS tasks was 144:16, which is a medium-imbalanced dataset. To address the imbalanced data problem in classification, we extended the original SLR and applied the formulation using weighted logistic regression (King and Zeng, 2001; Maalouf and Siddiqi, 2014). The likelihood function of the logistic regression can be rewritten as follows:

$$\boldsymbol{P}\left(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta}\right)=\prod_{\boldsymbol{i}=1}^{\boldsymbol{N}_{\text{input}}}\boldsymbol{\sigma}_{\boldsymbol{i}}^{w_1 y_i}\left(1-\boldsymbol{\sigma}_{\boldsymbol{i}}\right)^{w_0\left(1-y_i\right)},$$

where $\boldsymbol{X}=\left(x_1,\cdots,x_{N_{input}}\right)$ is an input feature vector, $\boldsymbol{\beta}$ is a weight vector including a bias term, $\boldsymbol{y}$ is the outcome vector (either $y_i=1$ or $y_i=0$), $w_1=N_{\text{trial\_all}}/\left(N_{\text{class}}*N_{\text{trial\_DMTS}}\right)$, $w_0=\left(N_{\text{trial\_all}}\right)/\left(N_{\text{class}}*N_{\text{trial\_nback}}\right)$, and $\sigma_i=1/\left(1+\exp\left(-x\right)\right)$.

To improve computational efficiency, we used Z currents as cortical currents to calculate the sum of the current amplitudes and power spectral densities (Morishige et al., 2021).

3. Results

3.1. Behavior

All participants performed both modified 2-back and DMTS tasks with high success rates (mean success rate ± standard deviation, 94.3 ± 3.3% and 98.7 ± 2.7%, respectively). The response times for the two conditions were 0.83 ± 0.22 and 0.79 ± 0.24 s, respectively. There was no significant difference in response time [paired *t*-test: $t(11)=1.6657$, $p=0.1240$]. However, the success rate of the modified 2-back task was significantly lower than that of the DMTS condition [paired t-test: $t(13) = 3.2412$, $p = 0.006$], indicating that the EEG comparison among the different conditions could be influenced by the difficulty of the task.

3.2. Cortical and extra-brain source currents

The cortical current in each participant was estimated using the extra-dipole method. We calculated the trial-averaged values from the estimated current densities and plotted the absolute and maximum values on the cortical surface model. The cortical regions of the dorsolateral prefrontal cortex (DLPFC), posterior parietal cortex (PPC), and early visual areas showed large current intensities. These areas are related to visual working memory processes (Figure 3).

We also searched for the maximum current densities across all dipoles on the cortical surface of each participant and calculated the mean values and standard deviations. The values were 133.2 ± 99.9 pAm/mm^2. In previous electrophysiological studies, the estimated current densities ranged from 25 to 250 pAm/mm^2(Hämäläinen et al., 1993). The values calculated in this study were within these ranges. We also calculated the mean values of the absolute eye currents from single-trial data. These amplitudes ranged from 0.12 to 53.2 nAm, and these estimated values were similar to those of previous research with respect to the order of magnitude (Katila et al., 1981; Morishige et al., 2014).

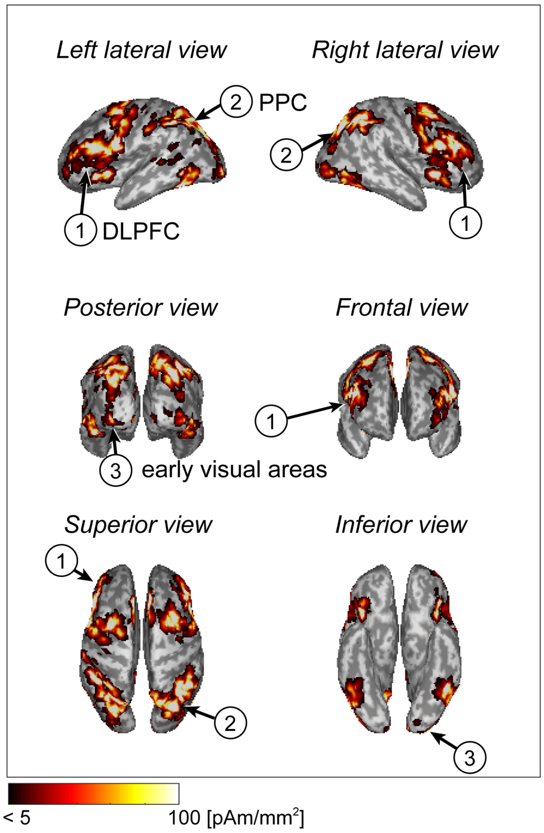


FIGURE 3
Cortical current distribution using a statistical map generated by Neurosynth (example of a typical subject).

If memory retention is achieved through sustained neuronal firing patterns, there should be differences in the intensity of the estimated current at each dipole. However, if the function is implemented in periodic patterns, the spectral features of the estimated currents should differ. We examined whether there were differences in the magnitude of the estimated currents in response to different memory loads and found significant differences in the encoding and retention subperiods ([0.2–1.0 s]: $p = 0.001$, FDR-corrected, paired t-test; [1.0–1.8 s]: $p = 0.002$, FDR-corrected, paired t-test; [1.8–2.6 s]: $p = 0.01$, FDR-corrected, paired t-test; [4.2–5.0 s]: $p = 0.04$, FDR-corrected, paired t-test; [5.0–5.5 s]: $p < 0.0001$, FDR-corrected, paired t-test; [5.5–6.0 s]: $p = 0.004$, FDR-corrected, paired t-test; Figures 4A,B). Additionally, spectral features of beta and gamma waves had significant differences in several cortical regions ([1.8–2.6 s]: (high beta) p = 0.001, (gamma) $p = 0.04$, FDR-corrected, paired t-test; [2.6–3.4 s]: (low beta) $p = 0.02$, (high beta) $p = 0.02$, (gamma) $p = 0.04$, FDR-corrected, paired t-test; [3.4–4.2 s]: (low beta) $p = 0.02$, (high beta) $p = 0.01$, (gamma) $p = 0.02$, FDR-corrected, paired t-test; [4.2–5.0 s]: (high beta) $p = 0.04$, (gamma) $p = 0.03$, FDR-corrected, paired t-test; [5.0–5.5 s]: (high beta) $p = 0.01$, (gamma) $p = 0.03$, FDR-corrected, paired t-test; [6.0–6.5 s]: (gamma) $p < 0.0001$, FDR-corrected, paired t-test; [7.0–7.5 s]: (gamma) $p = 0.04$, FDR-corrected, paired t-test; Figures 4A,B).

3.3. Classification

If the estimated cortical currents contain information about visual working memory, the task conditions must be predicted from the currents or power spectra during the encoding and retention periods. Considering the results of the group analysis in the previous subsection, we investigated the representation of working memory task conditions using the current amplitudes and power spectral densities during each encoding/retention subperiod by computing the sum of the absolute current amplitude in a subperiod and the average power spectral densities in each significant frequency band (low beta, high beta, and gamma waves) using Welch's method. We used weighted sparse logistic regression to reduce the input dimension of the power spectrum densities and classified the trials as modified n-back or DMTS tasks. The classification accuracies in six encoding and six retention subperiods were 84.8 ± 5.1%, 84.0 ± 4.0%, 84.1 ± 5.0%, 85.3 ± 3.5%, 85.8 ± 3.6%, 84.3 ± 3.0%, 84.4 ± 3.7%, 85.1 ± 4.0%, 84.7 ± 5.2%, 82.9 ± 5.3%, 85.9 ± 3.5%, and 84.5 ± 3.7%, respectively (Figure 5A). The precisions were 90.0 ± 1.5%, 89.4 ± 1.4%, 89.5 ± 1.4%, 90.2 ± 1.2%, 90.3 ± 1.4%, 89.7 ± 0.9%, 89.7 ± 1.1%, 89.9 ± 1.0%, 89.4 ± 1.8%, 89.4 ± 2.0%, 90.2 ± 1.0%, and 90.0 ± 1.4%, respectively. The recalls were 93.3 ± 4.9%, 93.1 ± 4.0%, 93.1 ± 5.1%, 93.7 ± 3.5%, 94.3 ± 3.3%, 93.0 ± 3.3%, 93.2 ± 3.8%, 93.8 ± 4.4%, 94.0 ± 4.9%, 91.7 ± 4.8%, 94.6 ± 3.8%, and 93.0 ± 3.3%, respectively. The F-measures were 91.6 ± 3.0%, 91.2 ± 2.4%, 91.2 ± 3.0%, 91.9 ± 2.0%, 92.2 ± 2.1%, 91.3 ± 1.8%, 91.4 ± 2.2%, 91.8 ± 2.4%, 91.6 ± 3.1%, 90.5 ± 3.2%, 92.3 ± 2.1%, and 91.5 ± 2.1%, respectively. The balanced accuracies were 52.0 ± 5.6%, 49.1 ± 3.8%, 49.4 ± 3.6%, 52.8 ± 5.3%, 53.2 ± 6.3%, 50.5 ± 4.1%, 50.3 ± 3.9%, 51.4 ± 4.1%, 49.4 ± 5.2%, 49.2 ± 5.9%, 52.5 ± 5.3%, and 51.5 ± 8.2%, respectively. In total, 72 of all 168

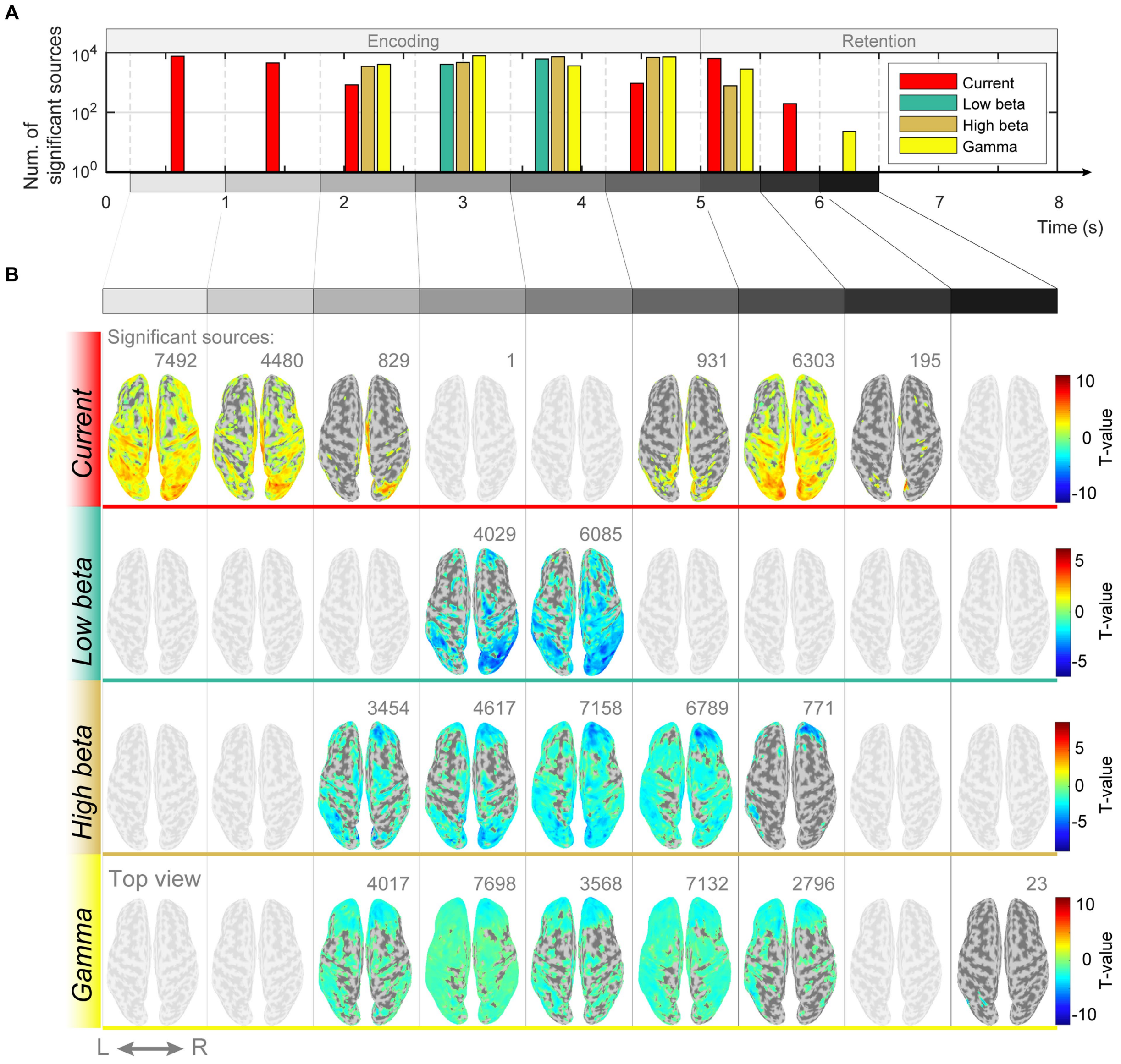


FIGURE 4
Differences in magnitudes of estimated source currents and power spectral densities between modified n-back and DMTS conditions. **(A)** Number of significant current sources for each subperiod of encoding and retention. **(B)** Significant current source locations on the cortical surface map for the subperiods of encoding and retention.

accuracies (= [12 subperiods] × [14 participants]), 64 of 168 precisions, 60 of 168 recalls, 66 of 168 F-measures, and 64 of 168 balanced accuracies reached significance ($p < 0.05$, permutation test, FDR-corrected; Supplementary Tables S2–S6).

To investigate whether the scores differed by time interval, a randomized block design one-way ANOVA was performed. The ANOVA results identified no significant differences among the scores of the subperiods [accuracy: $F(11,143) = 0.81$, $p = 0.63$; precision: $F(11,143) = 1.29$, $p = 0.33$; recall: $F(11,143) = 0.641$, $p = 0.77$; F-measure: $F(11,143) = 0.79$, $p = 0.65$; balanced accuracy: $F(11,143) = 1.23$, $p = 0.36$]. We also used weighted SLR to examine the frequency bands of the features used for identification and found that all types of dipoles, current amplitudes, and low beta, high beta, and gamma waves were selected as discrimination features for each subperiod (Figure 5B).

4. Discussion

In this study, we examined the brain mechanisms underlying executive control over memory retention in working memory to determine whether this was a simple persistent spiking or a periodic pattern. We measured the scalp EEG data, while the participants performed modified n-back working memory tasks and estimated the cortical currents from the EEG data by introducing a statistical map generated by Neurosynth as prior information. A group analysis of the cortical current level revealed that both the current amplitudes and power spectra were significantly different between the modified n-back and DMTS conditions. We classified information on the working memory task conditions using the power spectrum of the currents during the encoding and retention periods. Our results indicate that executive control over memory retention may

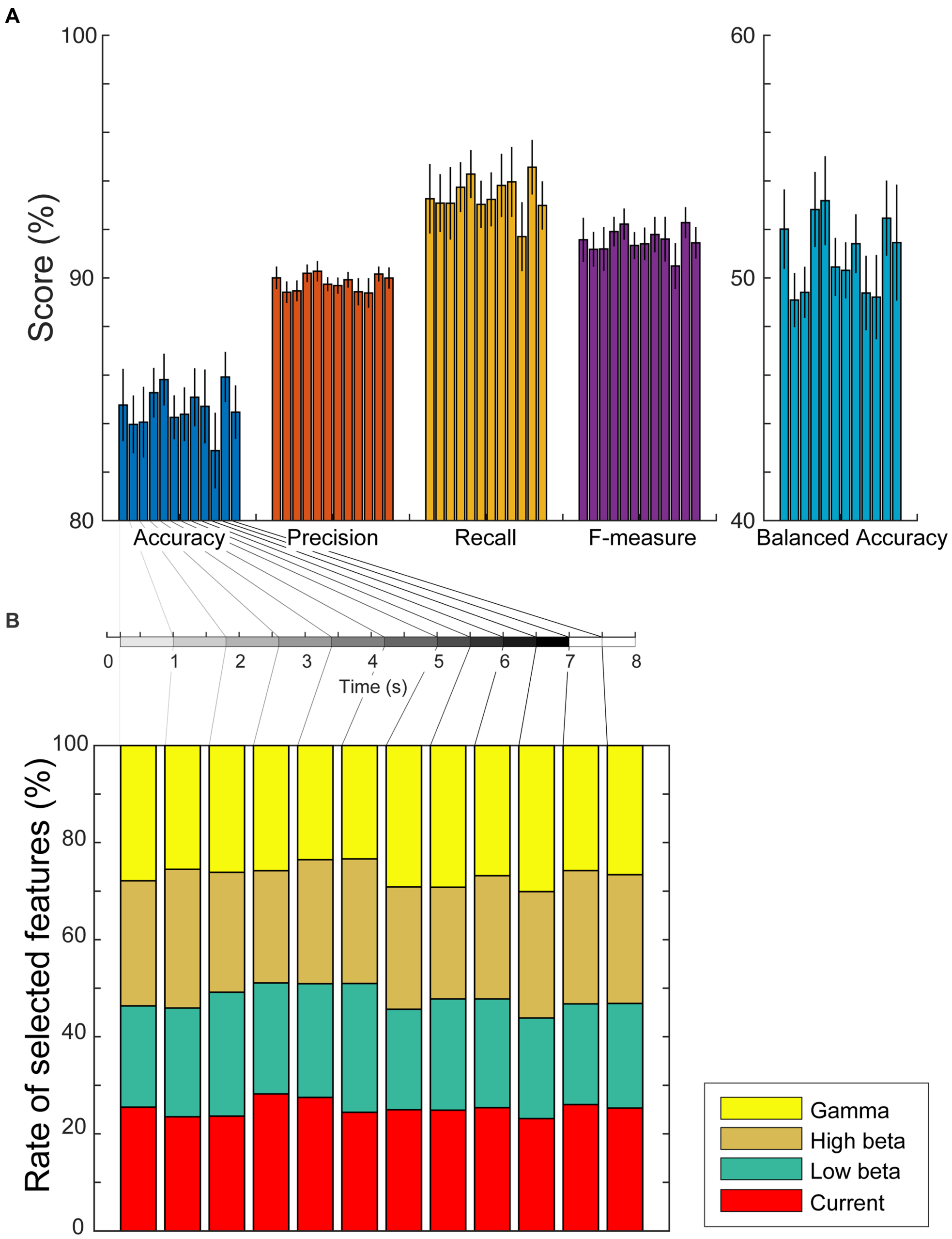


FIGURE 5
(A) Mean values and standard errors of scores (accuracy, precision, recall, F-measure, and balanced accuracy) for each subperiod of encoding and retention using the weighted sparse logistic regression method. **(B)** Ratios of types of selected dipole numbers. We counted the number of times it was selected as a weighted SLR feature for each trial, calculated the mean ratio for each participant, and plotted the average ratios as a stacked bar chart. The rate of selected dipole for currents, low beta, high beta, and gamma waves are shown as red, green, brown, and yellow bars, respectively.

be represented by both current amplitudes and oscillatory representations in the beta and gamma bands over multiple cortical regions that contribute to visual working memory function.

Although group analysis methods are commonly used in the analysis of fMRI data, they have not been previously applied to whole-brain cortical currents estimated from observed data owing to technical difficulties. In this study, using the method by Takeda et al. (2019) in combination with the extra-dipole method (Morishige et al., 2014), eye artifacts can be effectively removed at the current estimation stage and examined using the obtained cortical currents with high temporal–spatial resolution. It is particularly significant that we investigated the changes in brain activities during a short time interval (0.8 s visual cue repetition on encoding period and 3 s retention) of memory encoding and retention by performing a time-frequency analysis with high spatial resolution.

In the original version of n-back task, the overlap between the encoding and retention periods prevented a clear separation of the functional roles of the two for discussion. We revised the experimental paradigm and established separate retentions to allow for a clear separation from the encoding period.

During the retention period, both modified n-back task and the DMTS task required participants to temporarily remember one (or a few) of the stimuli repeatedly presented seven times. When comparing the cortical currents in the modified n-back and DMTS task conditions over the retention period, if there was evidence of behaviors in which working memory was used more strongly during this period under the modified n-back task condition, significant differences in the retention period would be expected; however, there was no evidence of such a behavior. Behavioral performance (success rate) in the modified n-back task condition was lower than that in the DMTS task condition because it only represented the difficulty of encoding. Therefore, the difference between the two groups with respect to working memory should be investigated during the period of encoding rather than retention.

In our experiment, the modified n-back and DMTS tasks were presented randomly without any additional instructions. In the flow of the DMTS task, the same arrows were presented repeatedly. The participant becomes intuitive about the third arrow and is convinced that this is a DMTS task through the presentation of the red stimulus. Therefore, before presenting the first or second stimulus, the participants did not realize that it was a DMTS task or a modified n-back task. In our group analysis, we investigated the differences between the modified n-back and DMTS tasks, so the significant differences in the low/high beta and gamma bands were found in the time intervals from the third subperiod of encoding to the first subperiod of the retention, which were also reasonable results.

Pesonen et al. examined event-related desynchronization (ERD) and event-related synchronization (ERS) responses for targets and non-targets under four different memory load conditions (0-, 1-, 2-, and 3-back) from EEG data (Pesonen et al., 2007). They found that the early-appearing beta rhythm (14–30 Hz) decreased with an increasing memory load. Additionally, the beta rhythms increased under the 0- and 1-back memory load conditions. Our group analysis results correspond to the differences in the power of the beta frequency band calculated from the 2-back and 0-back tasks. Therefore, the finding that beta is significantly negative is consistent with the results of Pesonen et al.

In addition, event-related brain oscillatory responses in the beta frequency range are associated with cognitive processing and motor cortex activity. In the original version of the n-back task, participants were required to respond by pressing a button immediately after the presentation of the visual stimulus. The encoding period of working memory and the period of motor preparation overlap, making it difficult to distinguish between the beta waves originating from both. By contrast, in our modified n-back task, the button was pressed after the retention period. Therefore, the effects of oscillations on motor planning and cognitive memory processes should be discussed separately. Our results suggest that beta oscillations mainly reflect the influence of cognitive and memory processing and that the effect of motor planning is small.

The subperiods with significant differences in gamma oscillations overlapped with those in beta oscillations. It has been hypothesized that gamma and beta oscillations may be synchronized. Lundqvist et al. examined brief bursts of high gamma (50–120 Hz) and high beta (20–35 Hz) oscillations in monkeys (Lundqvist et al., 2018). Beta bursts are associated with suppressing gamma bursts and object information during spiking. Gamma and beta bursting were anti-correlated over time but only at recording sites where spiking carried information about objects to be remembered. The interplay between beta and gamma bursts suggests a potential mechanism for controlling working memory. The relationship between high gamma and high beta oscillations should also be investigated.

Pesonen et al. showed that the magnitude of alpha oscillations decreases with memory load (Krause et al., 2000; Pesonen et al., 2007). However, in this study, no significant differences between the modified n-back and DMTS task conditions were observed in any subperiod of encoding and retention (all subperiods and dipoles of theta and alpha oscillations, $p > 0.05$, FDR-corrected, paired *t*-test). Haegens et al. suggested that alpha oscillations have similar inhibitory roles in sensory-motor areas in DMTS tasks. In general, sensory alpha has been suggested to have inhibitory functions, and it might be that beta has a similar role, but the frequency is shifted upward in the higher-order cortex. Interactions between the mediodorsal thalamus and prefrontal cortex likely produce beta oscillations. Thus, Lundqvist et al. hypothesized that the network between the mediodorsal thalamus and prefrontal cortex might be involved in regulating working memory activity. In contrast, the superficial layers of the prefrontal cortex may contain the contents themselves (Lundqvist et al., 2018). Therefore, there may have been no significant difference between the alpha oscillations of the modified n-back and DMTS conditions in this study. However, there is another possibility that these discrepancies between previous studies and our results may be at least partially explained by different task flows. The original version of the n-back task required a constant memory load because encoding and retention were repeated simultaneously. In contrast, in the modified n-back task used in this study, encoding and retention were sequential and repeated with a short break after each trial. Therefore, the effect of memory load varies among subperiods, and its effect may be relatively small.

The potential increased in the parietal region 300 ms after the visual stimulus presentation. Moreover, it is also known that the potential varies with the magnitude of the memory load (McEvoy, 1998; Segalowitz et al., 2001). The time interval during which the seventh visual stimulus was presented was the time of the greatest memory load in the 2-back task. In the current study, the estimated currents were significantly larger during the time range in which the sixth and seventh visual stimuli were presented, possibly for these reasons. However, the estimated currents were also significantly larger during the encoding subperiods when the first and second visual stimuli were presented. The main reason for this was presumably an imbalance in the number of trials in the modified n-back and DMTS tasks. The modified n-back task had a larger proportion of trials; therefore, the participants tended to expect the modified n-back task to start before each trial began. Because we are not certain if this is the main reason, we should review the observed data to clarify the cause.

Attempts to decode working memory contents have been made by many researchers using various measurement techniques such as neural activities, scalp surface EEG, MEG, and fMRI (Harrison and Tong, 2009; Christophel et al., 2012; Syrjälä et al., 2021). Many studies have reported that periodic components of theta/alpha bandwidths contribute to the representation of memory content and task conditions (Kawasaki et al., 2010; Sauseng et al., 2010; Akiyama et al., 2017) and that beta and gamma bandwidths contribute to their realization (Howard et al., 2003; Lundqvist et al., 2016; Daume et al., 2017; Lundqvist et al., 2018). The ability to classify memory content by using fMRI suggests the presence of specific activity patterns. Although various ways of representing the contents of working memory have been proposed, there are too few methods that discuss them in a unified manner. By combining methods of estimating cortical currents from EEG data and classifying brain information from the estimated currents using the SLR, it is possible to examine brain activity related to working memory with higher temporal and spatial resolutions than that associated with conventional methods. Our results indicate that both persistent neural and oscillatory activities in specific brain regions contribute to the retention of memory task conditions, but both contribute to its realization in a wide range of brain regions.

The time intervals with significant differences varied widely among the participants. Individual differences may be large because of differences in information processing abilities and strategies among participants. Classification may be significant in the subperiods of encoding after the presentation of the first and second visual stimuli. This finding may also be explained by an imbalance in the number of trials required for the modified n-back and DMTS tasks.

In this study, we analyzed the estimated cortical currents only during the encoding and retention periods. However, using our method of analysis, it is also possible to analyze the retrieval periods. Therefore, in future, we would like to clarify how working memory task conditions and their contents are represented not only in the encoding and retention periods but also in the retrieval periods. In addition, we conduct experiments not only on the modified 2-back task but also on the modified 3-back task, which is more difficult. We compute the current amplitudes and power spectra and compare them and classification of correct and incorrect response items in the modified n-back task to confirm that both persistent neural and oscillatory activities are associated with working memory contents and loads.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of Toyama Prefectural University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

HT, KI, MK, and K-iM designed the experiments. SY and K-iM performed the acquisition, wrote the code for data analysis, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Support Center for Advanced Telecommunications Technology Research Foundation (SCAT) and contract with the National Institute of Information and Communications Technology entitled 'Analysis of multi-modal brain measurement data and development of its application for BMI open innovation' (Grant No. 209). K-iM was partially supported by the Toyama First Bank Scholarship Foundation. MK was supported by AMED under grant number JP18dm0307008.

Acknowledgments

We thank Yudai Gamano and Ryota Sakauchi for their collaboration in the early stages of this study. We also thank Yoshiki Nagata, Koji Shima, and Nana Hiroki of the Hokuriku Health Service Association for their technical assistance with MRI experiments during the COVID-19 pandemic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1222749/full#supplementary-material

References

Akiyama, M., Tero, A., Kawasaki, M., Nishiura, Y., and Yamaguchi, Y. (2017). Theta-alpha EEG phase distributions in the frontal area for dissociation of visual and auditory working memory. *Sci. Rep.* 7:42776. doi: 10.1038/srep42776

Baddeley, A. (2010). Working Memory. *Curr. Biol.* 20, R136–R140. doi: 10.1016/j.cub.2009.12.014

Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc. Series B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Chai, W. J., Hamid, A. I. A., and Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: a review. *Front. Psychol.* 9:401. doi: 10.3389/fpsyg.2018.00401

Christophel, T. B., Hebart, M. N., and Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *J. Neurosci.* 32, 12983–12989. doi: 10.1523/JNEUROSCI.0184-12.2012

Cohen, M. X. (2014). Analyzing Neural Time Series Data: theory and Practice. Massachusetts, MA: The MIT Press.

Constantinidis, C., and Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nat. Rev. Neurosci.* 17, 438–449. doi: 10.1038/nrn.2016.43

D'Esposito, M., and Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142. doi: 10.1146/annurev-psych-010814-015031

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. *Neuroimage* 9, 179–194. doi: 10.1006/nimg.1998.0395

Daume, J., Gruber, T., Engel, A. K., and Friese, U. (2017). Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory. *J. Neurosci.* 37, 313–322. doi: 10.1523/JNEUROSCI.2130-16.2016

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 65, 413–497. doi: 10.1103/RevModPhys.65.413

Harrison, S. A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635. doi: 10.1038/nature07832

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cereb. Cortex* 13, 1369–1374. doi: 10.1093/cercor/bhg084

Katila, T., Maniewski, R., Poutanen, T., Varpula, T., and Karp, P. J. (1981). Magnetic fields produced by the human eye (invited). *J. Appl. Phys.* 52, 2565–2571. doi: 10.1063/1.329000

Kawasaki, M., Kitajo, K., and Yamaguchi, Y. (2010). Dynamic links between theta executive functions and alpha storage buffers in auditory and visual working memory. *Eur. J. Neurosci.* 31, 1683–1689. doi: 10.1111/j.1460-9568.2010.07217.x

King, G., and Zeng, L. (2001). Logistic Regression in Rare Events Data. *Polit. Anal.* 9, 137–163. doi: 10.1093/oxfordjournals.pan.a004868

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* 55, 352–358. doi: 10.1037/h0043688

Krause, C. M., Sillanmäki, L., Koivisto, M., Saarela, C., Häggqvist, A., Laine, M., et al. (2000). The effects of memory load on event-related EEG desynchronization and synchronization. *Clin. Neurophysiol.* 111, 2071–2078. doi: 10.1016/S1388-2457(00)00429-6

Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L., and Miller, E. K. (2018). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* 9:394. doi: 10.1038/s41467-017-02791-8

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L. L., Buschman, T. J. J., and Miller, E. K. K. (2016). Gamma and Beta bursts underlie working memory. *Neuron* 90, 152–164. doi: 10.1016/j.neuron.2016.02.028

Maalouf, M., and Siddiqi, M. (2014). Weighted logistic regression for large-scale imbalanced and rare events data. *Knowl. Based Syst.* 59, 142–148. doi: 10.1016/j.knosys.2014.01.012

McEvoy, L. (1998). Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice. *Cereb. Cortex* 8, 563–574. doi: 10.1093/cercor/8.7.563

Miller, E. K., Lundqvist, M., and Bastos, A. M. (2018). Working memory 2.0. *Neuron* 100, 463–475. doi: 10.1016/j.neuron.2018.09.023

Miltner, W. H. R., Braun, C., Arnold, M., Witte, H., and Taub, E. (1999). Coherence of gamma-band EEG activity as a basis for associative learning. *Nature* 397, 434–436. doi: 10.1038/17126

Morishige, K.-i., Hiroe, N., Sato, M.-a., and Kawato, M. (2021). Common cortical areas have different neural mechanisms for covert and overt visual pursuits. *Sci. Rep.* 11:13933. doi: 10.1038/s41598-021-93259-9

Morishige, K.-i., Yoshioka, T., Kawawaki, D., Hiroe, N., Sato, M.-a., and Kawato, M. (2014). Estimation of hyper-parameters for a hierarchical model of combined cortical and extra-brain current sources in the MEG inverse problem. *Neuroimage* 101, 320–336. doi: 10.1016/j.neuroimage.2014.07.010

Mosher, J. C., Leahy, R. M., and Lewis, P. S. (1999). EEG and MEG: forward solutions for inverse methods. *I.E.E.E. Trans. Biomed. Eng.* 46, 245–259. doi: 10.1109/10.748978

Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., et al. (2003). The neural basis of individual differences in working memory capacity: an FMRI study. *Neuroimage* 18, 789–797. doi: 10.1016/S1053-8119(02)00032-0

Pesonen, M., Hämäläinen, H., and Krause, C. M. (2007). Brain oscillatory 4-30 Hz responses during a visual n-Back memory task with varying memory load. *Brain Res.* 1138, 171–177. doi: 10.1016/j.brainres.2006.12.076

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38. doi: 10.1016/j.neuroscience.2005.06.005

Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Curr. Opin. Behav. Sci.* 1, 40–46. doi: 10.1016/j.cobeha.2014.08.004

Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G. L., and Von Stein, A. (1998). Synchronization between prefrontal and posterior association cortex during human working memory. *Neurobiology* 95, 7092–7096. doi: 10.1073/pnas.95.12.7092

Sato, M.-a., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., et al. (2004). Hierarchical Bayesian estimation for MEG inverse problem. *Neuroimage* 23, 806–826. doi: 10.1016/j.neuroimage.2004.06.037

Sauseng, P., Griesmayr, B., Freunberger, R., and Klimesch, W. (2010). Control Mechanisms in Working Memory: A Possible Function of EEG Theta Oscillations. *Neurosci. Biobehav. Rev.* 34, 1015–1022. doi: 10.1016/j.neubiorev.2009.12.006

Segalowitz, S. J., Wintink, A. J., and Cudmore, L. J. (2001). P3 topographical change with task familiarization and task complexity. *Cogn. Brain Res.* 12, 451–457. doi: 10.1016/S0926-6410(01)00082-9

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23:S208. doi: 10.1016/j.neuroimage.2004.07.051

Suzuki, K., and Yamashita, O. (2021). MEG current source reconstruction using a Meta-analysis FMRI prior. *Neuroimage* 236:118034. doi: 10.1016/j.neuroimage.2021.118034

Syrjälä, J., Basti, A., Guidotti, R., Marzetti, L., and Pizzella, V. (2021). Decoding Working memory task condition using magnetoencephalography source level long-range phase coupling patterns. *J. Neural Eng.* 18:016027. doi: 10.1088/1741-2552/abcefe

Takeda, Y., Suzuki, K., Kawato, M., and Yamashita, O. (2019). MEG source imaging and group analysis using VBMEG. *Front. Neurosci.* 13:241. doi: 10.3389/fnins.2019.00241

WU-Minn HCP Consortium. (2015). WU-Minn HCP 900 subjects data release: reference manual. Available at: www.humanconnectome.org/storage/app/media/documentation/s900/HCP_S900_Release_Reference_Manual.pdf.

Yamashita, O., Sato, M. A., Yoshioka, T., Tong, F., and Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of FMRI activity patterns. *Neuroimage* 42, 1414–1429. doi: 10.1016/j.neuroimage.2008.05.050

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635

Yoshioka, T., Toyama, K., Kawato, M., Yamashita, O., Nishina, S., Yamagishi, N., et al. (2008). Evaluation of hierarchical Bayesian method through Retinotopic brain activities reconstruction from FMRI and MEG signals. *Neuroimage* 42, 1397–1413. doi: 10.1016/j.neuroimage.2008.06.013

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 27 October 2023
DOI 10.3389/fnins.2023.1277501

The autism spectrum disorder risk gene *NEXMIF* over-synchronizes hippocampal CA1 network and alters neuronal coding

Rebecca A. Mount[1†], Mohamed Athif[1†], Margaret O'Connor[2], Amith Saligrama[1,3], Hua-an Tseng[1], Sudiksha Sridhar[1], Chengqian Zhou[1], Emma Bortz[1], Erynne San Antonio[1], Mark A. Kramer[4], Heng-Ye Man[2*] and Xue Han[1*]

[1]Department of Biomedical Engineering, Boston University, Boston, MA, United States, [2]Department of Biology, Boston University, Boston, MA, United States, [3]Commonwealth School, Boston, MA, United States, [4]Department of Mathematics, Boston University, Boston, MA, United States

OPEN ACCESS

EDITED BY
Reza Rastmanesh,
American Physical Society, United States

REVIEWED BY
Jason Wester,
The Ohio State University, United States
Le Wang,
Rutgers, The State University of New Jersey, United States

*CORRESPONDENCE
Xue Han
✉ xuehan@bu.edu
Heng-Ye Man
✉ hman@bu.edu

†These authors have contributed equally to this work

RECEIVED 14 August 2023
ACCEPTED 10 October 2023
PUBLISHED 27 October 2023

CITATION
Mount RA, Athif M, O'Connor M, Saligrama A, Tseng H, Sridhar S, Zhou C, Bortz E, San Antonio E, Kramer MA, Man H-Y and Han X (2023) The autism spectrum disorder risk gene *NEXMIF* over-synchronizes hippocampal CA1 network and alters neuronal coding.
Front. Neurosci. 17:1277501.
doi: 10.3389/fnins.2023.1277501

Mutations in autism spectrum disorder (ASD) risk genes disrupt neural network dynamics that ultimately lead to abnormal behavior. To understand how ASD-risk genes influence neural circuit computation during behavior, we analyzed the hippocampal network by performing large-scale cellular calcium imaging from hundreds of individual CA1 neurons simultaneously in transgenic mice with total knockout of the X-linked ASD-risk gene *NEXMIF* (neurite extension and migration factor). As *NEXMIF* knockout in mice led to profound learning and memory deficits, we examined the CA1 network during voluntary locomotion, a fundamental component of spatial memory. We found that *NEXMIF* knockout does not alter the overall excitability of individual neurons but exaggerates movement-related neuronal responses. To quantify network functional connectivity changes, we applied closeness centrality analysis from graph theory to our large-scale calcium imaging datasets, in addition to using the conventional pairwise correlation analysis. Closeness centrality analysis considers both the number of connections and the connection strength between neurons within a network. We found that in wild-type mice the CA1 network desynchronizes during locomotion, consistent with increased network information coding during active behavior. Upon *NEXMIF* knockout, CA1 network is over-synchronized regardless of behavioral state and fails to desynchronize during locomotion, highlighting how perturbations in ASD-implicated genes create abnormal network synchronization that could contribute to ASD-related behaviors.

KEYWORDS

autism spectrum disorder, network analysis, E/I balance, functional connectivity, network closeness centrality, pairwise correlation, GCaMP6f

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects 1 in 36 children (by the age of 8) in the United States (Maenner et al., 2023). ASD is characterized by three core behavioral symptoms: impairments in communication, restrictive and repetitive behaviors, and difficulty with social interactions (American Psychiatric Association, 2013). As one of the most heritable neuropsychiatric disorders, the genetic basis of ASD are widely heterogeneous and often polygenic (Satterstrom et al., 2020). Human genomic studies have identified numerous genes implicated in ASD risk. To understand the contribution of these genes to ASD pathophysiology, transgenic mice (Crawley, 2012; Hulbert and Jiang, 2016) and non-human primates (Zhou et al., 2019) containing such gene disruptions have been developed to model aspects of the behavioral, molecular, and cellular phenotypes seen in individuals with ASD.

Many ASD risk genes are thought to disrupt neural circuit development, leading to elevated network excitability through increasing synaptic-level excitatory/inhibitory (E/I) balance (Gonçalves et al., 2013). While it is unclear how increased synaptic E/I ratio alters network dynamics in vivo, computational modeling has revealed that the E/I balance is critical for maintaining proper asynchrony within a network (Litwin-Kumar et al., 2011) and that an increased E/I ratio elevates neural synchrony (Litwin-Kumar and Doiron, 2012; Middleton et al., 2012). Thus, it has been hypothesized that ASD risk gene mutations over-synchronize neural networks, leading to a reduction in network information encoding that disturbs cognitive performance (Zohary et al., 1994; Cohen and Maunsell, 2009; Rubin et al., 2017). Consistent with this theoretical framework, ASD animal models with an increased E/I balance exhibit increased neuronal correlations, as well as deficits in social interaction (Yizhar et al., 2011; Selimbeyoglu et al., 2017) and sensory discrimination (Chen et al., 2020). While lacking single neuron resolution, EEG variability analysis in humans has allowed the estimation of neural synchrony. As EEG provides an aggregate measure of neural activity-dependent extracellular electrical currents, lower EEG variability is indicative of greater neural synchrony. One study showed that ASD individuals without detectable EEG epileptiform activity exhibited lower EEG variability and higher functional E/I ratios than typically developing children (Bruining et al., 2020). Lower EEG variability is associated with decreased accuracy on a facial recognition task in typically developing children (Mcintosh et al., 2008). Finally, a low-dose ketamine infusion in healthy adults, thought to increase the E/I ratio, creates specific deficits in a spatial working memory task (Murray et al., 2014). Together, these computational and experimental evidence, in both animal models and human subjects, indicate that E/I imbalance and neural synchrony contribute to ASD network pathophysiology which ultimately results in behavioral disruptions.

Mutations in an X-linked gene, *NEXMIF* (neurite extension and migration factor, also known as *KIDLIA*, *KIAA2022*, or *Xpn*) were first discovered in several males with ASD, intellectual disability, and other co-morbidities (Cantagrel et al., 2004; Van Maldergem et al., 2013). Since then, several studies have reported additional ASD individuals with mutations or deletions in the *NEXMIF* gene (Lim et al., 2013; Iossifov et al., 2014; Charzewska et al., 2015; Kuroda et al., 2015; De Lange et al., 2016; Farach and Northrup, 2016; Webster et al., 2017; Yuen et al., 2017; Lambert et al., 2018; Lorenzo et al., 2018; Panda et al., 2020; Stamberger et al., 2020; Wang et al., 2020). *NEXMIF* is now recognized as a Category 1 gene in the Simons Foundation Autism Research Initiative (SFARI) database, further implicating it as an ASD-risk gene. NEXMIF protein is expressed exclusively in neuronal nuclei and loss of *NEXMIF* expression leads to aberrant neuronal migration and reduced dendritic growth due to a dysregulation in actin dynamics in neurite tips (Gilbert and Man, 2016). Thus, *NEXMIF* is critical for proper dendritic extension and neuronal migration in the developing mouse cortex (Gilbert and Man, 2016). Additionally, *NEXMIF* knockdown results in a significant loss of synapses with a twofold greater loss of GABAergic synapses compared to glutamatergic synapses in cultured neurons (Gilbert et al., 2020), suggesting an increased synaptic E/I balance. *NEXMIF* knockout (NEXMIF KO) mice demonstrate a variety of behavioral deficits, most notably reduced social interaction, impaired communication vocalizations, and increased self-grooming (indicative of repetitive behavior).

We analyzed the publicly available atlas of gene expression in adult mice available from the Allen Brain Institute, and found that NEXMIF expression is the highest in the hippocampus (Allen Institute for Brain Science, 2004a; Hawrylycz et al., 2007). As hippocampal structure (Dager et al., 2007; Groen et al., 2010; Chaddad et al., 2017; English et al., 2017; Reinhardt et al., 2020) and function (Just et al., 2007; Green et al., 2013; Gu et al., 2015; Krach et al., 2015) are often disrupted in individuals with ASD, we examined the hippocampal network in NEXMIF KO mice to understand how ASD-implicated *NEXMIF* gene mutations alter hippocampal function at both the cellular and network levels. Because *NEXMIF* KO leads to profound learning and memory deficits (Gilbert et al., 2020), it is extremely difficult to train these animals on hippocampal-dependent learning and memory tasks. Thus, we examined how *NEXMIF* KO altered CA1 cellular dynamics and network connectivity patterns during locomotion, an important aspect of spatial memory, by performing cellular calcium imaging from tens to hundreds of individual CA1 neurons simultaneously in NEXMIF KO male mice and wild-type (WT) male littermates during locomotion. We found that KO of *NEXMIF* did not alter calcium event shape and frequency in individual neurons but increased behaviorally specific neuronal responses during locomotion. We then characterized network effects of *NEXMIF* KO using Pearson correlation and network closeness centrality and discovered that loss of *NEXMIF* creates over-synchronization of the CA1 network during locomotion.

Results

NEXMIF WT and KO mice exhibit similar locomotor behavior

Because of the various behavioral deficits observed in adult NEXMIF KO mice (Gilbert et al., 2020), we first examined NEXMIF expression profiles by analyzing the mouse cortex and hippocampus RNA-Seq data from the Allen Brain Institute's Cell Types Database (Allen Institute for Brain Science, 2004b;

Yao et al., 2021) and the RNA In-Situ Hybridization data from the Allen Brain Institute's Mouse Brain Atlas (Allen Institute for Brain Science, 2004a; Hawrylycz et al., 2007). Interestingly, we found that NEXMIF expression is most prominent in the hippocampus (**Figures 1B, C**) without obvious difference between excitatory versus inhibitory neurons (**Figure 1A**), consistent with the observation that NEXMIF KO mice exhibit severe learning and memory deficits (Allen Institute for Brain Science, 2004a; Gilbert et al., 2020). To understand how *NEXMIF* contributes to hippocampal circuit functions, we then characterized CA1 neural responses using calcium imaging while mice were head-fixed and navigating freely on a spherical treadmill (**Figure 1D**). Since it is difficult for NEXMIF KO mice to perform hippocampal-dependent learning and memory tasks as observed in our previous study (Gilbert et al., 2020), we examined how *NEXMIF* KO changes hippocampal circuity during locomotion, a fundamental component of spatial memory.

We performed wide-field calcium imaging from hundreds of individual dorsal CA1 neurons simultaneously in both WT and KO animals during voluntary locomotion, comparing homozygous NEXMIF KO male mice with complete deletion of *NEXMIF* and their WT male littermates. Since *NEXMIF* is an X-linked gene and NEXMIF KO male mice are infertile, homogenous female mice cannot be generated. Thus, we used only male KO mice that have a complete deletion of *NEXMIF*. Briefly, we first injected AAV-Synapsin-GCaMP6f into the CA1 to label neurons specifically, and then surgically removed the overlying cortex and implanted an imaging window above CA1. The imaging field of view was centered on the stratum pyramidale, about 100 μm below the imaging window, though it is possible some interneurons in the stratum oriens were in the field of view as well. Each mouse was recorded for 10 minutes per day every other day over a 5-day period (**Figure 1E**). We did not detect noticeable differences across the three calcium imaging sessions from the same mouse, and thus all recording sessions from each mouse were grouped for further analysis.

We first examined voluntary movement kinematics between KO mice (n = 8 mice) and WT littermates (n = 7 mice). "Resting" and "running" bouts were identified based on movement speed (details in section "Materials and methods," **Figures 1F, G**) simultaneously recorded with each imaging session. WT and KO mice exhibited a similar number of running bouts (periods of continuous running) within each 10-minutes session (**Figure 1H**), with similar running bout duration (**Figure 1I**) and speed (**Figure 1J**), and overall speed across the entire session (**Figure 1K**). Furthermore, these movement kinematic measures were not correlated with the age of the mice in either WT or KO groups (**Supplementary Figure 1**). Thus, *NEXMIF* KO does not alter overall movement kinematics in our experimental setting, allowing us to examine NEXMIF-induced changes in neuronal responses independent of behavioral alterations.

Calcium event shape and frequency are undisturbed in NEXMIF KO mice

We next examined calcium events across individual neurons recorded in WT versus KO mice. The recorded calcium fluorescence videos were first motion corrected and individual cells were segmented (Shen et al., 2018; **Figure 1L**). A GCaMP6 fluorescence trace was then extracted for each cell and normalized to its peak fluorescence to account for variation in GCaMP6f expression between neurons (**Figures 1M, N**). We then identified individual calcium events as described previously (Zemel et al., 2022) (see section "Materials and methods," **Figures 2A–D**). The identified calcium events occurred at a rate of 2.24 ± 0.50 events/min over the entire imaging session, similar to previous CA1 recordings using GCaMP6 (Mount et al., 2021), and there was no difference between WT and KO mice (WT: 2.10 ± 0.35 events/min, mean ± standard deviation (SD), n = 18 sessions from 7 mice; KO: 2.32 ± 0.53 events/min, n = 24 sessions from 8 mice, Wilcoxon rank sum test, p = 0.12). Thus, GCaMP6 calcium imaging is capable of capturing neural activity dynamics in both mice groups.

We estimated neural activity using both the rise time and the frequency of individual calcium events, as the rising phase of calcium events captures the sharp increases in intracellular calcium that are common during spike bursts (Huang et al., 2021). We found that calcium event rise time was longer during running than resting in both WT and KO mice, but there was no difference between KO and WT during either behavioral condition (**Figure 2F**). Additionally, we calculated full width at half-maximum amplitude (FWHM) to estimate calcium buffering capacity, as the duration of a calcium event captures overall intracellular calcium change (McMahon and Jackson, 2018; Huang et al., 2021). FWHM was similar regardless of behavioral condition or genotype (**Figure 2G**). Thus, *NEXMIF* KO does not affect the overall activity or calcium buffering capacity of CA1 neurons.

NEXMIF KO increases the fraction of movement-modulated neurons

Since CA1 neurons are known to increase their activity during locomotion (Vanderwolf, 1969; Fuhrmann et al., 2015), we next compared calcium event rates during resting versus running. We found that calcium event rate across the entire population increased from resting to running in both WT and KO animals, but there was no difference between WT and KO (**Figure 3A**). After observing this population-level change in neural activity during locomotion, we next evaluated how individual CA1 neurons are modulated by movement. To determine whether a neuron is modulated by movement, we binarized the GCaMP6f dF/F trace (**Figures 3C, D, J, K**) to calcium event trace with ones assigned to the entire rising phase of each calcium event and zeros everywhere else (**Figures 3F, G, M, N**). We then computed the difference in event density during running versus resting and compared it to a shuffled null distribution. In each shuffle, we circularly shifted each binarized calcium trace by a random temporal offset relative to movement and calculated the difference in activity between the running periods and resting periods (**Figures 3I, P**). This procedure was repeated 1,000 times to form the null distribution. A cell was deemed to be movement-modulated if the observed neural activity difference was greater than the 97.5th percentile of the shuffled null distribution for that cell. Using this analysis, we found that 31.0% of neurons were movement-modulated in KO animals, significantly higher than the 25.7% observed in WT (**Figure 3B**). As

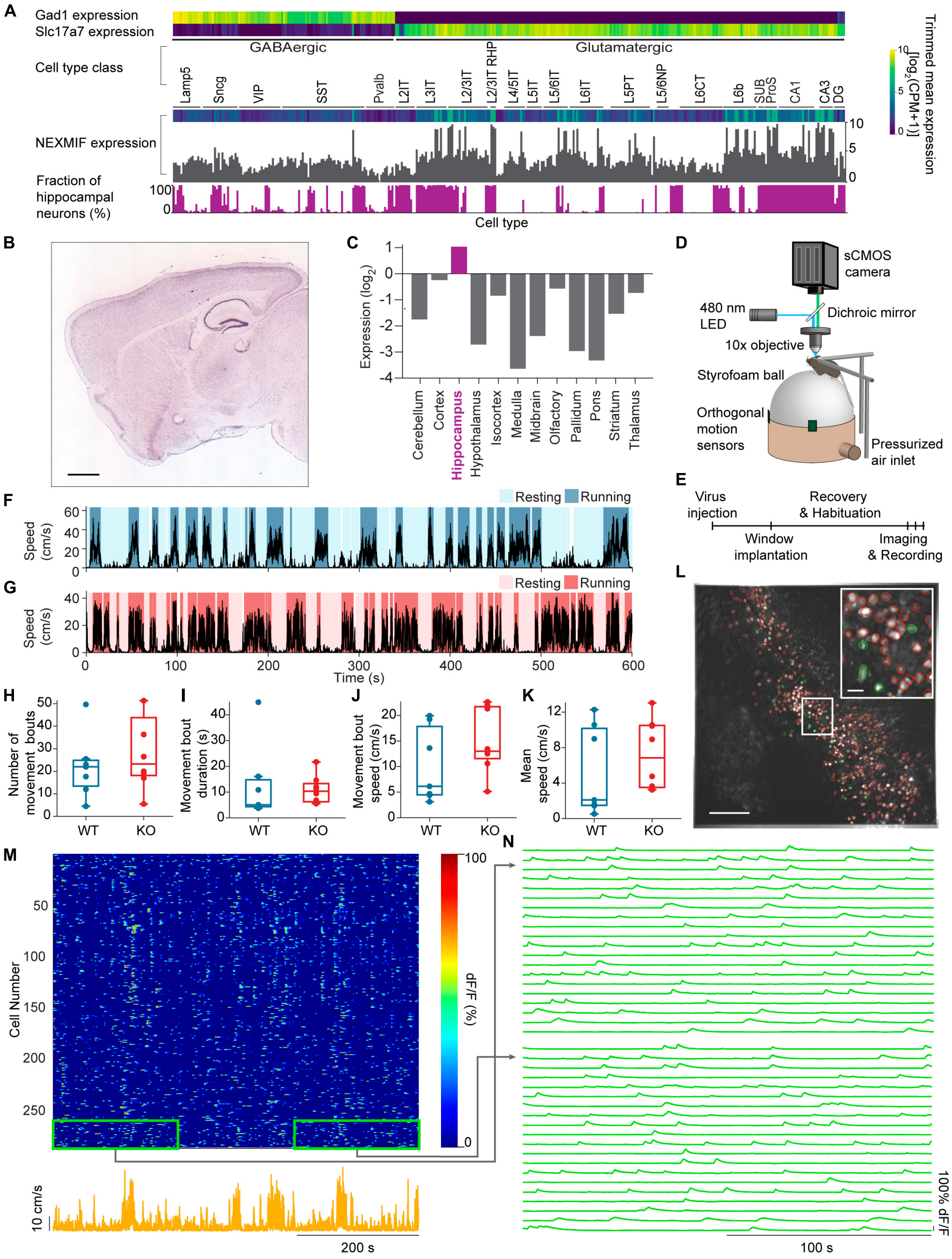


FIGURE 1

Experimental set-up and movement behavior. **(A)** Mouse cortex and hippocampus single cell RNA-Seq data from the Allen Brain Institute's Cell Types Database showing expression levels for the GABAergic marker gene Gad1, the glutamatergic marker gene Slc17a7 and the *NEXMIF* gene in each transcriptomic cell type. Bottom, the fraction of hippocampal neurons among the total sequenced cells in each transcriptomic cell type. **(B)** RNA In-Situ Hybridization data showing *NEXMIF* expression in a sagittal slice including the hippocampus, from the Allen Brain Institute's Mouse Brain Atlas, https://mouse.brain-map.org/experiment/show?id=69531127. Scale bar: 839 μm. **(C)** Quantification of relative expression levels of *NEXMIF* in panel **(B)**. **(D)** The experimental setup illustrating a mouse head-fixed under a custom wide-field microscope, voluntarily navigating a spherical treadmill. **(E)** Experimental timeline. Animal's movement speed during an example experimental session in a WT animal **(F)** and a KO animal **(G)**. Identified resting (light blue and light pink) and running (dark blue and dark pink) bouts are overlaid on the movement speed traces. **(H)** Average number of movement bouts per 10-min session (WT: 22.0 ± 14.2 bouts, mean ± SD, n = 7 WT mice; KO: 28.4 ± 16.6 bouts, n = 8 KO mice, Wilcoxon rank sum test p = 0.44). **(I)** Average movement bouts duration (WT: 12.8 ± 14.9 s, KO: 10.9 ± 5.43 s, Wilcoxon rank sum test, p = 0.34). **(J)** Mean speed during movement bouts (WT: 10.16 ± 7.3 cm/s, KO: 15.04 ± 6.33 cm/s, Wilcoxon rank sum test, p = 0.19). **(K)** Average speed over the entire imaging session (WT: 5.36 ± 5.0 cm/s, KO: 7.25 ± 3.92 cm/s, Wilcoxon rank sum test, p = 0.34). Example maximum-minus-minimum projection fluorescence image across the entire recording session. All selected ROIs are outlined in red, with highlighted cells in panel **(L)** shown in green. Scale bar: 200 μm. Inset: zoom-in of white box. Scale bar: 40 μm. **(M)** Heat map of GCaMP6f dF/F traces for the ROIs shown in panel **(L)** during an example session (top) and animal's corresponding movement speed (bottom). **(N)** Zoom-ins of the heat map regions outlined in green in panel **(M)**, showing the fluorescence traces for 20 representative cells from the beginning and the end of the imaging session. In panels **(H–K)**, each dot corresponds to an individual session (box: interquartile range, whiskers: 1.5 ± interquartile range, middle line: median).

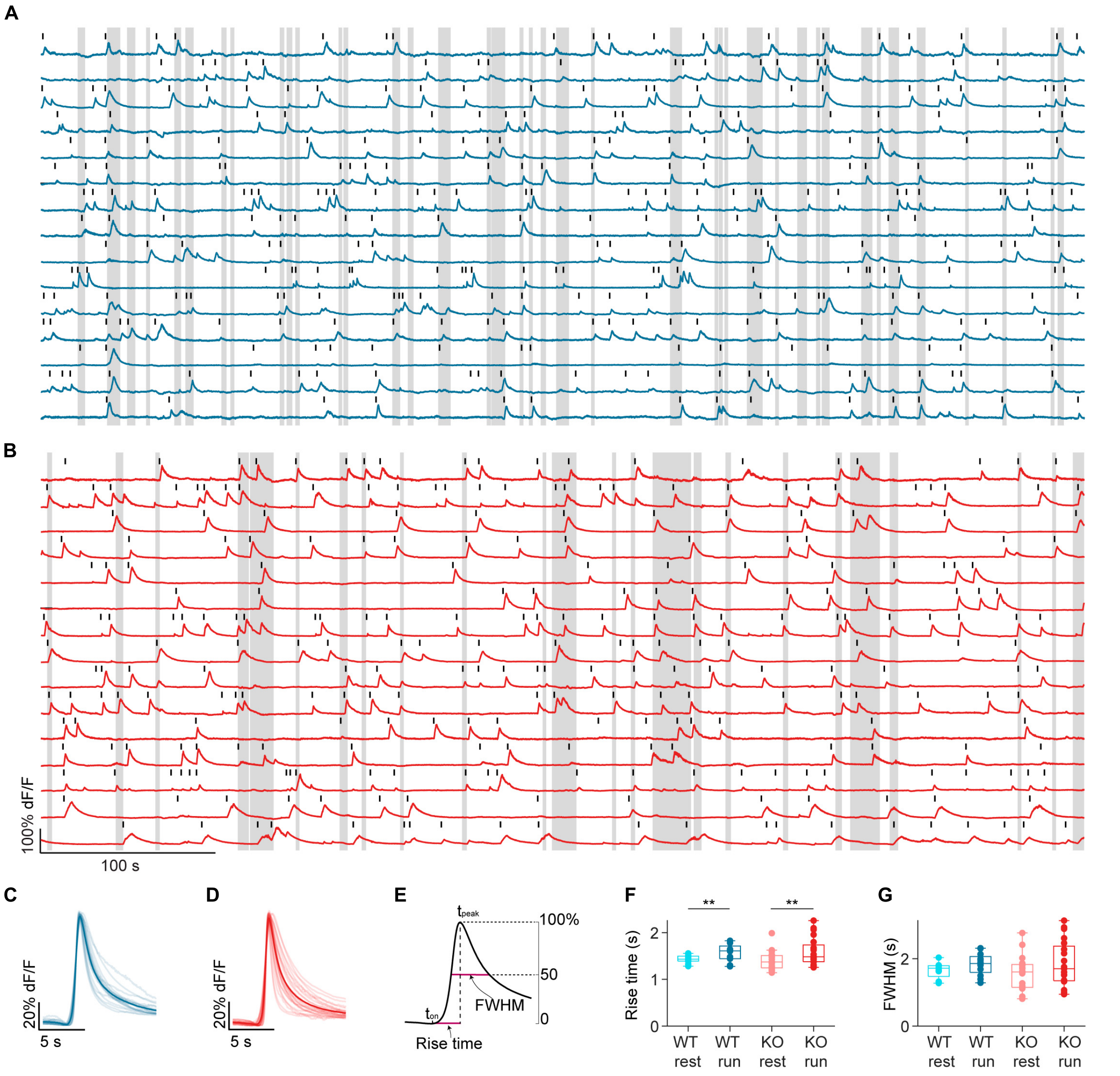


FIGURE 2
NEXMIF knockout does not change calcium event shape. Example fluorescence traces from **(A)** a WT animal and **(B)** a KO animal, with movement bouts shown as gray shading. Fifteen cells are shown from each mouse. Each detected calcium event is marked with a black line. Average calcium event shape from WT sessions (**C**, dark blue) and KO sessions (**D**, dark red). Events were first averaged within each cell, and then averaged across all cells in a session. Each session average is shown as a light line, and the population average is shown as the solid line. **(E)** Schematic calcium event rise time and full width at half-maximum (FWHM) calculation. **(F)** Mean calcium event rise time in WT mice during rest (light blue) and run (dark blue), and in KO during rest (light red) and run (dark red). (WT rest: 1.43 ± 0.08 s, mean ± SD, n = 12 sessions from 6 mice, WT run: 1.57 ± 0.19 s, KO rest: 1.40 ± 0.20 s, n = 20 sessions from 8 mice; KO run: 1.59 ± 0.30 s, Linear Model, behavioral condition: p = 0.008, WT/KO genotype: p = 0.71, interaction: p = 0.66.) **(G)** Mean FWHM in WT mice during rest (light blue) and run (dark blue), and in KO during rest (light red) and run (dark red). (WT rest: 1.65 ± 0.23 s, WT run: 1.84 ± 0.33 s, KO rest 1.56 ± 0.51 s, KO run: 1.86 ± 0.33 s, Linear Model, significance against intercept-only model: p = 0.21.) In panels **(F,G)**, each dot corresponds to an individual session (box: interquartile range, whiskers: 1.5 × interquartile range, middle line: median). $^{**}p$ < 0.01.

expected, the movement-modulated cell population increased total dF/F during running, whereas the non-modulated cell population showed no difference between behavioral conditions (Figures 3E, L). Accordingly, event rate increased during running in the movement-modulated cell population, but did not change in non-modulated cells (Figures 3H, O). The percentage of cells that were movement-modulated in each session did not depend on the time the animal spent running or the animal's average speed during the session for either mouse group or behavioral condition (Supplementary Figure 2). This increase in the proportion of movement-modulated cells in KO mice suggests that *NEXMIF* KO increases behavioral responses of the CA1 circuit. As *NEXMIF* KO increases E/I synaptic ratio of individual cells (Gilbert et al., 2020), our results support the hypothesis that increased synaptic level E/I ratio by ASD risk gene mutation increases behaviorally evoked network responses, consistent with the observation that sensory

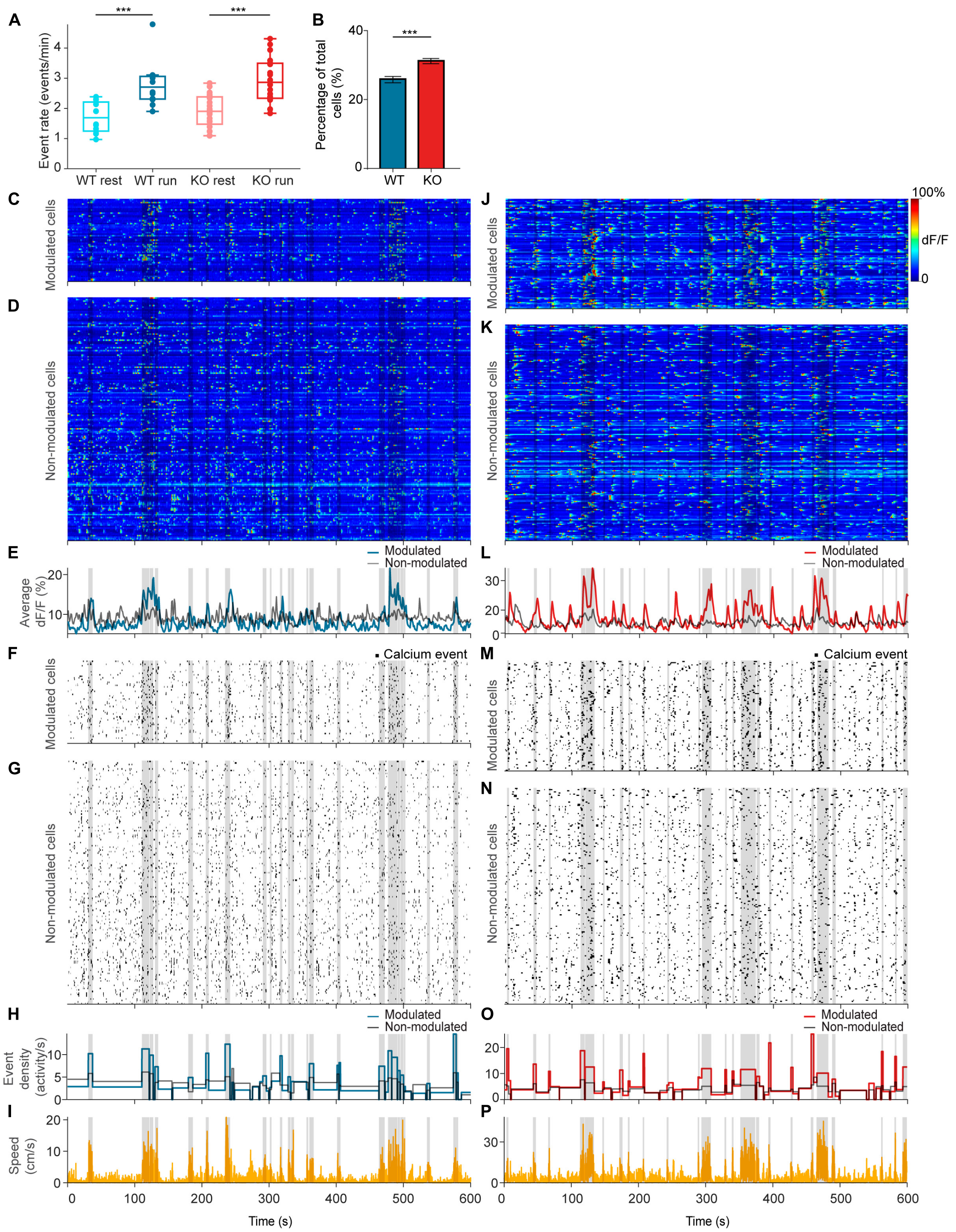


FIGURE 3

NEXMIF knockout increases fraction of movement-modulated cells. **(A)** Calcium event rate in WT mice during rest (light blue) versus run (dark blue), and in KO during rest (light red) versus run (dark red). (WT rest: 1.71 ± 0.51 events/min, mean ± SD, *n* = 12 sessions in 6 mice; WT run: 2.79 ± 0.75 events/min; KO rest: 1.90 ± 0.54 events/min, *n* = 20 sessions in 8 mice; KO run: 2.93 ± 0.75 events/min, Linear Model, behavioral condition: $p = 4.91 \times 10^{-6}$, WT/KO genotype: $p = 0.41$, interaction: $p = 0.86$.) Each dot corresponds to an individual session (box: interquartile range, whiskers: 1.5 × interquartile range, middle line: median). **(B)** Fraction of all neurons that are movement modulated in WT (blue) versus KO (red) mice (WT: 25.7 ± 2.01%, proportion ± 95% confidence interval, *n* = 1,805 cells from 6 mice; KO: 30.1 ± 1.8%, *n* = 2,530 cells from 8 mice, Fisher's exact test, $p = 1.6 \times 10^{-4}$). Example sessions from a **(C–I)** WT animal and a **(J–P)** KO animal. **(C,J)** Heat map of GCaMP6f dF/F traces for movement-modulated cells and **(D,K)** non-movement-modulated cells in WT and KO mice. Average dF/F across movement-modulated cells in WT (**E**, blue) and KO (**L**, red), and non-movement-modulated cells (**E,L**, black). **(F,M)** Binarized calcium traces for all movement-modulated cells and **(G,N)** all non-movement-modulated cells in the example sessions. Average calcium event density across all movement-modulated cells in WT (**H**, blue) and KO (**O**, red), and non-movement-modulated cells (**H,O**, black). **(I,P)** Corresponding movement speed (orange) for the session. All plots are overlaid with movement bouts in gray. ***$p < 0.001$.

stimuli lead to an over-activation of the hippocampus in individuals with ASD (Green et al., 2013).

NEXMIF KO increases functional connectivity between neuron pairs during running

Computational studies have shown that increased synaptic E/I ratio increases network synchrony measured as population pairwise correlations, thus decreasing network information coding capability (Zohary et al., 1994; Litwin-Kumar and Doiron, 2012; Middleton et al., 2012). Additionally, several animal models with deletions of ASD risk genes exhibit increased neuronal correlations (Selimbeyoglu et al., 2017; Chen et al., 2020). As *NEXMIF* KO increases E/I ratio, like many other ASD risk gene mutations, we next examined whether *NEXMIF* KO influences CA1 network synchrony by calculating Pearson correlation between the binarized traces of simultaneously recorded neuron pairs (**Figures 4A, B, G, H**). The binarized traces include only the rising phase of calcium events to avoid overestimation of correlation due to the slow decay kinetics of GCaMP6f. To account for variations in event rate, we determined whether the measured correlation between each neuron pair was significantly greater than chance observation given the event rates of the neurons in the pair. To estimate chance observations, we shifted the binarized traces of two neurons relative to one another with a random time lag and obtained a shuffled Pearson correlation coefficient. We repeated this shuffling procedure 2,000 times to create a shuffled null distribution. If the observed correlation coefficient was greater than the 95th percentile of the shuffled null distribution, the neuron pair was deemed significantly correlated (correlated pair). If the observed correlation coefficient was below the 95th percentile of the shuffled distribution, the correlation was deemed non-significant (random pair) (**Figure 4C**).

As many neurons exhibited elevated event rate during running (**Figure 3A**), we first identified correlated pairs during running (running-relevant pairs) versus resting (resting-relevant pairs) separately to account for variation in event rates during these periods. Specifically, to identify running-relevant pairs, we only considered the calcium event traces from neuron pairs when animals were running. Similarly, for resting-relevant pairs, only data during resting was considered. We found that the fraction of pairs that are correlated during running is smaller than during resting in both WT and KO mice (**Figure 4D**), and KO animals contained more correlated cell pairs compared to WT mice during both resting and running (**Figure 4D**). When we compared correlation coefficients between correlated pairs, we found no difference between WT and KO mice during both resting and running (**Figure 4E**). The correlation coefficients of random pairs were also similar between WT and KO during both behavioral conditions (**Figure 4F**). Thus, running desynchronizes the overall CA1 neural network by reducing the fraction of functionally connected neurons without altering connectivity strength between neuron pairs in both WT and KO groups. *NEXMIF* KO increases CA1 synchronization by increasing the fraction of functionally connected neurons without altering the connectivity strength during either resting or running.

Since the running-relevant pairs and resting-relevant pairs are often not the same neuron pairs, we could not directly compare how connectivity changes relevant neuron pairs during resting versus running. Thus, we next identified correlated pairs using calcium event traces throughout the entire session (session-relevant pairs) (**Figures 4G, H**). To identify session-relevant pairs, we compared the observed correlation between a neuron pair to the shuffled distribution using the entire recording period (**Figure 4I**). We found that the fraction of session-relevant pairs was increased in *NEXMIF* KO (**Figure 4J**). Interestingly, in WT mice, the correlation strengths of these session-relevant pairs were slightly higher during running, but were not significantly different between resting and running, indicating that when an animal switches between the two behavioral states, the relevant CA1 network connectivity remains largely stable (**Figure 4K**). In KO mice, however, correlation strength among session-relevant cells is significantly higher during running than resting (**Figure 4K**). In contrast, random pairs decreased their correlation strength during running in both WT and KO animals (**Figure 4L**).

To further investigate whether movement-modulated cells contribute to the increase of correlation strength in KO during running, we separately examined the correlation strength between two movement-modulated cells, between a movement-modulated and a non-modulated cell, and between two non-modulated cells (**Figure 4M**). In WT mice, correlation coefficient is significantly different only between two movement-modulated cells (**Figure 4N**), which may contribute to the small but non-significant increase across all pairs as shown in **Figure 4K**. However, in KO mice, correlation coefficients between two modulated neurons, and between a modulated and a non-modulated neuron pair were both higher during running than resting (**Figure 4O**). Thus, the increase in correlation coefficients in KO mice during running is a result of connectivity strength involving movement-modulated cells.

Since running increases event rates, we next evaluated how event rate impacts Pearson correlation coefficient measures. Under the condition of very sparse event rates observed in our study (WT: 2.12 ± 1.36 events/min, mean ± SD, n = 1817 neurons from 12 sessions in 6 mice, KO: 2.35 ± 1.48 events/min, n = 2845 neurons from 20 sessions in 8 mice), Pearson correlation coefficients decreased as event rate increases (**Supplementary Figure 3**). Thus, as running increased event rates, the observed increase in correlation coefficients cannot be explained by increased activity of individual neurons. Together, these results demonstrate that *NEXMIF* KO leads to over-synchronization of the CA1 network, particularly during running, by increasing the strength of pairwise correlations and synchronizing a larger fraction of CA1 neurons.

Overall network connectivity is exaggerated during locomotion in *NEXMIF* KO

After establishing functional connectivity changes between neuron pairs using Pearson correlation, we further characterized connectivity of the CA1 network as a whole using graph theory analysis. We first created network maps using the correlated cell pairs during either resting or running. Each cell is a node in the map, and a correlated cell pair is connected by an edge between

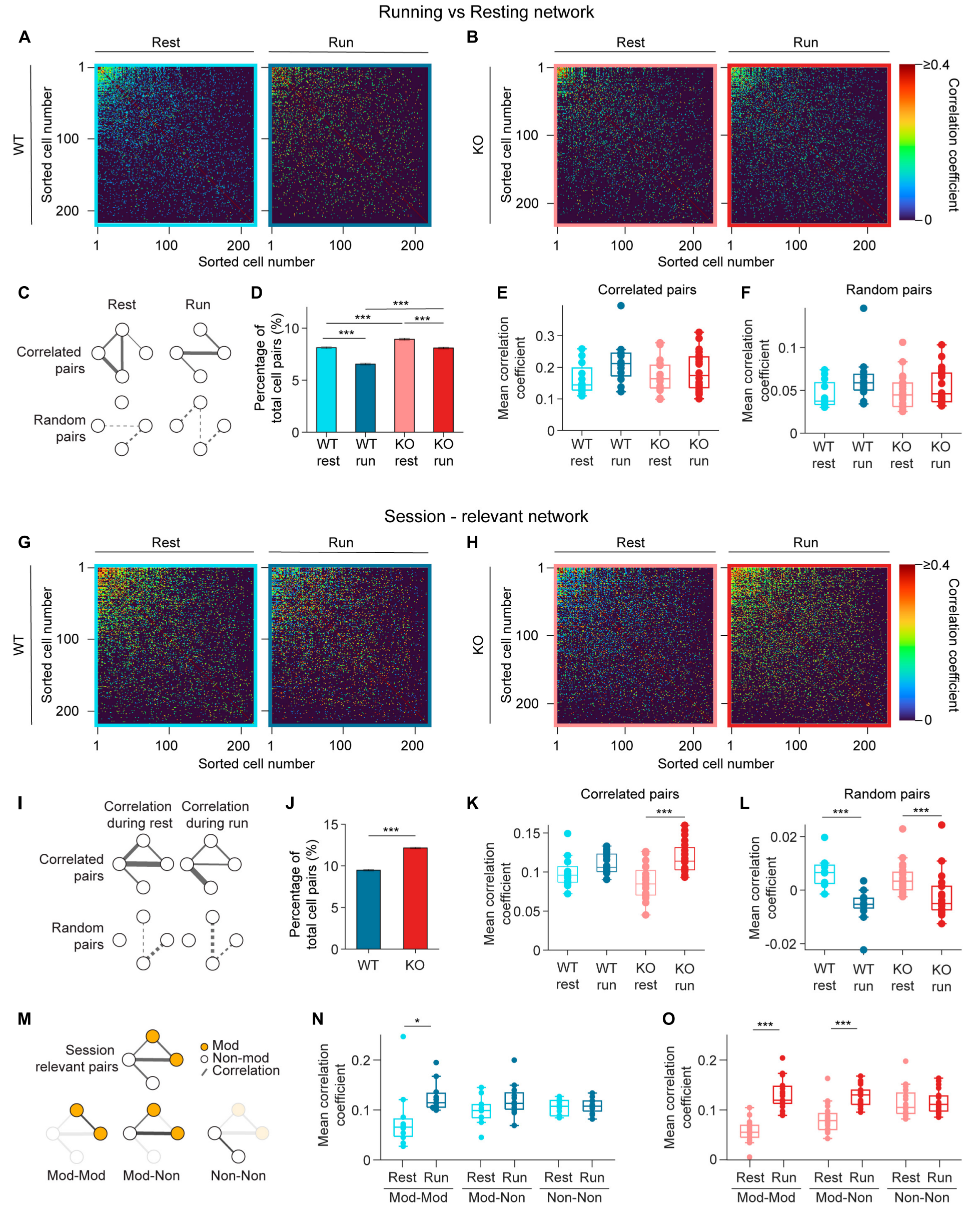


FIGURE 4

Pairwise correlation increases during running in NEXMIF KO mice. **(A)** Correlation matrices of pairwise Pearson correlation coefficient during resting (left) and running (right) for rest-relevant (left) and run-relevant neurons (right) from an example WT animal. Within each matrix, the most correlated cell pairs are sorted to the top left corner. Correlated pairs are colored corresponding to their correlation coefficient, random pairs are colored black. **(B)** Same as in panel **(A)**, but for an example KO animal. **(C)** Schematic networks showing (top left) resting-relevant and (top right) running-relevant pairs, and (bottom) the corresponding random pairs. **(D)** Fraction of neuron pairs that are correlated in WT mice during rest (light blue) and run (dark blue), and in KO during rest (light red) and run (dark red). (WT rest: 8.12 ± 0.10%, proportion ± 95% confidence interval, 301,335 neuron pairs, WT run: 6.55 ± 0.09%; KO rest: 8.93 ± 0.09%, 361,687 neuron pairs, KO run: 8.09 ± 0.09%. Fisher's exact test, WT rest vs. WT run: $p = 2.1 \times 10^{-121}$, KO rest vs. KO run: $p = 1.76 \times 10^{-37}$, WT rest vs. KO rest: $p = 7.7 \times 10^{-32}$, WT run vs. KO run: $p = 3.4 \times 10^{-127}$.) **(E)** Pearson correlation coefficients during resting for rest-relevant and during running for run-relevant cell pairs (WT rest: 0.16 ± 0.05, mean ± SD, n = 12 sessions in 6 WT mice, WT run: 0.22 ± 0.07; KO rest: 0.17 ± 0.05, n = 20 sessions in 8 KO mice, KO run: 0.19 ± 0.06, Linear Model, significance against intercept-only model: p = 0.10). **(F)** Same as panel **(E)**, for random cell pairs. (WT rest: 0.05 ± 0.02, mean ± SD, n = 12 sessions in 6 WT mice, WT run: 0.06 ± 0.03; KO rest: 0.05 ± 0.02, n = 20 sessions in 8 KO mice, KO run: 0.05 ± 0.02, Linear Model, significance against intercept-only model: p = 0.14.) **(G)** Same as in panel **(A)**, for session-relevant neurons from the same WT animal. **(H)** Same as in panel **(B)**, for session-relevant neurons from the same KO animal. **(I)** Schematic of a session-relevant network with line widths denoting the correlation strength of correlated pairs (top left) during rest and (top right) during running, and (bottom) the corresponding random pairs. **(J)** Fraction of neuron pairs that are correlated in WT mice (blue) and in KO (red) mice during the entire session (WT: 9.48 ± 0.10%, proportion ± 95% confidence interval, 301,335 neuron pairs, KO: 12.15 ± 0.11%, 361,687 neuron pairs; Fisher's exact test, $p = 4.1 \times 10^{-266}$). **(K)** Pearson correlation coefficients during resting or running of session-relevant cell pairs. (WT rest: 0.10 ± 0.02, mean ± SD, n = 12 sessions in 6 WT mice, WT run: 0.11 ± 0.01; KO rest: 0.09 ± 0.02, n = 20 sessions in 8 KO mice, KO run: 0.12 ± 0.02, Linear Model, interaction: p = 0.04, post-hoc Linear Model, WT rest vs. WT run: p ± 0.12; KO rest vs. KO run: $p = 2.33 \times 10^{-6}$.) **(L)** Same as panel **(K)**, for random cell pairs. (WT rest: $6 \times 10^{-3} \pm 5 \times 10^{-3}$, mean ± SD, n = 12 sessions in 6 WT mice, WT run: $-6 \times 10^{-3} \pm 6 \times 10^{-3}$; KO rest: $4 \times 10^{-3} \pm 6 \times 10^{-3}$, n = 20 sessions in 8 KO mice, KO run: $-2 \times 10^{-3} \pm 8 \times 10^{-3}$, Linear Model, behavioral condition: $p = 4.29 \times 10^{-3}$, WT/KO genotype: p = 0.40, interaction: p = 0.10.) **(M)** Schematic of (top) the session-relevant network in I and (bottom) the same network decomposed into

(Continued)

FIGURE 4 (Continued)
three sub-networks in which both cells are modulated (Mod-Mod, left), one modulated and one non-modulated (Mod-Non, middle), or both are non-modulated (Non-Non, right). Session-relevant pairs are connected by a line, with modulated cells in orange and non-modulated cells in white. **(N)** Mean correlation coefficient of the three session-relevant sub-networks in WT animals (Mod-Mod rest: 0.08 ± 0.03, mean ± SD, n = 12 sessions in 6 mice, Mod-Mod run: 0.13 ± 0.03; Mod-Non rest: 0.10 ± 0.03, Mon-Non run: 0.12 ± 0.03; Non-Non rest: 0.11 ± 0.02, Non-Non run: 0.11 ± 0.02; Linear Model, interaction between Mod-Mod network and Behavior: $p = 3.11 \times 10^{-3}$, *post-hoc* Linear Model, rest vs. run: $p = 3.67 \times 10^{-2}$.) **(O)** Same as in panel **(N)**, for KO animals (Mod-Mod rest: 0.05 ± 0.03, n = 20 sessions in 8 KO mice, Mod-Mod run: 0.13 ± 0.03; Mon-Non rest: 0.08 ± 0.03, Mon-Non run: 0.12 ± 0.02; Non-Non rest: 0.11 ± 0.03, Non-Non run: 0.11 ± 0.03, Linear Model, interaction between Mod-Mod network and Behavior: $p = 1.44 \times 10^{-8}$, *post-hoc* Linear Model, rest vs. run: $p = 6.69 \times 10^{-10}$; Linear Model, interaction between Mod-non network and Behavior: $p = 3.0 \times 10^{-4}$, *post-hoc* Linear Model, rest vs. run: $p = 8.53 \times 10^{-7}$). In panels **(E,F,K,L,N,O)**, each dot corresponds to an individual session (box: interquartile range, whiskers: 1.5 ± interquartile range, middle line: median). $^*p < 0.05$, $^{***}p < 0.001$.

those two nodes. To quantify the connectivity of each network graph, we calculated the closeness centrality of each neuron. Closeness centrality is a metric used in graph theory to measure node importance, which takes both number of connections and connection strength (correlation coefficient) into account (details in section "Materials and methods"). Briefly, a greater closeness centrality value for a neuron indicates that the neuron is connected, both directly and indirectly, to a greater number of nodes in the network (Figure 5E).

As fluorescence imaging allowed us to visualize the anatomical relationship between recorded neurons, we first arranged the network graph using the anatomical position of each cell (Figures 5A, C). We did not observe any obvious spatial patterns in the closeness centrality within CA1 networks in the anatomical maps. Consequently, to better visualize the strength of network connectivity, we arranged each map as a force-directed graph where cells are positioned closer if their functional connectivity is higher regardless of their absolute anatomical location (Figures 5B, D). In WT force-directed maps, cells were more tightly clustered during resting than running, indicating higher connectivity during resting. However, KO force-directed maps showed similar amounts of clustering between resting and running (Figure 5B). We also noted that the change in each neuron's closeness centrality varied widely from resting to running (Figures 5A–D). Thus, to quantify the changes in overall network connectivity between behavioral conditions, we computed the difference in average closeness centrality between the resting graph and running graph for each recording session (Figure 5F). We found that WT mice showed a significant decrease in closeness centrality during running compared to resting, demonstrating that the CA1 network is desynchronized during locomotion. This network-level observation is consistent with our pairwise Pearson correlation analysis showing that in WT mice, fewer cell pairs were correlated during locomotion while correlation strength remained constant (Figures 4D, K). In contrast, KO mice showed similar closeness centrality during running and resting, suggesting that KO network fails to desynchronize during locomotion (Figures 5C, D). This lack of overall network desynchronization in KO mice measured with closeness centrality could be due to the opposing effects we observed with Pearson correlation analysis, which showed the fraction of correlated cells in KO mice is lower during running (Figure 4D) while correlation strength is higher (Figure 4K). Further, WT mice exhibited a greater decrease in closeness centrality than KO mice, consistent with the higher fraction of correlated cells in KO mice compared to WT mice (Figure 4D). Together, these results confirm that while the WT CA1 network desynchronizes during locomotion, *NEXMIF* KO impairs CA1 network desynchronization.

Discussion

In this study, we examined how loss of *NEXMIF*, an ASD risk gene highly expressed in the hippocampus, influences individual CA1 neurons' responses and CA1 network functional connectivity using large-scale single-cell resolution calcium imaging. As NEXMIF KO mice exhibit profound learning and memory deficits as indicated by Barnes maze and novel object tests (Gilbert et al., 2020), we probed the hippocampal network during voluntary locomotion, a fundamental aspect of spatial memory. We compared the patterns of neural activation between NEXMIF KO and WT littermates during quiescent immobility versus active locomotion. We found that spontaneous calcium event rate is similar between WT and KO mice, but a larger percentage of CA1 neurons are activated during movement in KO mice. Furthermore, a greater fraction of neuron pairs is correlated in KO animals, and the KO network is overly synchronized during locomotion. Overall, our results demonstrate that loss of *NEXMIF* leads to increased behaviorally evoked responses and elevated network synchronization, both of which could contribute to the disruption of CA1 network coding ability during behavior.

Our previous work has shown that in an open-field task, NEXMIF KO mice are more active than WT mice, reflecting higher levels of anxiety in NEXMIF KO mice (Gilbert et al., 2020). We did not find a difference in locomotion kinematics between KO and WT mice, likely due to differences in experimental conditions. However, the average running speed observed is comparable to those reported previously. Further, locomotion behavior did not vary with the age of the mice. Thus, our experimental paradigm allows us to probe the impact of *NEXMIF* KO on neural circuits during locomotion in the absence of behavioral changes.

As increased cellular and synaptic level E/I ratio in ASD can lead to increased neuronal excitability, epilepsy occurs in about 10% of people with ASD (Lukmanji et al., 2019) [about 15 times higher than incidence in the general population (Fiest et al., 2017)] and is particularly prevalent in individuals with *NEXMIF* mutations (Tye et al., 2019; Stamberger et al., 2020). We did not observe differences in basal calcium event rate in NEXMIF KO mice, but we detected significantly more neurons that selectively increased their activity during movement in KO animals. These observations provide evidence that in *NEXMIF* KO conditions, the elevated synaptic E/I ratio is not correlated to a broad increase in spontaneous neuronal activity, but rather a selective increase in

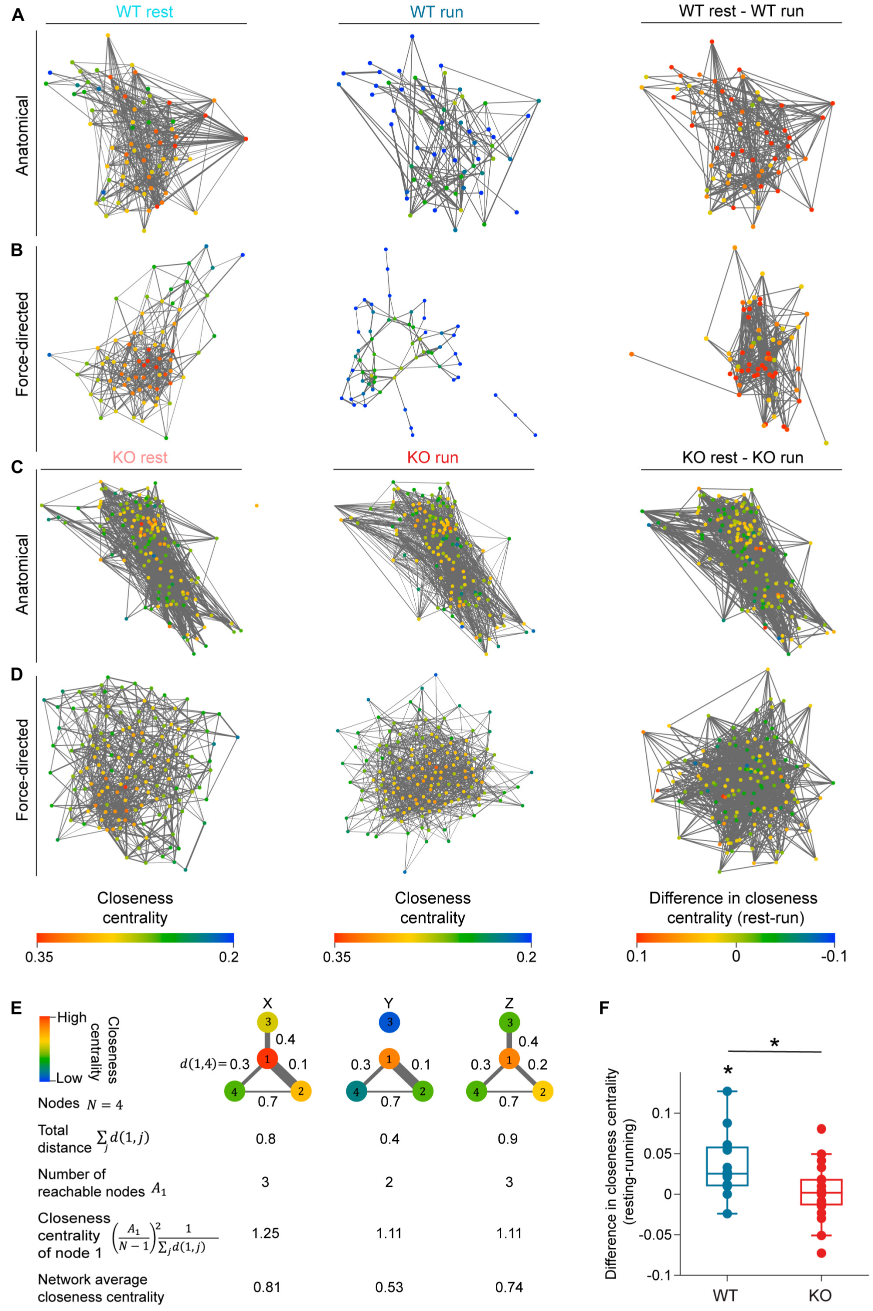


FIGURE 5

NEXMIF knockout increases overall functional connectivity of the CA1 network. Anatomical network maps of closeness centrality during (left) resting, (middle) running, and (right) the change between resting and running from example **(A)** WT and **(C)** KO animals. In left and middle columns, each cell is color coded based on its closeness centrality measure and correlated pairs in the corresponding behavioral condition are connected by an edge. In the right column, each cell is color coded based on the change in its closeness centrality from resting to running, and session-relevant pairs are connected by an edge. Edge width represents normalized correlation coefficient. **(B,D)** Similar to panels **(A,C)**, but shown as a force-directed graph. **(E)** A schematic of closeness centrality computation. An example network of four nodes is shown in three different network states (X, Y, Z) with the distance of the edge shown between each pair of nodes. Closeness centrality of node 1 in the three states are 1.25, 1.11, and 1.11, respectively. The reduction of node 1's closeness centrality in state Y compared to state X is due to the loss of reachable node 3. The decrease of node 1's closeness centrality in state Z relative to state X is due to the increased distance to node 2. **(F)** Average closeness centrality difference (resting-running) in WT (blue) and KO (red) mice. (WT: 0.036 ± 0.041, mean ± SD, n = 12 sessions from 6 mice, significantly greater than 0, Wilcoxon signed rank test, $p = 6.8 \times 10^{-3}$; KO: 0.002 ± 0.034, n = 20 sessions from 8 mice not significantly different from 0, Wilcoxon sign rank test, p = 0.77, Wilcoxon rank sum test between WT and KO, p = 0.02.) Each dot corresponds to an individual session (box: interquartile range, whiskers: 1.5 < interquartile range, middle line: median). *$p < 0.05$.

responding during relevant behavior. This behavioral state-specific increase in neuronal activity in the CA1 could be due to a global increase in synaptic inputs to the CA1 during movement, in which an increased E/I synaptic ratio leads to a greater excitatory drive to CA1 neurons. However, it is also possible that the observed increase in neuronal activity reflects movement-dependent changes in intrinsic neuronal excitability, in addition to altered synaptic inputs.

Another leading hypothesis of ASD pathophysiology argues for overconnectivity within local brain regions and underconnectivity between interconnected brain regions, supported by several exciting human studies (Casanova et al., 2002; Casanova, 2004; Girgis et al., 2007; Kennedy and Courchesne, 2008; Wang et al., 2013; Zikopoulos and Barbas, 2013; Di Martino et al., 2014). We observed increased fractions of functionally correlated CA1 neuron pairs in *NEXMIF* KO animals during both immobility and active locomotion, as well as increased correlation strength during running in session-relevant cell pairs from KO mice, particularly with higher contribution from movement-modulated cells. Additionally, while the WT network was desynchronized during running, the KO network failed to desynchronize. Interestingly, we also detected a reduction in correlation coefficients across random pairs during running in both WT and KO, consistent with an overall network desynchronization effect, even though the correlation coefficients between random pairs were not deemed significantly higher than chance observation given their event rate. Each of these observations is consistent with abnormally increased functional connectivity within the CA1 circuit of NEXMIF KO mice during locomotion, supporting the local overconnectivity hypothesis. The elevated E/I synaptic ratio could contribute to this increased functional connectivity (Litwin-Kumar and Doiron, 2012; Middleton et al., 2012), but again, we cannot rule out the possibility that *NEXMIF* KO also changes intrinsic biophysical properties that lead to the observed over-synchronization. While CA1 pyramidal cells are known to have limited lateral connections, the elevated E/I synaptic ratio could result from reduced inhibitory inputs from local interneurons or increased excitatory inputs from upstream areas. Further work is needed to better understand the exact mechanisms by which *NEXMIF* alters both cellular biophysical properties such as ion channel expression and functional connectivity between the hippocampus and its interconnected areas. Additionally, intracellular calcium signaling is known to be important for neuronal morphogenesis and migration during development. While this study is limited to adult animals, future studies using similar calcium imaging approaches during development could provide insights into how changes in intracellular calcium dynamics upon *NEXMIF* KO may influence neurite extension and migration and contribute to the connectivity changes observed here.

We probed network functional connectivity using two measures, Pearson correlation between pairs of neurons and network closeness centrality. In WT animals, the number of correlated cell pairs decreased during locomotion while correlation strength of session-relevant correlated pairs was stable, ultimately resulting in decreased closeness centrality of the WT network during running. These results indicate decreased functional connectivity in the WT CA1 network during movement. Such network desynchronization would lead to increased information encoding capability, consistent with the idea that the CA1 network encodes relevant information during active movement (Colgin, 2013). As locomotion is a fundamental component of spatial navigation and memory, this dynamic change in information coding capability would allow for flexible and efficient encoding of a WT animal's current environment for spatial memory.

In NEXMIF KO animals, however, a larger number of cell pairs are significantly correlated in both behavioral conditions than in WT animals, and session-relevant cell pairs are dominated by stronger correlations during running. Additionally, network closeness centrality of the KO CA1 network failed to decrease during movement, in sharp contrast to the reduction seen in WT networks. These different measures all support the consequence of *NEXMIF* KO in exaggerating network synchrony and preventing network desynchronization during active behavior.

Our previous study revealed that loss of NEXMIF led to a reduction in mature functional spines, leading to reduced excitatory synaptic strength in NEXMIF KO mice. While there was a reduction of both glutamatergic and GABAergic synaptic proteins in KO mice, the reduction in GABAergic synaptic density was double the loss of glutamatergic synapses in cultured neurons (Gilbert et al., 2020). This increase in synaptic E/I ratio in KO mice (Gilbert et al., 2020) likely contributes to the observed network over-synchronization in KO mice, which would lead to a decreased information encoding capacity in the CA1 network of NEXMIF KO mice. The higher percentage of movement-modulated cells observed in KO mice could reflect a compensatory mechanism in the CA1, to homeostatically increase information encoding capability throughout development. Alternatively, this higher percentage could be due to the increased number of correlated cells during running, as these correlations could arise from common inputs to these cell pairs that are activated upon movement. Overall, our observations of increased functional connectivity indicate a reduced ability to process spatial information and spatial encoding that could lead to the impaired spatial memory and contextual fear memory observed in NEXMIF KO mice (Gilbert et al., 2020).

Materials and methods

Animal surgery and recovery

All animal procedures were approved by the Boston University Institutional Animal Care and Use Committee. Eight homozygous *NEXMIF* KO (maintained on a C57Bl/6 genetic background) male mice and seven WT male littermates were used in this study (Gilbert et al., 2020). Mice were 7–34 weeks old at the start of experiments. Animals first underwent stereotaxic viral injection surgery, targeting the hippocampus (anterior/posterior: −2.0 mm, medial/lateral: +1.4 mm, dorsal/ventral: −1.6 mm from bregma). Mice were injected with 500–750 nl of AAV9-synapsin-GCaMP6f.WPRE.SV40 virus, obtained from the University of Pennsylvania Vector Core (titer ∼6e12 GC/ml). Injections were performed with a blunt 33-gauge stainless steel needle (NF33BL-2, World Precision Instruments) and a 10 μl microinjection syringe (Nanofil, World Precision Instruments), using a microinjector pump (UMP3 UltraMicroPump, World Precision Instruments). The needle was lowered over 1 min and remained in place for 1 min before infusion. The rate of infusion was 50 nl/min. After infusion, the needle remained in place for 7–10 min before being withdrawn over 1 min. The skin was then sutured closed with a tissue adhesive (Vetbond, 3M). After complete recovery (7+ days after virus injection), animals underwent a second surgery to implant a sterilized custom imaging cannula (outer diameter:

3.17 mm, inner diameter: 2.36 mm, height: 2 mm). The imaging cannula was fitted with a circular coverslip (size 0, outer diameter: 3 mm, Deckgläser Cover Glasses, Warner Instruments), adhered to the bottom using a UV-curable optical adhesive (Norland Optical Adhesive 60, P/N 6001, Norland Products). During surgery, an approximately 3.2 mm craniotomy was created (centered at anterior/posterior: −2.0 mm, medial/lateral: +1.7 mm) and the cortical tissue overlaying the hippocampus was aspirated away to expose the corpus callosum. The corpus callosum was then thinned until the underlying CA1 became visible. The imaging cannula was then tightly fit over the hippocampus and sealed in place using a surgical silicone adhesive (Kwik-Sil, World Precision Instruments). The imaging window was secured in place using bone adhesive (C&B Metabond, Parkell) and dental cement (Stoelting). A custom aluminum head-plate was also affixed to the skull anterior to the imaging window. Analgesic was provided for at least 48 h after each surgery, and mice were single-housed after window implantation surgery to prevent damage to the head-plate and imaging window.

Calcium imaging and movement data acquisition

After complete recovery from window implantation surgery (7+ days), animals were habituated to experimenter handling and head fixation on the spherical treadmill. Each animal was habituated to running on the spherical treadmill while head-fixed for at least 3 days prior to the first recording day. During each recording session, animals were positioned under a custom wide-field microscope and allowed to run freely on the spherical treadmill. The spherical treadmills consisted of a three-dimensional printed plastic housing and a Styrofoam ball supported by air (Dombeck et al., 2007). The imaging microscope was equipped with a scientific complementary metal oxide semiconductor (sCMOS) camera (ORCA-Flash4.0 LT Digital CMOS camera C11440-42U, Hamamatsu) and a 10 × 0.28 M Plan Apo objective (Mitutoyo). GCaMP6f excitation was accomplished with a 5 W light emitting diode (M470L4, ThorLabs). The microscope included an excitation filter (No. FF01-468/553-25, Semrock), a dichroic mirror (No. FF493/574-Di01-25 × 36, Semrock), and an emission filter (No. FF01-512/630-25, Semrock). The imaging field of view was 1.343 × 1.343 mm (1,024 × 1,024 pixels). Image acquisition was performed using HC Image Live (Hamamatsu), and images were stored offline as multi-page tagged image files (TIFs) for further analysis.

Each animal underwent three 10-min recording sessions, one per day, every other day over 5 days (**Figure 1E**). A total of 21 recording sessions were collected from 8 KO mice and 16 sessions were collected from 7 WT mice. In 24 recording sessions (from 4 WT mice and 8 KO mice), a custom MATLAB script was used to trigger image frame capture at 20 Hz and to synchronize image acquisition with movement tracking. Digital transistor-transistor logic (TTL) pulses were delivered to the camera via a common input/output interface (No. USB-6259, National Instruments), and TTL pulses were also recorded using a commercial system (RZ5D, Tucker Davis Technologies). Motion data was collected using a modified ViRMEn system (Gritton et al., 2019). Movement was tracked using two computer universal serial bus mouse sensors affixed to the plastic housing at the equator of the Styrofoam ball, 78° apart. The mouse sensors' x- and y-surface displacement data were acquired at 100 Hz on a separate computer, and a multi-threaded Python script was used to send packaged <dx, dy> data to the image acquisition computer via a RS232 serial link. Packaged motion data was recorded on the image acquisition computer using a modified ViRMEn MATLAB script and synchronized to each acquired imaging frame.

In the remaining 13 sessions (from 4 WT mice and 2 KO mice), image acquisition was triggered using a Teensy microcontroller system (Romano et al., 2019), and experiments were performed using an identical spherical treadmill. Digital pulses were sent from a Teensy 3.2 (TEENSY32, PJRC) to the sCMOS camera via SMA connectors and coaxial cables to trigger frame capture at 20 Hz. TTL pulses were recorded using the same TDT commercial system. Movement was tracked using two computer mouse sensors (ADNS-9800 laser motion sensors, Tindie) affixed to the plastic housing at the equator of the Styrofoam ball, about 75° apart. The x- and y-surface displacement was collected by the Teensy at 20 Hz and sent to the image acquisition computer via a standard USB-microUSB cable.

Movement analysis

As both movement data acquisition systems collect the same numerical data, linear velocity can be calculated the same way for all sessions. Linear velocity in perpendicular X and Y directions was calculated as:

$$X = \frac{L - R\,cos\theta}{\cos\theta(\frac{\pi}{2} - \theta)}$$

$$Y = R$$

where L is the vertical reading from the left sensor, R is the vertical reading from the right sensor, and θ is the angle between the sensors. Linear velocity V was then calculated as:

$$V = \sqrt{X^2 + Y^2}$$

Linear velocity values were then interpolated at 20 Hz.

To identify sustained periods of movement with high linear velocity (running bouts), we used a Fuzzy logic-based thresholding algorithm. We first assigned each velocity data point a Fuzzy membership value using a sigmoidal membership function F:

$$F(V,\ a, c) = \frac{1}{1 + e^{-a(V-c)}}$$

where the threshold c is the 20th percentile of the velocity of that session or 5 cm/s, whichever is higher. a is set at 0.8. The resulting velocity trace was then smoothed using a moving average filter of 1.5 s. Next, the smoothed trace was thresholded at 10% of its maximum value. Time periods with velocity higher than this threshold that were at least 2 s long were considered high velocity periods ("running"). Time periods with velocity lower than this threshold that were at least 2 s long were considered low velocity periods ("resting"). Periods that did not satisfy either of these requirements were not considered for locomotion analysis (**Figures 1F, G**). Recording sessions in which the mouse spent less than 60 s (10% of the session) in one behavioral condition and sessions with less than 5 running bouts were not included for locomotion-related analysis (4 sessions from 2 WT mice and 1 session from 1 KO mouse).

Calcium imaging video motion correction

Videos were first motion corrected using a custom Python script as described previously (Keaveney et al., 2020). For each imaging session, we first generated a reference image by calculating the mean value of each pixel across the first 2,047 frames. We then performed a series of contrast-enhancing procedures to highlight image features as follows. We used a Gaussian filter (Python SciPy package, ndimage.Gaussian_filter, sigma = 50) to remove the low-frequency component, which represents the potential non-uniform background. We then captured the edges of the high-intensity area by calculating the differences between two Gaussian filtered images (sigma = 2 and 1). To obtain the edge-enhanced image, the edges were multiplied by 100 and added back to the first filtered image (sigma = 2). Finally, to prevent a potential overall intensity shift caused by photobleaching, we normalized the intensity of each image by subtracting the mean intensity of the image from each pixel and dividing by the SD of the intensity. We then calculated the cross-correlations between the processed reference image and each processed image frame, and obtained the displacement between the peak of the coefficient and the center of the image. We then applied a horizontal shift, opposite to the displacement, to the original frame to finalize the motion correction.

Cell segmentation

From the motion corrected video, a projection image was generated across all frames by subtracting the minimum fluorescence from the maximum fluorescence of each pixel (max-min projection image), and regions of interest (ROIs) corresponding to fluorescent cell bodies were automatically identified in the max-min projection image using a deep-learning algorithm based on U-Net (Ronneberger et al., 2015; Falk et al., 2019; Xiao et al., 2020). We first trained the deep-learning algorithm with manually curated data, containing the datasets reported in our previous studies (Shen et al., 2018; Zemel et al., 2022). For each training dataset, a max-min projection image was calculated as described above. We then divided the projection images and their corresponding ROI masks into small patches of 32 × 32 pixels as our training dataset. We also normalized each patch by shifting its mean intensity to zero and dividing the intensity of each pixel by the SD of the patch intensity. During training, each pixel was further augmented by randomly flipping vertically and/or horizontally, and rotating 90°C, 180°C, or 270°C.

To segment ROIs for the datasets in this study, the max-min projection image for each dataset was divided into 32 × 32 patches with 50% of each patch overlapping with its neighboring patches. Each patch was normalized as described above. As a result, each pixel was inferred four times, and we averaged the results from four inferences as the prediction score. The connected pixels with a predication score >0.5 were segmented as a potential ROI, and the set of segmentations was further refined with watershed transformation to obtain the ROIs representing single neurons. ROIs were then overlaid on the max-min projection image and manually inspected. ROIs that were identified by the machine learning algorithm but were not identified as a cell by the observer were manually removed. ROIs were then manually added to select cells that the machine learning algorithm did not properly identify. ROIs were added as a circle with a radius of 6 pixels (7.8 μm) based on morphology present in the max-min projection image, using the previously reported semi-automated custom MATLAB software called SemiSeg[1] (Mount et al., 2021).

GCaMP6f fluorescence trace extraction and normalization

We obtained the GCaMP6f fluorescence for each cell as the average fluorescence intensity across all pixels in that ROI. We then subtracted background fluorescence from each ROI, where the background fluorescence is the average pixel intensity across the pixels located within a ring centered at the corresponding cell ROI with an outer radius of 50 pixels and an inner radius of 15 pixels. The areas corresponding to other cell ROIs were excluded from this background ROI. Because the motion correction procedure introduces strips with high pixel intensities along the edges of the max-min projection image, 25 pixels along each edge of the image were also excluded from the calculation of background fluorescence. The resulting fluorescence trace for each cell was then interpolated at 20 Hz, linearly detrended (MATLAB function detrend), and normalized between 0 and 1. All traces were then manually inspected. Traces with large artifacts were removed.

Calcium event detection

Onsets of calcium events were identified in each fluorescence trace similarly to previous descriptions (Mount et al., 2021; Zemel et al., 2022). Briefly, we first applied a moving average filter of 1 s to smoothen each trace and calculated the spectrogram [MATLAB chronux, mtspecgramc with tapers = (2, 3) and window = (1, 0.05)], and averaged the power below 2 Hz. We then calculated the change in power at each time point (powerdiff) and identified outliers (3 median absolute deviations away from the median power) in powerdiff (MATLAB function isoutlier) to detect all significant changes in trace power. When multiple outliers occurred at consecutive time points, they were classified as a potential calcium event. We then calculated the rise time and amplitude (the difference in fluorescence value between the peak and the event onset) for each potential event and used an iterative process to include only true events and exclude incorrect potential events. Within each iteration, an amplitude threshold was calculated for each potential event [iteration 1: 7 SDs of the trace in the 10 s (200 data points) prior to calcium event onset]. Potential events with a rise time greater than 150 ms (3 data points) and an

[1] github.com/HanLabBU/SemiSeg

amplitude above the calculated threshold were marked as correctly identified events for analysis. All the data points corresponding to these events (from beginning of event rise to end of event fall) were removed prior to the next iteration. We then repeated this process by re-calculating the amplitude threshold for the remaining potential events and again marking correctly identified events for analysis using the same criterion for rise time and the new amplitude threshold. For each successive iteration, the amplitude threshold was decreased by 40% and the duration to inspect prior to calcium event onset was increased by 75%. The iterative process stops once no events are marked as correctly identified events. This iterative method is more robust in capturing events that occur close together, while only minimally increasing identification of false positives. The preceding event in a sequence will incorrectly bias the SD of the trace in the window prior to a following event in the sequence, and this bias is removed when the preceding event is not included in the window prior to the event onset. All traces were then manually inspected to confirm event detection accuracy.

Calcium event parameter and event rate analysis

For each detected calcium event, the rise time is defined as the duration from the calcium event onset, t_{on} to its peak t_{peak} (**Figure 2E**). To determine the full width at half-maximum amplitude (FWHM), we first calculated event height as the fluorescence different between t_{on} and t_{peak}. FWHM was determined as the duration between the two points at 50% of the event height (**Supplementary Figure 4A**). If a subsequent calcium event was detected before the end point of the FWHM, the given event was excluded from FWHM analysis (**Supplementary Figure 4B**). Because it is difficult to reliably estimate FWHM if an event contains multiple small peaks, we further examined the number of peaks above 75% of the event height, and if more than one peak was identified (**Supplementary Figure 4C**), the given event was also excluded from FWHM analysis.

Total event rate was calculated across the entirety of each trace, counting each identified calcium event as one event. Event rate during either running or resting was calculated by counting the number of calcium events in all bouts of the relevant behavioral condition and dividing by the total time that the mouse spent in that behavioral condition.

Determination of movement-modulated cells

To determine movement-modulated cells, we binarized each fluorescence trace by assigning ones to the entire rising phase (t_{on} to t_{peak}) of each calcium event and zeros to the rest of the trace. We then concatenated all of the resting or running bouts separately, and summated the binarized trace among each concatenated period ("total activity"). We then subtracted the total activity during resting from the total activity during running to create an activity metric A. The calculation can be summarized as:

$$A = \left(\frac{\sum_{run} x}{\sum_{run} t} - \frac{\sum_{rest} x}{\sum_{rest} t} \right) \times 100\%$$

where x is the binarized calcium trace, and t is time. Next, we created a shuffled distribution of the activity metric for each cell by circularly shifting the binarized trace relative to the movement trace by a uniformly distributed random time lag 1,000 times and calculating A for each shuffle. If the true (non-shifted) A for a neuron was greater than the 97.5th percentile of the shuffled distribution, the cell was considered movement-modulated. Cells that did not meet this criterion were considered non-movement-modulated.

Pairwise Pearson correlation analysis

For pairwise correlation analysis, we calculated the Pearson correlation coefficient between the binarized traces for each pair of neurons. Each binarized trace was the same as that used in determination of movement-modulated cells [ones to the entire rising phase (t_{on} to t_{peak}) of each calcium event and zeros to the rest of the trace]. Only neuron pairs that were at least 20 pixels (26.2 μm) apart were included in all correlation analysis to eliminate potential fluorescence cross-contamination. We calculated pairwise correlation during resting alone, during running alone, or during the entire duration of the session. To calculate pairwise correlation during resting alone or running alone, we concatenated the calcium activity during all resting or running periods. To calculate pairwise correlation during the entire duration of the session, we used the full length of the calcium traces for each cell pair.

To determine whether the correlation coefficient for each cell pair was above chance level for each behavioral condition, we created a shuffled distribution of correlation by circularly shifting one trace relative to the other trace by a uniformly distributed random time lag 2,000 times and calculating the Pearson correlation coefficient for each shuffle. If the empirical (non-shifted) Pearson correlation for a pair of neurons was greater than the 95th percentile of the shuffled distribution, the cell pair was considered correlated. Positive correlation coefficients between neuron pairs that were not greater than the 95th percentile were not considered significant (random pairs). Negative correlations were not included in any analyses due to the sparseness of GCaMP6f events.

To estimate connectivity among modulated cells, we calculated the number of correlated movement-modulated cell pairs as a fraction of all movement-modulated cell pairs. Similarly, we also calculated the number of correlated non-movement-modulated cell pairs as a fraction of all non-movement-modulated cell pairs.

Relationship between Pearson correlation coefficient and event rate analysis

As the calcium event rates detected in our study were sparse, we investigated the relationship between event rate and Pearson correlation coefficient. Specifically, to determine whether increasing event rate leads to an increase in pairwise Pearson correlation coefficient values by chance, we examined the relationship of the mean event rate of a neuron pair versus their shuffled correlation coefficient values. To calculate the shuffled correlation coefficient values of a neuron pair, we circularly shifted the calcium event vector of one neuron relative to the other by a random time lag uniformly distributed over the entire length of the trace, so that the temporal structure between the calcium event rates of the two neurons was destroyed. Next, Pearson correlation coefficient was calculated between the shuffled calcium event vectors of the neuron pair. This procedure was repeated 100 times for each cell pair, using either resting or running periods separately for each imaging session. These shuffled correlation coefficient values were then plotted against the mean event rate of the pair (**Supplementary Figure 3A**). The observed (true) correlation coefficients of session-relevant correlated pairs (**Supplementary Figure 3B**) and random pairs (**Supplementary Figure 3C**) were similarly plotted against the average event rate of each pair of neurons.

Network closeness centrality analysis

To quantify network connectivity patterns, we calculated closeness centrality similarly to the description in Wuchty and Stadler (2003). Specifically, for each session, we created an undirected graph using correlated cell pairs during running and an undirected graph using correlated cell pairs during resting. Each cell was considered as a node and each correlated cell pair was connected by an edge. Edge weight was the Pearson correlation coefficient (calculated in the appropriate state, rest or run) between the binarized calcium traces of the cell pair. For each node i, closeness centrality $c\left(i\right)$ is defined as:

$$c\left(i\right) = \left(\frac{A_i}{N-1}\right)^2 \frac{1}{C_i}$$

where A_i is the number of nodes reachable to node i and C_i is the sum of distances from node i to all reachable nodes. The distance $d\left(i,j\right)$ between nodes i and j is defined as:

$$d\left(i,j\right) = \sqrt{\log\left(\frac{1}{w_{i,j}}\right)}$$

where $w_{i,j}$ is the edge weight. Closeness centralities of all the nodes were averaged within each network and multiplied by the number of nodes for normalization across networks with different numbers of nodes. Force-directed networks were created using a MATLAB implementation of a force-directed node placement algorithm that spatially clusters nodes proportional to $d\left(i,j\right)$ (Fruchterman and Reingold, 1991).

Statistical analysis

Statistical analyses were performed using MATLAB. Using Shapiro-Wilk's normality test, we determined that most of our datasets do not follow normal distribution. Thus, non-parametric tests were used. Specifically, Wilcoxon rank sum test was used for comparisons between two groups (**Figures 1H–K**, **5F**) and Linear Models (LMs) were used for comparisons between three or more groups. LMs were used to test whether the independent variables (WT/KO genotype, rest/run behavioral conditions, or different types of cell groups) were significant predictors for the dependent variable Y of interest using the following models:

For **Figures 2F, G, 3A, 4E, F, K, L**:

$$Y \sim 1 + genotype + behavioral\ condition\ + genotype \times behavioral\ condition$$

For **Figure 4N**:

$$Y_{WT} \sim 1 + behavioral\ condition + cell\ group + behavioral\ condition \times cell\ group$$

For **Figure 4O**:

$$Y_{KO} \sim 1 + behavioral\ condition + cell\ group + behavioral\ condition \times cell\ group$$

Maximum likelihood estimation was used to estimate coefficients for the selected model. First, we compared the model's fit against an intercept-only model using a deviance test. If the model was significantly different from the intercept-only model, p values were calculated for the coefficient of each independent variable (genotype, behavioral condition, and cell group), and the interaction term (genotype × behavioral condition or behavioral condition × cell group). If the coefficient of the interaction term was significant, separate post-hoc Linear Models were used to test whether behavioral condition (**Figures 2F, G, 3A, 4E, F, K, L**: $Y_{WT} \sim 1 + behavioral\ condition$, and $Y_{KO} \sim 1 + behavioral\ condition$, **Figure 4N**: $Y_{WT,\ cell\ group} \sim 1 + behavioral\ condition$, **Figure 4O**: $Y_{KO,\ cell\ group} \sim 1 + behavioral\ condition$) was a significant predictor of the dependent variable. If the interaction term was not significant, independent variables with significant coefficient terms ($p < 0.05$) were considered as the significant predictors of the dependent variable. Further, because the variables in our study (genotypes, cell groups, and behavioral conditions) have only two levels (WT vs. KO, rest vs. run, modulated versus non-modulated cell pairs), a post-hoc test was not required when interaction term was not significant. Wilcoxon signed rank tests were used to test if medians were significantly different from 0 (**Figure 5F**). Finally, when comparing proportions (**Figures 3B**, **4D, J**), Fisher's exact test was used. Error bars show the 95% confidence interval defined as follows.

$$Confidence\ Interval = P \pm 1.96\sqrt{\frac{P(1-P)}{n}}$$

where P denotes the percentages, and n denotes the number of samples. Simple linear regression was used to compare the percentage of movement-modulated cells versus movement bout duration (**Supplementary Figure 2A**) or average speed (**Supplementary Figure 2B**), and kinematic measures vs. age (**Supplementary Figures 1A–D**).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal study was approved by the Boston University Institutional Animal Care and Use Committee. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

RM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review and editing. MA: Data curation, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing—review and editing. MO'C: Methodology, Resources, Writing—review and editing. AS: Methodology, Software, Writing—review and editing. H-AT: Methodology, Software, Writing—review and editing. SS: Methodology, Writing—review and editing. CZ: Methodology, Software, Writing—review and editing. MK: Methodology, Writing—review and editing. EB: Software, Writing—review and editing. EA: Software, Writing—review and editing. H-YM: Methodology, Writing—review and editing, Resources, Validation, Writing—review and editing. XH: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. XH acknowledged funding from NSF (CBET-1848029, CIF-1955981, and DIOS-2002971). H-YM acknowledged funding from NIH R01 MH130600, R21 MH133014, and the Harvard/MIT Joint Research Grants Program in Basic Neuroscience. RM acknowledged funding from NIH F31 MH123008-03 and NIH/NIGMS Quantitative Biology and Physiology Fellowship (T32 GM008764) through the Boston University Biomedical Engineering Department. EB acknowledges funding from NSF GRFP DGE-1840990 and NIH Translational Research in Biomaterials Training Grant T32-EB006359. The authors acknowledged computational work done on the Shared Computing Cluster in Boston University's Research Computing Services. The authors acknowledged research support from the Boston University Micro and Nano Imaging Facility (S10OD024993).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1277501/full#supplementary-material

References

Allen Institute for Brain Science (2004a). *Allen Mouse Brain Atlas ISH*. Washington, DC: Allen Institute for Brain Science.

Allen Institute for Brain Science (2004b). *Cell Types Database: RNA-Seq Data*. Washington, DC: Allen Institute for Brain Science.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596

Bruining, H., Hardstone, R., Juarez-Martinez, E., Sprengers, J., Avramiea, A., Simpraga, S., et al. (2020). Measurement of excitation-inhibition ratio in autism spectrum disorder using critical brain dynamics. *Sci. Rep.* 10:9195.

Cantagrel, V., Lossi, A., Boulanger, S., Depetris, D., Mattei, M., Gecz, J., et al. (2004). Disruption of a new X linked gene highly expressed in brain in a family with two mentally retarded males. *J. Med. Genet.* 41, 736–742. doi: 10.1136/jmg.2004.021626

Casanova, M. F. (2004). White matter volume increase and minicolumns in autism. *Ann. Neurol.* 56:453.

Casanova, M. F., Buxhoeveden, D. P., Switala, A. E., and Roy, E. (2002). Minicolumnar pathology in autism. *Neurology* 58, 428–432.

Chaddad, A., Desrosiers, C., Hassan, L., and Tanougast, C. (2017). Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder. *BMC Neurosci.* 18:52. doi: 10.1186/s12868-017-0373-0

Charzewska, A., Rzońca, S., Janeczko, M., Nawara, M., Smyk, M., Bal, J., et al. (2015). A duplication of the whole KIAA2022 gene validates the gene role in the pathogenesis of intellectual disability and autism. *Clin. Genet.* 88, 297–299. doi: 10.1111/cge.12528

Chen, Q., Deister, C., Gao, X., Guo, B., Lynn-Jones, T., Chen, N., et al. (2020). Dysfunction of cortical GABAergic neurons leads to sensory hyper-reactivity in a Shank3 mouse model of ASD. *Nat. Neurosci.* 23, 520–532. doi: 10.1038/s41593-020-0598-6

Cohen, M. R., and Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600. doi: 10.1038/nn.2439

Colgin, L. L. (2013). Mechanisms and functions of theta rhythms. *Annu. Rev. Neurosci.* 36, 295–312.

Crawley, J. N. (2012). Translational animal models of autism and neurodevelopmental disorders. *Dialogues Clin. Neurosci.* 14, 293–305.

Dager, S., Wang, L., Friedman, S., Shaw, D., Constantino, J., Artru, A., et al. (2007). Shape mapping of the hippocampus in young children with autism spectrum disorder. *Am. J. Neuroradiol.* 28, 672–677.

De Lange, I., Helbig, K., Weckhuysen, S., Møller, R., Velinov, M., Dolzhanskaya, N., et al. (2016). De novo mutations of KIAA2022 in females cause intellectual disability and intractable epilepsy. *J. Med. Genet.* 53, 850–858.

Di Martino, A., Yan, C., Li, Q., Denio, E., Castellanos, F., Alaerts, K., et al. (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78

Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L., and Tank, D. W. (2007). Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* 56, 43–57. doi: 10.1016/j.neuron.2007.08.003

English, D., McKenzie, S., Evans, T., Kim, K., Yoon, E., and Buzsáki, G. (2017). Pyramidal cell-interneuron circuit architecture and dynamics in hippocampal networks. *Neuron* 96, 505–520.e7. doi: 10.1016/j.neuron.2017.09.033

Falk, T., Bensch, R., Çiçek, Ö, Abdulkadir, A., Marrakchi, Y., Böhm, A., et al. (2019). U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16, 67–70.

Farach, L. S., and Northrup, H. (2016). KIAA2022 nonsense mutation in a symptomatic female. *Am. J. Med. Genet. A* 170, 703–706. doi: 10.1002/ajmg.a.37479

Fiest, K., Sauro, K., Wiebe, S., Patten, S., Kwon, C., Dykeman, J., et al. (2017). Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. *Neurology* 88, 296–303.

Fruchterman, T. M. J., and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Softw. Pract. Exp.* 21, 1129–1164.

Fuhrmann, F., Justus, D., Sosulina, L., Kaneko, H., Beutel, T., Friedrichs, D., et al. (2015). Locomotion, theta oscillations, and the speed- correlated firing of hippocampal neurons are controlled by a medial septal glutamatergic circuit. *Neuron* 86, 1253–1264. doi: 10.1016/j.neuron.2015.05.001

Gilbert, J., and Man, H.-Y. (2016). The X-linked autism protein KIAA2022/KIDLIA regulates neurite outgrowth via N-Cadherin and d-Catenin Signaling. *eNeuro* 3:ENEURO.0238-16.2016 doi: 10.1523/ENEURO.0238-16.2016

Gilbert, J., O'Connor, M., Templet, S., Moghaddam, M., Di Via Ioschpe, A., Sinclair, A., et al. (2020). NEXMIF/KIDLIA knock-out mouse demonstrates autism-like behaviors, memory deficits, and impairments in synapse formation and function. *J. Neurosci.* 40, 237–254. doi: 10.1523/JNEUROSCI.0222-19.2019

Girgis, R., Minshew, N., Melhem, N., Nutche, J., Keshavan, M., and Hardan, A. (2007). Volumetric alterations of the orbitofrontal cortex in autism. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 31, 41–45.

Gonçalves, J. T., Anstey, J. E., Golshani, P., and Portera-Cailliau, C. (2013). Circuit level defects in the developing neocortex of Fragile X mice. *Nat. Neurosci.* 16, 903–909. doi: 10.1038/nn.3415

Green, S., Rudie, J., Colich, N., Wood, J., Shirinyan, D., Hernandez, L., et al. (2013). Overreactive brain responses to sensory stimuli in youth with autism spectrum disorders. *J. Am. Acad. Child Adolesc. Psychiatry* 52, 1158–1172.

Gritton, H., Howe, W., Romano, M., DiFeliceantonio, A., Kramer, M., Saligrama, V., et al. (2019). Unique contributions of parvalbumin and cholinergic interneurons in organizing striatal networks during movement. *Nat. Neurosci.* 22, 586–597. doi: 10.1038/s41593-019-0341-3

Groen, W., Teluij, M., Buitelaar, J., and Tendolkar, I. (2010). Amygdala and hippocampus enlargement during adolescence in Autism. *J. Am. Acad. Child Adolesc. Psychiatry* 49, 552–560.

Gu, X., Eilam-Stock, T., Zhou, T., Anagnostou, E., Kolevzon, A., Soorya, L., et al. (2015). Autonomic and brain responses associated with empathy deficits in autism spectrum disorder. *Hum. Brain Mapp.* 36, 3323–3338.

Hawrylycz, M., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176.

Huang, L., Ledochowitsch, P., Knoblich, U., Lecoq, J., Murphy, G., Reid, R., et al. (2021). Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *Elife* 10:e51675. doi: 10.7554/eLife.51675

Hulbert, S. W., and Jiang, Y.-H. (2016). Monogenic mouse models of autism spectrum disorders: Common mechanisms and missing links. *Neuroscience* 321, 3–23. doi: 10.1016/j.neuroscience.2015.12.040

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221.

Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., and Minshew, N. J. (2007). Functional and anatomical cortical underconnectivity in autism: Evidence from an fMRI study of an executive function task and corpus callosum morphometry. *Cereb. Cortex* 17, 951–961. doi: 10.1093/cercor/bhl006

Keaveney, M. K., Rahsepar, B., Tseng, H., Fernandez, F., Mount, R., Ta, T., et al. (2020). CaMKIIa-positive interneurons identified via a microRNA-based viral gene targeting strategy. *J. Neurosci.* 40, 9576–9588. doi: 10.1523/JNEUROSCI.2570-19.2020

Kennedy, D. P., and Courchesne, E. (2008). The intrinsic functional organization of the brain is altered in autism. *Neuroimage* 39, 1877–1885.

Krach, S., Kamp-Becker, I., Einhäuser, W., Sommer, J., Frässle, S., Jansen, A., et al. (2015). Evidence from pupillometry and fMRI indicates reduced neural response during vicarious social pain but not physical pain in autism. *Hum. Brain Mapp.* 36, 4730–4744. doi: 10.1002/hbm.22949

Kuroda, Y., Ohashi, I., Naruto, T., Ida, K., Enomoto, Y., Saito, T., et al. (2015). Delineation of the KIAA2022 mutation phenotype: Two patients with X-linked intellectual disability and distinctive features. *Am. J. Med. Genet. A* 167A, 1349–1353. doi: 10.1002/ajmg.a.37002

Lambert, N., Dauve, C., Ranza, E., Makrythanasis, P., Santoni, F., Gimelli, S., et al. (2018). Novel NEXMIF pathogenic variant in a boy with severe autistic features, intellectual disability, and epilepsy, and his mildly affected mother. *J. Hum. Genet.* 63, 847–850. doi: 10.1038/s10038-018-0459-2

Lim, E., Raychaudhuri, S., Sanders, S., Stevens, C., Sabo, A., MacArthur, D., et al. (2013). Rare complete knockouts in humans: Population distribution and significant role in autism spectrum disorders. *Neuron* 77, 235–242. doi: 10.1016/j.neuron.2012.12.029

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* 15, 1498–1505.

Litwin-Kumar, A., Oswald, A. M., Urban, N. N., and Doiron, B. (2011). Balanced synaptic input shapes the correlation between neural spike trains. *PLoS Comput. Biol.* 7:e1002305. doi: 10.1371/journal.pcbi.1002305

Lorenzo, M., Stolte-Dijkstra, I., van Rheenen, P., Smith, R., Scheers, T., and Walia, J. (2018). Clinical spectrum of KIAA2022 pathogenic variants in males: Case report of two boys with KIAA2022 pathogenic variants and review of the literature. *Am. J. Med. Genet. A* 176, 1455–1462. doi: 10.1002/ajmg.a.38667

Lukmanji, S., Manji, S., Kadhim, S., Sauro, K., Wirrell, E., Kwon, C., et al. (2019). The co-occurrence of epilepsy and autism: A systematic review. *Epilepsy Behav.* 98, 238–248.

Maenner, M., Warren, Z., Williams, A., Amoakohene, E., Bakian, A., Bilder, D., et al. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 Sites, United States, 2020. *MMWR Surv. Summ.* 72, 1–14.

Mcintosh, A. R., Kovacevic, N., and Itier, R. J. (2008). Increased brain signal variability accompanies lower behavioral variability in development. *PLoS Comput. Biol.* 4:e1000106. doi: 10.1371/journal.pcbi.1000106

McMahon, S. M., and Jackson, M. B. (2018). An inconvenient truth: Calcium sensors are calcium buffers. *Trends Neurosci.* 41, 880–884. doi: 10.1016/j.tins.2018.09.005

Middleton, J. W., Omar, C., Doiron, B., and Simons, D. J. (2012). Neural correlation is stimulus modulated by feedforward inhibitory circuitry. *J. Neurosci.* 32, 506–518. doi: 10.1523/JNEUROSCI.3474-11.2012

Mount, R., Sridhar, S., Hansen, K., Mohammed, A., Abdulkerim, M., Kessel, R., et al. (2021). Distinct neuronal populations contribute to trace conditioning and extinction learning in the hippocampal CA1. *Elife* 10:e56491. doi: 10.7554/ELIFE.56491

Murray, J., Anticevic, A., Gancsos, M., Ichinose, M., Corlett, P., Krystal, J., et al. (2014). Linking microcircuit dysfunction to cognitive impairment: Effects of disinhibition associated with schizophrenia in a cortical working memory model. *Cereb. Cortex* 24, 859–872. doi: 10.1093/cercor/bhs370

Panda, P. K., Sharawat, I. K., Joshi, K., Dawman, L., and Bolia, R. (2020). Clinical spectrum of KIAA2022/NEXMIF pathogenic variants in males and females: Report of three patients from Indian kindred with a review of published patients. *Brain Dev.* 42, 646–654. doi: 10.1016/j.braindev.2020.06.005

Reinhardt, V., Iosif, A., Libero, L., Heath, B., Rogers, S., Ferrer, E., et al. (2020). Understanding hippocampal development in young children with autism spectrum disorder. *J. Am. Acad. Child Adolesc. Psychiatry* 59, 1069–1079.

Romano, M., Bucklin, M., Gritton, H., Mehrotra, D., Kessel, R., and Han, X. A. (2019). Teensy microcontroller-based interface for optical imaging camera control during behavioral experiments. *J. Neurosci. Methods* 320, 107–115. doi: 10.1016/j.jneumeth.2019.03.019

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *ArXiv[Preprint].* doi: 10.1109/ACCESS.2021.3053408

Rubin, R., Abbott, L. F., and Sompolinsky, H. (2017). Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9366–E9375. doi: 10.1073/pnas.1705841114

Satterstrom, F., Kosmicki, J., Wang, J., Breen, M., De Rubeis, S., An, J., et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* 180, 568–584.e23. doi: 10.1016/j.cell.2019.12.036

Selimbeyoglu, A., Kim, C., Inoue, M., Lee, S., Hong, A., Kauvar, I., et al. (2017). Modulation of prefrontal cortex excitation/inhibition balance rescues social behavior in CNTNAP2-deficient mice. *Sci. Transl. Med.* 9:eaah6733. doi: 10.1126/scitranslmed.aah6733

Shen, S., Tseng, H., Hansen, K., Wu, R., Gritton, H., Si, J., et al. (2018). Automatic cell segmentation by adaptive thresholding (ACSAT) for large-scale calcium imaging datasets. *eNeuro* 5:ENEURO.0056-18.2018. doi: 10.1523/ENEURO.0056-18.2018

Stamberger, H., Hammer, T., Gardella, E., Vlaskamp, D., Bertelsen, B., Mandelstam, S., et al. (2020). NEXMIF encephalopathy: An X-linked disorder with male and female phenotypic patterns. *Genet. Med.* 23, 363–373.

Tye, C., Runicles, A. K., Whitehouse, A. J. O., and Alvares, G. A. (2019). Characterizing the interplay between autism spectrum disorder and comorbid medical conditions: An integrative review. *Front. Psychiatry* 9:751. doi: 10.3389/fpsyt.2018.00751

Van Maldergem, L., Hou, Q., Kalscheuer, V., Rio, M., Doco-Fenzy, M., Medeira, A., et al. (2013). Loss of function of KIAA2022 causes mild to severe intellectual disability with an autism spectrum disorder and impairs neurite outgrowth. *Hum. Mol. Genet.* 22, 3306–3314. doi: 10.1093/hmg/ddt187

Vanderwolf, C. (1969). Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr. Clin. Neurophysiol.* 26, 407–428.

Wang, J., Barstein, J., Ethridge, L., Mosconi, M., Takarae, Y., and Sweeney, J. (2013). Resting state EEG abnormalities in autism spectrum disorders. *J. Neurodev. Disord.* 5:24.

Wang, T., Hoekzema, K., Vecchio, D., Wu, H., and Eichler, E. E. (2020). Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nat. Commun.* 11:4932.

Webster, R., Cho, M., Retterer, K., Millan, F., Nowak, C., Douglas, J., et al. (2017). De novo loss of function mutations in KIAA2022 are associated with epilepsy and neurodevelopmental delay in females. *Clin. Genet.* 91, 756–763. doi: 10.1111/cge.12854

Wuchty, S., and Stadler, P. F. (2003). Centers of complex networks. *J. Theor. Biol.* 223, 45–53.

Xiao, S., Gritton, H., Tseng, H., Zemel, D., Han, X., and Mertz, J. (2020). High-contrast multifocus microscopy with a single camera and z-splitter prism. *Optica* 7:1486. doi: 10.1364/optica.404678

Yao, Z., van Velthoven, C., Nguyen, T., Goldy, J., Sedeno-Cortes, A., Baftizadeh, F., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* 184, 3222–3241.e26.

Yizhar, O., Fenno, L., Prigge, M., Schneider, F., Davidson, T., O'Shea, D., et al. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* 477, 171–178. doi: 10.1038/nature10360

Yuen, R., Merico, D., Bookman, M., Thiruvahindrapuram, B., Patel, R., Whitney, J., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* 20, 602–611.

Zemel, D., Gritton, H., Cheung, C., Shankar, S., Kramer, M., and Han, X. (2022). Dopamine depletion selectively disrupts interactions between striatal neuron subtypes and LFP oscillations. *Cell Rep* 38:110265 doi: 10.1016/j.celrep.2021.110265

Zhou, Y, Sharma, J, Ke, Q, Landman, R, Yuan, J, Chen, H, et al. (2019). Atypical behaviour and connectivity in SHANK3-mutant macaques. *Nature* 570, 326–331. doi: 10.1038/s41586-019-1278-0

Zikopoulos, B., and Barbas, H. (2013). Altered neural connectivity in excitatory and inhibitory cortical circuits in autism. *Front. Hum. Neurosci.* 7:609. doi: 10.3389/fnhum.2013.00609

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143.

TYPE Original Research
PUBLISHED 04 January 2024
DOI 10.3389/fnins.2023.1234215

Immature functional development of lumbar locomotor networks in adult *Irf8*$^{-/-}$ mice

Itaru Yazawa[1,2]*, Yuko Yoshida[3], Ryusuke Yoshimi[3,4] and Keiko Ozato[3]

[1]Department of Food and Nutrition, Kyushu Nutrition Welfare University, Kitakyushu, Japan, [2]Laboratory of Neural Control, National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health (NIH), Bethesda, MD, United States, [3]Division of Developmental Biology, National Institute of Child Health and Human Development (NICHD), NIH, Bethesda, MD, United States, [4]Department of Stem Cell and Immune Regulation, Graduate School of Medicine, Yokohama City University, Yokohama, Japan

OPEN ACCESS

EDITED BY
Jacob Raber,
Oregon Health and Science University,
United States

REVIEWED BY
Laura Ferrucci,
Sapienza University of Rome, Italy
Patrick John Whelan,
University of Calgary, Canada

*CORRESPONDENCE
Itaru Yazawa
✉ iyninds@gmail.com

RECEIVED 03 June 2023
ACCEPTED 11 December 2023
PUBLISHED 04 January 2024

CITATION
Yazawa I, Yoshida Y, Yoshimi R and Ozato K (2024) Immature functional development of lumbar locomotor networks in adult *Irf8*$^{-/-}$ mice.
Front. Neurosci. 17:1234215.
doi: 10.3389/fnins.2023.1234215

To date, research on the role of the brainstem and spinal cord in motor behavior has relied on *in vitro* preparations of the neonatal rodent spinal cord, with or without the brainstem; their spatial and temporal scope are subject to technical limitations imposed by low oxygen tension in deep tissues. Therefore, we created an arterially perfused *in situ* preparation that allowed us to investigate functional interactions in the CNS from the neonatal to adult period. Decerebrated rodents were kept alive via total artificial cardiopulmonary bypass for extracorporeal circulation; the plasma oxygen and ion components needed for survival were supplied through the blood vessels. Interferon regulatory factor 8 (IRF8) is a transcription factor that promotes myeloid cell development and stimulates innate immune responses. In the brain, IRF8 is expressed only in microglia and directs the expression of many genes that serve microglial functions. Recent evidence indicates that IRF8 affects behavior and modulates Alzheimer's disease progression in a mouse model. However, whether this immune deficiency arising from the absence of IRF8 influences the development of the neuronal network in the spinal cord is unknown. We applied the above methodology to mice of all ages and electrophysiologically explored whether the absence of IRF8 influences the development of lumbar central pattern generator (CPG) networks. In mice of all ages, bilateral neuronal discharges by the normal CPG networks activated by the modulated sympathetic tone via descending pathways at high flow rates became organized into discharge episodes punctuated by periods of quiescence. Similar discharge episodes were generated by the adult CPG networks (≥P14 days) activated by drug application. However, discharge episodes elicited by activating the neonatal-juvenile CPG networks (<P14 days) occurred alternately on the left and right sides. Interestingly, discharge episodes elicited by the CPG networks in adult IRF8 knockout mice (P11–12 weeks) consisted of those elicited by the CPG networks of both periods. Thus, it was suggested that growing up with immunodeficiency due to loss of IRF8 might interfere with the normal development of functions exerted by the lumbar CPG network because IRF8 plays a role in the normal development of the lumbar CPG network.

KEYWORDS

adult interferon regulatory factor 8 (IRF8)-deficient mice, decerebrate and arterially perfused *in situ* preparations, immature adult lumbar CPG, discharge episodes, rhythmic discharge

Introduction

To date, most research on the role of the brainstem and spinal cord in motor behavior has relied on *in vitro* preparations of the neonatal rodent spinal cord, with or without the brainstem; their spatial and temporal scope are subject to technical limitations imposed by low oxygen tension in deep tissues (St John 1985; Brockhaus et al., 1993; Wang et al., 1996; Wilson et al., 2003; Fong et al., 2008). To overcome this difficulty, we modified the arterially perfused *in situ* preparation, originally developed by Pickering and Paton (2006), which allows us to investigate functional interactions in the central nervous system (CNS), especially between the brainstem and the lower spinal cord, from the neonatal to adult period. This preparation can be used to explore unknown autonomous functions and provide clues to their mechanisms, as well as to track functional changes in the CNS around critical periods. In this methodology, decerebrated mice were kept alive via total artificial cardiopulmonary bypass for extracorporeal circulation; the plasma oxygen and ion components needed for survival were supplied by blood vessels (Yazawa, 2014).

Interferon regulatory factor 8 (IRF8) is a transcription factor that promotes myeloid cell development and stimulates innate immune responses (Tamura et al., 2000; McLellan et al., 2002; Tamura and Ozato, 2002; Yamanaka et al., 2008). In the brain, IRF8 is expressed only in microglia and directs the expression of many genes that serve microglial functions (Kierdorf et al., 2013). Microglia play a role in regulating the number of neural stem cells from the embryo to the postnatal stage by inducing the apoptosis of unnecessary neural stem cells and neurons and engulfing them (Cunningham et al., 2013; Brown and Neher, 2014). In the process of neural circuit formation after the postnatal stage, microglia contribute to the functional maturation of neural circuit formation by retaining only the necessary synapses among the excess synapses and eliminating the unnecessary synapses (Schafer et al., 2012; Brown and Neher, 2014). IRF8-deficient mice (*Irf8*$^{-/-}$ mice), in which macrophages and microglia are defective in functions, including cytokine production, are known as an animal model for human chronic myeloid leukemia, in which granulocytes (neutrophils) are systemically increased (Holtschke et al., 1996); these mice are recognized as a vital tool for studying the immunological events related to the disease. Masuda et al. showed that microglia expressing IRF8 in the lumbar cord dorsal horn increase after peripheral nerve injury and that IRF8 is needed for cutaneous tactile allodynia, and the perception of pain, revealing that IRF8 in microglia affects neuronal morphology and function (Masuda et al., 2012). Furthermore, it has been shown that IRF8 is expressed in microglia from the embryonic stage and throughout adulthood at similar levels and is thought to direct the development and maintenance of neuronal networks (Kierdorf et al., 2013; Saeki et al., 2023).

However, whether the loss of IRF8-related microglia resulting from the absence of IRF8 influences the development of the neuronal network in the lumbar spinal cord is unknown.

In this study, the above methodology was applied to mice of all ages, and the interplay of the discharge episodes from the left and right peripheral motor nerves resulting from the activation of the lumbar central pattern generator (CPG) networks was examined using electrophysiological techniques to explore whether the absence of IRF8 influences the development of lumbar CPG networks.

Materials and methods

Subjects

Ten female wild-type (WT) mice and 10 female *Irf8*$^{-/-}$ mice on a C57BL/6 background (Jackson Laboratories), aged 11–12 weeks and weighing 15.5–21.4 g, were used in this study, along with 20 male Swiss Webster mice (Taconic Laboratory) aged 5–51 days and weighing 4.1–45.2 g. The experimental protocols were approved by the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute of Child Health and Human Development (NICHD)/National Institutes of Health (NIH) Animal Care and Use Committee.

Decerebrate and arterially perfused *in situ* mouse preparation

Experiments were performed on 5 female WT and 5 female *Irf8*$^{-/-}$ mice on the C57BL/6 background (Jackson Laboratories) aged 11–12 weeks and 10 male Swiss Webster mice (Taconic Laboratory) aged 5–21 days. Mice were sedated by inhalation anesthesia with 5.0% halothane and were intraperitoneally injected with an anesthetic combination of ketamine and xylazine (0.5–1.0 μL/g; ketamine:xylazine ratio = 7:1). The concentration of inhaled halothane was maintained at 1.5–2.0% during surgery. The depth of anesthesia was assessed by respiratory rate and responsiveness to tail pinch.

The same surgical procedure as described in our previous study (see Yazawa, 2014; Yazawa and Shioda, 2015) was then performed to prepare the decerebrate and arterially perfused *in situ* preparation. Mice were fixed in a supine position in a dissection chamber, and a median laparotomy was performed from the xiphoid to the lower abdomen. The stomach, small and large intestines, spleen, and pancreas as well as their dominant vessels were ligated and removed. Then, a thoracotomy was performed to allow us to directly visualize the heart and lungs, and both the pleura and the pericardium were removed after an intracardiac injection of 10 U/L heparin. The preparation was immediately submerged in Ringer's solution infused with a 95% O_2–5% CO_2 gas mixture and maintained at 5–10°C to induce suspended animation. Ringer's solution consisted of the following (in mM): 125 NaCl, 3 KCl, 24 $NaHCO_3$, 1.25 KH_2PO_4, 1.25 $MgSO_4$, 2.5 $CaCl_2$, and 10 D-glucose, equilibrated with 95% O_2–5%

CO_2 at pH 7.40–7.45 at room temperature (Chizh et al., 1997). After confirmation of cardiac arrest, a craniotomy was performed. Decerebration was performed with suction at the precollicular level. To prevent fluid from accumulating in the subcutaneous tissue, the skin was removed from the entire body. The bilateral lungs were cut at the level of the lobar bronchi and the apex of the left ventricle was incised.

After the mouse was transported to the recording chamber and then held in a supine position, a double-lumen catheter (Φ 1.0 mm, DL-AS-040; Braintree Scientific, MA, USA) was inserted into the heart through the incision in the left ventricle. To ensure that the perfusate entered the ascending aorta without backing up into the left ventricle, we modified the outer diameter of the tip of the catheter to be slightly larger than the inner diameter. Arterial perfusion was immediately started with carbogen-bubbled Ringer's solution containing an oncotic agent (1.25–1.28% Ficoll-70), heparin (10–20 U/L), and penicillin–streptomycin–neomycin (50 U/L) at room temperature. Finally, the right atrium was incised to maintain the internal pressure of the heart at atmospheric pressure, and the incised part of the left ventricle was then sutured to secure the catheter in the ascending aorta.

After the preparation resumed spontaneous breathing, the muscle relaxant *d*-tubocurarine (2 μM) was added to the perfusate to induce immobilization. The left phrenic nerve (PHN) was identified at the diaphragm level, detached from blood vessels and connective tissues, and severed at the distal end. The left and right peripheral motor nerves were carefully detached at the knee level and severed at their distal ends. Although there was pronounced bradycardia at the initiation of perfusion, ventricular fibrillation never developed.

The same perfusion circuit system as described in our previous studies (see Yazawa, 2014; Yazawa and Shioda, 2015) was used to keep the preparations alive at room temperature. The perfusate, equilibrated with 95% O_2–5% CO_2 at the reservoir, was circulated via the perfusion circuit with a peristaltic pump (model 323 U pump, model 318MC pump head; Watson-Marlow, Wilmington, MA, USA), transfused into the aortic arch of the preparation through bubble traps and net filters (nylon net pore size, 20 μm), and then recycled from the recording chamber back to the reservoir. After the preparation resumed spontaneous breathing, the perfusion flow was always set to >5× the total blood volume (TBV) per minute at room temperature. TBV was calculated as 1/13 (g) of body weight according to the calculation methods described by Mitruka and Rawnsley (1981) and by Harkness and Wagner (1989). In addition, systemic blood pressure was monitored via the second lumen of a double-lumen catheter using a strain-gauge pressure transducer (Pressure Monitor BP-1, WPI, FL, USA). All chemicals used in this study were purchased from Sigma (St. Louis, MO, USA).

Hindlimb preparation

Five female WT mice and 5 female *Irf8*$^{-/-}$ mice on a C57BL/6 background (Jackson Laboratories; aged 11–12 weeks), along with 10 male Swiss Webster mice (Taconic Laboratory; aged 6–51 days), were used to produce hindlimb preparations, which were obtained by transecting decerebrate and arterially perfused *in situ* preparations at the level of the fifth thoracic vertebra. In this preparation, a double-lumen catheter (NCV25GW-200 W; CMP Inc., Tokyo, Japan) was inserted into the descending aorta from the severed end of the thoracic aorta, and the thoracic aorta was ligated at the level of the 6th thoracic vertebra to prevent leakage of perfusate. Arterial perfusion was initiated at 5× TBV/min at room temperature. After spontaneous alternating and synchronous movements were observed in the left and right hindlimbs of the preparation, 1–2 μM ᴅ-tubocurarine was added to the perfusate, and the peripheral motor nerves were detached as described above.

Extracellular recordings

Suction electrodes constructed of polyethylene tubing (PE 50; Becton, Dickinson and Company, Franklin Lakes, NJ, USA) were used to record neuronal discharge from the left PHN, left (L-PN), and right peroneal (R-PN), and left tibial nerves (L-TN) at room temperature. PHN discharge is an indicator of the output derived from the brainstem respiratory center (Barman and Gebber, 1976). PN and TN discharges are indicators of the outputs produced by the CPG network formed between the fourth lumbar and third sacral spinal segments and by the CPG network formed between the fourth lumbar and second sacral spinal segments, respectively. The change in systemic pressure is an indicator of changes in sympathetic tone derived from the cardiovascular center of the brainstem (Coleridge and Coleridge, 1980; Julius and Nesbitt, 1996). The resultant neurograms were amplified ×1,000, filtered at 1–3000 Hz, and digitized using a Digidata 1320A and a Clampex (Axon Instruments, Union City, CA, USA) at sampling rates of 10,000 Hz. All data were saved on the hard disk of a compatible computer for further analysis. Lab Chart 7 software (AD Instruments Inc., Colorado Springs, CO, USA) was used for analysis in this study.

Data analysis

In this study, we used the same data analysis methods as described in our previous studies (see Yazawa and Shioda, 2015). L-PN, R-PN, and L-TN discharges were selected from a recorded sequence, and the integrated waveforms were used to evaluate the phase difference between the two motor nerves. Using methods of circular statistics described by Batschelet (1981), the phase difference between the peak amplitudes of the two neuronal discharges during discharge episodes in the L-PN and R-PN and the L-PN and L-TN were determined. In the phase-shift analysis, each cycle period of L-PN discharge during the discharge episode was measured. Subsequently, the time lag between L-PN and R-PN or L-TN discharges in the cycle period of the L-PN discharge was measured. The phase value was obtained by dividing the time lag between the L-PN and R-PN or L-TN discharges in the cycle period of the L-PN discharge. Each phase value was then multiplied by 360. The values were then plotted on a circle representing the phase difference of possible phases from zero to 360°. Phase values of zero and 360° are equivalent and reflect synchrony; in contrast, 180° represents alternation. The mean phase and the coupling ratio (*r*), which indicates the concentration of phase values around the mean, are shown by the direction and the length of the vector originating from the center of the circle. If the phases of two discharges are strongly coupled, then the phase values will be expected to be highly concentrated around the mean phase. The coupling was considered significant when the Rayleigh test, which determines whether the concentration *r* is sufficiently high to state that coupling was present (Batschelet, 1981), had a *p*-value <0.001. All data compressed to a sample rate of 20 Hz were used.

Results

Perfusion flow dependence of systemic pressure (black), L-PN (Red), R-PN (green), and PHN (gray) discharge

From the results regarding the dependence of systemic pressure and phrenic and peripheral motor nerve discharges under room temperature on the perfusion flow rate in a decerebrate and arterially perfused *in situ* preparation of Swiss Webster mice aged 14–31 days described in a previous work by one of the authors (Yazawa, 2014), the following phenomena were found to be induced: (i) Resumption of spontaneous breathing occurred within 15 min after the onset of perfusion at room temperature. (ii) If the perfusion flow rate was high enough to generate a systemic pressure of >30 mmHg, spontaneous respiration was initiated. Additionally, when the flow rate was set at >5× TBV/min, PHN discharge showed a pattern of increasing amplitude, and its frequency displayed a regular rhythm. (iii) As the flow rate increased further, each neuronal discharge transformed into a discharge episode of increasing frequency and duration, which occurred periodically. (iv) All discharge episodes derived from the three nerves were produced at the same time. (v) When the flow rate was set at >10× TBV/min, three neuronal discharge episodes clearly showed rhythmic discharge patterns. (vi) Small changes in systemic pressure were elicited during and after discharge episodes. In addition, (vii) although an increase in perfusion flow volume caused an increase in oxygen consumption as described in human extracorporeal circulation (Fox et al., 1982; Kirklin and Barratt-Boyes, 1993), increased metabolism also caused an increase in neural activity. The present study is the first to investigate whether the above phenomena, especially (iii) to (vii), are produced even in decerebrate and arterially perfused *in situ* preparations made from 11- to 12-week-old mice on a C57BL/6 background.

Figure 1 represents typical examples of recordings showing the perfusion flow dependence of systemic pressure, L-PN, R-PN, and PHN discharge at room temperature in decerebrate and arterially perfused *in situ* preparations made from adult C57BL/6 mice aged 11–12 weeks. At a high flow rate (>10× TBV/min), each nerve discharge transformed into a discharge episode of increasing frequency and duration, which occurred periodically. Figures 1A,B show the data collected on perfusion flow dependence at 10× and 14× TBV/min, respectively. Simultaneously, the systemic pressure was monitored (upper). The integrated waveforms of the L-PN ($\int$L-PN), R-PN ($\int$R-PN), and PHN ($\int$PHN) discharges are shown in the lower panel. All data were obtained from the same preparation. Asterisks display discharge episodes (yellow-shaded region). The three nerve discharge episodes were produced at approximately the same time. At flow rates of >10× TBV/min, they showed rhythmic discharge patterns during discharge episodes. Several small systemic pressure changes were elicited during discharge episodes. In addition, the frequency of occurrence of the L-PN, R-PN, and PHN discharge episodes increased with increasing flow rates. Similar results to those shown in Figure 1 were reproduced in all the preparations made from mice from the C57BL/6 background aged 11–12 weeks (*n* = 5), indicating that a certain sympathetic tone resulting from an increase in flow rate activated the lumbar CPG network via descending pathways and initiated discharge episodes (see "A decerebrate and arterially perfused *in situ* preparation" section of Discussion).

Discharge episodes in peripheral motor nerves and phase relationships between the L-PN (red) and R-PN (green) and L-PN (red) and L-TN (blue) rhythmic discharge episodes induced at high flow rates in decerebrate and arterially perfused *in situ* preparations made from adult C57BL/6 mice aged 11–12 weeks

In the decerebrate and arterially perfused *in situ* mouse preparations, a certain sympathetic tone resulting from an increase in flow rate is modulated when using high flow rates (>10× TBV/min) at room temperature because the preparation is exposed to a hyperoxic/normocapnic state. Modulated sympathetic tone activates the lumbar CPG network via descending pathways and generates discharge episodes and rhythmic neuronal discharge episodes, and locomotor-like activity is autonomously generated in the hindlimbs of the preparation (Yazawa, 2014). We next investigated the occurrence pattern of the discharge episodes in peripheral motor nerves and phase relationships between the L-PN/R-PN and L-PN/L-TN rhythmic discharge episodes, induced at high flow rates, in preparations made from adult C57BL/6 mice aged 11–12 weeks.

Figures 2A,B show the instances of discharge episodes and rhythmic neuronal discharge episodes from the L-PN and R-PN, induced at flow rates of 14× and 16× TBV/min at room temperature, in decerebrate and arterially perfused *in situ* preparations made from adult WT and *Irf8*$^{-/-}$ C57BL/6 mice. In the preparations from WT mice, the L-PN and R-PN discharge episodes became organized into 'discharge episodes (episodic periods; yellow-shaded regions)' consisting of rhythmic and burst-like discharges punctuated by periods of quiescence (silent periods; blue-shaded regions), with simultaneously repeated episodic and silent periods on both sides (Figures 2A1,2). In the preparations made from adult *Irf8*$^{-/-}$ mice, although the L-PN and R-PN discharge episodes also became organized into 'discharge episodes (episodic periods; yellow-shaded regions)' consisting of rhythmic and burst-like discharges punctuated by periods of quiescence (silent periods; blue-shaded regions), the bilateral neuronal discharge episodes were not necessarily simultaneously repeated episodic and silent periods (Figures 2B1,2. Figures 2A3,B3 display the integrated waveforms of the L-PN ($\int$ L-PN) and R-PN ($\int$R-PN) discharges in regions ⓐ and ⓑ surrounded by dashed lines in Figures 2A2,B2, where rhythmic rather than burst-like discharges occurred. The phase difference between the peak of the integrated waveforms of the L-PN ($\int$ L-PN) and R-PN ($\int$ R-PN) rhythmic discharges in the preparations made from adult WT and *Irf8*$^{-/-}$ mice was approximately 230° (*r* = 0.752) and approximately 240° (*r* = 0.782), respectively. In both cases, the rhythm frequency of elicited left–right alternating discharges remained constant at 1–2 Hz. Similar results to those shown in Figure 2 were reproduced in all preparations made from adult WT (*n* = 5) and *Irf8*$^{-/-}$ mice (*n* = 5).

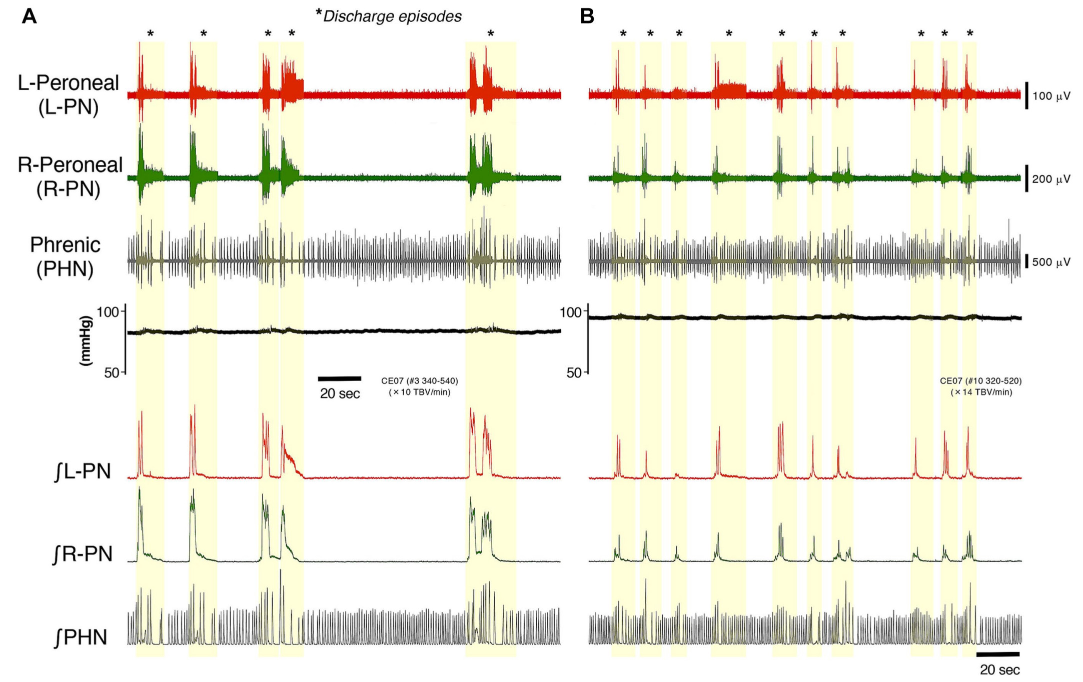


FIGURE 1
Figure represents typical examples of recordings showing the perfusion flow dependence of systemic pressure, L-PN (red) and R-PN (green), and PHN (gray) discharge at room temperature in decerebrate and arterially perfused *in situ* preparations made from mice on the C57BL/6 background aged 11–12 weeks. **(A,B)** Show the data collected on perfusion flow dependence at 10× and 14× TBV/min, respectively. Simultaneously, the systemic pressure (black) was monitored (upper). The integrated waveforms of the L-PN (∫L-PN; red), R-PN (∫R-PN; green), and PHN (∫PHN; gray) discharges are shown in the lower panel. Asterisks show discharge episodes (yellow-shaded region). All data were obtained from the same preparation.

From the above, it was indicated that modulated sympathetic tone activated the lumbar CPG network via descending pathways and generated discharge episodes and rhythmic neuronal discharge episodes and that locomotor-like activity was autonomously generated in the hindlimbs of the preparations made from adult WT and *Irf8*$^{-/-}$ mice aged 11–12 weeks.

Occurrence pattern of discharge episodes in peripheral motor nerves and phase relationships between L-PN (red) and R-PN (green) and between L-PN (red) and L-TN (blue) rhythmic discharge episodes induced by the application of rhythmogenic drugs at a certain flow rate in hindlimb preparations made from adult C57BL/6 mice aged 11–12 weeks

We applied rhythmogenic drugs such as serotonin (5-HT), N-methyl-D, L-aspartate (NMDA), dopamine (DA) and/or noradrenaline (NA) to hindlimb preparations at a certain flow rate and explored whether discharge episodes and neuronal discharge episodes resulting from lumbar CPG network activation, as shown in Figures 2A,B, were induced.

Figures 3A1,B1 show typical examples of neuronal discharges from the L-TN, L-PN, and R-PN induced by the application of 20 μM 5-HT + 10 μM NMDA +1 μM NA to hindlimb preparations made from adult WT and *Irf8*$^{-/-}$ C57BL/6 mice. The perfusion flow rate was set at 7.5× and 8× TBV/min. Asterisks show discharge episodes. The lower panels present expanded views of neuronal discharge episodes of the L-PN, R-PN, and L-TN in the underlined parts of Figures 3A1,B1. It was found that discharge episodes induced in the three nerves repeatedly displayed episodic periods with discharge episodes (yellow-shaded region) and silent periods without discharge episodes (blue-shaded region) and that each occurrence pattern of discharge episodes in the L-PN and R-PN in Figures 3A1,B1 resembled that of discharge episodes shown in Figures 2A2,B2.

Figure 3A2 presents the integrated waveforms of the L-PN (∫L-PN) and L-TN (∫L-TN) discharges in regions ⓐ and ⓑ of the lower panel of Figure 3A1 and shows the L-PN (∫L-PN) and R-PN (∫R-PN) discharges in region © of the same lower panel. Figure 3B2 displays the integrated waveforms of the L-PN (∫L-PN) and L-TN (∫L-TN) discharges in region ⓐ of the lower panel of Figure 3B1 and shows the L-PN (∫L-PN) and R-PN (∫R-PN) discharges of regions ⓑ and © of the same lower panel. The phase difference between the rhythmic discharges in the L-PN and R-PN of the preparations made from adult WT and *Irf8*$^{-/-}$ C57BL/6 mice was approximately 335° ($r = 0.983$) and 260° ($r = 0.677$), respectively. In both cases, the rhythm frequency of elicited left–right alternating discharges remained constant at <4 Hz. Similar results to those shown in Figure 3 were reproduced in all preparations made from adult WT ($n = 5$) and *Irf8*$^{-/-}$ C57BL/6 mice ($n = 5$). On the other hand, the phase difference between the rhythmic discharges in the L-PN and L-TN of the preparations made from adult WT and *Irf8*$^{-/-}$ C57BL/6 mice was approximately 325° ($r = 0.987$) and 320° ($r = 0.987$), respectively. In both cases, the rhythm frequency of elicited flexion-extension-like discharges remained constant at <4 Hz.

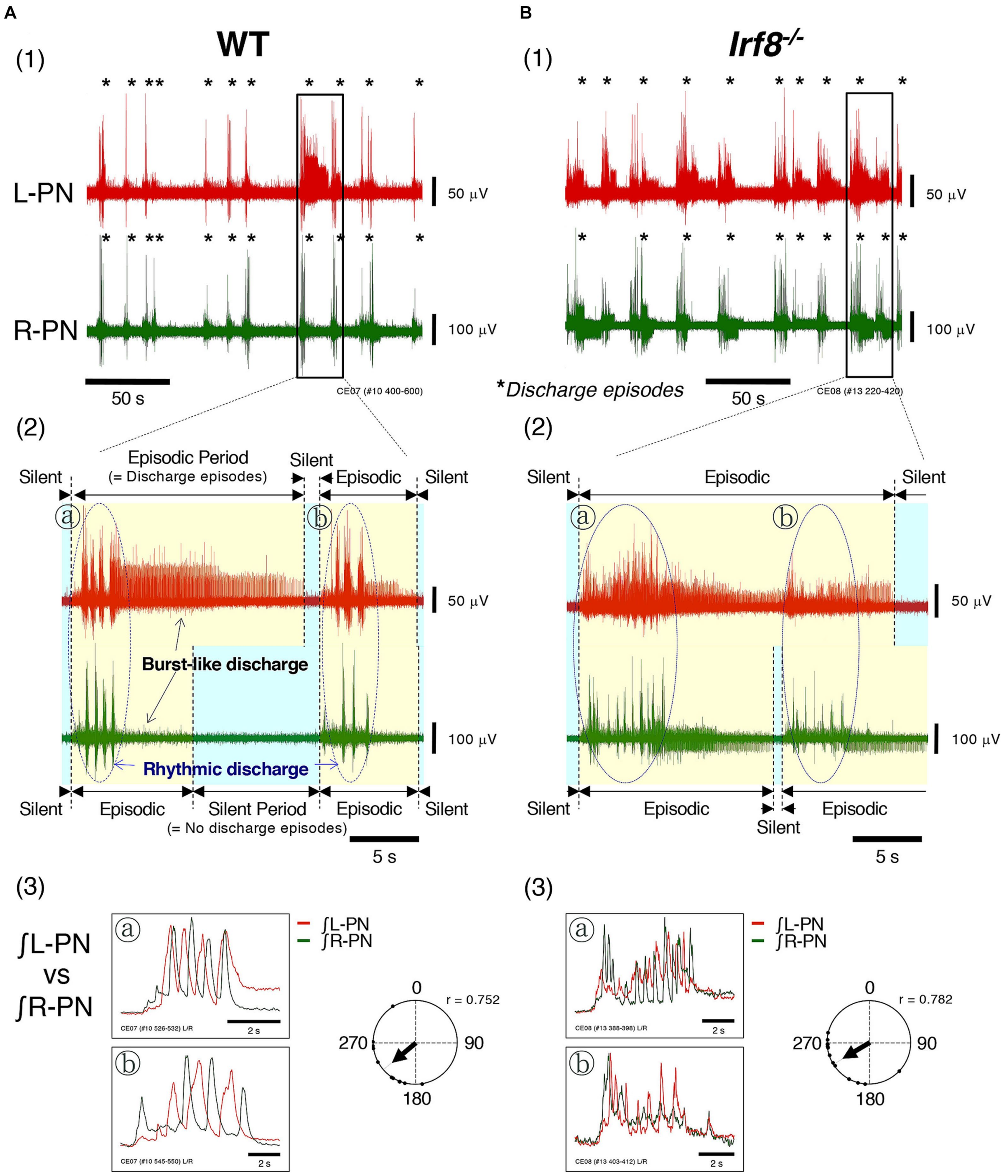


FIGURE 2
Figure shows the instances of discharge episodes and rhythmic neuronal discharge episodes from the bilateral peripheral motor nerves, induced at high flow rates (>10× TBV/min) at room temperature in decerebrate and arterially perfused *in situ* preparations made from adult C57BL/6 mice aged 11–12weeks. **(A1,B1)** Show typical examples of neuronal discharges recorded from L-PN (red) and R-PN (green) at room temperature, induced at 14× and 16× TBV/min, in preparations made from adult WT and *Irf8*$^{-/-}$ C57BL/6 mice aged 11–12weeks. Asterisks show discharge episodes. **(A2,B2)** present enlarged views of the L-PN (red) and R-PN (green) discharges surrounded by the rectangular regions of **(A1,B1)**, where neuronal discharge episodes on both sides became organized into 'discharge episodes consisting of rhythmic and burst-like discharges (episodic periods; yellow-shaded regions)' punctuated by periods of quiescence (silent periods; blue-shaded regions), and the bilateral neuronal discharge episodes were simultaneously repeated episodic and silent periods. **(A3,B3)** Display the integrated waveforms of the L-PN (∫L-PN; red) and R-PN (∫R-PN; green) discharges in regions ⓐ and ⓑ surrounded by dashed lines of **(A2,B2)**, where rhythmic rather than burst-like discharges occur. Each data point shown in **(A,B)** was obtained from the same preparation. Circular statics were used to determine the phase difference from 0 to 360° between the instances of the rhythmic discharges in the L-PN and R-PN discharge episodes (*n* =5). The phase difference between the rhythmic discharges in the L-PN (red) and R-PN (green) of preparations made from WT and *Irf8*$^{-/-}$ mice was approximately 230° (*r* =0.752) and approximately 240° (*r* =0.782), respectively.

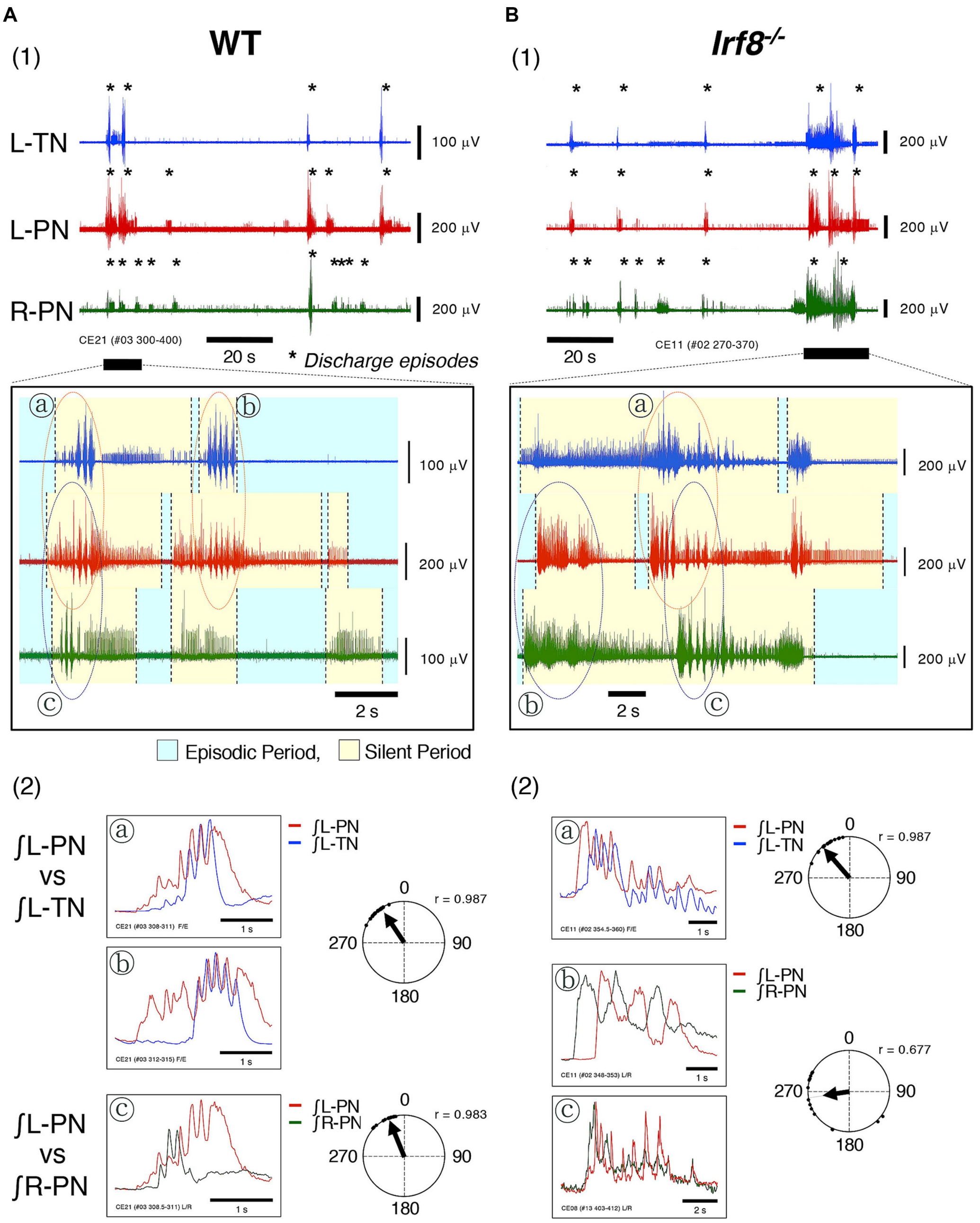


FIGURE 3
Figure shows typical examples of discharge episodes and rhythmic neuronal discharge episodes from the peripheral motor nerves induced by applying rhythmogenic drugs to hindlimb preparations made from adult C57BL/6 mice aged 11–12 weeks at room temperature. Each data point shown in **(A,B)** was obtained from the same preparation. **(A1,B1)** Show typical examples of neuronal discharges from L-TN (blue), L-PN (red), and R-PN (green), induced by applying 20 μM 5-HT + 10 μM NMDA +1 μM noradrenaline (NA), in hindlimb preparations made from adult WT and *Irf8*$^{-/-}$ C57BL/6 mice, in which the perfusion flow rate was set to 7.5× and 8× TBV/min, respectively. Asterisks show discharge episodes. The lower panels present expanded views of neuronal discharge episodes of the L-PN (red), R-PN (green) and L-TN (blue) in the underlined parts of **(A1,B1)**. Discharge episodes induced in these three nerves repeatedly displayed episodic periods with discharge episodes (yellow-shaded region) and silent periods without discharge episodes (blue-shaded region). Each occurrence pattern of discharge episodes in the L-PN (red) and R-PN (green) in **(A1,B1)** resembled that of discharge episodes shown in **(A2,B2)**. ⓐ and ⓑ, surrounded by dashed lines (red) in the lower panel of **(A1)**, show rhythmic discharges in the L-TN (blue) and L-PN (red). ©, surrounded by dashed lines (purple) in the lower panel of **(A1)**, shows rhythmic discharges in the L-PN (red) and R-PN (green). However, ⓐ, surrounded by dashed lines (red) in the lower panel of **(B1)**, shows rhythmic discharges in the L-TN (blue) and L-PN (red). ⓑ and ©, surrounded by dashed lines (purple) in the lower panel of **(B1)**, show rhythmic discharges in the L-PN (red) and R-PN (green). **(A2)** presents the integrated waveforms of the L-PN (∫L-PN; red) and L-TN (∫L-TN; blue) discharges in regions ⓐ and ⓑ in the lower panel of **(A1)** and the L-PN (∫L-PN; red) and R-PN (∫R-PN; green) discharges in region © of the same lower panel. **(B2)** Displays the integrated waveforms of the L-PN (∫L-PN; red) and L-TN (∫L-TN; blue) discharges in region ⓐ in the lower panel of **(B1)** and the L-PN (∫L-PN; red) and R-PN (∫R-PN; green) discharges in regions ⓑ and © of the same lower panel. Circular statics were used to determine the phase difference from 0 to 360° between the instances of rhythmic discharges in the L-PN (red) and L-TN (blue) discharge episodes and the L-PN (red) and R-PN (green) discharge episodes (*n* = 5). The phase difference between the rhythmic discharges in the L-PN (red) and R-PN (green) of preparations made from adult WT and Irf8$^{-/-}$ C57BL/6 mice was approximately 335° (*r* = 0.983) and approximately 260° (*r* = 0.677), respectively. The phase difference between the rhythmic discharges in the L-PN (red) and L-TN (blue) of preparations made from adult WT and *Irf8*$^{-/-}$ C57BL/6 mice was approximately 325° (*r* = 0.987) and approximately 320° (*r* = 0.987), respectively.

Similar results to those shown in Figure 3 were reproduced in all preparations made from adult WT ($n = 5$) and *Irf8*$^{-/-}$ C57BL/6 mice ($n = 5$).

Based on the results shown in Figures 2, 3, the pattern of occurrence of discharge episodes generated by lumbar CPG network activation differed in adult WT and *Irf8*$^{-/-}$ C57BL/6 mice.

Developmental changes in the pattern of occurrence of discharge episodes caused by activation of the lumbar CPG network from the neonatal-juvenile stage (postnatal day < 14) to adulthood (postnatal day ≥ 14)

During the first 2 weeks of life, rodents acquire motor behaviors such as weight bearing and postural reflexes (Clarac et al., 1998). Mice can support their body weight by postnatal day 9 (P9), and many of the walking gait characteristics of mice at postnatal day 14 (P 14) are qualitatively similar to those of adult mice. In addition, the CPG network in the lumbar spinal cord is functionally mature by postnatal days 10–12 (P10–12) and is capable of generating locomotor-like activity (Jiang et al., 1999). During the postnatal period, microglia, which are the major phagocytes in the CNS promote apoptosis, eliminate apoptotic cells, prevent the overproduction of neurons by phagocytosing synapses and neurites, and contribute to the refinement of neuronal circuits (Salter and Stevens, 2017). As microglia mature, they alter their own transcriptional and functional identity as a result of changes in their density and morphology (Zusso et al., 2012). Brain microglia play a specialized role in microglial phagocytosis during development (Matcovitch-Natan et al., 2016; Hammond et al., 2019). However, IRF8-related microglia are normally absent in the lumbar cord dorsal horn of adult *Irf8*$^{-/-}$ mice (Masuda et al., 2012).

To understand the development of functional aspects of the lumbar CPG network caused by the absence of IRF8-related microglia, we examined developmental changes in the pattern of occurrence of discharge episodes caused by activation of the lumbar CPG network from the neonatal-juvenile period to adulthood using Swiss Webster mice (Taconic Laboratory) from 5 to 51 days of age.

Figure 4 shows schematics of the pattern of occurrence of discharge episodes (left) and the discharge patterns during the episodes (right panels) recorded from the L-PN and R-PN in the preparations made from mice after postnatal day five at high flow rates (> 10× TBV/min) at room temperature. In the decerebrate and arterially perfused *in situ* preparations made from mice aged 5–21 days, the bilateral neuronal discharge cycled between episodic periods with discharge episodes and silent periods without discharge episodes. They clearly showed rhythmic discharge episodes and represented a left–right alternating rhythmic discharge pattern beginning with synchronous discharge patterns, and the frequency of elicited left–right alternating rhythmic discharges remained constant at 1–2 Hz (Figure 4A). Similar results to those shown in Figure 4A were reproduced in all 10 preparations (raw data not shown). In hindlimb preparations made from mice aged 14–51 days, after administration of 20–140 μM 5-HT, 10–70 μM NMDA, and 1–5 μM NA, each neuronal discharge transformed into a discharge episode of increasing frequency and duration, which occurred periodically, although bilateral neuronal discharge episodes did not occur at the same time. However, once the neuronal discharge episodes were initiated on both sides, they periodically and repeatedly generated episodic and silent periods. The frequency of left/right alternating discharge episodes in the L-PN and R-PN was <5 Hz (Figure 4B). Similar results were reproduced in all hindlimb preparations ($n = 5$) (raw data not shown). In hindlimb preparations made from mice aged 6–8 days, after administration of 40–120 μM 5-HT, 20–60 μM NMDA, and 40–450 μM DA or 1–3 μM NA, each neuronal discharge transformed into a discharge episode of increasing frequency and duration, which occurred periodically. The neuronal discharge episodes consisted of a rhythmic, burst-like, and then rhythmic discharge (episodic periods) and were always generated on either side. Neuronal discharge episodes on one side displayed a burst-like discharge whenever silent periods were produced on the other side. The frequency of left/right alternating discharge episodes in the L-PN and R-PN was <2 Hz (Figure 4C). Similar results were reproduced in all hindlimb preparations ($n = 5$) (raw data not shown).

Based on the results shown in Figures 2A2,B2, 3A1,B1, 4, the discharge episodes caused by lumbar CPG network activation in adult *Irf8*$^{-/-}$ C57BL/6 mice consisted of discharge episodes caused by activation of the newborn/juvenile and adult lumbar CPG networks, indicating that early-life immunodeficiency due to loss of IRF8 might interfere with the normal development of functions of the lumbar CPG network.

Discussion

In this study, to understand the development of functional aspects of the lumbar CPG network in adult IRF8-deficient mice developing in the absence of IRF8-related microglia in the dorsal horn of the spinal cord, we used decerebrated and arterially perfused *in situ* preparations and extracellular recordings, investigated the developmental changes in the pattern of occurrence of discharge episodes generated by activation of the lumbar CPG network in Swiss Webster mice from the neonatal-juvenile stage to adulthood, and examined the pattern of occurrence of discharge episodes generated by activation of the lumbar CPG network in adult WT and *Irf8*$^{-/-}$ mice on the C57BL/6 background. The results indicated that the discharge episodes exerted by activation of the lumbar CPG network in adult *Irf8*$^{-/-}$ C57BL/6 mice consisted of the discharge episodes exerted by activation of the newborn-juvenile and adult lumbar CPG networks, suggesting the possibility that early-life immunodeficiency due to loss of IRF8 might interfere with the normal development of functions of the lumbar CPG network.

Mechanism(s) of left and right rhythmic activity induced at high flow rates (≥10× TBV/min) at room temperature in the hindlimbs of decerebrated and arterially perfused *in situ* preparations

The decerebrate and arterially perfused *in situ* preparations survived via total artificial cardiopulmonary bypass for extracorporeal circulation, and the oxygen and ion components in the plasma needed for survival were supplied by blood vessels at room temperature.

In this preparation, afferent inputs from mechanosensors of the heart wall (Bishop et al., 1983; Hainsworth, 1991; Hines et al., 1994)

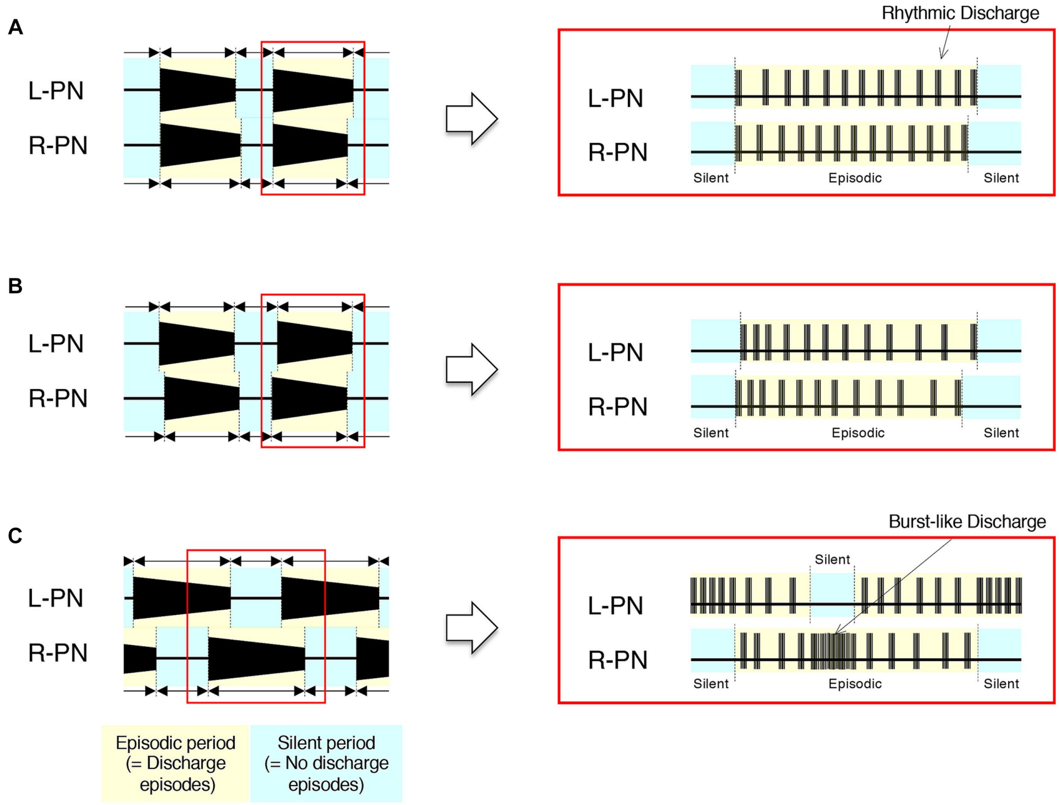


FIGURE 4
Schematics of developmental changes in the pattern of occurrence of discharge episodes (left) and neuronal discharge episodes (right panels) caused by activation of the lumbar CPG network from the neonatal-juvenile stage to adulthood at room temperature. **(A)** Schematics of the typical pattern of occurrence of discharge episodes (left) and the discharge patterns during the discharge episode (right panels) recorded from the L-PN and R-PN in decerebrate and arterially perfused *in situ* preparations made from mice after postnatal day five at high flow rates (>10× TBV/min). In this preparation, neuronal discharges became organized into 'discharge episodes' of increasing frequency and duration, punctuated by periods of quiescence as the flow rate increased, and the bilateral neuronal discharge episodes repeated episodic periods with discharge episodes and silent periods without discharge episodes. At a flow rate of <10× TBV/min, neuronal discharges during discharge episodes showed a burst-like discharge. However, at a flow rate of ≥10× TBV/min, they clearly showed rhythmic discharge episodes and represented a left–right alternating rhythmic discharge pattern beginning with synchronous discharge patterns. Regardless of age, the rhythm frequency of elicited left–right alternating discharges remained constant at 1–2 Hz. Similar results were reproduced in all 10 preparations made from Swiss Webster mice (Taconic Laboratory) aged 5–21 days (raw data not shown). **(B,C)** Represent schematics of the typical pattern of occurrence of discharge episodes (left) and the neuronal discharge patterns during the discharge episode (right panels) recorded from the L-PN and R-PN in hindlimb preparations made on postnatal days 14–51 and 6–8, respectively. In this preparation, neuronal discharge episodes consisting of a rhythmic and burst-like discharge (episodic periods) were generated by applying serotonin (5-HT), *N*-methyl-D, L-aspartate (NMDA), and dopamine (DA) or noradrenaline (NA) to the preparation at a flow rate of 5–7× TBV/min. In **(B)**, after administration of 20–140 μM 5-HT, 10–70 μM NMDA, and 1–5 μM NA, neuronal discharges became organized into episodes punctuated by periods of quiescence. Neuronal discharge episodes did not simultaneously occur on both sides. However, once the neuronal discharge episodes were initiated on both sides, they periodically and repeatedly generated episodic periods with discharge episodes and silent periods without discharge episodes. The frequency of left/right alternating discharge episodes in the L-PN and R-PN was <5 Hz. Similar results were reproduced in all hindlimb preparations (*n* = 5) (raw data not shown). In **(C)**, after administration of 40–120 μM 5-HT, 20–60 μM NMDA, and 40–450 μM DA or 1–3 μM NA, neuronal discharges became organized into episodes punctuated by periods of quiescence. In the preparations made from mice aged 6–8 days, the neuronal discharge episodes consisted of a rhythmic, burst-like, and then rhythmic discharge (episodic periods) and were always generated on either side. Neuronal discharge episodes on one side displayed a burst-like discharge whenever silent periods were produced on the other side. The frequency of left–right alternating discharge episodes in the L-PN and R-PN was <2 Hz. Similar results were reproduced in all hindlimb preparations (*n* = 5) (raw data not shown).

to the cardiovascular center of the brainstem can be ignored because the right atrium was incised to maintain the internal pressure of the heart at atmospheric pressure. Afferent inputs from the stretch receptors of the lungs (Kalia and Sullivan, 1982; Hines et al., 1994) to the respiratory center of the brainstem can be ignored because of the removal of the lungs. In addition, afferent inputs from glomus type I cells on the carotid body that sense thermal changes (Alcayaga et al., 1993) to the respiratory center of the brainstem can be ignored because the preparation was maintained at room temperature. A peristaltic pump was used to provide pressure pulse waves to the baroreceptors of the preparation, as parasympathetic/sympathetic control of vascular resistance via the baroreflex is affected specifically by pulsatile rather than non-pulsatile flow (James and de Burgh Daly, 1970; Chapleau et al., 1989). Furthermore, the effect of the impulse from the central chemoreceptor, the pH/PCO_2 sensor, on the respiratory center of the brainstem can be ignored because the pH of the perfusate was maintained within the physiological range before and after systemic perfusion (Loeschcke, 1982; O'Regan and Majcherczyk, 1982; Nattie, 1998; Ballantyne and Scheid, 2001). Therefore, the homeostasis of this preparation was maintained under the influence of afferent inputs from baroreceptors and peripheral chemoreceptors in the aortic arch and carotid sinus, along with central chemoreceptors distributed on the ventral medullary surface.

After the resumption of spontaneous breathing in the preparation, PHN discharge occurred in a pattern of increasing amplitude, while peripheral motor nerve discharge occurred in a pattern of decreasing amplitude hundreds of milliseconds after the occurrence of PHN discharges. When the flow rate was set at >10× TBV/min, each neuronal discharge transformed into a discharge episode of increasing frequency and duration, which occurred periodically. Discharge episodes in peripheral motor nerves on both sides displayed an alternating pattern of left–right discharge. The physiological condition of the preparation under this flow rate setting was considered to be as follows: Although the sympathetic tone of the preparation increased with increasing perfusion flow volume, the sympathetic tone of the preparation maintained in the hypothermic state was extremely low compared with that of animals maintained at normothermia. The preparation was susceptible to a hyperoxic state due to the high flow rate at room temperature. Thus, when a high flow rate was set, the sympathetic tone seen at the high flow rate (≥10× TBV/min) was easily modulated by afferent input from the peripheral chemoreceptors. Locomotor-like activity, produced by modulated sympathetic tone activating the lumbar CPG network via the spinal descending pathway, was observed in the hindlimbs of the preparation.

Neural networks comprising the lumbar CPG network caused by IRF8-related microglial cell deficiency

Microglia, which are the major phagocytes in the CNS, contribute to the postnatal refinement of neuronal circuits by promoting apoptosis, eliminating apoptotic cells, and preventing the overproduction of neurons (Salter and Stevens, 2017). Brain microglia mature while altering their own transcriptional and functional identity as a result of changes in their density and morphology, and mature microglia play specialized phagocytic roles during development (Zusso et al., 2012; Matcovitch-Natan et al., 2016; Hammond et al., 2019). In addition, microglia settle in different brain regions at varying rates during development and express specific local gene profiles and phenotypes in adulthood. Thus, microglia display spatial heterogeneity in the brain (Schwarz et al., 2012; De Biase et al., 2017; Ayata et al., 2018). IRF8-related microglia in the lumbar cord dorsal horn were found to be absent in adult *Irf8*$^{-/-}$ mice, whereas they were low in adult wild-type mice (Masuda et al., 2012). We speculate, based on the results of the studies described above, that the absence of IRF8-related microglia in the dorsal horn of the spinal cord inhibited the postnatal refinement of the lumbar CPG network and interfered with the normal functional development of the lumbar CPG network. The candidates for the interneurons composing the CPG network that produces locomotor-like activity are a group of interneurons located in L1-L6 near the central canal and the medial middle zone (Kjaerulff et al., 1994). Some dorsally derived interneurons originating from the dorsal horn area migrate ventrally during development, and others migrate ventrally after development (Gross et al., 2002; Lu et al., 2015). Based on this finding and the results shown in Figures 2–4, we propose that immunodeficiency due to loss of IRF8 interferes with the normal development of the inhibitory and excitatory neural circuits and dorsally derived interneurons connecting the bilateral neural networks that constitute the lumbar CPG network.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal study was approved by the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute of Child Health and Human Development (NICHD)/National Institutes of Health (NIH) Animal Care and Use Committee. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

IY conceived and initiated the project, performed the experiments with mice, analyzed the data, steered the entire project, and wrote the manuscript. IY, YY, RY, and KO approved the final version to be published. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the intramural program of the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute of Child Health and Human Development (NICHD)/National Institutes of Health (NIH).

Acknowledgments

The authors wish to thank Shigefumi Yokota (Dept. of Anatomy and Neuroscience, Shimane Univ. School of Med.) for his technical support with the data analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alcayaga, J., Sanhueza, Y., and Zapata, P. (1993). Thermal dependence of chemosensory activity in the carotid body superfused in vitro. *Brain Res.* 600, 103–111. doi: 10.1016/0006-8993(93)90407-e

Ayata, P., Badimon, A., Strasburger, H. J., Duff, M. K., Montgomery, S. E., Loh, Y. H. E., et al. (2018). Epigenetic regulation of brain region-specific microglia clearance activity. *Nat. Neurosci.* 21, 1049–1060. doi: 10.1038/s41593-018-0192-3

Ballantyne, D., and Scheid, P. (2001). Central chemosensitivity of respiration: a brief overview. *Respir. Physiol.* 129, 5–12. doi: 10.1016/s0034-5687(01)00297-3

Barman, S. M., and Gebber, G. L. (1976). Basis for synchronization of sympathetic and phrenic nerve discharges. *Am. J. Phys.* 231, 1601–1607. doi: 10.1152/ajplegacy.1976.231.5.1601

Batschelet, E. (1981). "Circular statistics in biology" in *Mathematics in biology*. eds. R. Sibson and J. E. Cohen (London, UK: Academic Press)

Bishop, V. S., Malliani, A., and Thorén, P. (1983). "Cardiac mechanoreceptors" in *Handbook of physiology*. eds. E. Page, H. A. Fozzard and R. J. Solaro, vol. *2* (MD: American Physiological Society, Bethesda), 497–555.

Brockhaus, J., Ballanyi, K., Smith, J. C., and Richter, D. W. (1993). Microenvironment of respiratory neurons in the in vitro brainstem-spinal cord of neonatal rats. *J. Physiol.* 462, 421–445. doi: 10.1113/jphysiol.1993.sp019562

Brown, G. C., and Neher, J. J. (2014). Microglia phagocytosis of live neurons. *Nat. Rev. Neurosci.* 15, 209–216. doi: 10.1038/nrn3710

Chapleau, M. W., Hajduczok, G., and Abboud, F. M. (1989). Pulsatile activation of baroreceptors causes central facilitation of baroreflex. *Am. J. Phys.* 256, H1735–H1741. doi: 10.1152/ajpheart.1989.256.6.H1735

Chizh, B. A., Headley, P. M., and Paton, J. F. (1997). An arterially-perfused trunk-hindquarters preparation of adult mouse in vitro. *J. Neurosci. Meth.* 76, 177–182. doi: 10.1016/s0165-0270(97)00096-4

Clarac, F., Vinay, L., Cazalets, J. R., Fady, J. C., and Jamon, M. (1998). Role of gravity in the development of posture and locomotion in the neonatal rat. *Brain Res. Brain Res. Rev.* 28, 35–43. doi: 10.1016/s0165-0173(98)00024-1

Coleridge, H. M., and Coleridge, J. C. (1980). Cardiovascular afferents involved in regulation of peripheral vessels. *Annu. Rev. Physiol.* 42, 413–427. doi: 10.1146/annurev.ph.42.030180.002213

Cunningham, C. L., Martínes-Cerdeño, V., and Noctor, S. C. (2013). Microglia regulate the number of neural precursor cells in the developing cerebral cortex. *J. Neurosci.* 33, 4216–4233. doi: 10.1523/JNEUROSCI.3441-12.2013

De Biase, L. M., Schuebel, K. E., Fusfeld, Z. H., Jair, K., Hawes, I. A., Cimbro, R., et al. (2017). Local cues establish and maintain region-specific phenotypes of basal ganglia microglia. *Neuron* 95, 341–356.e6. doi: 10.1016/j.neuron.2017.06.020

Fong, A. Y., Corcoran, A. Z., Zimmer, M. B., Andrade, D. V., and Milsom, W. K. (2008). Respiratory rhythm of brainstem-spinal cord preparations: effects of maturation, age, mass and oxygenation. *Respir. Physiol. Neurobiol.* 164, 429–440. doi: 10.1016/j.resp.2008.09.008

Fox, L. S., Blackstone, E. H., Kirklin, J. W., Stewart, R. W., and Samuelson, P. N. (1982). Relationship of whole body oxygen consumption to perfusion flow rate during hypothermic cardiopulmonary bypass. *J. Thorac. Cardiovasc. Surg.* 83, 239–248. PMID: 6977073. doi: 10.1016/S0022-5223(19)37303-9

Gross, M. K., Dottori, M., and Goulding, M. (2002). Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* 34, 535–549. doi: 10.1016/s0896-6273(02)00690-6

Hainsworth, R. (1991). Reflexes from the heart. *Physiol. Rev.* 71, 617–658. doi: 10.1152/physrev.1991.71.3.617

Hammond, T. R., Dufort, C., Dissing-Olesen, L., Giera, S., Young, A., Wysoker, A., et al. (2019). Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. *Immunity* 50, 253–271.e6. doi: 10.1016/j.immuni.2018.11.004

Harkness, J. E., and Wagner, J. E. (1989). "Biology and husbandry" in *The biology and medicine of rabbits and rodents*. eds. J. E. Harkness and J. E. Wagner (Philadelphia: Lea & Febiger), 372.

Hines, T., Toney, G. M., and Mifflin, S. W. (1994). Responses of neurons in the nucleus tractus solitarius to stimulation of heart and lung receptors in the rat. *Circ. Res.* 74, 1188–1196. doi: 10.1161/01.res.74.6.1188

Holtschke, T., Löhler, J., Kanno, Y., Fehr, T., Giese, N., Rosenbauer, F., et al. (1996). Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. *Cells* 87, 307–317. doi: 10.1016/s0092-8674(00)81348-3

James, J. E., and de Burgh Daly, M. (1970). Comparison of the reflex vasomotor responses to separate and combined stimulation of the carotid sinus and aortic arch baroreceptors by pulsatile and non-pulsatile pressures in the dog. *J. Physiol.* 209, 257–293. doi: 10.1113/jphysiol.1970.sp009165

Jiang, Z., Carlin, K. P., and Brownstone, R. M. (1999). An in vitro functionally mature mouse spinal cord preparation for the study of spinal motor networks. *Brain Res.* 816, 493–499. doi: 10.1016/s0006-8993(98)01199-8

Julius, S., and Nesbitt, S. (1996). Sympathetic overactivity in hypertension. A moving target. *Am. J. Hypertens.* 9, 113S–120S. doi: 10.1016/0895-7061(96)00287-7

Kalia, M., and Sullivan, J. M. (1982). Brainstem projections of sensory and motor components of the vagus nerve in the rat. *J. Comp. Neurol.* 211, 248–264. doi: 10.1002/cne.902110304

Kierdorf, K., Erny, D., Goldmann, T., Sander, V., Schulz, C., Perdiguero, E. G., et al. (2013). Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* 16, 273–280. doi: 10.1038/nn.3318

Kirklin, J. W., and Barratt-Boyes, B. G. (1993). "Hypothermia, circulatory arrest, and cardiopulmonary bypass" in *Cardiac surgery* (New York: Churchill Livingstone), 61–127.

Kjaerulff, O., Barajon, I., and Kiehn, O. (1994). Sulphorhodamine-labelled cells in the neonatal rat spinal cord following chemically induced locomotor activity in vitro. *J. Physiol.* 478, 265–273. doi: 10.1113/jphysiol.1994.sp020248

Loeschcke, H. H. (1982). Central chemosensitivity and the reaction theory. *J. Physiol.* 332, 1–24. doi: 10.1113/jphysiol.1982.sp014397

Lu, D. C., Niu, T., and Alaynick, W. A. (2015). Molecular and cellular development of spinal cord locomotor circuitry. *Front. Mol. Neurosci.* 8:25. doi: 10.3389/fnmol.2015.00025

Masuda, T., Tsuda, M., Yoshinaga, R., Tozaki-Saitoh, H., Ozato, K., Tamura, T., et al. (2012). IRF8 is a critical transcription factor for transforming microglia into a reactive phenotype. *Cell Rep.* 1, 334–340. doi: 10.1016/j.celrep.2012.02.014

Matcovitch-Natan, O., Winter, D. R., Giladi, A., Aguilar, S. V., Spinrad, A., Sarrazin, S., et al. (2016). Microglia development follows a stepwise program to regulate brain homeostasis. *Science* 353:6301. doi: 10.1126/science.aad8670

McLellan, A. D., Kapp, M., Eggert, A., Linden, C., Bommhardt, U., Bröcker, E. B., et al. (2002). Anatomic location and T-cell stimulatory functions of mouse dendritic cell subsets defined by CD4 and CD8 expression. *Blood* 99, 2084–2093. doi: 10.1182/blood.V99.6.2084

Mitruka, B. M., and Rawnsley, H. M. (1981). *Clinical biochemical and hematological reference values in normal experimental animals and normal humans*. Masson: New York, 413.

Nattie, E. E. (1998). "Central chemoreceptors, pH, and respiratory control" in *pH and brain function*. eds. K. Kaila and B. R. Ransom (New York, NY: Wiley-Liss, Inc.), 535–560.

O'Regan, R. G., and Majcherczyk, S. (1982). Role of peripheral chemoreceptors and central chemosensitivity in the regulation of respiration and circulation. *J. Exp. Biol.* 100, 23–40. doi: 10.1242/jeb.100.1.23

Pickering, A. E., and Paton, J. F. (2006). A decerebrate, arterially-perfused *in situ* preparation of rat: utility for the study of autonomic and nociceptive processing. *J. Neurosci. Meth.* 155, 260–271. doi: 10.1016/j.jneumeth.2006.01.011

Saeki, K., Pan, R., Lee, E., Kurotaki, D., and Ozato, K. (2023). IRF8 configures enhancer landscape in postnatal microglia and directs microglia specific transcriptional programs. *BioRxiv*. doi: 10.1101/2023.06.25.546453

Salter, M. W., and Stevens, B. (2017). Microglia emerge as central players in brain disease. *Nat. Med.* 23, 1018–1027. doi: 10.1038/nm.4397

Schafer, D. P., Lehrman, E. K., Kautzman, A. G., Koyama, R., Mardinly, A. R., Yamasaki, R., et al. (2012). Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron* 74, 691–705. doi: 10.1016/j.neuron.2012.03.026

Schwarz, J. M., Sholar, P. W., and Bilbo, S. D. (2012). Sex differences in microglial colonization of the developing rat brain. *J. Neurochem.* 120, 948–963. doi: 10.1111/j.1471-4159.2011.07630.x

St John, W. M. (1985). Medullary regions for neurogenesis of gasping: noeud vital or noeuds vitals? *J. Appl. Physiol.* 81, 1865–1877. doi: 10.1152/jappl.1996.81.5.1865

Tamura, T., Nagamura-Inoue, T., Shmeltzer, Z., Kuwata, T., and Ozato, K. (2000). ICSBP directs bipotential myeloid progenitor cells to differentiate into mature macrophages. *Immunity* 13, 155–165. doi: 10.1016/s1074-7613(00)00016-9

Tamura, T., and Ozato, K. (2002). ICSBP/IRF-8: its regulatory roles in the development of myeloid cells. *J. Interf. Cytokine Res.* 22, 145–152. doi: 10.1089/107999002753452755

Wang, W., Fung, M. L., Darnall, R. A., and St John, W. M. (1996). Characterizations and comparisons of eupnoea and gasping in neonatal rats. *J. Physiol.* 490, 277–292. doi: 10.1113/jphysiol.1996.sp021143

Wilson, R. J., Chersa, T., and Whelan, P. J. (2003). Tissue P_{O2} and the effects of hypoxia on the generation of locomotor-like activity in the *in vitro* spinal cord of the neonatal mouse. *Neurosci.* 117, 183–196. doi: 10.1016/S0306-4522(02)00831-X

Yamanaka, K., Chun, S. J., Boillee, S., Fujimori-Tonou, N., Yamashita, H., Gutmann, D. H., et al. (2008). Astrocytes as determinants of disease progression in inherited amyotrophic lateral sclerosis. *Nat. Neurosci.* 11, 251–253. doi: 10.1038/nn2047

Yazawa, I. (2014). Reciprocal functional interactions between the brainstem and the lower spinal cord. *Front. Neurosci.* 8:124. doi: 10.3389/fnins.2014.00124

Yazawa, I., and Shioda, S. (2015). Reciprocal functional interactions between the respiration/circulation center, the upper spinal cord, and the trigeminal system. *Transl. Neurosci.* 6, 87–102. doi: 10.15151/tnsci-2015-0008

Zusso, M., Methot, L., Lo, R., Greenhalgh, A. D., David, S., and Stifani, S. (2012). Regulation of postnatal forebrain amoeboid microglial cell proliferation and development by the transcription factor runx1. *J. Neurosci.* 32, 11285–11298. doi: 10.1523/JNEUROSCI.6182-11.2012

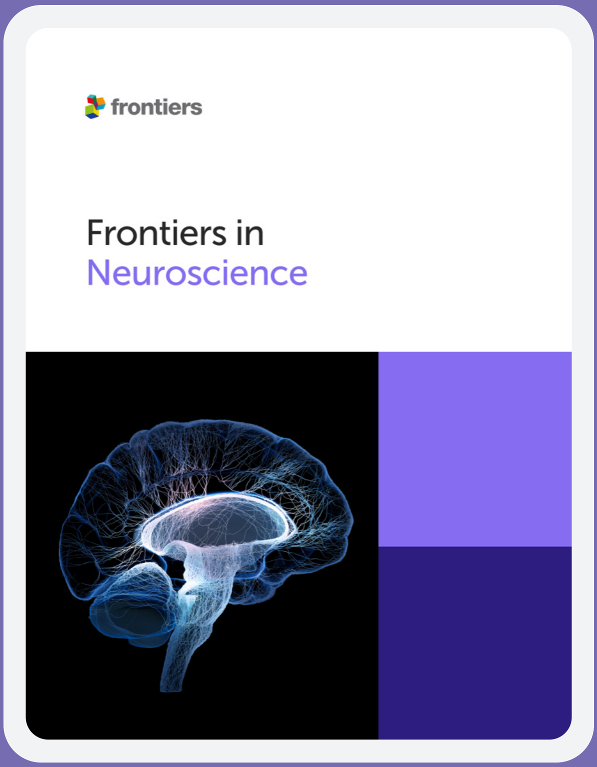
Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



frontiers

Frontiers in Neuroscience

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact