Resting-State fMRI and the Risk of Overinterpretation: Noise, Mechanisms, and a Missing Rosetta Stone

Gang Chen*[a], Zhengchen Cai[b], Konrad P. Kording[c], Thomas T. Liu[d], Joshua Faskowitz[e], Peter A. Bandettini[e], Bharat Biswal[f], and Paul A. Taylor[a]

[a]Scientific and Statistical Computing Core, National Institute of Mental Health, USA
[b]The Neuro (Montreal Neurological Institute-Hospital), McGill University, Canada
[c]Department of Bioengineering and Department of Neuroscience, University of Pennsylvania, USA
[d]Departments of Radiology, Psychiatry, and Bioengineering, University of California San Diego, USA
[e]Section on Functional Imaging Methods, National Institute of Mental Health, USA
[f]Department of Biomedical Engineering, New Jersey Institute of Technology, USA

Abstract

Resting-state fMRI is widely used to describe spontaneous neural activity via correlation-based synchronization measures (e.g., seed-based or network "connectivity", as well as principal and independent component analyses), yet it faces two fundamental obstacles: pervasive non-neural noise and the absence of a definitive "Rosetta Stone" linking the measured BOLD signals to underlying neural events. Although "correlation does not imply causation" has become a cliché, leveraging correlations effectively—and understanding their inferential pitfalls—remains a nuanced challenge. Correlation-based analyses are typically not able to yield causal conclusions, yet they are frequently used to underpin causal narratives in neuroscience research and especially in clinical contexts, which represent a problematic case of overinterpretation.

Using causal inference reasoning, simulations and analytic methods, we address three critical questions when performing resting-state fMRI:
(1) How reliable are correlation estimates for capturing cross-regional synchrony?
(2) What are the consequences of inaccuracies in estimated correlations?
(3) To what extent do estimated correlations reflect causal neural interactions?

We identify two principal pitfalls. First, correlation estimates are systematically distorted, at both the neural and hemodynamic response levels, by diverse noise sources, including variability in neurovascular coupling, producing spurious, suppressed, or even reversed effects. Second, graph-based approaches lack causal interpretability (in the standard counterfactual definition), meaning that even large samples and strong statistical evidence may obscure fundamental ambiguities in what the correlations represent. These underlying interpretational challenges extend to other noninvasive modalities that also utilize correlation-based analyses, such as EEG, MEG, and fNIRS. In light of these challenges, we advocate for three priorities: (1) cautious interpretation that avoids causal overreach, (2) multimodal validation to cross-check findings against independent measures, and (3) enhanced methodological rigor, particularly in biomarker discovery and clinical trials, to ensure that resting-state fMRI provides meaningful insights.

1 Introduction

To give scientific explanations is to show how events and statistical regularities fit into the causal structure of the world. Salmon (1984).

... to explain an event is to provide some information about its causal history. Lewis (1986).

Causal inference seeks to establish cause-and-effect relationships by distinguishing true causal effects from spurious associations, accounting for confounders and biases. Unlike correlations, which capture statistical associations, causal inference grounds its interpretations in a chain of ordered processes that can be directly tested

*Corresponding author. E-mail address: gangchen@mail.nih.gov

and manipulated in a study design. In observational studies, apparent relationships may stem from confounding variables or noise, complicating our ability to derive direct causal conclusions. Such relationships may also result from selection of data or strategies meant to minimize biases. In fields like neuroscience, where understanding mechanisms is fundamental, starting from a causal perspective is thus essential for reliable scientific inference.

In functional magnetic resonance imaging (fMRI), particularly resting-state fMRI, correlations between blood oxygenation level-dependent (BOLD) signals across brain regions are widely used to study brain data, neurological disorders, and cognitive function (Biswal et al., 1995; Lynch et al., 2024; Baldermann et al., 2024; Piramide et al., 2024; Larivière et al., 2020; Greicius et al., 2003). These correlations (often referred to as "functional connectivity"; Friston, 2011), whether derived from seed-based or network analyses or from principal- and independent-component–based methods, are commonly visualized as networks (Appendix A; Fig. 1), and help identify patterns like the default mode network or extract features such as hub structure and centrality (Sporns, 2018; Margulies et al., 2016). However, interpreting these correlations as evidence of neural mechanisms or causal relationships is fraught with challenges due to the indirect nature of the BOLD signal, which is mediated by neurovascular coupling and influenced by non-neural factors like hemodynamic response (HDR) variability and physiological noise (Logothetis, 2008).

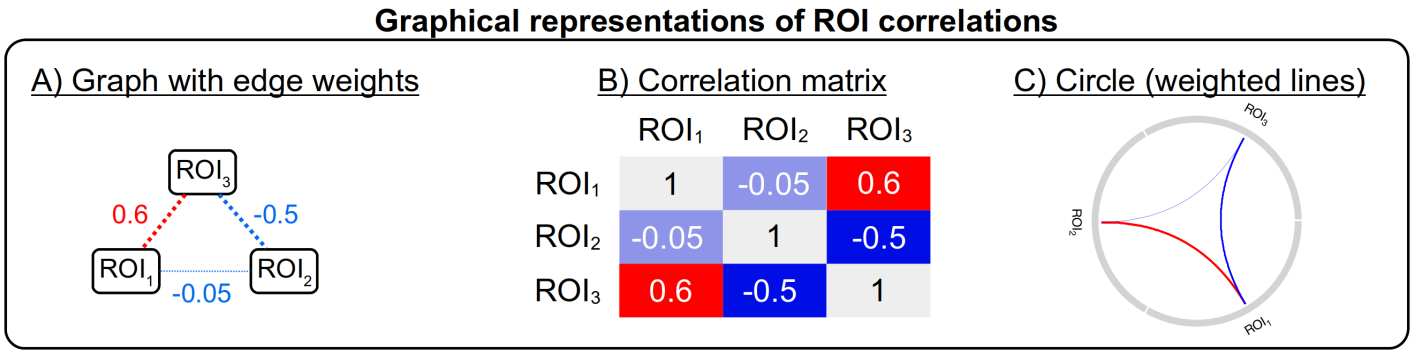

Figure 1: Schematic illustration of associations among three regions of interest (ROIs) in resting-state data analysis. (A) Relationships are depicted as a triangle, with dotted lines representing correlations between ROIs. (B) A correlation matrix among three ROIs is colored to show values and patterns. (C) Correlations are visualized via a Circos plot, with arc thickness indicating magnitude and colors (red/blue) representing positive/negative values.

Resting-state fMRI, with nearly 50,000 publications from 2003 to 2022 (Wei et al., 2024), has become very popular in neuroscience by enabling noninvasive study of brain activity through spontaneous BOLD fluctuations. Yet, its reliance on correlation-based analyses often overlooks critical limitations, such as the influence of HDR, which acts as a low-pass filter of typically undetermined shapes, and exogenous factors like motion or scanner noise. These factors, combined with cross-regional heterogeneity in neurovascular coupling, obscure the link between BOLD signals and neural activity, leading to potential misinterpretations, particularly when correlations are equated with causal or mechanistic insights (Mehler and Kording, 2020; Reid et al., 2019).

Causal inference provides a principled framework for addressing these challenges in neuroimaging by explicitly modeling cause–effect relationships, using tools such as directed acyclic graphs (DAGs) and counterfactual reasoning (Pearl, 2009). In fMRI, the BOLD signal is only an indirect marker of neural activity, mediated by neurovascular coupling, a complex and incompletely understood process involving cerebral blood flow (CBF) and oxygen metabolism ($CMRO_2$) (Fig. 2A). Although individual physiological components can be measured in separate studies (Bortel et al., 2020; Yuen et al., 2019), integrating them is far from straightforward. Measurements are often acquired at different bodily sites using distinct instruments (e.g., pulse oximetry or respiratory belts), and even when secondary signals such as breathing or heart rate are recorded during scanning, it is generally infeasible to capture all relevant physiological processes simultaneously.

As a result, inference must rely on a conceptual mapping between neural events and BOLD responses. We refer to this mapping as a "Rosetta Stone"—a hypothetical construct that, if specified in sufficient detail, would

provide a systematic translation from BOLD fluctuations to underlying average [1] neural activity (Fig. 2B). Crucially, this "Rosetta Stone" is not a one-to-one deterministic mapping from hemodynamic response to neural activity, but rather the missing integrative framework needed to translate between them. Its construction would require unifying biophysical principles, vascular variability, and intermediary physiological processes into a model that enables more grounded causal interpretation despite incomplete information.

In the absence of this Rosetta Stone, the interpretation of resting-state correlations remains intrinsically ambiguous: does a high correlation reflect synchronous neuronal firing, common hemodynamic response, or the influence of shared mediating or exogenous factors? Is it driven by common vascular architecture, or by quirks of the analysis pipeline? Such uncertainty renders interpretations contingent on assumptions that are often implicit and error-prone—even when participant motion is not a concern, the factor that has traditionally drawn the most scrutiny. At present, the field relies on heuristic approximations shaped by prior knowledge, modeling conventions, and biological plausibility, yet these are not anchored in a validated causal framework. Recognizing this limitation should encourage more cautious interpretation of resting-state patterns: acknowledging what cannot be inferred from correlational measures alone, avoiding premature mechanistic claims, and seeking convergent evidence from other modalities or experimental perturbations. It also underscores the need for deeper investigation into the relationship between neural activity and the HDR—in essence, the search for a Rosetta Stone that can translate the region-specific mapping from neural activity to the hemodynamic response (Drew, 2022; Bortel et al., 2020; Yuen et al., 2019).

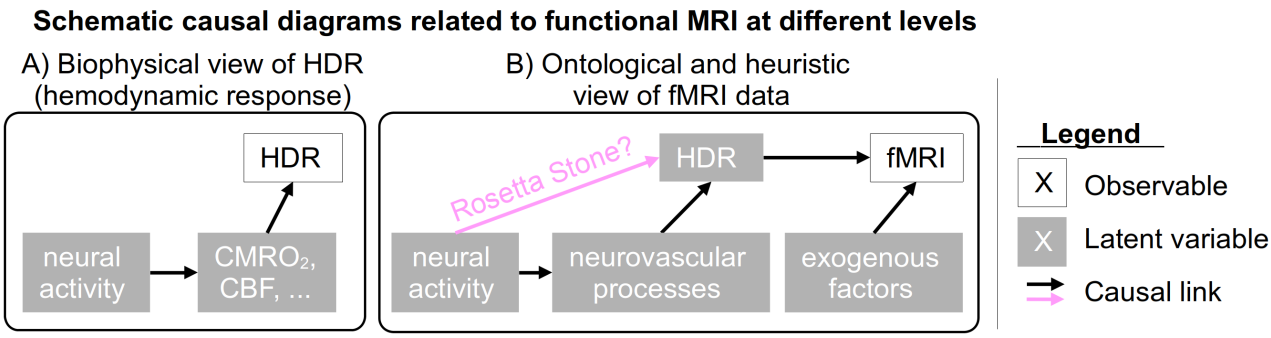

Figure 2: Illustrations of relationships among neural activity, HDR, and fMRI (or BOLD) signal. (A) The HDR reflects metabolic and physiological interactions, including $CMRO_2$ and CBF. (B) The fMRI signal arises from neural activity mediated by neurovascular processes and contaminated by exogenous factors (e.g., physiological fluctuations, motion artifacts). Deciphering this "Rosetta Stone" linking neural activity to fMRI signals remains a central challenge. It is important to recognize that this mapping is not strictly unidirectional: physiological factors such as arterial pressure pulsations may also influence neuronal activity (Jammal Salameh et al., 2024).

Thus, we view researchers and clinicians who utilize resting-state fMRI as balancing two important aspects. First, estimated correlations serve critical purposes, such as identifying consistent brain network patterns, generating hypotheses for biomarkers, and informing network-level analyses (Sporns, 2018). Second, however, the large assumptions about the BOLD signal as an indirect proxy for neural activity, coupled with unaccounted neurovascular heterogeneity and non-neural sources, necessarily limit causal interpretations. This commentary examines these challenges using causal reasoning, simulations, and analytical methods to address three questions:

(1) How reliable are fMRI BOLD correlations in capturing cross-regional neuronal synchrony?
(2) What are the consequences of inaccuracies in estimated correlations?
(3) How much insight can correlations provide about underlying neural relationships?

We focus our investigation on resting-state fMRI, rather than task-based paradigms. Although many of the issues we raise also apply to task-based designs, their impact is substantially smaller because the experimental structure is known and under the researcher's control. Naturalistic paradigms share several challenges with resting-state fMRI, yet their reliance on across-individual correlations reduces the degree to which noise sources are shared and thus limits some forms of confounding factors. Resting-state fMRI remains a valuable and widely

[1] Obviously, individual neurons' activities could never be estimated from low-pass fMRI data. The many unobserved activities will thus always remain confounders for any causal analysis.

used tool, and our aim is not to diminish its utility. Rather, we seek to clarify where interpretational caution is warranted and how unobserved variability can obscure or bias certain inferences, so that the strengths of resting-state data can be leveraged without overlooking their constraints.

All simulation code is available at: `https://github.com/afni/apaper_DAG2`.

2 Cross-region relationships: resting-state data analysis

A resting-state fMRI experiment is observational, unlike task-based experiments that involve stimulus-driven manipulation and randomization. Although correlations are estimated from fMRI signals, the intrinsic relationships occur at the neural level. While correlation analysis does not assume a complete absence of noise, it lacks built-in mechanisms to mitigate its effects, leaving results susceptible to distortion. Therefore, an important issue to resolve is how accurately the underlying neural correlations of interest can be estimated from the proxy fMRI signal.

Additionally, while a correlation value reflects the degree of temporal synchrony between two regions, it is often interpreted as a proxy for the strength of information flow despite lacking directionality and causal validity. It is therefore important to consider both the feasibility and the limitations of such an interpretation. In particular, we aim to address the following questions in relation to Fig. 1:

(1) Does a zero or practically small correlation (e.g., -0.05) imply *no* information flow?

(2) Does a positive (0.6) or negative (-0.05) correlation imply "excitatory" or "inhibitory" information flow?

(3) Does a large (0.6) or small (-0.05) correlation imply strong or weak information flow?

The graphical depiction of causal relationships in Fig. 2B underscores the central challenges in interpreting resting-state data. From the causal inference perspective, it is important to distinquish among the three related but distinct levels involved in the data-generating process: fMRI signal, HDR, and neural activity. While fMRI is often treated as a proxy for neural activity, the measured signal is shaped by additional layers of influence: (i) neurovascular coupling processes—mediators in the causal inference framework—along the path neural activity $\rightarrow$ HDR, and (ii) exogenous factors directly influencing the path exogenous factors $\rightarrow$ fMRI. Without fully modeling or accounting for these pathways, simplified assumptions can yield interpretations of fMRI correlations that diverge markedly from the underlying neuronal dynamics. In this sense, underspecified elements within the neural activity–fMRI causal chain constitute primary sources of potential distortion.

Both mediating and exogenous sources challenge resting-state analysis, but, as discussed below, the inability to characterize the neurovascular processes is especially problematic. Whereas exogenous factors typically contribute random variability that can often be treated as nuisance signals, the neurovascular processes and their spatially and temporally varying dynamics are endogenous[2], systematically shaping the fMRI signal in complex, nonlinear, and poorly understood ways. This challenge is what we refer to as the missing Rosetta Stone problem: the absence of a clear mapping between underlying neuronal activity and observed BOLD signals at fine spatial and temporal scales. In the following, we illustrate the biases that can arise from this gap.

2.1 Impact of noise on estimated correlation accuracy in resting-state fMRI

Resting-state fMRI relies on correlations between BOLD signals to infer cross-regional synchrony, yet these correlations are easily susceptible to non-neural influences, complicating their interpretation as genuine neural relationships (Reid et al., 2019). Such influences, which we refer to broadly as noise, arise from exogenous factors (e.g., head motion, scanner instabilities, physiological fluctuations from respiration and cardiac pulsations) and from neurovascular processes such as variability in neurovascular coupling (Fig. 2B). In this commentary, neurovascular processes further encompasses unmodeled nonlinearities, heterogeneity in vascular anatomy, and spatially variable hemodynamic responses—including their possible dependence on brain "state" (e.g., alertness,

[2]In the current context, an exogenous variable is a root node in the DAG with no parents while an endogenous variable is a node in the DAG that does have parents. Here "exogenous" and "endogenous" are used in the causal inference sense (root vs. non-root nodes), which differs from the econometric usage where exogeneity is defined relative to the model residual term.

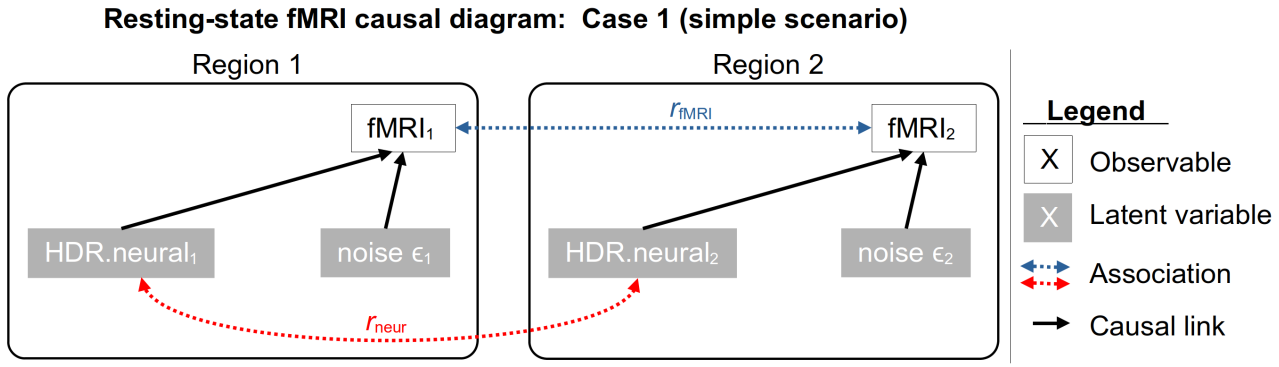

Figure 3: Causal diagram for a simple case of correlation between two regions in resting-state data analysis. In this heuristic model, the fMRI signal is decomposed into two components: a hemodynamic response directly associated with neural activity, and the resultant sum of miscellaneous sources of noise not associated with neural activity that correspond to the endogenous neurovascular processes and exogenous factors in Fig. 2B. The path diagram illustrates relationships among three components from each region: fMRI signal, neural HDR, and noise. A dashed, double-headed arrow represents a statistical association that does not imply causation. The association at the neural level, r_{neur} (shown in red), is unknown, while the estimated association based on the fMRI signal, r_{fMRI} (shown in blue), is our proxy for the neural association r_{neur} but is necessarily subject to inaccuracy due to the presence of the noise components (i.e., ϵ_1 and ϵ_2).

stress, caffeine levels)—all of which remain unknown to the researcher. Together, these noise sources distort the observed fMRI signal, which reflects both the neural component of the hemodynamic response and additive non-neural noise, thereby creating discrepancies between estimated BOLD correlations and the underlying neural correlations (Appendix B).

2.1.1 Case 1: simple noise scenario

We begin with a simple scenario in which noise is uncorrelated with both neural activity and across regions. In this case, the presence of noise systematically reduces the estimated fMRI correlation relative to the true neural correlation (Fig. 3; Table 1 summarizes parameters used here and in the next case). Simulations demonstrate that as the noise magnitude increases relative to the neural signal, the estimated correlation is attenuated in a sigmoid-like fashion (Fig. 4). For instance, when noise and neural signals are of comparable magnitude, the estimated correlation is reduced by about 50%; when noise dominates, correlations can approach zero, effectively obscuring genuine neural relationships. This attenuation effect, long recognized in statistics (Spearman, 1904), aligns with empirical observations of modest individual-level resting-state correlations, typically 0.0–0.3 for cortex–cortex pairs and 0.0–0.2 for subcortical–cortical pairs (Luckett et al., 2023).

Term	Description
HDR.neural_i	Neural component of the hemodynamic response in region i
ϵ_i	Noise component of the fMRI signal in region i (a mixture of exogenous and endogenous sources; if $\rho_{\text{mix}} = 0$, only exogenous sources remain)
r_{neur}	Correlation of neural components HDR.neural across regions; the true neural-level association
ρ_ϵ	Correlation of noise components ϵ across regions (correlation of exogenous sources)
ρ_{mix}	Correlation between HDR.neural and ϵ within or across regions (reflecting endogenous mixing of neural and noise signals)
R_v	Variance ratio of noise to neural components; relative scale of noise
fMRI_i	Observed fMRI signal in region i, combining HDR.neural_i and ϵ_i
r_{fMRI}	Correlation between observed fMRI signals across regions; empirical proxy for neural association

Table 1: Reference table of symbols used in the scenarios described in the main text (Figs. 3–6).

However, real-world fMRI signals are influenced by more complex noise structures, including shared noise sources, region-specific noise levels, and variability in the latency, amplitude, and shape of both neural and hemodynamic responses. As these realistic features are incorporated into data analysis, their heterogeneity can be expected not only to attenuate but also to distort correlation estimates in more intricate ways. The next

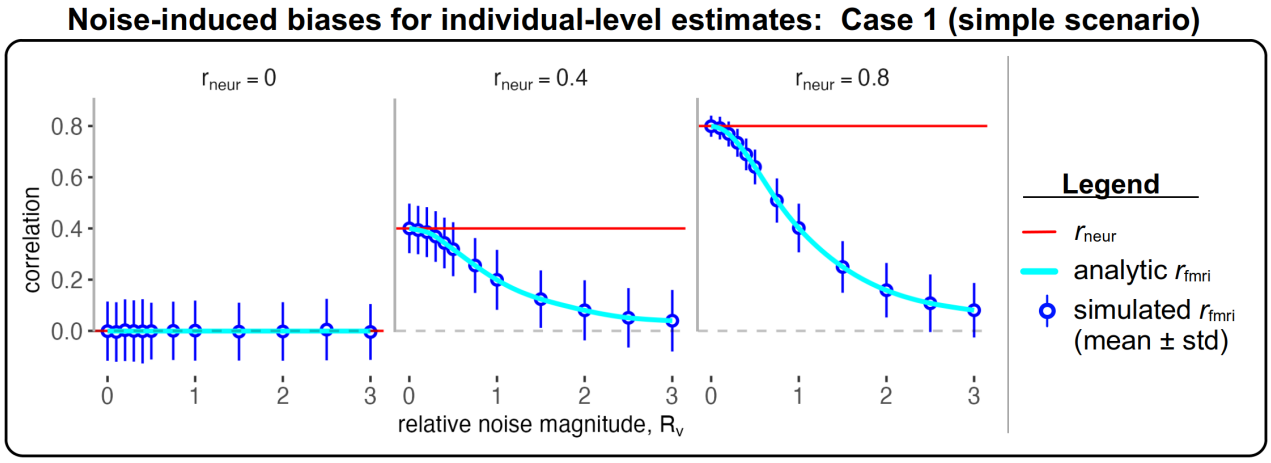

Figure 4: Simulation results for a simple causal scenario: Biases and uncertainties in estimated cross-region correlation r_{fMRI} due to noise, under the setup illustrated in Fig. 3. The three panels correspond to different assumed neural-level correlations r_{neur} (red line) across varying noise magnitudes R_v (x-axis; defined as the ratio of noise to signal variability, see Appendix B). Each fMRI-level simulation included 300 time points (equivalent to 10 minutes of scanning at a 2-second repetition time) and was repeated 500 times. The analytically predicted fMRI correlation r_{fMRI} is shown (cyan line), alongside simulated results (blue circles: mean; vertical lines: ± 1 standard deviation). Simulations and theory align closely. Across all cases, r_{fMRI} values are attenuated toward zero relative to r_{neur}, following an approximate sigmoid function of R_v, with uncertainty generally increasing as noise grows. When $r_{\text{neur}} = 0$, estimates are unbiased but exhibit relatively large, constant uncertainty.

scenario introduces these additional complexities.

2.1.2 Case 2: complexity due to neural-noise and noise-noise correlations

We now consider a more realistic scenario in which noise is correlated with the neural HDR both within and across regions (Fig. 5). Such correlations can readily arise from nonlinear neurovascular coupling or acquisition-induced spatial dependencies (e.g., Fourier transformations in k-space; Kellman and McVeigh, 2005). In this setting, we treat noise–neural "mixing" correlations (ρ_{mix}) as constant within and across regions, while noise–noise correlations (ρ_ϵ) are varied between 0 and 0.4. As shown in Fig. 6, shared noise synchrony produces complex bias patterns: estimated fMRI correlations may be underestimated, overestimated, or even spuriously high when the true neural correlation r_{neur} is zero. In Fig. 6A ($\rho_{\text{mix}} = 0$), r_{fMRI} estimates are biased toward ρ_ϵ. By contrast, in Fig. 6B ($\rho_{\text{mix}} \neq 0$), the direction of bias varies with the noise-to-signal ratio R_v. For example, when neural activity and noise are correlated, cross-regional noise synchrony can partially offset underestimation (upper-middle panel of Fig. 6B vs. Fig. 6A) or induce spurious correlations (upper-left panels), depending on their relative strengths. In addition, noise degrades precision, increasing uncertainty across correlation values, with standard deviations up to 0.1 (Liu, 2016).

These simulation results highlight the challenges of accurately estimating and interpreting resting-state correlations at the individual level. In more realistic scenarios with moderate cross- and within-region noise, bias patterns become complex: underestimation is common, but overestimation or inflation can also occur, consistent with empirical observations (Korponay et al., 2024). These difficulties in estimation extend to the group level, where resting-state analyses are typically applied, such as in studies of brain development, aging, and disorders including Alzheimer's disease and schizophrenia (Yamada et al., 2017; Ibrahim et al., 2021).

How does noise affect group-level comparisons of correlations? Fig. 7 presents group-level simulation results across varying neural–noise mixing (ρ_{mix}), cross-region noise correlations (ρ_ϵ), and noise magnitudes (R_v). In these scenarios, some parameters are identical across groups (black text) while others differ (colored text). The simulations illustrate how noise, and differences in its properties between groups, can produce spurious group differences, suppress true differences, or even reverse them. For instance, variations in noise magnitude, noise synchrony, or neural–noise coupling can generate false group differences, whereas high noise levels may obscure or invert genuine differences. These distortions are also relevant to associations with behavioral traits and to

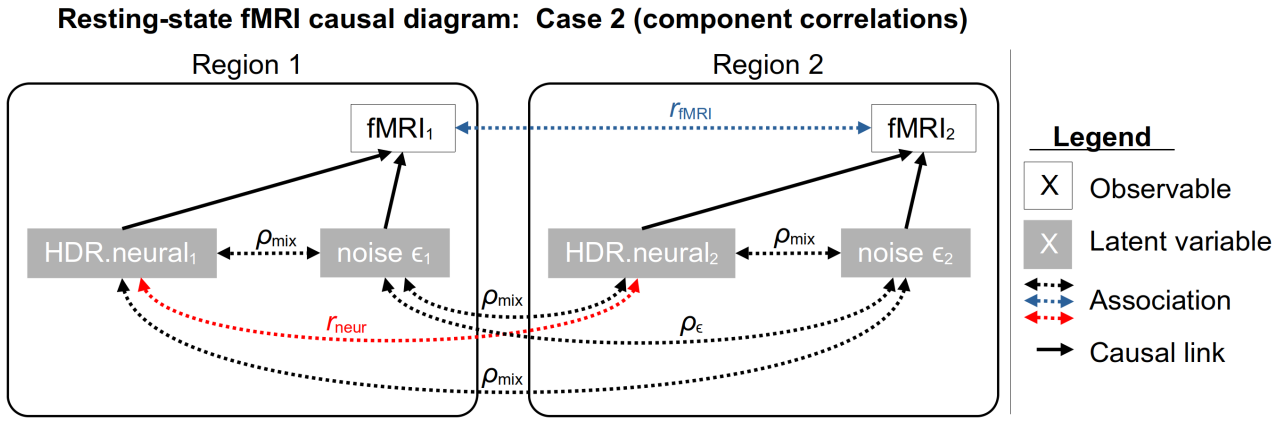

Figure 5: Causal diagram for resting-state fMRI analysis. The fMRI signal is decomposed into two components: the hemodynamic response driven by neural activity and miscellaneous sources of noise. Four types of region-pair correlations are illustrated among three variables from each region (fMRI signal, neural component, and noise): fMRI correlation r_{fMRI}, neural-level correlation r_{neur}, neural–noise correlation ρ_{mix}, and noise–noise correlation ρ_ϵ. Here, noise corresponds to both neurovascular processes and exogenous factors from Fig. 2B. Dashed double-headed arrows indicate statistical associations without implying causation. The true neural-level association, r_{neur} (red), is unknown, whereas the observed fMRI-based association, r_{fMRI} (blue), is affected by interactions with other associations, including neural–noise mixing (ρ_{mix}) and cross-region noise correlations (ρ_ϵ).

naturalistic fMRI data, where inter-individual correlations are subject to similar biases (Taylor et al., 2021).

Interestingly, there are special cases in which biased correlation estimates have a less severe impact. For instance, differences in noise magnitude, neural–noise correlations, or noise–noise correlations across groups do not always produce spuriously strong evidence for a group-level effect. Fig. 8A illustrates such a case: although noise biases the estimated correlation for each group, the pair of biases does not result in a false signal of a strong group difference. Conversely, Fig. 8B shows a scenario in which similar noise parameters introduce differential bias across groups, yet the estimated group-level correlations still preserve the correct direction of the underlying neural difference. It is important to note that these are exceptional cases; in all instances, the individual group estimates remain biased relative to the true neural-level correlations.

2.1.3 Challenges of noise in resting-state fMRI

The persistent presence of noise in resting-state fMRI reflects the complex interplay of neurovascular coupling, physiological processes, and participant state. While various techniques can attenuate some noise sources, complete elimination remains unattainable. Noise-reduction efforts often focus on exogenous factors (Fig. 2), such as head motion and physiological fluctuations, yet these remain difficult to fully mitigate. Progress in addressing the complications due to neurovascular processes (Fig. 2B) has been lacking; current methods cannot entirely decouple their influence on the fMRI signal. For instance, the relationship between neural activity and the BOLD response is demonstrably nonlinear (Zhang et al., 2019), and neural events account for only a small fraction ($\sim$10%) of resting-state BOLD variability (Lu et al., 2019). This establishes a practical lower bound on achievable noise reduction, further complicated by variability across brain regions, individuals, and experimental groups.

Physiological and group-related factors exacerbate these challenges. Age-related changes in neurovascular coupling and autonomic function (e.g., heart rate variability) can mimic or obscure developmental effects (Tsvetanov et al., 2020), while motion artifacts—particularly prevalent in populations such as ADHD or Parkinson's patients—and respiration-induced noise further confound interpretation (Tu and Zhang, 2022). Pharmacological manipulations, including psychedelics, alter vascular dynamics and complicate BOLD signal interpretation (Cardone et al., 2024). These latent sources underscore the need for a "Rosetta Stone" framework capable of translating fMRI correlations into neural activity by accounting for spatial and temporal heterogeneity of both endogenous and exogenous effects (Fig. 2B). Without such a framework, correlation-based analyses risk producing spurious, suppressed, or reversed interpretations, limiting their utility for biomarker discovery or clinical

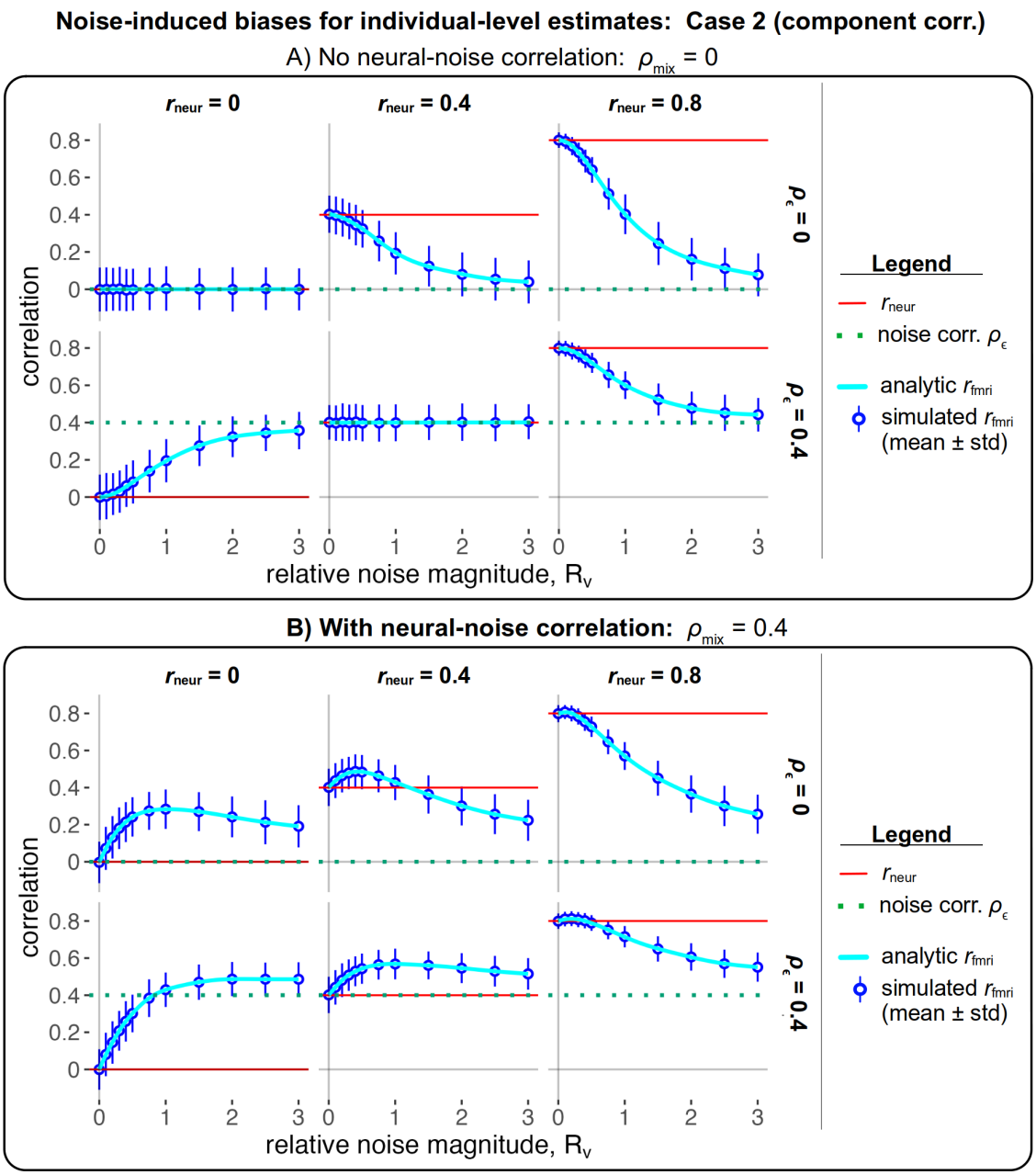

Figure 6: Simulation results for a more complex causal scenario incorporating neural–noise mixing (ρ_{mix}) and noise–noise correlations (ρ_ϵ): biases and uncertainties in cross-region fMRI correlation estimates (r_{fMRI}) under the scenario illustrated in Fig. 5. In each panel, assumed neural-level correlations (r_{neur}, red line) are shown across varying noise magnitudes R_v (x-axis; ratio of noise to signal variability). Each fMRI-level simulation included 300 time points (10 minutes of scanning at 2-second TR) and was repeated 500 times. Analytically predicted fMRI correlations (r_{fMRI}, cyan line) are shown alongside simulated results (blue circles: mean; vertical lines: ± 1 standard deviation); theory and simulation align closely. A) No neural–noise mixing ($\rho_{\text{mix}} = 0$). In the top row ($\rho_\epsilon = 0$, corresponding to Fig. 4), estimated r_{fMRI} values are monotonically biased toward zero. In the bottom row, biases shift monotonically toward the nonzero value of $\rho_\epsilon = 0.4$, consistent with the pattern seen in the top row. B) Neural–noise correlations are included ($\rho_{\text{mix}} = 0.4$), reflecting both neurovascular processes and exogenous factors in Fig. 2B. Biases now appear whenever $R_v > 0$ and follow complex, non-monotonic patterns across the range of noise magnitudes.

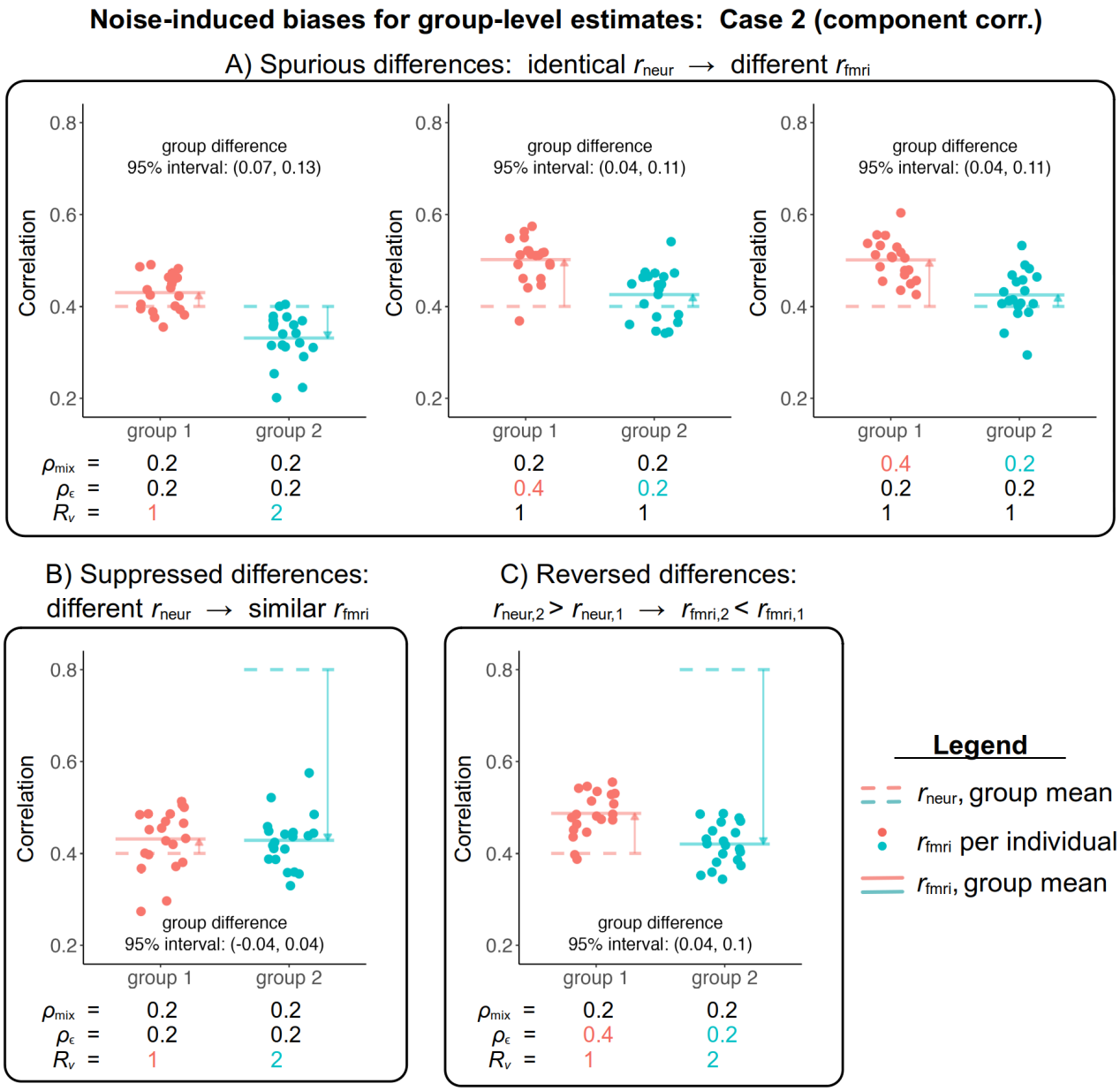

Figure 7: Examples of how underlying neural-level group differences can translate differently at the observed fMRI level. Simulated data for two groups (20 individuals each, 300 time points) illustrate how noise characteristics can produce misleading group comparisons. Individual fMRI correlation values are shown as color-coded dots, jittered for clarity. A) Spurious differences: Despite identical neural-level correlations (dashed line), apparent group differences emerge due to (i) differing noise magnitudes, (ii) differing noise–noise correlations, or (iii) differing neural–noise correlations. B) Suppressed difference: True differences at the neural level are attenuated in the observed fMRI correlations. C) Reversed difference: The observed group difference is inverted. Although neural-level correlations (r_{neur}) are higher on average for group 2, the estimated fMRI correlations (r_{fMRI}) are substantially higher for group 1, leading to a reversed conclusion.

applications (Tu et al., 2024).

The impacts of noise illustrated here—including those shown in Figs. 4, 6, and 7—are derived from hypothetical scenarios where both neural-level correlations and noise amplitudes are known. In real-world studies, however, the ground truth is unavailable. Investigators cannot determine whether a measured correlation (e.g., $r_{\text{fMRI}} = 0.2$) is inflated by noise-neural correlation, attenuated by high noise amplitude, or unbiased. Moreover, noise-neural correlation may vary across different region pairs, resulting in an effectively limitless set of unknown combinations.

Noise-reduction techniques such as bandpass filtering, motion censoring, and physiological signal regression aim to mitigate these biases (Caballero-Gaudes and Reynolds, 2017), but limitations remain. Even with advanced processing, the noise scale often remains comparable to or exceeds the neural-related BOLD signal, producing biased estimates. Methods like global signal removal can introduce further distortions (Murphy and Fox, 2017),

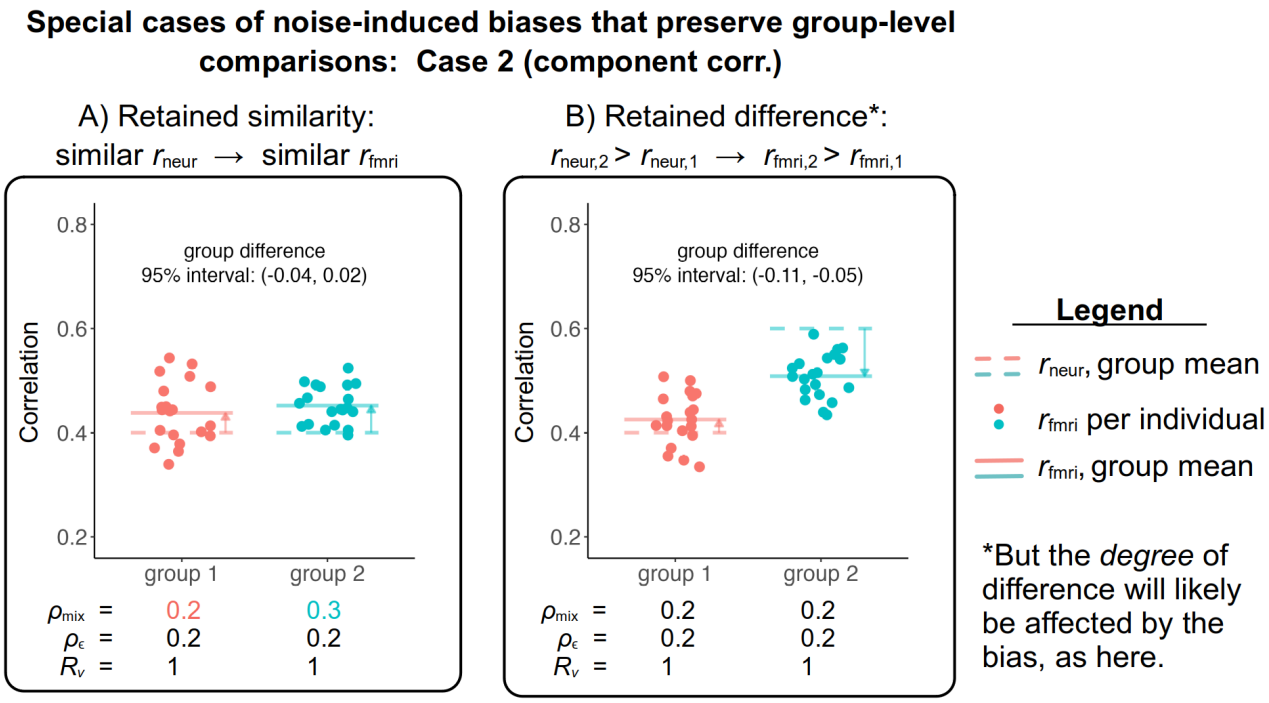

Figure 8: Additional group-level simulation results (see Fig. 7 for details) illustrating special cases in which neural-level group comparisons are preserved in statistical inference despite noise-induced biases. A) Retained similarity: Although neural–noise correlations ρ_{mix} differ, no strong statistical evidence of a group difference emerges at either the neural or fMRI level. B) Retained directionality: The direction (sign) of the group difference is consistent between neural- and fMRI-level estimates, although the magnitude of the difference is smaller for the fMRI-level correlations r_{fMRI}. In both cases, individual group estimates remain biased, and differences in group sizes could alter the statistical outcome.

and neurovascular coupling variability remains largely unaddressed (Rangaprakash et al., 2023). Acquisition parameters, including field strength and voxel size, further influence noise characteristics and inter-regional relationships, amplifying sensitivity to analysis pipeline choices.

An often underappreciated source of distortion arises from participant drowsiness and microsleeps. Even during relatively short scans (5–10 minutes), such events can occur multiple times and substantially distort correlation estimates, particularly in default mode and sensory networks (Tagliazucchi and Laufs, 2014). One proposed strategy is to derive resting-state-like signals from task-based fMRI, under the assumption that participants remain awake throughout. In this approach, modeled task responses, baseline, and motion effects are regressed out, leaving the residual time series as a putative "pure" resting-state signal. However, this procedure is likely unrealistic given the limits of current modeling. Typical approaches fail to fully capture task-related influences, including mis-specification of the hemodynamic response function, unmodeled cognitive demands, tonic baseline shifts associated with resting-state dynamics (as opposed to phasic task-related events) (Grimm et al., 2024), and disruption of spontaneous cognition or default mode activity. Consequently, residualization is unlikely to fully isolate true resting-state activity.

The central challenge in this aspect of resting-state analysis lies in the unknowable extent of these distortions. While noise and resulting bias may not always alter qualitative statistical inferences—as illustrated in Fig.8—there is currently no reliable method to identify which cases are affected or estimate the proportion of "safe" instances. The demonstrated statistical distortions (Fig. 7) underscore the need for caution when interpreting correlation-based analyses.

These concerns are magnified in whole-brain analyses, where parameters vary simultaneously across many regions. Any simplifying assumption that neglects such heterogeneity implicitly assumes uniform behavior across the brain, an unlikely scenario. The application of statistical inference adds another layer of complexity. For example, larger sample sizes are often pursued to detect small or subtle differences in correlation values (Marek et al., 2022). However, this pursuit carries pitfalls. With only 20 individuals per group, Fig. 8A illustrates a case where the lack of strong statistical evidence inferentially preserves the ground truth of no group difference at the

neural level. Yet as sample size increases, even trivial differences in correlation may generate strong statistical evidence, producing the illusion of a meaningful group difference where none exists.

2.2 Lost in causation: challenges of network-level inferences with correlations

Above, we considered typical group comparisons based on individual-level correlation values. We now turn to another widely used approach in resting-state fMRI: correlation-based graphs and networks, viewed through the lens of causal inference. Most causality methods applied to fMRI implicitly assume unconfoundedness or causal sufficiency; that is, all relevant variables are observed. In practice, this assumption is almost certainly violated in brain networks, given the multitude of unobserved neural processes and physiological confounds. Consequently, causal claims drawn from fMRI require considerable caution and ideally should be supported by interventions or convergent lines of evidence.

Graph-based analyses enable network-level inferences by representing the brain as nodes (regions) connected by edges (correlations between nodes). These methods often employ sophisticated mathematical metrics, many of which rely on binarized networks created using arbitrary thresholds. Network features are computed at micro, meso, and macro scales and include measures such as hub coefficient, node centrality, node efficiency, betweenness centrality, closeness centrality, cliques, modularity, community structure, holes or cavities, clustering coefficient (local efficiency), eigenvector centrality, paths, and shortcuts (Medaglia, 2017). Importantly, all these downstream features depend on the estimated r_{fMRI}, which itself is only a proxy for neuronal-level correlations r_{neur}. This raises a critical question: if the fundamental unit of the edge is unreliable, what are the implications for all derived topological measures?

For example, edges in a network are often defined using correlation thresholds (e.g., $r_{\text{fMRI}} \geq 0.1$ or $p \leq 0.001$) or density cutoffs (e.g., top 15% of correlations). Once thresholded, edges are typically discretized, disregarding their original correlation magnitude. Unlike standard model parameters, graph metrics rarely come with measures of uncertainty such as standard errors. This absence masks the precision of the underlying estimates, potentially fostering undue confidence in results derived from inherently noisy data. Together, these factors complicate the interpretation, reproducibility, and comparability of graph-based analyses.

A more fundamental challenge lies in the nature of the edges themselves, independent of noise. Heuristically, it has long been recognized that correlations are difficult to map onto cross-region neural interactions in the presence of confounders (Reid et al., 2019). Here, we extend this understanding with a quantitative examination of the underlying factors and explore additional scenarios that further highlight these limitations.

2.2.1 Case 3: No endogenous noise, yet interpretations remain ambiguous

When considering the correlation between a pair of brain regions, three aspects of their potential causal relationship are of interest: directionality, magnitude, and sign. Even with definitive "Rosetta Stone", establishing directionality is generally unattainable from correlation alone. This limitation raises an important point: even without access to directionality, the remaining components—magnitude and sign—may still carry meaningful information. The critical question, then, is to what extent correlation can provide valid insight into the magnitude and sign of the underlying neural relationship. Put differently, if directionality is set aside, how closely do observed cross-regional correlations reflect the "true" neural-level relationship?

In this case, we consider small networks for analytic simplicity. The associations among these regions, estimated from fMRI data, can be represented as a correlation matrix or a small undirected graph (cf. Fig. 1A–B). For clarity, we make deliberately idealized assumptions: the regions are linearly associated, there are no endogenous noise sources, and therefore the true fMRI correlations (r_{fMRI} in Fig. 5 and Table 1) are estimated with effectively perfect accuracy and precision relative to the underlying neural relationships. One might expect that, under such idealized conditions, fMRI-estimated correlations could be directly interpreted as reflecting neural associations or "connectivity strength." However, even in this simplified scenario, analytical reasoning reveals a degeneracy of interpretations, resulting in ambiguity for all but the smallest networks.

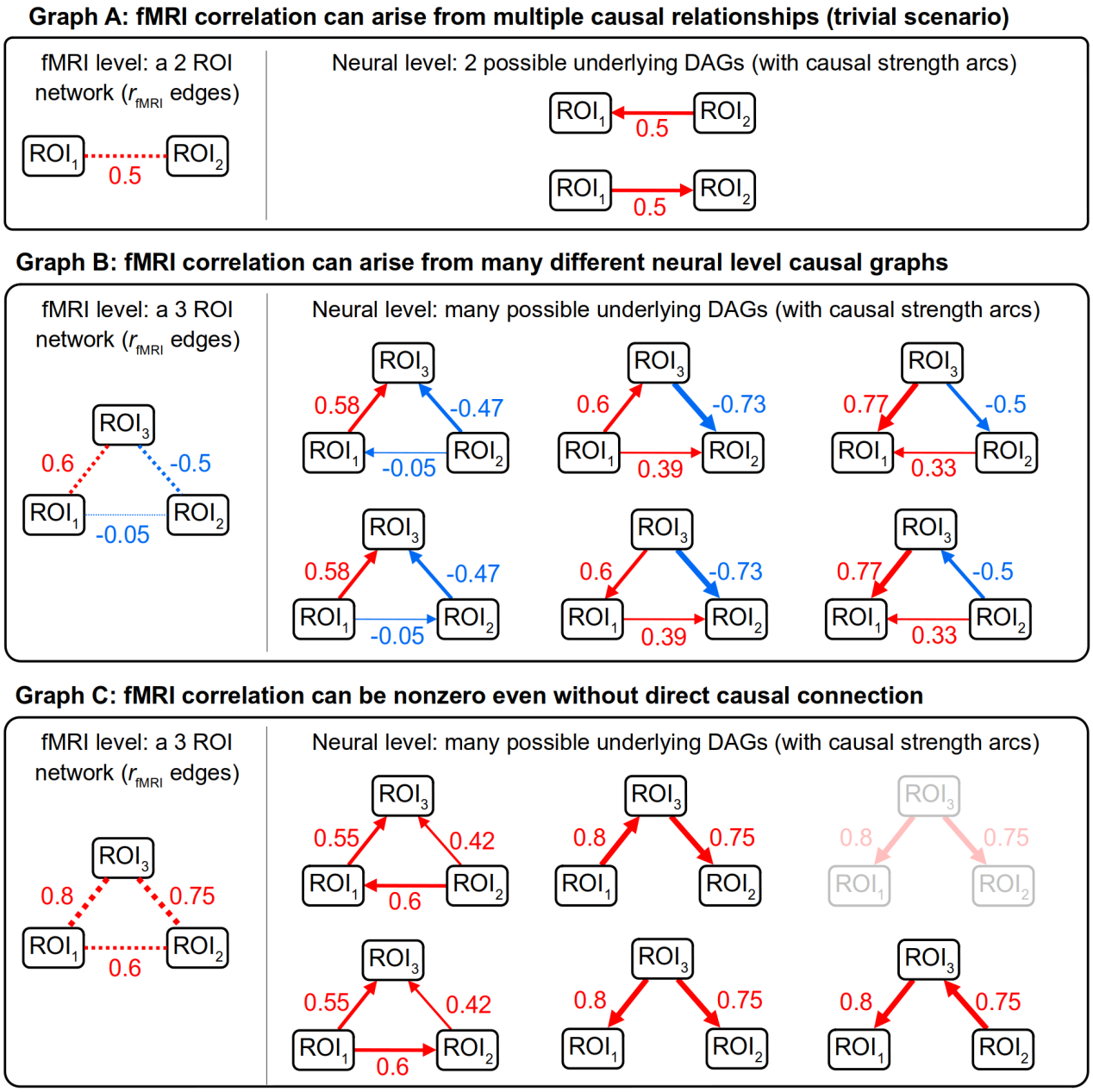

Figure 9: Cross-region correlation versus causation. A single observed correlation pattern in a network can arise from multiple underlying causal structures. In each panel, dotted lines indicate correlations, solid lines indicate directed influences, red denotes positive values, blue denotes negative values, and line thickness reflects magnitude. A) Correlation may reflect causation under special conditions. When there is no nonlinearity and no endogenous noise between two regions, their correlation can directly reflect the underlying association. B) Correlation magnitude does not equal causal strength. Even with the same set of correlations as in Fig. 1, multiple (here, six) acyclic causal graphs may be consistent with the observed correlations. C) Correlation does not imply direct causation. In some cases, two regions (e.g., ROI_1 and ROI_2) may be correlated entirely due to mediation or confounding by a third region (e.g., ROI_3), without any direct causal link. These special scenarios correspond to the heuristic examples discussed in Reid et al. (2019). NB: Here, two DAG solutions are degenerate (i.e., indistinguishable under the observed correlations); one of the equivalent solutions is shown faded.

Fig. 9A illustrates the simplest possible case: a network of two regions with no endogenous noise. For the region pair of ROI_1 and ROI_2, the estimated r_{fMRI} can only be explained by two possible neural-level structures: networks with identical causal strengths but opposite directions of influence. Here, the causal strength is not a correlation but shares its numerical value and reflects the true underlying influence between ROIs. Thus, while the direction of influence remains ambiguous, both the magnitude and sign are unambiguous and align with the observed r_{fMRI} (Chen et al., 2024). Given that directionless correlations are not expected to resolve directionality, many researchers would accept r_{fMRI} as a reasonable proxy for underlying causal strength in terms of magnitude and sign in this special case.

The situation grows more complex with the introduction of a third region. In general, correlations among

three regions can be generated by six distinct acyclic causal graphs (Appendix E). In Fig. 9B, the observed r_{fMRI} pattern includes one strong positive correlation, one strong negative correlation, and one near-zero correlation. Here, the challenge is not only directional ambiguity but also the potential mismatch in both sign and magnitude between correlations and true causal effects. For example, an observed correlation of $r_{\text{fMRI}} = 0.6$ between ROI_1 and ROI_3 could correspond to causal strengths of $0.57, 0.6$, or 0.77, whereas a near-zero correlation between ROI_1 and ROI_2 could still reflect a substantially positive causal effect.

Some special cases are worth noting. The near-zero correlation of –0.05 between ROI_1 and ROI_2 (Fig. 9B) illustrates the pitfalls of thresholding, a common practice in the field. Notably, even if this correlation were exactly zero—while the other two correlations remained unchanged—the set of six possible acyclic causal structures would persist, with only slight changes in strength on some causal effects (per the analytical formulas in Appendix E).

Negative correlations are also frequently filtered out, set to zero, or transformed via absolute-value operations in graph-based analyses. These practices largely reflect methodological constraints—many widely used graph metrics were originally defined for non-negative adjacency matrices—as well as interpretational uncertainty, rather than principled evidence that negative correlations are uninformative. Such conventions discard information rather than resolving its ambiguity, thereby embedding implicit assumptions about relevance and information flow into network construction. Moreover, as illustrated in Fig. 9B, these preprocessing maneuvers can further obscure, rather than clarify, the relationship between observed correlations and the underlying causal architecture.

Another illustrative case is shown in Fig. 9C, where all three pairwise correlations appear strong at the r_{fMRI} level. Yet such a pattern could arise from networks with far weaker direct influences. Strikingly, a correlation of $r_{\text{fMRI}} = 0.6$ between ROI_1 and ROI_3 may occur even in the absence of any direct causal link between them.

From a causal inference perspective, these examples underscore two key lessons. First, once a network involves more than two regions, a single observed correlation pattern can be consistent with multiple distinct causal structures. Second, both the sign and magnitude of observed correlations may differ markedly from the true causal effects driving them.

As demonstrated above, the ambiguity of correlation values, especially in a network-wide matrix, reflects a classical inverse problem, where one attempts to infer underlying neural structure from observed statistical associations. This makes it challenging to assign concrete meaning to individual graphical features derived from the correlation matrix (e.g., "hubness"). These examples also provide helpful evidence for addressing the three questions raised at the beginning of this section:

- *Does a zero or practically small correlation (e.g., -0.05 in Fig. 1) imply no information flow?* Surprisingly, the answer depends on other correlation values in the network. In Fig. 9B, columns 1 and 3 present four counterexamples.
- *Does a positive (0.6) or negative (-0.05) correlation imply excitatory or inhibitory information flow, respectively?* Again, the answer depends on the underlying causal relationships. Three cases in Fig. 9C and four cases in Fig. 9B are counterexamples.
- *Does a large (0.6) or small (-0.05) correlation imply strong or weak information flow, respectively?* Again, it depends: three cases in Fig. 9C and four cases in Fig. 9B present counterexamples.

The simple causal scenarios in Fig. 9 highlight a key limitation of graph-based analyses: even with perfect, noise-free neural-level correlations among just three regions, multiple, distinct causal structures can explain the same correlation data. Within a larger network, the number of compatible configurations for underlying causal structures grows quickly. That is, the difficulty with finding a unique underlying mapping of correlation at the r_{fMRI} level to the causal strength at the neural level increases. These methods have assumed known neural-level correlations and acyclic relationships to emphasize that ambiguity in causal interpretation is not just a product of noise or complexity; rather, it is inherent to the correlational framework itself.

In real fMRI correlation data, several additional challenges further complicate graph-based analyses. Measurement noise and error vary across regions, interactions are often nonlinear, and studies typically involve far more than three regions; each of these challenges magnifies interpretive ambiguity. Moreover, correlations capture only

instantaneous synchrony, failing to reveal causal directionality or time-lagged effects. Real cross-region causal relationships may also form dynamic feedback loops that a static, acyclic graph cannot represent. Altogether, these factors severely limit the interpretability of methods that rely solely on cross-region correlations.

These challenges undermine the reliability of correlation-based network inferences on underlying neural-level relationships. Cross-region interactions typically are not straightforwardly coded in the observed data or data-driven features. Rather, the underlying network structure is latent and must be inferred through theory-informed modeling approaches (Peel et al., 2022). While features such as edges, hubs, rich club, and communities are widely used, they lack meaningful biological interpretability without an understanding of the causal relationships between brain regions. Graph-based representations, therefore, remain susceptible to large variations due to acquisition parameters, preprocessing choices and sample sizes, limiting their explanatory power. Similarly, large-scale organization revealed by the principal gradient or geometry does not imply causality or directly explain the mechanisms underlying observed patterns. Beyond interpretational challenges, both approaches are sensitive to preprocessing decisions and susceptible to distortions from non-neural signals, further restricting their utility for mechanistic insights.

It is important to clarify the intent of the "three key questions" regarding correlation interpretation. Graph-based analyses—whether involving edges, hubs, rich-club structure, communities, or centrality measures—naturally invite interpretations related to influence, importance, coordination, or communication among brain regions. For such graph-derived features to admit mechanistic biological interpretation, however, some form of directional constraint, causal interaction, or information transfer must be assumed, even if not explicitly articulated. Otherwise, a correlation-derived edge reflects only statistical covariation at the fMRI signal level, without specifying excitation, inhibition, or direct dependence. The purpose of these questions is therefore to make these implicit assumptions explicit, and to clarify why correlation-based graphical features alone do not justify mechanistic conclusions in resting-state fMRI.

The ambiguities in interpreting graph-based analyses of resting-state fMRI reflect the classical Markov equivalence problem, in which different graphical models (e.g., DAGs) can imply identical sets of conditional independence relationships among variables. This is a well-established observation (e.g., Papo and Buldú, 2025; Zalesky et al., 2012; Rubinov, 2023). These observations fit well in the context of the issues we have raised here, and we reiterate them as relevant to and bolstering for them. Particularly related to resting-state graphs, we highlight the problematic implications, where in practice correlations are often interpreted mechanistically or directionally despite the absence of temporal structure or experimental manipulation. In this context, the combination of unanchored fMRI signals and widespread heuristic causal interpretation can amplify the risk of over-interpretation.

3 Discussion

This commentary is intended to clarify sources of potential misunderstanding in resting-state (and naturalistic) fMRI and to offer a framework for recognizing pitfalls in interpretation. We refer specifically to correlation-based analyses, including seed-based and network-level "connectivity" approaches, as well as methods that derive cross-region relationships from principal component analysis or independent component analysis. Since the primary interest of neuroscience lies in neural activity, it is critical to recognize the limitations of fMRI as a proxy for neural signals and the risks of drawing inferences based on correlations alone. This will not remove the utility of resting-state fMRI, but instead help to keep its applications from leaking into problematic terrains. Causal inference provides a useful and systematic framework for addressing these challenges.

3.1 Resting-state fMRI through the lens of causal inference

Our analysis builds on some key prior works. First, similar to Liu (2013) and Liu (2016), we started from an examination of noise sources and their impact on correlation estimation. Here we have integrated this analysis

within a causal inference perspective and explicitly modeled the noise couplings. Importantly, we have demonstrated the effects of these patterns on group or condition comparisons—often leading to distortions, spurious differences, difference suppression, sign flips, and more. Second, we have also extended the network-level focus of Reid et al. (2019) to a broader, quantitatively grounded discussion of biases in correlation estimation and their interpretive consequences. We derived general, analytical solutions for correlation-based causal structures, showing difficulties in interpreting cross-region correlations.

We have examined resting-state fMRI through a causal lens and identified two central pitfalls. First, pervasive noise sources systematically bias correlation estimates, complicating comparisons across groups or conditions. Second, the link between estimated cross-regional correlations and their true neural relationships is indirect and ambiguous, posing serious challenges for graph-based analyses (e.g., edges, hubs, communities). Together, these issues place fundamental limits on the mechanistic insights that rs-fMRI can provide.

The cases here illustrate several quantitative challenges in resting-state fMRI when viewed through a causal inference framework. Yet an even greater difficulty often lies in how these measures are interpreted, particularly in the widespread use of the term *functional connectivity.* At its core, the underlying measure is simply correlation—a construct that is, both axiomatically and idiomatically, non-causal. The shift from "correlation" to "connectivity," however, blurs the boundary between association and causation. By pairing "connectivity" with "functional," the term acquires connotations of deeper causal mechanisms, creating room for semantic slippage and over-interpretation. As a result, the phrase invites divergent readings—sometimes meant as strictly correlational, other times suggestive of causal processes—making it rhetorically appealing but conceptually precarious.

Thus, in many cases correlation is implicitly treated as a proxy for causality, serving more as a heuristic than a principled inference. This tendency, common not only in resting-state studies but across scientific disciplines, reflects a broader cognitive bias toward constructing plausible causal narratives (Hill et al., 2024). Such a pitfall frequently surfaces in methods development and is especially pervasive in the interpretation and discussion of resting-state findings. While heuristics play a role in developing insights toward scientific discovery, a deeper mechanistic understanding can only be achieved through rigorous modeling, which can itself be firmly grounded using causal principles. We emphasize, however, that this tendency toward causal overinterpretation is not unique to resting-state fMRI but is widespread across neuroscience and other observational sciences. Resting-state analyses simply make the issue especially visible because the data are weakly constrained by design, typically leaving many mechanistic explanations–not just neural synchrony–consistent with the same covariance patterns. Our aim is therefore to clarify interpretational boundaries and promote rigor, rather than to single out any particular paradigm or methodology.

The discussion here also applies to a large degree to correlation-based approaches in naturalistic fMRI data. For instance, in data collected under closer-to-real-life conditions (e.g., movie or audio clips), the causal relationships depicted in Fig. 6 remain relevant, though the region pair is replaced by a participant pair. In the naturalistic case, the correlation ρ_ϵ of noise terms between participants is likely lower than in resting-state data, while the correlation of the neurally-related HDRs, ρ_{mix}, is probably similar. Consequently, while overestimation is likely less severe, the underestimation of inter-individual correlations may persist.

In summary, the following are key points about correlation in fMRI, which forms the basis for resting-state and naturalistic paradigms:

1. *Omnipresent noise sources bias correlation estimates and comparisons.* In real data, the biases and uncertainties in estimation are difficult to quantify, but they can be substantial, particularly given the known noisiness of fMRI data. Thus, even consistent correlation patterns may be driven partly or predominantly by non-neural factors.
2. *The mapping between estimated cross-region correlations and their underlying neural-level relationships is not straightforward.* Unreliable correlation estimates can result in misleading inferences, and this presents interpretational difficulties for graph-based features (e.g., edges, hubs, communities) in particular.
3. *Correlations alone do not support causal interpretation.* Strength and differences in "functional connectiv-

ity," which is observed at the fMRI level, cannot necessarily be directly attributed to causal changes within individual causal strength at the neural level.

3.2 Interpretation difficulties

A single network of correlational relationships among brain regions, which can be summarized in a correlation matrix, is compatible with many different underlying causal structures. This ambiguity reflects what in causal inference is known as *Markov equivalence*: multiple causal graphs can generate the same pattern of statistical dependencies, rendering causal interpretation indeterminate (Spirtes et al., 2001; Pearl, 2009). From a biological perspective, this idea loosely resonates with the concept of degeneracy in systems neuroscience—the principle that distinct neural structures can give rise to the same function or output, thereby enhancing robustness and adaptability (Edelman and Gally, 2001). Evolution has indeed favored such redundancy in neural and other organ systems. Still, it is important to emphasize that the epistemic indeterminacy of causal modeling is not itself evidence of ontological degeneracy, but only an analogy. The actual mechanisms underlying this many-to-one mapping remain unknown and lie beyond the reach of fMRI observations.

It is important to emphasize that correlation-based analyses are valuable in their own right. Clinical or predictive utility does not require mechanistic causal interpretation: correlations can function as black-box biomarkers that forecast disease trajectories or treatment responsiveness, and they can uncover network-wide patterns with potential predictive utility, for example in identifying disease signatures (e.g., Lynch et al., 2024; Baldermann et al., 2024; Khoo et al., 2017; Larivière et al., 2020). Descriptive and organizational insights from correlations remain meaningful even when interpreted cautiously. For instance, recent work (e.g., Du et al., 2024) shows that certain correlation patterns correspond closely to task-evoked patterns at the individual level, highlighting that correlation can capture stable aspects of functional architecture independent of mechanistic claims. The challenge arises when such correlations are interpreted at a granular level without accounting for causality, confounding, or underlying mechanisms. Accordingly, their role should be understood as pattern recognition, rather than as evidence of direct functional relevance.

The issues around interpreting individual correlations within a network are difficult to resolve. By definition, correlation cannot distinguish between direct effects, confounding influences, or shared causes. In practice, estimates are further distorted by spurious associations from unmeasured variables and measurement error. In fMRI, these problems are amplified by the limited ability to model non-neuronal contributions in full detail and by noise sources that are rarely causally unrelated to neural signals, leading to systematic biases in correlation estimates. For example, the very small correlations (e.g., $\sim$0.01) often reported in large-scale brain-wide association studies (Marek et al., 2022) may achieve a designated level of statistical evidence with sufficient sample size, yet provide little biological or clinical insight, echoing longstanding concerns from genome-wide association studies (Turkheimer, 2024).

The absence of a true "Rosetta Stone" for neurovascular coupling underlies these interpretive challenges: the HDR reflects vascular and metabolic dynamics rather than direct neural firing, so the timing of its observed features can diverge from underlying neural events in multiple ways, which can vary across even an individual's brain. First, at a broad anatomical level, the frontal, temporal, parietal, and occipital lobes are each predominantly supplied by distinct branches of the cerebral vasculature. These vascular territories can contribute to regional differences in BOLD sensitivity and vascular reactivity. The differences in neurovascular coupling, capillary density, and baseline flow/metabolism can affect how signal changes are interpreted across lobes. This is part of why vascular effects are difficult to resolve with a one-size-fits-all model (Huck et al., 2023; Huck et al., 2019), such as the controversial method of global signal removal. Second, vascular transit delays vary across regions—differences in vessel density, baseline flow, and capillary architecture mean that identical neural inputs can produce BOLD peaks at different times; regions with longer arterial paths or sparser microvasculature show systematically delayed HDRs. Third, the neurovascular cascade itself is nonuniform: each step—from astrocyte signaling and endothelial response to smooth-muscle relaxation—has distinct kinetics, and factors such as age,

caffeine intake, or vascular health can speed up or slow down any link in this chain, reshaping the HDR's onset and peak. Fourth, baseline physiology and noise introduce further variability: regional differences in resting vascular tone and cerebrovascular reactivity shift HDR peak latency even when neural timing is identical.

Empirical work highlights the large amount of variability in these features in real data. Consider the variability of a single feature of the hemodynamic response, such as the time to peak. Regional HDR peak latencies can differ by 2–3 s within the same individuals—far exceeding plausible neural latency variations, hence suggesting strong variations in the responses themselves (Handwerker et al., 2004; Chen et al., 2023). Moreover, simultaneous EEG–fMRI studies show electrophysiological peaks preceding BOLD peaks by 5-6 s, only part of which reflects true neural timing differences (Rosa et al., 2010). Other features such as the duration of the peak, the time to nadir and time to baseline return can show similar, if not greater, variation.

Finally, a region can appear to have BOLD response even in the absence of any local neural spiking. Strong neural response in one area can passively redirect blood flow through shared arterial supplies—so-called arterial steal—increasing or decreasing BOLD signals in neighboring regions despite no local neural change (Suarez et al., 2021). Likewise, activity millimeters away can generate a BOLD signal in a voxel via its draining vein, falsely suggesting local activation (Drew, 2019). Factors such as global blood-pressure fluctuations, systemic CO_2 changes, or anticipatory vasodilation can all drive BOLD fluctuations with no direct neural correlate (Howarth et al., 2020; Tsvetanov et al., 2020; Mark et al., 2015; Ekstrom, 2010; Narain, 2006; Ekstrom, 2010; van Niftrik et al., 2023).

A critical question is whether correlation-based results can be meaningfully interpreted in the absence of an explicit causal framework. One proposed approach is to employ partial correlations (Smith et al., 2011), which aims to remove the influence of other regions or variables. However, the interpretability of partial correlations rests on the strong assumption that there is no endogenous noise or latent dependence structure. This assumption is rarely tenable in fMRI: both unmeasured confounders and mediating processes are likely to operate simultaneously across distributed brain regions. Such complexities fundamentally compromise the identifiability of true associations, thereby limiting the inferential validity of correlation-based analyses. Thus, replacing marginal with partial correlations does not, by itself, resolve the inherent interpretational challenges of these approaches.

It is important to emphasize that correlations derived from resting-state fMRI, while often reliable and reproducible, should not be conflated with one-to-one representations of local neural-level causal strengths. Most researchers appreciate that fMRI correlations are indirect measures, yet it is still common in the literature for interpretations to drift toward causal or mechanistic language. The reproducibility of large-scale networks observed across datasets and scanners worldwide clearly indicates that these patterns reflect robust properties of the brain's functional organization. At the same time, reproducibility alone does not guarantee that the observed correlations map directly onto neuronal interactions. Shared physiological artifacts, vascular dynamics, or preprocessing choices can also produce consistent patterns across studies. Particularly when studying group differences or longitudinal trajectories, knowing the relative contributions of the underlying components to the patterns is key to making an accurate evaluation and interpretation. The key distinction, therefore, is between reproducibility and mechanistic interpretation: robust correlation patterns are valuable and likely biologically meaningful at some level, but without external validation or causal manipulation, they should not be assumed to reflect direct neural communication.

Other association measures—signal coherence, mutual information, transfer entropy, or ICA—share similar limitations, capturing statistical relationships without causal clarity. Because the popular term "functional connectivity" can inadvertently imply causation, it is more accurate to interpret these measures simply as patterns of statistical association. The presence of substantial noise, confounders, and mediators further complicates interpretation, introducing biases such as sign reversals or spurious associations when evaluating individual edges or arcs within a network. Thus, inaccuracies in fMRI correlation estimates hinder reliable causal discovery (Spirtes and Zhang, 2016; Glymour et al., 2019). Without theory-driven causal frameworks, methods like brain-wide association studies require thousands of participants yet still fail to resolve ambiguities from unmodeled

confounders.

Other neuroimaging modalities rely on correlation patterns and face similar challenges. EEG and MEG avoid neurovascular coupling but are affected by partial volume conduction and the ill-posed nature of the inverse problem in source localization. PET, SPECT, and fNIRS rely on indirect neural proxies, while naturalistic fMRI retains resting-state's observational limitations. These issues underscore the need for cautious interpretation and validation beyond fMRI (Shahsavarani et al., 2023).

Overinterpretation in resting-state fMRI most commonly arises in the following situations:

- when correlation magnitude is equated with causal influence, a risk amplified by the language of "functional connectivity";
- when graph-theoretic edges, hubs, or communities are interpreted as functional or communicative pathways, despite the absence of a one-to-one or identifiable mapping between observed correlations and underlying causal connections (Fig. 9);
- when correlation patterns are taken to reveal causal network topology, even though causal relationships are directional, latent, and generally not identifiable from correlation structure alone;
- when group differences in correlations are interpreted as mechanistic explanations of disease, without accounting for noise sources and systematic biases (Fig. 7);
- when predictive performance is conflated with explanatory validity, effectively treating associations as evidence of causal mechanisms, as can occur in brainwide association analyses.

Such interpretations exceed what can be supported by observational fMRI correlations alone, particularly in the presence of latent confounding (e.g., motion, physiology, and preprocessing choices), indirect pathways, shared inputs, and unknown or regionally variable neurovascular mappings.

To recapitulate, our central points are:

1. Correlation-based analyses remain valuable tools for identifying statistical regularities and large-scale functional organization.
2. Mechanistic or directional interpretations, however, require external constraints that typical resting-state designs do not provide.
3. The core risk is not the use of correlations themselves, but the overinterpretation of correlations or changes in correlations as evidence for specific types of neural activity.

3.3 Potential improvements

Advancing fMRI research requires not only identifying limitations and flaws but also fostering collaboration and innovation to address them. While the challenges are complex, constructive criticism and shared efforts within the community can establish better practices and drive progress. Our goal here is to encourage open dialogue, to promote more careful interpretations, and to promote new ideas, rather than to diminish the value of current contributions.

One improvement when considering resting-state fMRI is to decrease the sensitivity of results to various acquisition and processing choices, as well as some of the biases observed here, by adopting transparent thresholding and the more general "highlight, don't hide" approach (Allen et al., 2012; Taylor et al., 2023). Transparent thresholding de-emphasizes the arbitrary cut-off of presented results, by showing sub-threshold results along with the primary suprathreshold ones, but de-emphasizing the former with an opacity that decreases as the statistic decreases. This has been shown to provide a more complete and understandable picture of neuroimaging results (Sundermann et al., 2024; Taylor et al., 2025). Given the biases and uncertainties that are known to be present in resting-state fMRI results, traditional all-or-nothing thresholding is a processing choice will likely discard meaningful results, as well as introduce further biases, particularly toward non-reproducibility at the meta-analysis level—see examples in Taylor et al. (2023) and Taylor et al. (2025). In general, extending this approach to reduce the role of thresholds and the sensitivity to arbitrary features in the data and processing produces more robust results.

Noise mitigation remains a major challenge in resting-state fMRI. Over three decades, numerous denoising strategies have been proposed—motion regression, physiological modeling, principal component approaches, and global signal removal—with continual methodological refinements (e.g., Kundu et al., 2017; Caballero-Gaudes and Reynolds, 2017). Independent component analysis (ICA) has become especially influential for data-driven decomposition, artifact separation, and dimensionality reduction (Calhoun et al., 2009; Pruim et al., 2015)..

However, denoising in resting-state fMRI faces fundamental limits, and genuinely new breakthroughs are rare for principled reasons. The challenges are not merely technical but stem from identifiability constraints, physiological entanglement, and the causal structure of the BOLD signal itself. Many nuisance processes—motion, respiration, cardiac pulsatility, scanner drift—operate on the same timescales and share covariance structure with neural fluctuations; global respiratory artifacts, for example, can mimic large-scale neural modes. Because BOLD mixes neural and vascular signals in a nonlinear, region-dependent manner, there is an inherent point beyond which removing noise inevitably removes some neural variance as well.

ICA illustrates these limitations. Its independence is statistical rather than mechanistic: neural and vascular contributions that are causally linked may still be separated (or conflated) depending on their statistical mixture properties. ICA thus imposes statistical independence, not causal structure. Moreover, ICA operates entirely on BOLD mixtures rather than on neural sources. Consequently, ICA does not supply the ingredients required for causal inference—such as interventions, structural constraints, or mechanistic priors—and therefore remains a powerful descriptive tool but fundamentally an observational, correlation-based method.

Nevertheless, there remains substantial room for improvement in resting-state fMRI research. Proactive strategies during data acquisition can enhance data quality at the source, reducing reliance on extensive post-hoc adjustments. For example, shortening TRs with multiband fMRI increases temporal resolution, allowing better capture of faster fluctuations in neural activity and physiological signals, such as cardiac and respiratory cycles. This also helps reduce the aliasing of physiological noise into the low-frequency band typically used for resting-state connectivity (0.01–0.1 Hz), potentially improving correlation estimates between regions. Nonetheless, these gains are context-dependent and come with trade-offs, including reduced signal-to-noise ratio per volume, increased preprocessing complexity, and greater demands on hardware.

Moreover, validating fMRI-derived findings increasingly depends on integrating complementary neuroimaging modalities. For instance, simultaneous EEG-fMRI studies have demonstrated a degree of correlation consistency between resting-state fMRI and specific EEG frequency bands (Xavier et al., 2025). In contrast, divergent correlation patterns have been reported when comparing resting-state fMRI with metabolic measures obtained from [^{18}F]-fluorodeoxyglucose functional positron emission tomography (Ionescu et al., 2021; Voigt et al., 2022; Deery et al., 2024), highlighting the need for multimodal validation.

Layer (laminar) fMRI can provide critical cross-validation for correlation-based analyses of resting-state fMRI by resolving activity across specific cortical layers (e.g., superficial, middle, deep). This layer specificity offers insights into the directionality and anatomical underpinnings of cross-region correlations. Techniques such as vascular space occupancy imaging enhance interpretability by measuring cerebral blood volume rather than BOLD signals, thereby reducing confounds from large draining veins and better localizing activity to neuronal sources (Lu and van Zijl, 2012; Huber et al., 2018).

Layer-specific analyses can also attenuate systemic physiological noise (e.g., respiration, cardiac pulsations) and minimize contamination from global vascular signals, particularly in deeper layers less affected by superficial veins. Laminar-level correlations reveal circuit-level organization and inter-regional dynamics, adding granularity to whole-brain connectivity analyses. For instance, while conventional correlation is symmetric and undirected, laminar fMRI can distinguish input and output patterns: feedforward projections typically target layer 4 (middle), whereas feedback projections terminate in superficial (layers 1–3) and deep (layers 5–6) layers. Thus, layer-specific resting-state signals can help infer whether cross-region correlation reflects feedforward or feedback interactions (Huber et al., 2017).

Layer-resolved resting-state correlations can also be compared with known anatomical projections from tracer

or diffusion imaging studies. When the laminar origin and termination of cross-region correlations align with established anatomical pathways, this adds biological plausibility to the observed resting-state patterns (Deshpande et al., 2022). Moreover, because vascular artifacts tend to dominate superficial layers, examining laminar signal profiles allows the identification and down-weighting of spurious correlations arising from non-neuronal sources. Finally, the sub-millimeter resolution of layer fMRI helps determine whether observed cross-region correlation arises from specific laminar circuits or is an artifact of spatial averaging, thus refining the interpretation of resting-state data.

Other imaging modalities, despite their respective limitations, may also help cross-validate resting-state fMRI findings:

- Arterial spin labeling (ASL): ASL directly measures cerebral blood flow (CBF), offering physiological insight distinct from BOLD signals, which reflect changes in deoxygenated hemoglobin. It is particularly useful in contexts where neurovascular coupling is altered, such as in aging or cerebrovascular disease (Liu and Brown, 2007), even while also requiring an understanding of its own hemodynamic changes.
- Transcranial magnetic stimulation (TMS): As a causal intervention technique, TMS perturbs cortical excitability in a time-locked and spatially targeted manner. When combined with resting-state fMRI, it creates a perturbation–response paradigm that can test causal hypotheses, providing a framework beyond the correlational nature of rs-fMRI (Bohning et al., 1998; Hampson and Hoffman, 2010; Beynel et al., 2020).
- Electroencephalography (EEG) and magnetoencephalography (MEG): These modalities offer millisecond-level temporal resolution, complementing fMRI's spatial precision. Simultaneous EEG–fMRI enables spatiotemporal alignment of neural activity, while multimodal fusion approaches enhance interpretability and may improve biomarker identification (Wirsich et al., 2021; Etkin and Mathalon, 2024).
- Diffusion tensor imaging (DTI): DTI maps white matter tracts, providing a structural backbone for evaluating the plausibility of fMRI-derived functional networks (Arbabyazd et al., 2021).
- Positron emission tomography (PET): PET provides metabolic and neurochemical data that, when integrated with fMRI, offer a richer understanding of the neural substrates underlying cross-region correlations.

In particular, ASL and TMS may offer the most promising avenues toward uncovering the long-sought Rosetta Stone for validating resting-state fMRI. In general, multimodal approaches—including wide-field fluorescence calcium imaging in animal models—can reveal both converging and diverging evidence across techniques, deepening our understanding of the neural basis of cross-region correlations (Vafaii et al., 2024).

The integration of multimodal studies may complement the limitations of each approach and enhance the validity of causal relationships. For example, incorporating multimodal data can mitigate issues like the low test-retest reliability of fMRI biomarkers (Tozzi et al., 2020; Wirsich et al., 2021; Tetereva et al., 2025). Combining EEG and fMRI can enhance the robustness of biomarkers by capturing both spatial and temporal dynamics of brain activity. Similarly, the integration of fMRI with ASL, DTI, or PET provides a multidimensional perspective, improving the interpretation of neural and vascular influences on brain function.

Importantly, no single modality can recover the full causal mapping between neural activity and the fMRI signal. ASL, TMS, EEG/MEG, PET, and layer-fMRI each shed light on different components of the signal-generation process, but none provides a complete solution given the bidirectional, nonlinear nature of neurovascular coupling (including blood-to-neuron feedback via mechanosensitive pathways). For example, ASL directly quantifies perfusion changes, helping dissociate vascular variability from fMRI covariance, while TMS introduces externally controlled perturbations that break the symmetry of observational data and allow testing whether fMRI correlations are directionally modifiable. Foundational causal insight may also come from invasive work in animal models (e.g., Jammal Salameh et al., 2024). The realistic goal is incrementally "carving" a Rosetta Stone using a diverse set of tools from different modalities constraint, not full identification. Finally, resting-state networks are not purely hemodynamic artifacts. Our focus is on interpretational caution: unobserved neurovascular variability and spontaneous BOLD fluctuations can complicate inference, even when the underly-

ing networks reflect true neural activity. Convergent evidence from EEG, MEG, and other modalities can thus play a complementary role.

3.4 The scope of our discussion

It should be noted that the absence of a definitive neural-to-hemodynamic mapping affects task-based fMRI as well as resting-state analyses, though its impact is substantially smaller in the task setting. In task paradigms, experimental structure (randomized trial sequences, inter-trial jitter, known event timing, repeated trials, and explicit HRF modeling) imposes strong constraints that reliably link designed inputs to estimated hemodynamic responses. These constraints support a reliable causal interpretation at the level of the hemodynamic response, even though the underlying neural activity remains indirectly inferred. In this setting, the primary ambiguity lies between neural activity and its hemodynamic expression rather than between the task and the measured response. It is also worth noting that while exogenous noise may increase estimation uncertainty, it does not introduce bias into the estimates.

In contrast, resting-state fMRI fluctuations arise spontaneously and are typically not anchored to an externally defined experimental reference. Without experimental structure, there is no principled way to separate neural and vascular contributions or to determine which shared fluctuations across regions reflect neural interactions rather than hemodynamic or noise correlations. When an unobserved neurovascular mapping is combined with unconstrained hemodynamic response dynamics, even associative interpretations of regional correlations become unreliable. Importantly, resting-state analyses are much more vulnerable to influences of exogenous noise (e.g., head motion) that, independent of the unresolved neurovascular coupling issue, can produce unreliable fMRI correlation estimates (Fig. 6) and distort group comparisons (Fig. 7).

The correlation-causation challenge is not unique to resting-state fMRI; it is a universal limitation of observational inference. Numerous fields, including economics, ecology, genetics, and social-network science, have developed domain-specific strategies to address it, typically by leveraging external sources of variation (e.g., instruments, natural experiments, exogenous shocks) or by adopting strong, field-specific assumptions. Standard resting-state fMRI lacks both experimental manipulation and identifiable exogenous variation, making it especially vulnerable to causal overinterpretation when cross-regional correlations are treated mechanistically. Our discussion focuses on this widely used subset of resting-state analyses, while acknowledging that other resting-state measures (e.g., ALFF, fALFF, ReHo, resting-state HRF estimation) probe local features of spontaneous activity and can complement correlation-based approaches. We emphasize that the uncertainty in neural drivers of spontaneous BOLD signals and the non-identifiability of causal structure from correlations are distinct issues, but in practice they compound interpretational uncertainty. This practical interaction motivates their joint discussion in the present framework.

We invoke here the Rosetta Stone metaphor in a limited conceptual sense: to illustrate that uncertainty in the translation between neural activity and the hemodynamic response should introduce intrinsic limits on inference. As with the historical Rosetta Stone, translation proceeds through partial, imperfect, and progressively refined correspondences rather than a deterministic lookup. Existing neurovascular models, multimodal measurements, and physiological data already serve as important fragments of this translation key. Resting-state analyses lack such externally controlled inputs, and thus rely heavily on assumptions in place of a biophysical translation layer, making its uncertainty particularly consequential. The metaphor is intended to highlight this structural dependency, not to suggest a binary or all-or-nothing solution.

It is important to distinguish between the descriptive and mechanistic roles of correlation-based analysis. Correlations can capture robust, reproducible statistical regularities in large-scale fMRI signal and have been productively used to identify organizational gradients, characterize network structure, predict behavior, and guide hypothesis generation. These contributions do not require causal identification. Our argument concerns a different interpretational layer: this fMRI covariation, on its own, does not specify the neural mechanisms or directional pathways that give rise to it. In this sense, a correlation may reflect reproducible and meaningful

functional co-involvement while still being "spurious" with respect to a proposed causal explanation. Importantly, due to the interaction between exogenous noise and the unknown neurovascular mapping, correlations may also be "spurious" in the narrower sense of not reflecting true dependencies even at the hemodynamic response level (Figs. 6-7). Recognizing this distinction allows correlation-based approaches to be used to their strengths, as descriptive, predictive, and organizational tools, while avoiding unwarranted mechanistic interpretations in the absence of an appropriate Rosetta Stone to translate.

The noise sources we have discussed (motion, physiology, sampling variability, and heterogeneity in neurovascular coupling) affect all forms of fMRI, including task-based paradigms. However, the interpretational challenge is amplified in resting-state fMRI because, in the absence of experimentally controlled inputs, noise-induced covariance cannot be anchored to externally constrained models of neural events. This distinction does not imply that resting-state correlations are unreproducible; indeed, large-scale patterns replicate robustly. Rather, reproducible statistical dependencies may arise from stable mixtures of neural and non-neural factors, and their robustness does not confer mechanistic specificity. Our goal is therefore not to single out resting-state fMRI but to clarify the interpretational limits that follow from unconstrained data structure, while acknowledging the substantial descriptive and predictive value of the resulting correlation patterns.

Looking forward, a constructive framework emphasizes appropriate uses and methodological priorities for resting-state fMRI. Descriptively, correlation structure is well-suited for mapping large-scale functional architecture, evaluating individual differences, and building predictive biomarkers. Mechanistically, progress will likely come from integrating resting-state fMRI with modalities or designs that impose additional constraints such as TMS–fMRI perturbations, ASL-based perfusion measurements, EEG/MEG-based temporal disambiguation, laminar fMRI, and generative biophysical models. These complementary approaches can narrow the plausible mechanism space even if they cannot fully resolve it. Priorities for future work include improving physiological characterization, formalizing uncertainty in neurovascular mapping, and developing causal modeling frameworks that explicitly incorporate unobserved confounding and nonlinear hemodynamics. Together, these directions offer a practical path toward using resting-state fMRI constructively while remaining within the limits imposed by its observational nature.

Many large-scale cohort studies and clinical projects face practical constraints that preclude multimodal imaging, interventional designs, or high-resolution protocols. It is neither feasible to impose idealized standards that would render most studies unattainable, nor desirable to discourage well-designed research conducted within conventional resting-state fMRI settings. Even within standard single-modality fMRI, substantial progress can be achieved through methodological transparency, and appropriately calibrated interpretational claims. Large cohorts also offer distinctive strengths such as statistical strength, reliability estimation, out-of-sample prediction, and reproducible descriptive characterizations that remain essential assets even without multimodal augmentation.

4 Conclusions

Association metrics in resting-state fMRI, such as correlations, may provide valuable insights into relationships among brain regions. However, their accuracy is often compromised by the proxy nature of the fMRI signal, by mediation and confounding effects, by pervasive noise, and imprecise noise removal during preprocessing. This leads to significant ambiguities in regional localization and network-level inferences, often failing to clarify underlying causal relationships.

Opportunities for progress remain. Improved preprocessing, biophysically informed modeling, and integrative use of complementary modalities may incrementally reduce these ambiguities. Constructing a true "Rosetta Stone" that reliably links fMRI signal fluctuations to their neural generators remains a central challenge for the field. Although no clear solution exists at present, converging evidence from electrophysiology, vascular physiology, and multimodal imaging offers the most promising path forward. The potential gains—greater

validity, reliability, and depth of inference—are essential for advancing neuroimaging toward a more mechanistic understanding of brain function.

Acknowledgments

PT and GC were supported by the NIMH Intramural Research Program (ZICMH002888) of NIH/HHS, USA. JF and PAB were supported by NIMH IRP (ZIAMH002783) of NIH/HHS, USA. ZC was funded by the FRQS Postdoctoral Fellowship (doi.org/10.69777/336081), Fonds de recherche du Québec – Santé (FRQS) and the MNI Jeanne Timmins Costello Fellowship, Canada.

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Appendices

A Typical resting-state fMRI analysis methods

Resting-state fMRI studies focus on cross-region associations, typically quantified using correlation values. Unlike stimulus-driven task-based experiments, which involve active modulation through carefully designed protocols—including randomization, and complex modeling—resting-state fMRI passively observes brain activity without external intervention. The key distinction is that task-based studies intentionally manipulate the data-generating process to support causal inference, whereas resting-state studies rely on observational data alone.

ROI-based resting-state fMRI analysis typically involves the following steps:

(a) Acquisition – Selecting scanner strength, voxel size, TR, duration of scan, number of runs, number of sessions, number of subjects, potential co-acquisition of measures such as breathing, heart rate, alertness and more.

(b) Processing – Motion adjustment, spatial alignment, spatial smoothing, and regression-based denoising (some combination of respiration, cardiac activity, motion profiles, etc.) are performed to reduce noise and artifacts.

(c) Correlation estimation – Correlations for fMRI signal fluctuations are quantified at the individual level to infer cross-region synchronization (and region definition itself, through atlas selection, has significant effects on outcomes, as in Li et al. (2024)).

(d) Statistical inference – Group-level comparisons of correlation patterns across individuals, conditions, or clinical groups reveal differences in cross-region associations.

(e) Network-based analysis – Graph theoretical methods or independent component analysis are used to infer large-scale features.

B Inaccuracy of estimated correlations: modeling details and formulas

This appendix provides the technical details, including model formulas and parameter notations, for the correlation estimation scenarios discussed in Subsection 2.1.

For a sample of size n with true correlation r, the expected value of the Pearson correlation estimator is approximately (Fisher, 1915; Zimmerman et al., 2003):

$$r - r(1 - r^2)/2n, \tag{1}$$

assuming Gaussian data. The standard error is approximately (Bonett, 2008; Gnambs, 2023):

$$(1 - r^2)/\sqrt{n - 3}. \tag{2}$$

These formulas indicate minimal bias for large n (e.g., $n \geq 100$), typical in fMRI, but uncertainty persists,

Term	Description
$\mathsf{HDR.neural}_i$	Neural component of hemodynamic response for region i
σ_{neur}	Standard deviation of $\mathsf{HDR.neural}$ signal (assumed homogeneous across regions)
ϵ_i	Noise component for region i ($i = 1, 2$)
σ_ϵ	Standard deviation of noise (assumed homogeneous across regions)
fMRI_i	Observed fMRI signal for region i, containing both $\mathsf{HDR.neural}_i$ and ϵ_i ($i = 1, 2$)
R_v	Variability ratio: ratio of noise variability to $\mathsf{HDR.neural}$ variability; defined as $\sigma_\epsilon/\sigma_{\text{neur}}$
r_{neur}	Actual underlying correlation of neural component $\mathsf{HDR.neural}$ between two regions
r_{fMRI}	Estimated correlation of observed fMRI signal between two regions
ρ_{mix}	Correlation characterizing the inhomogeneous mixing of neural component $\mathsf{HDR.neural}$ and noise ϵ (within or between regions)
ρ_ϵ	Correlation of noise ϵ between two regions

Table 2: A reference table of symbols used in model formulas and data simulations; this expands the list and descriptions of some variables used in the main paper and listed there in Table 1.

especially for small correlations.

In the very simple scenario of Case 1 (Sec. 2.1.1), the fMRI signal is comprised of $\mathsf{HDR.neural}_i$ (neural component) and ϵ_i (noise) for region i, which have respective variances σ^2_{neur} and σ^2_ϵ (each assumed to be Gaussian and equal across regions). Here, noise is uncorrelated with neural signals and across regions. The variability ratio, $R_v = \sigma_\epsilon/\sigma_{\text{neur}}$, quantifies noise relative to the neural signal (Chen et al., 2022). The estimated fMRI correlation r_{fMRI} relates to the neural-level correlation r_{neur} as:

$$r_{\text{fMRI}} = r_{\text{neur}}/(1 + R_v^2). \tag{3}$$

This shows that this simple noise attenuates the estimated r_{fMRI}, with the attenuation rate $1/(1 + R_v^2)$.

In the more realistic scenario of Case 2 (Sec. 2.1.2), noise correlates with neural signals with a mixing fraction ρ_{mix} and across regions with a fraction ρ_ϵ. The resulting fMRI correlation is:

$$r_{\text{fMRI}} = (r_{\text{neur}} + 2\rho_{\text{mix}} R_v + \rho_\epsilon R_v^2)/(1 + 2\rho_{\text{mix}} R_v + R_v^2). \tag{4}$$

This model accounts for the complex bias patterns, including both under- and overestimation, observed in the simulations shown in Fig. 6.

Simulations for the group comparisons shown in Fig. 7 used the model:

$$\mathsf{fMRI}_i = \mathsf{HDR.neural}_i + \epsilon_i\,.$$

The parameters are specified in the figure caption, generating 300 time points for 20 individuals per group.

C Correlation: misrepresentation due to nonlinearity

Nonlinearity can introduce inaccuracies in characterizing cross-region associations, as Pearson correlation only captures linear relationships. To illustrate this, consider a hypothetical scenario where the signals for two regions, fMRI_1 and fMRI_2, are assumed to have finite variances and a quadratic relationship:

$$\mathsf{fMRI}_2 = \alpha + \beta\,\mathsf{fMRI}_1^2 + \epsilon,$$

where α and β are constants, and ϵ is a random process independent of fMRI_1. If fMRI_1 has a symmetric distribution with $E(\mathsf{fMRI}_1) = 0$, the theoretical correlation between fMRI_1 and fMRI_2 becomes zero:

$$\text{Corr}(\mathsf{fMRI}_1,\ \mathsf{fMRI}_2) = 0.$$

(In practice, for finite-length time series, the empirical correlation values are small, approaching 0 over many iterations.) Thus, despite a strong nonlinear association between the signals, no correlation is detected. This example highlights how correlation analysis can fail to capture or misrepresent nonlinear relationships.

D Model formulas for cross-region correlations (Case 1 and Case 2)

Under the causal assumptions in Fig. 3, a simple model can be formulated for two regions i and j the following region-level model,

$$\begin{aligned} \mathsf{fMRI}_i &= \mathsf{HDR.neural}_i + \epsilon_i, \\ \mathsf{HDR.neural}_i &\sim \mathcal{N}(0,\ \sigma^2_{\text{neur}}),\ \epsilon_i \sim \mathcal{N}(0,\ \sigma^2_\epsilon),\ i = 1,\ 2. \end{aligned} \tag{5}$$

The residuals ϵ_i symbolize the combined effects of neurovascular processes and exogenous factors in Fig. 2B (described further, below). Given that neural activity and the hemodynamic response operate on different timescales, neurovascular coupling likely introduces a nonlinear transformation that varies across brain regions. As a result, modeling it along with exogenous factors as simple additive noise constitutes a substantial oversimplification. The present approach is not intended to ignore the underlying complexity (Drew, 2019). Rather, the deliberate simplification serves as a heuristic device to highlight how misleading correlation estimates can be, even under highly idealized assumptions.

The challenge of resting-state fMRI analysis arises from two broad categories of noise, both embedded in the residual term ϵ_i in model (5). The exogenous component, illustrated in Fig. 2, is statistically independent of (or uncorrelated with) the neurally related signal $\mathsf{HDR.neural}_i$. In other words, its stochastic variation is driven entirely by processes external to the neural mechanism. In contrast, the endogenous component, represented as neurovascular processes in Fig. 2, captures variability introduced by neurovascular coupling across the brain. Because this variability lies on the causal pathway from neural activity to the observed fMRI signal, its stochastic fluctuations share a common source with the neural signal, making it endogenous and inherently correlated with $\mathsf{HDR.neural}_i$.

For analytic convenience and without loss of generality, we first set the mean of the neurally-evoked BOLD to zero. We also make a simplifying assumption that the variances of σ^2_{neur} and σ^2_ϵ are homogeneous between the two regions. This is the setup for Case 1 (Sec. 2.1.1 and Fig. 3). As a result, the correlation r_{fMRI} of the fMRI signal between the two regions, as shown in (3), can be derived as follows:

$$\begin{aligned} r = \text{Corr}(\mathsf{fMRI}_1,\ \mathsf{fMRI}_2) &= \text{Cov}(\mathsf{fMRI}_1,\ \mathsf{fMRI}_2)/\sqrt{\text{Var}(\mathsf{fMRI}_1)\text{Var}(\mathsf{fMRI}_2)} \\ &= r_{\text{neur}}\sigma^2_{\text{neur}}/(\sigma^2_{\text{neur}} + \sigma^2_\epsilon) \\ &= r_{\text{neur}}/(1 + (\sigma_\epsilon/\sigma_{\text{neur}})^2) \\ &= r_{\text{neur}}/(1 + R^2_v). \end{aligned}$$

A more realistic model can be constructed under the causal assumptions in Case 2 (Sec. 2.1.2 and Fig. 5). The model (5) above relies on two implicit independence assumptions: first, the neurally-evoked BOLD responses ($\mathsf{HDR.neural}_i$) are assumed to be independent, and second that the noise (ϵ_i) sources are independent as well. These assumptions are likely quite unrealistic. To address this limitation, we consider a quadrivariate distribution:

$$\begin{aligned} &\begin{bmatrix} \mathsf{HDR.neural}_1 \\ \mathsf{HDR.neural}_2 \\ \epsilon_1 \\ \epsilon_2 \end{bmatrix} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{SKS}),\ \boldsymbol{S} = \text{diag}(1,\ 1,\ R_v,\ R_v),\ \boldsymbol{K} = \begin{bmatrix} 1 & r_{\text{neur}} & \rho_{c11} & \rho_{c12} \\ r_{\text{neur}} & 1 & \rho_{c21} & \rho_{c22} \\ \rho_{c11} & \rho_{c21} & 1 & \rho_\epsilon \\ \rho_{c12} & \rho_{c22} & \rho_\epsilon & 1 \end{bmatrix}; \\ &\mathsf{fMRI}_i = \mathsf{HDR.neural}_i + \epsilon_i,\ i = 1, 2. \end{aligned} \tag{6}$$

where $\boldsymbol{K}$ represents the correlation matrix while $\boldsymbol{S}$ is a diagonal matrix that contains the standard deviations. Without loss of generality, we assume zero means for all four variables and a standard deviation of one for the neurally-evoked BOLD responses.

The formulation (6) provides valuable insights into the accuracy of estimated fMRI-level correlations. To simplify the complexity of the correlation matrix $\boldsymbol{K}$, we introduce the assumption of homogeneity among the four correlations between neurally-evoked BOLD responses and noise: $\rho_{c11} = \rho_{c12} = \rho_{c21} = \rho_{c22} = \rho_{\text{mix}}$. Thus, the correlation r_{fMRI} of the fMRI signal between the two regions, as shown in (4), can be derived as,

$$\begin{aligned} r_{\text{fMRI}} = \text{Corr}(\mathsf{fMRI}_1,\ \mathsf{fMRI}_2) &= \text{Cov}(\mathsf{fMRI}_1,\ \mathsf{fMRI}_2)/\sqrt{\text{Var}(\mathsf{fMRI}_1)\text{Var}(\mathsf{fMRI}_2)} \\ &= (r_{\text{neur}}\sigma^2_{\text{neur}} + 2\rho_{\text{mix}}\sigma_{\text{neur}}\sigma_\epsilon + \rho_\epsilon\sigma^2_\epsilon)/(\sigma^2_{\text{neur}} + 2\rho_{\text{mix}}\sigma_\epsilon + \sigma^2_\epsilon) \\ &= (r_{\text{neur}} + 2\rho_{\text{mix}}(\sigma_{\text{neur}}/\sigma_\epsilon) + \rho_\epsilon(\sigma_{\text{neur}}/\sigma_\epsilon)^2)/(1 + 2\rho_{\text{mix}}(\sigma_{\text{neur}}/\sigma_\epsilon) + (\sigma_\epsilon/\sigma_{\text{neur}})^2) \\ &= (r_{\text{neur}} + 2\rho_{\text{mix}}R_v + R^2_v)/(1 + 2\rho_{\text{mix}}R_v + R^2_v). \end{aligned}$$

E Causal relationships among three regions (in Case 3)

Here, we analytically demonstrate that a set of three correlations for three regions can correspond to multiple configurations of causal relationships, as illustrated in Section 2.2. Previously, Reid et al. (2019) provided conceptual demonstrations of these features. In this section, we derive explicit formulas that show the causal relationships.

Assume that the linear relationships among three random variables X, Y, and Z from which the neural-level signal at three brain regions are characterized with correlation values r_{xy}, r_{yz}, and r_{xz} (Fig. 10A); we also assume here that no correlation values are the special cases of 1 and -1. Without loss of generality, X, Y, and Z are standardized with zero mean and unit variance.

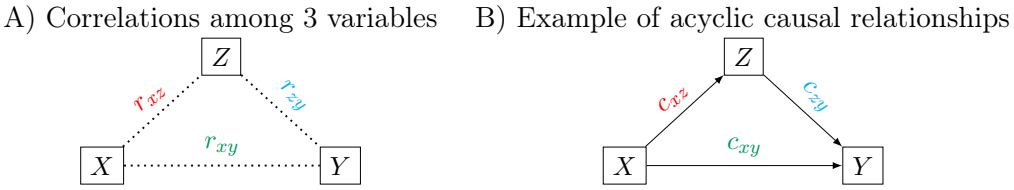

Figure 10: Mapping pairwise correlations to causation relationships. (A) The pairwise correlations among three random variables X, Y, and Z are r_{xy}, r_{xz}, and r_{zy}. (B) One specific case of acyclic causal relationships among the three variables is shown with Z as a mediator for causal relationship $Z \to Y$ and with X as a confounder for the causal relationship $Z \to Y$. The parameters c_{xy}, c_{xz}, and c_{zy} are the causal counterparts of the pairwise correlations r_{xy}, r_{xz}, and r_{zy}.

The six acyclic causal relationships depicted in Fig. 9B are graphically isomorphic in the sense that they share the same causal structure through variable rotations. Therefore, we illustrate the derivation only for the case in Fig. 10B that corresponds to the first scenario on row 1 of Fig. 9B, while all other cases can be similarly derived through variable rotations. Our goal is to express the causal effects c_{xy}, c_{xz}, and c_{zy} among the three variables in terms of their pairwise correlations r_{xy}, r_{xz}, and r_{zy}.

We construct a model for the linear relationships based on the data-generating process depicted in Fig. 10B,

$$\begin{aligned} Z &= c_{xz}X + \delta_1, \\ Y &= c_{xy}X + c_{zy}Z + \delta_2. \end{aligned} \tag{7}$$

where the zero-mean "noise" δ_1 is independent of X while the zero-mean "noise" δ_2 is independent of X and Z. The parameter c_{xy} represents the direct causal effect of X on Y, controlling for the mediator Z, while c_{zy} captures the causal effect of Z on Y, accounting for the influence of the confounder X. With the relationships below for the model (7),

$$\begin{aligned} r_{xz} &= Corr(X,\ Z) = c_{xz}, \\ r_{xy} &= Corr(X,\ Y) = c_{xy} + c_{zy}r_{xz}, \\ r_{zy} &= Corr(Z,\ Y) = c_{xy}r_{xz} + c_{zy}, \end{aligned}$$

we obtain the following causal effects,

$$\begin{aligned} c_{xz} &= r_{xz}, \\ c_{xy} &= (r_{xy} - r_{xz}r_{zy})/(1 - r_{xz}^2), \\ c_{zy} &= (r_{zy} - r_{xy}r_{xz})/(1 - r_{xz}^2). \end{aligned} \tag{8}$$

The causal effects illustrated in Fig. 9B-C are calculated using the formula (8).

References

Allen, E.A., Erhardt, E.B., Calhoun, V.D., 2012. Data visualization in the neurosciences: Overcoming the curse of dimensionality. Neuron 74, 603–608.

Arbabyazd, L., Shen, K., Wang, Z., Hofmann-Apitius, M., Ritter, P., McIntosh, A.R., Battaglia, D., Jirsa, V., 2021. Virtual connectomic datasets in alzheimer's disease and aging using whole-brain network dynamics modelling. eneuro 8, ENEURO.0475–20.2021.

Baldermann, J.C., Petry-Schmelzer, J.N., Schüller, T., Mahfoud, L., Brandt, G.A., Dembek, T.A., van der Linden, C., Krauss, J.K., Szejko, N., Müller-Vahl, K.R., Ganos, C., Al-Fatly, B., Heiden, P., Servello, D.,

Galbiati, T., Johnson, K.A., Butson, C.R., Okun, M.S., Andrade, P., Domschke, K., Fink, G.R., Fox, M.D., Horn, A., Kuhn, J., Visser-Vandewalle, V., Barbe, M.T., 2024. A critical role of action-related functional networks in gilles de la tourette syndrome. Nature Communications 15.

Beynel, L., Powers, J.P., Appelbaum, L.G., 2020. Effects of repetitive transcranial magnetic stimulation on resting-state connectivity: A systematic review. NeuroImage 211, 116596.

Biswal, B., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic Resonance in Medicine 34, 537–541.

Bohning, D.E., Shastri, A., Nahas, Z., Lorberbaum, J.P., Andersen, S.W., Dannels, W.R., Haxthausen, E.U., Vincent, D.J., George, M.S., 1998. Echoplanar BOLD fMRI of brain activation induced by concurrent transcranial magnetic stimulation. Invest. Radiol. 33, 336–340.

Bonett, D.G., 2008. Meta-analytic interval estimation for bivariate correlations. Psychological Methods 13, 173–181.

Bortel, A., Pilgram, R., Yao, Z.S., Shmuel, A., 2020. Dexmedetomidine – commonly used in functional imaging studies – increases susceptibility to seizures in rats but not in wild type mice. Frontiers in Neuroscience 14.

Caballero-Gaudes, C., Reynolds, R.C., 2017. Methods for cleaning the BOLD fMRI signal. NeuroImage 154, 128–149.

Calhoun, V.D., Liu, J., Adalı, T., 2009. A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. NeuroImage 45, S163–S172.

Cardone, P., Alnagger, N., Annen, J., Bicego, A., Gosseries, O., Martial, C., 2024. Psychedelics and disorders of consciousness: the current landscape and the path forward. Neuroscience of Consciousness 2024.

Chen, G., Cai, Z., Taylor, P.A., 2024. Through the lens of causal inference: Decisions and pitfalls of covariate selection. Aperture Neuro 4.

Chen, G., Pine, D.S., Brotman, M.A., Smith, A.R., Cox, R.W., Taylor, P.A., Haller, S.P., 2022. Hyperbolic trade-off: The importance of balancing trial and subject sample sizes in neuroimaging. NeuroImage 247, 118786.

Chen, G., Taylor, P.A., Reynolds, R.C., Leibenluft, E., Pine, D.S., Brotman, M.A., Pagliaccio, D., Haller, S.P., 2023. BOLD Response is more than just magnitude: Improving detection sensitivity through capturing hemodynamic profiles. NeuroImage 277, 120224.

Deery, H.A., Liang, E.X., Moran, C., Egan, G.F., Jamadar, S.D., 2024. Metabolic connectivity has greater predictive utility for age and cognition than functional connectivity. Brain Communications 7.

Deshpande, G., Wang, Y., Robinson, J., 2022. Resting state fmri connectivity is sensitive to laminar connectional architecture in the human brain. Brain Informatics 9.

Drew, P.J., 2019. Vascular and neural basis of the BOLD signal. Current Opinion in Neurobiology 58, 61–69.

Drew, P.J., 2022. Neurovascular coupling: motive unknown. Trends in Neurosciences 45, 809–819.

Du, J., DiNicola, L.M., Angeli, P.A., Saadon-Grosman, N., Sun, W., Kaiser, S., Ladopoulou, J., Xue, A., Yeo, B.T.T., Eldaief, M.C., Buckner, R.L., 2024. Organization of the human cerebral cortex estimated within individuals: networks, global topography, and function. Journal of Neurophysiology 131, 1014–1082.

Edelman, G.M., Gally, J.A., 2001. Degeneracy and complexity in biological systems. Proceedings of the National Academy of Sciences 98, 13763–13768.

Ekstrom, A., 2010. How and when the fMRI BOLD signal relates to underlying neural activity: The danger in dissociation. Brain Research Reviews 62, 233–244.

Etkin, A., Mathalon, D.H., 2024. Bringing imaging biomarkers into clinical reality in psychiatry. JAMA Psychiatry .

Fisher, R.A., 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10, 507.

Friston, K.J., 2011. Functional and effective connectivity: A review. Brain Connectivity 1, 13–36.

Glymour, C., Zhang, K., Spirtes, P., 2019. Review of causal discovery methods based on graphical models. Frontiers in Genetics 10.

Gnambs, T., 2023. A brief note on the standard error of the pearson correlation. Collabra: Psychology 9.

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proceedings of the National Academy of Sciences 100, 253–258.

Grimm, C., Duss, S.N., Privitera, M., Munn, B.R., Karalis, N., Frässle, S., Wilhelm, M., Patriarchi, T., Razansky, D., Wenderoth, N., Shine, J.M., Bohacek, J., Zerbi, V., 2024. Tonic and burst-like locus coeruleus stimulation distinctly shift network activity across the cortical hierarchy. Nature Neuroscience 27, 2167–2177.

Hampson, M., Hoffman, R.E., 2010. Transcranial magnetic stimulation and connectivity mapping: tools for studying the neural bases of brain disorders. Front. Syst. Neurosci. 4.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21, 1639–1651.

Hill, J., Perrett, G., Hancock, S., Win, L., Bergner, Y., 2024. Causal language and statistics instruction: Evidence from a randomized experiment. Statistics Education Research Journal 23, 5–5.

Howarth, C., Mishra, A., Hall, C.N., 2020. More than just summed neuronal activity: how multiple cell types shape the bold response. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20190630.

Huber, L., Handwerker, D.A., Jangraw, D.C., Chen, G., Hall, A., Stüber, C., Gonzalez-Castillo, J., Ivanov, D., Marrett, S., Guidi, M., Goense, J., Poser, B.A., Bandettini, P.A., 2017. High-resolution cbv-fmri allows mapping of laminar activity and connectivity of cortical input and output in human m1. Neuron 96, 1253–1263.e7.

Huber, L., Ivanov, D., Handwerker, D.A., Marrett, S., Guidi, M., Uludağ, K., Bandettini, P.A., Poser, B.A., 2018. Techniques for blood volume fmri with vaso: From low-resolution mapping towards sub-millimeter layer-dependent applications. NeuroImage 164, 131–143.

Huck, J., Jäger, A., Schneider, U., Grahl, S., Fan, A.P., Tardif, C., Villringer, A., Bazin, P., Steele, C.J., Gauthier, C.J., 2023. Modeling venous bias in resting state functional <scp>mri</scp> metrics. Human Brain Mapping 44, 4938–4955.

Huck, J., Wanner, Y., Fan, A.P., Jäger, A.T., Grahl, S., Schneider, U., Villringer, A., Steele, C.J., Tardif, C.L., Bazin, P.L., Gauthier, C.J., 2019. High resolution atlas of the venous brain vasculature from 7 t quantitative susceptibility maps. Brain Structure and Function 224, 2467–2485.

Ibrahim, B., Suppiah, S., Ibrahim, N., Mohamad, M., Hassan, H.A., Nasser, N.S., Saripan, M.I., 2021. Diagnostic power of resting-state fmri for detection of network connectivity in alzheimer's disease and mild cognitive impairment: A systematic review. Human Brain Mapping 42, 2941–2968.

Ionescu, T.M., Amend, M., Hafiz, R., Biswal, B.B., Wehrl, H.F., Herfert, K., Pichler, B.J., 2021. Elucidating the complementarity of resting-state networks derived from dynamic [18f]fdg and hemodynamic fluctuations using simultaneous small-animal pet/mri. NeuroImage 236, 118045.

Jammal Salameh, L., Bitzenhofer, S.H., Hanganu-Opatz, I.L., Dutschmann, M., Egger, V., 2024. Blood pressure pulsations modulate central neuronal activity via mechanosensitive ion channels. Science 383.

Kellman, P., McVeigh, E.R., 2005. Image reconstruction in snr units: A general method for snr measurement†. Magnetic Resonance in Medicine 54, 1439–1447.

Khoo, H.M., von Ellenrieder, N., Zazubovits, N., Dubeau, F., Gotman, J., 2017. Epileptic networks in action: Synchrony between distant hemodynamic responses. Annals of Neurology 82, 57–66.

Korponay, C., Janes, A.C., Frederick, B.B., 2024. Brain-wide functional connectivity artifactually inflates throughout functional magnetic resonance imaging scans. Nature Human Behaviour 8, 1568–1580.

Kundu, P., Voon, V., Balchandani, P., Lombardo, M.V., Poser, B.A., Bandettini, P.A., 2017. Multi-echo fmri: A review of applications in fmri denoising and analysis of bold signals. NeuroImage 154, 59–80.

Larivière, S., Weng, Y., Vos de Wael, R., Royer, J., Frauscher, B., Wang, Z., Bernasconi, A., Bernasconi, N., Schrader, D.V., Zhang, Z., Bernhardt, B.C., 2020. Functional connectome contractions in temporal lobe epilepsy: Microstructural underpinnings and predictors of surgical outcome. Epilepsia 61, 1221–1233.

Lewis, D., 1986. Causal Explanation, in: Lewis, D. (Ed.), Philosophical Papers Vol. Ii. Oxford University Press, pp. 214–240.

Li, X., Bianchini Esper, N., Ai, L., Giavasis, S., Jin, H., Feczko, E., Xu, T., Clucas, J., Franco, A., Sólon Heinsfeld, A., Adebimpe, A., Vogelstein, J.T., Yan, C.G., Esteban, O., Poldrack, R.A., Craddock, C., Fair, D., Satterthwaite, T., Kiar, G., Milham, M.P., 2024. Moving beyond processing- and analysis-related variation in resting-state functional brain imaging. Nat. Hum. Behav. 8, 2003–2017.

Liu, T.T., 2013. Neurovascular factors in resting-state functional mri. NeuroImage 80, 339–348.

Liu, T.T., 2016. Noise contributions to the fMRI signal: An overview. NeuroImage 143, 141–151.

Liu, T.T., Brown, G.G., 2007. Measurement of cerebral perfusion with arterial spin labeling: Part 1. methods. Journal of the International Neuropsychological Society 13.

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.

Lu, H., Jaime, S., Yang, Y., 2019. Origins of the resting-state functional mri signal: Potential limitations of the "neurocentric" model. Frontiers in Neuroscience 13.

Lu, H., van Zijl, P.C., 2012. A review of the development of vascular-space-occupancy (vaso) fmri. NeuroImage 62, 736–742.

Luckett, P.H., Lee, J.J., Park, K.Y., Raut, R.V., Meeker, K.L., Gordon, E.M., Snyder, A.Z., Ances, B.M., Leuthardt, E.C., Shimony, J.S., 2023. Resting state network mapping in individuals using deep learning. Frontiers in Neurology 13.

Lynch, C.J., Elbau, I.G., Ng, T., Ayaz, A., Zhu, S., Wolk, D., Manfredi, N., Johnson, M., Chang, M., Chou, J., Summerville, I., Ho, C., Lueckel, M., Bukhari, H., Buchanan, D., Victoria, L.W., Solomonov, N., Goldwaser, E., Moia, S., Caballero-Gaudes, C., Downar, J., Vila-Rodriguez, F., Daskalakis, Z.J., Blumberger, D.M., Kay, K., Aloysi, A., Gordon, E.M., Bhati, M.T., Williams, N., Power, J.D., Zebley, B., Grosenick, L., Gunning, F.M., Liston, C., 2024. Frontostriatal salience network expansion in individuals in depression. Nature 633, 624–633.

Marek, S., Tervo-Clemmens, B., Calabro, F.J., et al., 2022. Reproducible brain-wide association studies require thousands of individuals. Nature , 1–7.

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., Smallwood, J., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences 113, 12574–12579.

Mark, C.I., Mazerolle, E.L., Chen, J.J., 2015. Metabolic and vascular origins of the bold effect: Implications for imaging pathology and resting-state brain function: Understanding bold in disease and rest. Journal of Magnetic Resonance Imaging 42, 231–246.

Medaglia, J.D., 2017. Graph theoretic analysis of resting state functional mr imaging. Neuroimaging Clinics of North America 27, 593–607.

Mehler, D.M.A., Kording, K.P., 2020. The lure of misleading causal statements in functional connectivity research. `arXiv:1812.03363`.

Murphy, K., Fox, M.D., 2017. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. NeuroImage 154, 169–173.

Narain, C., 2006. Going down boldly. Nature Neuroscience 9, 474–474.

van Niftrik, C.H.B., Hiller, A., Sebök, M., Halter, M., Duffin, J., Fisher, J.A., Mikulis, D.J., Regli, L., Piccirelli, M., Fierstra, J., 2023. Heterogeneous motor bold-fmri responses in brain areas exhibiting negative bold cerebrovascular reactivity indicate that steal phenomenon does not always result from exhausted cerebrovascular reserve capacity. Magnetic Resonance Imaging 103, 124–130.

Papo, D., Buldú, J.M., 2025. Biological detail and graph structure in network neuroscience. Frontiers in Network Physiology 5.

Pearl, J., 2009. Causality: Models, Reasoning and Inference. 2nd edition ed., Cambridge University Press, Cambridge New York, NY Port Melbourne New Delhi Singapore.

Peel, L., Peixoto, T.P., De Domenico, M., 2022. Statistical inference links data and theory in network science. Nature Communications 13.

Piramide, N., De Micco, R., Siciliano, M., Silvestro, M., Tessitore, A., 2024. Resting-state functional mri approaches to parkinsonisms and related dementia. Current Neurology and Neuroscience Reports 24, 461–477.

Pruim, R.H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. Ica-aroma: A robust ica-based strategy for removing motion artifacts from fmri data. NeuroImage 112, 267–277.

Rangaprakash, D., Barry, R.L., Deshpande, G., 2023. The confound of hemodynamic response function variability in human resting-state functional mri studies. Frontiers in Neuroscience 17.

Reid, A.T., Headley, D.B., Mill, R.D., Sanchez-Romero, R., Uddin, L.Q., Marinazzo, D., Lurie, D.J., Valdés-Sosa, P.A., Hanson, S.J., Biswal, B.B., Calhoun, V., Poldrack, R.A., Cole, M.W., 2019. Advancing functional connectivity research from association to causation. Nature Neuroscience 22, 1751–1760.

Rosa, M., Kilner, J., Blankenburg, F., Josephs, O., Penny, W., 2010. Estimating the transfer function from neuronal activity to bold using simultaneous eeg-fmri. NeuroImage 49, 1496–1509.

Rubinov, M., 2023. Circular and unified analysis in network neuroscience. eLife 12.

Salmon, W.C., 1984. Scientific Explanation and the Causal Structure of the World. Princeton University Press, Princeton, New Jersey.

Shahsavarani, S., Thibodeaux, D.N., Xu, W., Kim, S.H., Lodgher, F., Nwokeabia, C., Cambareri, M., Yagielski, A.J., Zhao, H.T., Handwerker, D.A., Gonzalez-Castillo, J., Bandettini, P.A., Hillman, E.M., 2023. Cortex-wide neural dynamics predict behavioral states and provide a neural basis for resting-state dynamic functional connectivity. Cell Reports 42, 112527.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for fmri. NeuroImage 54, 875–891.

Spearman, C., 1904. The proof and measurement of association between two things. American Journal of Psychology 15, 72–101.

Spirtes, P., Glymour, C., Scheines, R., 2001. Causation, Prediction, and Search. The MIT Press.

Spirtes, P., Zhang, K., 2016. Causal discovery and inference: concepts and recent methodological advances. Applied Informatics 3.

Sporns, O., 2018. Graph theory methods: applications in brain networks. Dialogues in Clinical Neuroscience 20, 111–121.

Suarez, A., Valdes-Hernandez, P.A., Moshkforoush, A., Tsoukias, N., Riera, J., 2021. Arterial blood stealing as a mechanism of negative bold response: From the steady-flow with nonlinear phase separation to a windkessel-based model. Journal of Theoretical Biology 529, 110856.

Sundermann, B., Pfleiderer, B., McLeod, A., Mathys, C., 2024. Seeing more than the tip of the iceberg: Approaches to subthreshold effects in functional magnetic resonance imaging of the brain. Clin. Neuroradiol. 34, 531–539.

Tagliazucchi, E., Laufs, H., 2014. Decoding wakefulness levels from typical fmri resting-state data reveals reliable drifts between wakefulness and sleep. Neuron 82, 695–708.

Taylor, J.J., Kurt, H.G., Anand, A., 2021. Resting state functional connectivity biomarkers of treatment response in mood disorders: A review. Frontiers in Psychiatry 12.

Taylor, P.A., Aggarwal, H., Bandettini, P., Barilari, M., Bright, M., Caballero-Gaudes, C., Calhoun, V., Chakravarty, M., Devenyi, G., Evans, J., Garza-Villarreal, E., Rasgado-Toledo, J., Gau, R., Glen, D., Goebel, R., Gonzalez-Castillo, J., Gulban, O.F., Halchenko, Y., Handwerker, D., Hanayik, T., Lauren, P., Leopold, D., Lerch, J., Mathys, C., McCarthy, P., McLeod, A., Mejia, A., Moia, S., Nichols, T., Pernet, C., Pessoa, L., Pfleiderer, B., Rajendra, J., Reyes, L., Reynolds, R., Roopchansingh, V., Rorden, C., Russ, B., Sundermann, B., Thirion, B., Torrisi, S., Chen, G., 2025. Go figure: Transparency in neuroscience images preserves context and clarifies interpretation `arXiv:2504.07824`.

Taylor, P.A., Reynolds, R.C., Calhoun, V., Gonzalez-Castillo, J., Handwerker, D.A., Bandettini, P.A., Mejia, A.F., Chen, G., 2023. Highlight results, don't hide them: Enhance interpretation, reduce biases and improve reproducibility. NeuroImage 274, 120138.

Tetereva, A., Knodt, A.R., Melzer, T.R., van der Vliet, W., Gibson, B., Hariri, A.R., Whitman, E.T., Li, J., Lal Khakpoor, F., Deng, J., Ireland, D., Ramrakha, S., Pat, N., 2025. Improving predictability, reliability, and generalizability of brain-wide associations for cognitive abilities via multimodal stacking. PNAS Nexus 4.

Tozzi, L., Fleming, S.L., Taylor, Z.D., Raterink, C.D., Williams, L.M., 2020. Test-retest reliability of the human functional connectome over consecutive days: identifying highly reliable portions and assessing the impact of methodological choices. Network Neuroscience 4, 925–945.

Tsvetanov, K.A., Henson, R.N.A., Rowe, J.B., 2020. Separating vascular and neuronal effects of age on fmri bold signals. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20190631.

Tu, W., Cramer, S.R., Zhang, N., 2024. Disparity in temporal and spatial relationships between resting-state electrophysiological and fmri signals. eLife 13.

Tu, W., Zhang, N., 2022. Neural underpinning of a respiration-associated resting-state fmri network. eLife 11.

Turkheimer, E., 2024. Understanding the Nature–Nurture Debate. Cambridge University Press.

Vafaii, H., Mandino, F., Desrosiers-Grégoire, G., O'Connor, D., Markicevic, M., Shen, X., Ge, X., Herman, P., Hyder, F., Papademetris, X., Chakravarty, M., Crair, M.C., Constable, R.T., Lake, E.M.R., Pessoa, L., 2024. Multimodal measures of spontaneous brain activity reveal both common and divergent patterns of cortical functional organization. Nature Communications 15.

Voigt, K., Liang, E.X., Misic, B., Ward, P.G.D., Egan, G.F., Jamadar, S.D., 2022. Metabolic and functional connectivity provide unique and complementary insights into cognition-connectome relationships. Cerebral Cortex 33, 1476–1488.

Wei, W., Zhang, K., Chang, J., Zhang, S., Ma, L., Wang, H., Zhang, M., Zu, Z., Yang, L., Chen, F., Fan, C., Li, X., 2024. Analyzing 20 years of resting-state fmri research: Trends and collaborative networks revealed. Brain Research 1822, 148634.

Wirsich, J., Jorge, J., Iannotti, G.R., Shamshiri, E.A., Grouiller, F., Abreu, R., Lazeyras, F., Giraud, A.L., Gruetter, R., Sadaghiani, S., Vulliémoz, S., 2021. The relationship between eeg and fmri connectomes is reproducible across simultaneous eeg-fmri studies from 1.5t to 7t. NeuroImage 231, 117864.

Xavier, M., Esteves, I., Jorge, J., Abreu, R., Giraud, A.L., Sadaghiani, S., Wirsich, J., Figueiredo, P., 2025. Consistency of resting-state correlations between fmri networks and eeg band power. Imaging Neuroscience 3.

Yamada, T., Hashimoto, R.i., Yahata, N., Ichikawa, N., Yoshihara, Y., Okamoto, Y., Kato, N., Takahashi, H., Kawato, M., 2017. Resting-state functional connectivity-based biomarkers and functional mri-based neurofeedback for psychiatric disorders: A challenge for developing theranostic biomarkers. International Journal of Neuropsychopharmacology 20, 769–781.

Yuen, N.H., Osachoff, N., Chen, J.J., 2019. Intrinsic frequencies of the resting-state fmri signal: The frequency dependence of functional connectivity and the effect of mode mixing. Frontiers in Neuroscience 13.

Zalesky, A., Fornito, A., Bullmore, E., 2012. On the use of correlation as a measure of network connectivity. NeuroImage 60, 2096–2106.

Zhang, X., Pan, W.J., Keilholz, S., 2019. The relationship between local field potentials and the blood-oxygenation-level dependent mri signal can be non-linear. Frontiers in Neuroscience 13.

Zimmerman, D.W., Zumbo, B.D., Williams, R.H., 2003. Bias in estimation and hypothesis testing of correlation. Psicológica 24, 133–158.