A Causal Formulation of Spike-Wave Duality

Kasra Jalaldoust[*]
Department of Computer Science
Columbia University
kasra@cs.columbia.edu

Erfan Zabeh[*†]
Mortimer B. Zuckerman Mind Brain Behavior Institute
Columbia University
erfan.zabeh@columbia.edu

Abstract

Understanding the relationship between brain activity and behavior is a central goal of neuroscience. Despite significant advances, a fundamental dichotomy persists: neural activity manifests as both discrete spikes of individual neurons and collective waves of populations. Both neural codes correlate with behavior, yet correlation alone cannot determine whether waves exert a causal influence or merely reflect spiking dynamics without causal efficacy. According to the Causal Hierarchy Theorem, no amount of observational data—however extensive—can settle this question; causal conclusions require explicit structural assumptions or careful experiment designs that directly correspond to the causal effect of interest. We develop a formal framework that makes this limitation precise and constructive. Formalizing epiphenomenality via the invariance of interventional distributions Structural Causal Models (SCMs), we derive a certificate of sufficiency from Pearl's do-calculus that specifies when variables can be removed from the model without loss of causal explainability and clarifies how interventions should be interpreted under different causal structures of spike-wave duality. The purpose of this work is not to resolve the spike–wave debate, but to reformulate it. We shift the problem from asking which signal matters most to asking under what conditions any signal can be shown to matter at all. This reframing distinguishes prediction from explanation and offers neuroscience a principled route for deciding when waves belong to mechanism and when they constitute a byproduct of underlying coordination

Introduction and background

Neuroscience seeks to explain how diverse forms of neural activity give rise to behavior. These forms appear in many layers: the discrete action potentials of *spikes*, the collective oscillations of *waves*, and slower modulatory processes such as neuromodulator release, glial signaling, or hemodynamic changes. Each can be measured, each can be correlated with behavior, but correlation alone does not reveal causal force. A variable may appear essential while in fact being *epiphenomenal*: a secondary effect that mirrors true causes without altering them [1, 2]. Oscillations have long stood at the center of this dilemma [3, 4]. They are celebrated as carriers of coordination and dismissed as echoes of spiking activity, both at once [4, 5, 6, 7, 8, 9, 10]. To call waves epiphenomenal is not a small claim—it suggests that what looks like the music of the brain may in fact be only its resonance. To know the difference requires more than observation; it demands a causal lens that can separate what drives from what follows.

Causal inference has been long developed in computer science and statistics, and has proven its application in fields such as economics and social sciences[11, 12, 13, 14]. In neuroscience, causality has traditionally been interpreted through stimulation and perturbation, though recent work has begun

[*]The authors contributed equally.
[†]Corresponding author.

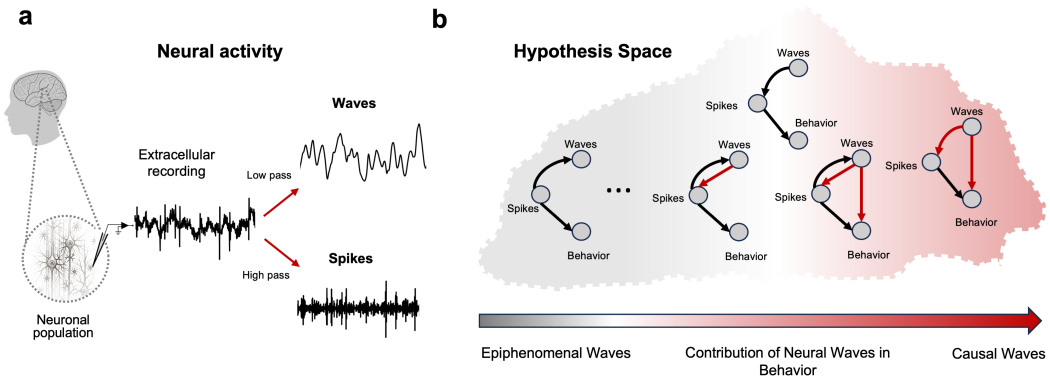

Figure 1: **Neural activity and hypothesis space for spike–wave–behavior interactions. (a)** Raw neuronal recording can be decomposed into population-level waves (low-pass filtered) and single-neuron spikes (high-pass filtered). These two levels of description coexist within the same signal but may play distinct roles in shaping behavior. **(b)** Candidate causal graphs span a spectrum from epiphenomenal to causal waves. Each node represents waves, spikes, or behavior; arrows denote possible causal dependencies. Red edges highlight putative direct influences of waves. Graphs on the left illustrate epiphenomenal waves, where waves carry no causal force once spikes are accounted for.

to examine it as a conceptual problem in its own right [15, 16, 17, 18]. Yet, no formal framework has been proposed to analyze the causal structure of brain waves and spikes. This absence underscores the need for a theoretical foundation—complementary to experimental approaches—that clarifies under what conditions waves can be said to exert causal influence on spiking activity and behavior.

The predominant view in neuroscience holds that spiking activity constitutes the fundamental language of computation between neurons [19, 20]. This perspective is grounded in several compelling arguments that suggest neural oscillations may be epiphenomenal rather than causally relevant to behavior. The spatial resolution limitations of field potential recordings raise fundamental questions about the precision with which population-level signals can interact with individual neurons. Neural oscillations are typically measured using relatively low-resolution recording electrodes, which aggregate activity across large populations of neurons [21, 22]. This coarse spatial sampling brings into question how fine-grained these population effects can be when interacting with individual neurons, potentially limiting their capacity for precise computational influence [23, 24]. Furthermore, the most persistent neuronal oscillations are prominently observed during sleep and unconscious conditions, suggesting their potential irrelevance to active behavioral processes in the brain [25, 26, 27, 28]. The prevalence of oscillatory activity during states of reduced consciousness has led some researchers to view these rhythms as byproducts of neural network dynamics rather than active contributors to cognitive function [29, 30].

However, extensive evidence supports the view that neural waves may indeed exert causal influence on behavior and neural dynamics. Multiple contributing factors beyond spiking activity, including neuromodulatory effects and subthreshold activity, are known to contribute to neuronal dynamics [31, 32, 33, 34]. Field potentials are thought to include components reflecting the aggregated effects of this subthreshold, non-spiking activity, potentially providing a mechanism for causal influence [35, 36]. The clinical applicability of brain waves provides compelling evidence for their functional relevance. Neural oscillations are successfully used in medical contexts for therapeutic monitoring, such as tracking anesthesia states in surgery rooms through field potential measurements [37, 38, 39]. This level of clinical applicability reinforces the functional relevance of brain waves beyond their well-documented behavioral associations across cognitive domains such as attention, memory, and navigation [40, 41, 42, 43, 44, 45].

From a biophysical perspective, even the most conservative view acknowledges that field potentials represent summations of population-level spiking activity and carry information about system states that can interact with individual neurons [46, 47]. The presence of gap junctions in the brain is well-established, and accordingly, the effect of population-level voltage signals on individual neurons through electrical coupling is biophysically plausible [48, 49, 50]. Furthermore, field potentials contain both oscillatory and non-oscillatory components, and much of the debate focuses on the

oscillatory component while less attention has been paid to non-periodic field potentials [51, 52, 53]. This distinction is important because different components of the field potential may have different causal roles in neural computation.

Analyses of spike–wave relationships have typically relied on correlation-based approaches, which are inherently limited. According to the Causal Hierarchy Theorem [54], making causal claims requires one of two elements beyond observational data: (1) explicit causal assumptions, or (2) interventional data from randomized controlled trials. This fundamental limitation means that no amount of observational data, regardless of sample size, can definitively resolve questions of causality without additional assumptions or experimental interventions. This theoretical constraint is particularly relevant to neuroscience, where perturbations are not necessarily equivalent to causal interventions [55]. The complex, interconnected nature of neural systems means that experimental manipulations may have cascading effects that obscure rather than clarify causal relationships.

This paper addresses these fundamental challenges by proposing a formal causal framework for analyzing spike-wave duality. Rather than attempting to resolve the spike-wave debate directly, our contribution is to provide a causal paradigm for reasoning about epiphenomenality in this context. This framework specifies how one might formally test when wave-behavior interactions are causally sufficient versus epiphenomenal, leaving the ultimate empirical conclusion to the neuroscience community. We present our results through a structured analysis that formalizes the conditions under which neural waves can be considered epiphenomenal or causally relevant, providing both theoretical foundations and practical tools for empirical investigation.

Problem statement and causal formulation

We present the problem statment and results in form of a debate between two fictional neuroscientists; John, who believes that waves are epiphenomenal, and Earl, who believes in causal relevance of waves to the behavior.

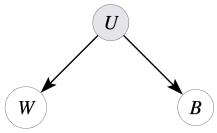

Figure 2: Confounded graph

Earl has conducted a passive (i.e., non-interventional) brain recording that involves behavior and waves simultaneously. He observes that waves and behavior are statistically dependent, i.e., $P(\mathbf{B} \mid \mathbf{W}) \neq P(\mathbf{B})$. Based on this observation, Earl hypothesizes that waves causally influence the behavior, and presents the findings to John. John disagrees with the causal conclusion, and speculates that the statistical dependence might be solely due to unobserved confounding, without any causal influence from waves to behavior. John's intuition is represented by the *confounded* graph shown in Figure 2; here, $\mathbf{B}$ obtains it's value as a function of unobserved confounder $\mathbf{U}$, without any mechanistic influence from $\mathbf{W}$. Earl realizes that his passively collected data can not refute the possibility presented by John.

In an attempt to formalize the causal claim, Earl designs an imaginary controlled experiment. He would perturb the waves physically to override the influence from the confounders $\mathbf{U}$ on $\mathbf{W}$, and at the same time samples from behavior under this intervention. The corresponding distribution is denoted by $P(\mathbf{B}; do(\mathbf{W}))$ in causal inference notation. Next, Earl would compare it with the passive/observational distribution of behavior denoted as $P(\mathbf{B})$. This comparison is a sufficient criterion to assess the causal claim:

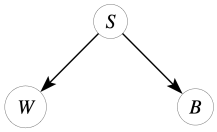

Figure 3: Fork graph

- If $P(\mathbf{B}; do(\mathbf{W})) \neq P(\mathbf{B})$, then there exists *some* causal influence from waves on behavior, and more refined measures are needed to quantify this causal effect.
- If $P(\mathbf{B}; do(\mathbf{W})) = P(\mathbf{B})$, then there is no causal influence from waves on behavior.

As the debate is unsettled, John contemplates the idea that the spikes generated by certain neurons are the sole determinant of behavior. If true, this deems waves *epiphenomenon*, with no causal relevance to behavior given the spikes are measured; the graph in Figure 3 summarizes this narrative. Notably, considering waves alongside spikes offers a better prediction for behavior, thus, simply discarding the waves is unjustified as long as behavior prediction is a scientific objective.[56, 57]. In probabilistic terms we write,

$$P(\text{behavior} \mid \text{spikes}, \text{waves}) \neq P(\text{behavior} \mid \text{spikes}). \quad (1)$$

Notably, Equation (1) is inconsistent with the fork graph; rule 1 of do-calculus implies that for all structural causal model that induce the fork graph, we have equality instead of inequality.

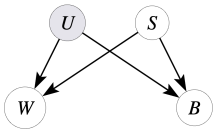

Figure 4: Confounded fork graph

If both the fork graph (Figure 3) and confounded graph (Figure 2) are not consistent with the brain data, what would be compatible graph? A plausible scenario is summarized in Figure 4; there are observed and unobserved phenomenon both causing the waves and behavior, simultaneously. John argues that if this graph is true, then the waves are epiphenomenon, since there is no direct causal influence from $\mathbf{W}$ to $\mathbf{B}$.

Again, to reject epiphenomenality, Earl must conduct the wave intervention experiment, and to justify epiphenomenality, John must observe all unobserved factors affecting the waves and behavior simultaneously, i.e., absorbing all of $\mathbf{U}$ into $\mathbf{S}$ in the confounded-fork graph.

John proposes that the waves influence some spikes and also get influence by some spikes, without any direct effect on behavior, as shown schematically in Figure 5. To investigate this further, Earl decomposes spikes (and all other factors) into subsets:

- **Confounders** $\mathbf{C}$, such that $\mathbf{C} \to \mathbf{W}$ and $\mathbf{C} \to \mathbf{B}$.
- **Mediators** $\mathbf{M}$, such that $\mathbf{W} \to \mathbf{M} \to \mathbf{B}$.
- **Exogenous variables** $\mathbf{E}$, such as $\mathbf{E}_W \to \mathbf{W}$ and $\mathbf{E}_B \to \mathbf{B}$.

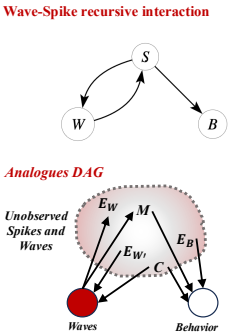

Figure 5: Systematic decomposition of unobserved spikes and waves.

Figure 5 shows this decomposition. In each experimental setting we may observe the above spikes, possibly in partial. Below, we consider different observability scenarios, and discuss epiphenomenality in each instance. In each example, the set $\mathbf{S}$ is the observed spikes and $\mathbf{U}$ contains all other factors, including the unobserved spikes.

Partially observed mediators. Suppose $\mathbf{M} \not\subset \mathbf{S}$, meaning that there exists some mediators that are not observed in $\mathbf{S}$. Assuming all other factors are observed, we describe this situation with the graph in Figure 6. There exists direct causal paths $\mathbf{W} \to M \to \mathbf{B}$ for every $M \in \mathbf{M} \setminus \mathbf{S}$. Therefore, interventions on $\mathbf{W}$ affect M and the changes in M carries over to $\mathbf{B}$, which implies that waves are causal to behavior in this case. This intuition is justified below.

Proposition 1. *[From causal paths to causal waves] Consider an SCM over the sets of variables* $\mathbf{W}, \mathbf{B}, \mathbf{S}, \mathbf{U}$. *If there exists a directed path from a wave* $W \in \mathbf{W}$ *to a behavior* $B \in \mathbf{B}$*, such as,*

$$W \to U_1 \to U_2 \to \ldots \to U_n \to B,$$

where $U_1, U_2, \ldots U_n \in \mathbf{U}$*, then,*

$$P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W})) \neq P(\mathbf{B} \mid \mathbf{S}). \quad (2)$$

In words, waves exert a causal effect on behavior whenever not all directed paths from waves to behavior are blocked by observation of at least one variable along each path (i.e., included in $\mathbf{S}$). Below we formally define this notion as causal relevance of a scientific phenomenon.

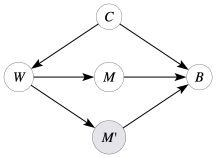

Figure 6: Partially observed mediators.

Definition 1 (Epiphenomenality)**.** *A set of variables* $\mathbf{Z}$ *is epiphenomenal to the outcome* Y *given the predictors* $\mathbf{X}$ *if,*

$$P(\mathbf{Y} \mid \mathbf{X}; do(\mathbf{Z})) = P(\mathbf{Y} \mid \mathbf{X}), \quad (3)$$

i.e., the passive distribution of outcome $\mathbf{Y}$ *conditional on the value of predictors* $\mathbf{X}$ *remains invariant under interventions on the variables* $\mathbf{Z}$*.*

Front-door graph. Suppose all mediators are observed (i.e., $\mathbf{M} \subset \mathbf{S}$), but $\mathbf{C} \subset \mathbf{U}$, meaning that none of the common causes of waves and behavior are observed. The graph in Figure 7 describes this situation. One might claim that since all of the causal influence of waves on behavior is mediated by observed spikes, it deems waves epiphenomenal. This interpretation is not accurate, as stated formally below.

Proposition 2. *[Causal waves in front-door graph] If the graph in Figure 7 is induced by the true SCM that governs the brain activities, then the waves would have causal influence on behavior. Formally speaking,*

$$P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W})) \neq P(\mathbf{B} \mid \mathbf{S}) \tag{4}$$

holds for all SCMs that induces the front-door graph, except for a measure zero subset.

proof sketch. Using do-calculus we can not derive the r.h.s. from the l.h.s., and since do-calculus is a complete derivation system ([58]), the above holds for all nondegenerate SCMs.

Propositions 1 and 2 apply to only specific graphs (or class of graphs), while the next result is a general graphical characterization of epiphenomenality which applies to any causal graph that we have not considered here.

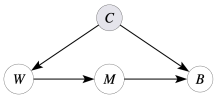

Figure 7: Front-door graph

Theorem 1. *Let Z, X, Y be disjoint sets of variables in a causal diagram $\mathcal{G}$. A set of variables Z is epiphenomenal to Y given X for all SCMs that induce the graph $\mathcal{G}$, if,*

$$Z \perp\!\!\!\perp_d Y \mid X \text{ in } \mathcal{G}_{\overline{Z(X)}}, \tag{5}$$

where $Z(X)$ are nodes in Z that are not ancestor to any nodes in X, and $\mathcal{G}_{\overline{Z(X)}}$ denotes the graph obtained from removing the arrows that point to the set of variables $Z(X)$.

Attested by Theorem 1, a certain case of epiphenomenality is when all confounders and mediators from Figure 5 are observed (Figure 8), which implies $P(\mathbf{B} \mid \mathbf{C}, \mathbf{M}; do(\mathbf{W})) = P(\mathbf{B} \mid \mathbf{C}, \mathbf{M})$.

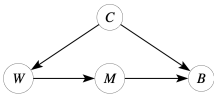

Figure 8: Epiphenomenality graph.

Discussion

This paper reframes the spike–wave debate in causal terms. Rather than asking whether waves matter more or less than spikes, we show how to decide when any signal matters at all. We define epiphenomenality as invariance of interventional distributions, provide a graph-based certificate for collapsing variables without causal loss, and demonstrate that the causal role of waves depends on the assumed graph. The key message is twofold. First, decoding success does not establish mechanism: prediction and explanation live at different levels. Second, causal claims require explicit diagrams and identification targets. Our framework clarifies when waves can be treated as summaries of spikes and when such abstraction erases mechanism. Practically, this shifts multiscale neuroscience from post-hoc debates to explicit causal programs: state assumptions, design experiments to test $P(B; \mathrm{do}(W))$ versus $P(B)$, and report results relative to the graph. While our formulation treats hidden causes abstractly within the exogenous space of the structural model, future theoretical extensions could expand this treatment by formalizing the role of latent confounders more explicitly. Approaches such as Instrumental Variable models or latent factor frameworks [59, 60] could, in principle, be embedded within our causal framing to represent structured hidden variables rather than unmodeled noise. Such extensions would not alter the conceptual component of our argument but would refine how unobserved causal structure is represented, bridging our formalism with probabilistic models of latent neural dynamics. Challenges such as hidden confounders, recursive coupling, and nonstationarity persist, yet within this framework they appear not as ambiguities but as declared assumptions, making the boundaries of inference explicit rather than contested.

In closing, the goal is not to decide whether waves cause behavior. It is to supply the tools for asking that question rigorously. By grounding the debate in causal model, interventions, and graphs, we return to the abstract's claim: the value of a signal lies not in its correlates, but in the causal answers it makes possible.

References

[1] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

[2] James Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, 2003.

[3] Sander van Bree, Daniel Levenstein, Matthew R Krause, Bradley Voytek, and Richard Gao. Processes and measurements: a framework for understanding neural oscillations in field potentials. *Trends in cognitive sciences*, 29(5):448–466, 2025.

[4] György Buzsáki. Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7:446–451, 2004. doi: 10.1038/nn1233.

[5] György Buzsáki. *Rhythms of the Brain*. Oxford University Press, New York, USA, 2006. ISBN 978-0195301069.

[6] Pascal Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–480, 2005.

[7] Pascal Fries. Rhythms for cognition: communication through coherence. *Neuron*, 88(1):220–235, 2015.

[8] Kai J Miller, Lise B Sorensen, Jeffrey G Ojemann, and Marcel Den Nijs. Power-law scaling in the brain surface electric potential. *PLoS Computational Biology*, 5(12):e1000609, 2009.

[9] Supratim Ray and John HR Maunsell. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biology*, 9(4):e1000610, 2011.

[10] Steven L Bressler and Carsten G Richter. Interareal oscillatory synchronization in top-down neocortical processing. *Current Opinion in Neurobiology*, 31:62–66, 2015.

[11] Guido W Imbens and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[12] Miguel A Hernán and James M Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.

[13] J. Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009.

[14] Paul Hünermund and Elias Bareinboim. Causal inference and data fusion in econometrics. *The Econometrics Journal*, 28(1):41–82, 2025.

[15] Konrad Körding and Joshua Tenenbaum. Causal inference in sensorimotor integration. *Advances in neural information processing systems*, 19, 2006.

[16] Ioana E Marinescu, Patrick N Lawlor, and Konrad P Kording. Quasi-experimental causality in neuroscience and behavioural research. *Nature human behaviour*, 2(12):891–898, 2018.

[17] Drew H Bailey, Alexander J Jung, Adriene M Beltz, Markus I Eronen, Christian Gische, Ellen L Hamaker, Konrad P Kording, Catherine Lebel, Martin A Lindquist, Julia Moeller, et al. Causal inference on human behaviour. *Nature human behaviour*, 8(8):1448–1459, 2024.

[18] Shan H Siddiqi, Konrad P Kording, Josef Parvizi, and Michael D Fox. Causal mapping of human brain function. *Nature reviews neuroscience*, 23(6):361–375, 2022.

[19] Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: Exploring the Neural Code*. MIT Press, 1999.

[20] Peter Dayan and Laurence F Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[21] György Buzsáki, Costas A. Anastassiou, and Christof Koch. The origin of extracellular fields and currents — eeg, ecog, lfp and spikes. *Nature Reviews Neuroscience*, 13:407–420, 2012. doi: 10.1038/nrn3241.

[22] Gaute T Einevoll, Christoph Kayser, Nikos K Logothetis, and Stefano Panzeri. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11): 770–785, 2013.

[23] Yoshinao Kajikawa and Charles E Schroeder. How local is the local field potential? *Neuron*, 72(5): 847–858, 2011.

[24] Henrik Lindén, Tom Tetzlaff, Tobias C Potjans, Klas H Pettersen, Sonja Grün, Markus Diesmann, and Gaute T Einevoll. Modeling the spatial reach of the lfp. *Neuron*, 72(5):859–872, 2011.

[25] Mircea Steriade, David A McCormick, and Terrence J Sejnowski. Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262(5134):679–685, 1993.

[26] J Allan Hobson and Edward F Pace-Schott. The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. *Nature Reviews Neuroscience*, 3(9):679–693, 2002.

[27] Lyle Muller, Giovanni Piantoni, Dominik Koller, Sydney S Cash, Eric Halgren, and Terrence J Sejnowski. Rotating waves during human sleep spindles organize global patterns of activity that repeat precisely through the night. *Elife*, 5:e17267, 2016.

[28] Susanne Diekelmann and Jan Born. The memory function of sleep. *Nature Reviews Neuroscience*, 11(2): 114–126, 2010.

[29] Giulio Tononi and Chiara Cirelli. Sleep function and synaptic homeostasis. *Sleep Medicine Reviews*, 10 (1):49–62, 2006.

[30] Vladyslav V Vyazovskiy and Kenneth D Harris. Sleep and the single neuron: the role of global slow oscillations in individual cell rest. *Nature Reviews Neuroscience*, 14(6):443–451, 2013.

[31] Eve Marder. Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1):1–11, 2012.

[32] Cornelia I Bargmann. Beyond the connectome: how neuromodulators shape neural circuits. *Bioessays*, 34 (6):458–465, 2012.

[33] Alain Destexhe, Michelle Rudolph, and Denis Paré. The high-conductance state of neocortical neurons in vivo. *Nature Reviews Neuroscience*, 4(9):739–751, 2003.

[34] Tamara Gedankien, Jennifer Kriegel, Erfan Zabeh, David McDonagh, Bradley Lega, and Joshua Jacobs. Cholinergic blockade reveals role for human hippocampal theta in encoding but not retrieval. *bioRxiv*, pages 2025–05, 2025.

[35] Bilal Haider, Alvaro Duque, Andrea R Hasenstaub, and David A McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, 26 (17):4535–4545, 2006.

[36] Michael Okun and Ilan Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, 11(5):535–537, 2008.

[37] Patrick L Purdon et al. Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proceedings of the National Academy of Sciences*, 110(12):E1142–E1151, 2013.

[38] Emery N Brown, Ralph Lydic, and Nicholas D Schiff. General anesthesia, sleep, and coma. *New England Journal of Medicine*, 363(27):2638–2650, 2010.

[39] Hadi Choubdar, Mahdi Mahdavi, Zahra Rostami, Erfan Zabeh, Martin J Gillies, Alexander L Green, Tipu Z Aziz, and Reza Lashgari. Neural oscillatory characteristics of feedback-associated activity in globus pallidus interna. *Scientific Reports*, 13(1):4141, 2023.

[40] Earl K Miller, Mikael Lundqvist, and André M Bastos. Working memory 2.0. *Neuron*, 100(2):463–475, 2018.

[41] Erfan Zabeh, Nicholas C Foley, Joshua Jacobs, and Jacqueline P Gottlieb. Beta traveling waves in monkey frontal and parietal areas encode recent reward history. *Nature Communications*, 14(1):5428, 2023.

[42] György Buzsáki and Andreas Draguhn. Neuronal oscillations in cortical networks. *Science*, 304(5679): 1926–1929, 2004. doi: 10.1126/science.1099745.

[43] Attila Losonczy, Boris V Zemelman, Alipasha Vaziri, and Jeffrey C Magee. Network mechanisms of theta related neuronal activity in hippocampal ca1 pyramidal neurons. *Nature neuroscience*, 13(8):967–972, 2010.

[44] Honghui Zhang, Andrew J Watrous, Ansh Patel, and Joshua Jacobs. Theta and alpha oscillations are traveling waves in the human neocortex. *Neuron*, 98(6):1269–1281, 2018.

[45] Bradley C Lega, Joshua Jacobs, and Michael Kahana. Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*, 22(4):748–761, 2012.

[46] Costas A Anastassiou, Rodrigo Perin, Henry Markram, and Christof Koch. Ephaptic coupling of cortical neurons. *Nature Neuroscience*, 14(2):217–223, 2011.

[47] Flavio Fröhlich and David A McCormick. Endogenous electric fields may guide neocortical network activity. *Neuron*, 67(1):129–143, 2010.

[48] Barry W Connors and Michael A Long. Electrical synapses in the mammalian brain. *Annual Review of Neuroscience*, 27:393–418, 2004.

[49] Michael VL Bennett and R Suzanne Zukin. Electrical coupling and neuronal synchronization in the mammalian brain. *Neuron*, 41.

[50] Sheriar G Hormuzdi, Michael A Filippov, Georgia Mitropoulou, Hannah Monyer, and Roberto Bruzzone. Electrical synapses: a dynamic signaling system that shapes the activity of neuronal networks. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1564(1):197–215, 2001.

[51] Biyu J He. The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66(3): 353–369, 2014.

[52] Kai J Miller, Lise B Sorensen, Jeffrey G Ojemann, and Marcel Den Nijs. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *NeuroImage*, 85:711–720, 2009.

[53] Thomas Donoghue, Matar Haller, Erik J Peterson, Paroma Varma, Priyadarshini Sebastian, Richard Gao, Torben Noto, Antonio H Lara, Joni D Wallis, Robert T Knight, et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nature neuroscience*, 23(12):1655–1665, 2020.

[54] Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. *On Pearl's Hierarchy and the Foundations of Causal Inference*, page 507–556. Association for Computing Machinery, New York, NY, USA, 1 edition, 2022. ISBN 9781450395861. URL `https://doi.org/10.1145/3501714.3501743`.

[55] John W Krakauer, Asif A Ghazanfar, Alex Gomez-Marin, Malcolm A MacIver, and David Poeppel. Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, 93(3):480–490, 2017.

[56] Omar G. Sani et al. Dissociative and prioritized modeling of behaviorally relevant neural activity from raw lfp. *Nature Neuroscience*, 2024. doi: 10.1038/s41593-024-01731-2.

[57] H.-L. Hsieh, B. Pesaran, and M. Shanechi. Multiscale modeling and decoding algorithms for spike- and field-based neural activity. *Journal of Neural Engineering*, 16(1):016018, 2019. doi: 10.1088/1741-2552/aaefc4.

[58] Yimin Huang and Marco Valtorta. Pearl's calculus of intervention is complete. *arXiv preprint arXiv:1206.6831*, 2012.

[59] Yixin Wang and David M Blei. The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528):1574–1596, 2019.

[60] Yixin Wang and David M Blei. Towards clarifying the theory of the deconfounder. *arXiv preprint arXiv:2003.04948*, 2020.

[61] Judea Pearl. Probabilistic reasoning in intelligent systems: networks of plausible inference. 1988.

[62] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

[63] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. *Proceedings of the National Conference on Artificial Intelligence*, 2:1219–1226, 2006.

[64] Lisa M Pfadenhauer, Ansgar Gerhardus, Kati Mozygemba, Kristin Bakke Lysdahl, Andrew Booth, Bjørn Hofmann, Philip Wahlster, Stephanie Polus, Jacob Burns, Louise Brereton, et al. Making sense of complexity in context and implementation: the context and implementation of complex interventions (cici) framework. *Implementation science*, 12:1–17, 2017.

[65] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

[66] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marí, et al. Inferring causation from time series in earth system sciences. *Nature Communications*, 10(1):2553, 2019.

[67] A. R. McIntosh. Nonlinear dynamics of neural systems. *Annual Review of Psychology*, 51:591–621, 2000.

[68] E. M. Izhikevich. Synchronization and phase resetting in networks of excitatory and inhibitory neurons. *IEEE Transactions on Neural Networks*, 18(3):617–627, 2007.

[69] P. J. Hyafil and colleagues. Cross-frequency coupling: a functional mechanism of neural oscillations. *Neuron*, 88:111–123, 2015.

[70] Anne Beuter, Leon Glass, Michael C Mackey, and Michele S Titcombe. Nonlinear dynamics in physiology and medicine. *Interdisciplinary Applied Mathematics*, 25, 2003.

[71] H. Markram and colleagues. Synaptic plasticity and spike-timing-dependent plasticity. *Biological Cybernetics*, 87:346–358, 2012.

[72] C. Glymour and colleagues. A review of causal discovery methods for time series analysis. *Journal of Machine Learning Research*, 18:1–29, 2019.

[73] L. Bressler and A. Seth. Evaluating causal relationships in neural systems: Granger causality, directed transfer function and other methods. *Biological Cybernetics*, 106:1–2, 2011.

[74] Anthony R McIntosh, Nada Kovacevic, and Roxane J Itier. Increased cerebro-cerebellar functional connectivity is associated with saccadic eye movements. *Cerebral Cortex*, 18(1):152–158, 2008. doi: 10.1093/cercor/bhm037.

[75] Michael Breakspear and Cornelis J Stam. Dynamics of a neural system with a multiscale architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1051–1074, 2005. doi: 10.1098/rstb.2005.1651.

[76] Ronan Perry, Julius Von Kügelgen, and Bernhard Schölkopf. Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. *Advances in Neural Information Processing Systems*, 35:10904–10917, 2022.

[77] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):947–1012, 2016.

[78] Amin Jaber, Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Causal discovery from soft interventions with unknown targets: Characterization and learning. *Advances in neural information processing systems*, 33:9551–9561, 2020.

[79] AmirEmad Ghassami, Saber Salehkaleybar, Negar Kiyavash, and Kun Zhang. Learning causal structures using regression invariance. *Advances in Neural Information Processing Systems*, 30, 2017.

[80] Joris M Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. *Journal of machine learning research*, 21(99):1–108, 2020.

[81] Biwei Huang, Kun Zhang, Jiji Zhang, Joseph Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous/nonstationary data. *Journal of Machine Learning Research*, 21(89):1–53, 2020.

[82] Adam Li, Amin Jaber, and Elias Bareinboim. Causal discovery from observational and interventional data across multiple environments. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 16942–16956. Curran Associates, Inc., 2023. URL `https://proceedings.neurips.cc/paper_files/paper/2023/file/368cba57d00902c752eaa9e4770bbbbe-Paper-Conference.pdf`.

[83] Kasra Jalaldoust, Saber Salehkaleybar, and Negar Kiyavash. Multi-domain causal discovery in bijective causal models. In *Fourth Conference on Causal Learning and Reasoning*, 2025. URL `https://openreview.net/forum?id=Li07fCvEhw`.

[84] James J Jun, Nicholas A Steinmetz, Joshua H Siegle, Daniel J Denman, Marius Bauza, Brian Barbarits, Albert K Lee, Costas A Anastassiou, Alexandru Andrei, Çağatay Aydın, et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236, 2017.

[85] Taesung Jung, Nanyu Zeng, Jason D Fabbri, Guy Eichler, Zhe Li, Erfan Zabeh, Anup Das, Konstantin Willeke, Katie E Wingel, Agrita Dubey, et al. Stable, chronic in-vivo recordings from a fully wireless subdural-contained 65,536-electrode brain-computer interface device. *bioRxiv*, pages 2024–05, 2025.

A Background

This section provides the necessary background on causal inference concepts that underpin our formal analysis of spike-wave duality. We introduce Structural Causal Models (SCMs), causal graphs, interventional distributions, and do-calculus as the foundational tools for reasoning about causality in neuroscientific contexts.

Notation: Throughout this paper, we use bold capital letters (e.g., $\mathbf{W}, \mathbf{S}, \mathbf{B}$) to denote sets of variables, regular capital letters (e.g., W, S, B) for individual variables, and lowercase letters (e.g., w, s, b) for specific values of variables. For notational simplicity in the main text, we write $do(\mathbf{W})$ as shorthand for $do(\mathbf{W} \leftarrow \mathbf{w})$ when the specific intervention values are clear from context.

A.1 Structural Causal Models

A Structural Causal Model (SCM) provides a mathematical framework for representing causal relationships between variables [1]. An SCM is defined as a tuple $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$ where:

- $\mathbf{U}$ is a set of exogenous (unobserved) variables that are determined by factors outside the model
- $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$ is a set of endogenous (observed) variables that are determined by variables in the model
- $\mathbf{F} = \{f_1, f_2, \ldots, f_n\}$ is a set of functions, where each f_i determines the value of V_i in terms of its parents: $V_i \leftarrow f_i(PA_i, U_i)$, where $PA_i \subset \mathbf{V} \setminus \{V_i\}$ are the parents of V_i and $U_i \subset \mathbf{U}$
- $P(\mathbf{U})$ is a probability distribution over the exogenous variables

The structural equations encode the causal mechanisms of the system. For instance, in the context of neural activity, we might have:

$$W \leftarrow f_W(\mathbf{S}_{\text{upstream}}, U_W) \tag{6}$$

$$S \leftarrow f_S(W, \mathbf{S}_{\text{other}}, U_S) \tag{7}$$

$$B \leftarrow f_B(\mathbf{S}, W, U_B) \tag{8}$$

where W represents a wave variable, S represents a spike variable, B represents behavior, and U_W, U_S, U_B are exogenous noise terms.

A.2 Causal Graphs and d-separation

Every SCM induces a directed acyclic graph (DAG) $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ where vertices correspond to endogenous variables and directed edges represent direct causal relationships. An edge $V_i \to V_j$ exists if and only if $V_i \in PA_j$.

The concept of d-separation (directional separation) provides a graphical criterion for determining conditional independence relationships encoded by the causal graph [61].

Definition 2 (d-separation)**.** *A path p between nodes X and Y in a DAG $\mathcal{G}$ is blocked by a set of nodes $\mathbf{Z}$ if and only if:*

1. *p contains a chain $i \to m \to j$ or a fork $i \leftarrow m \to j$ such that the middle node m is in $\mathbf{Z}$, or*
2. *p contains a collider $i \to m \leftarrow j$ such that the collision node m is not in $\mathbf{Z}$ and no descendant of m is in $\mathbf{Z}$.*

If all paths between X and Y are blocked by $\mathbf{Z}$, then X and Y are d-separated by $\mathbf{Z}$, denoted $X \perp\!\!\!\perp_d Y \mid \mathbf{Z}$.

The fundamental connection between d-separation and probabilistic independence is:

Theorem 2 (Global Markov Property)**.** *If $\mathcal{G}$ is the causal graph induced by SCM $\mathcal{M}$, then for any disjoint sets of variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$:*

$$\mathbf{X} \perp\!\!\!\perp_d \mathbf{Y} \mid \mathbf{Z} \text{ in } \mathcal{G} \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \text{ in } P \tag{9}$$

where P is the probability distribution induced by $\mathcal{M}$.

A.3 Interventions and the $do(x)$ operator

The key innovation of the causal inference framework is the formal treatment of interventions. An intervention on a variable X involves setting its value externally, overriding the natural causal mechanism that would normally determine X.

Definition 3 (Intervention and do-operator)**.** *An intervention $do(X \leftarrow x)$ on variable X in SCM $\mathcal{M}$ produces a new model $\mathcal{M}_x$ where:*

- *The structural equation for X is replaced by $X \leftarrow x$ (a constant)*
- *All other structural equations remain unchanged*
- *The corresponding graph $\mathcal{G}_x$ is obtained by removing all incoming edges to X*

The interventional distribution $P(Y; do(X \leftarrow x))$ is the distribution of Y under this modified model.

For our application, $P(\mathbf{B}; do(\mathbf{W} \leftarrow \mathbf{w}))$ represents the distribution of behavior when waves are experimentally set to specific values $\mathbf{w}$, breaking any natural dependencies that would normally determine wave activity.

The fundamental distinction is that generally:

$$P(Y \mid X = x) \neq P(Y; do(X \leftarrow x)) \tag{10}$$

The left side represents seeing $X = x$ (passive observation), while the right side represents making $X \leftarrow x$ (active intervention).

A.4 do-calculus

do-calculus provides three rules for deriving interventional distributions from observational data, given a causal graph [62, 63]:

Rule 1 (Insertion/deletion of observations):

$$P(\mathbf{Y}; do(\mathbf{X} \leftarrow \mathbf{x}), \mathbf{Z} = \mathbf{z}, \mathbf{W} = \mathbf{w}) = P(\mathbf{Y}; do(\mathbf{X} \leftarrow \mathbf{x}), \mathbf{W} = \mathbf{w}) \tag{11}$$

if $(\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$

Rule 2 (Action/observation exchange):

$$P(\mathbf{Y}; do(\mathbf{X} \leftarrow \mathbf{x}), do(\mathbf{Z} \leftarrow \mathbf{z}), \mathbf{W} = \mathbf{w}) = P(\mathbf{Y}; do(\mathbf{X} \leftarrow \mathbf{x}), \mathbf{Z} = \mathbf{z}, \mathbf{W} = \mathbf{w}) \tag{12}$$

if $(\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}},\underline{\mathbf{Z}}}}$

Rule 3 (Insertion/deletion of actions):

$$P(\mathbf{Y}; do(\mathbf{X} \leftarrow \mathbf{x}), do(\mathbf{Z} \leftarrow \mathbf{z}), \mathbf{W} \leftarrow \mathbf{w}) = P(\mathbf{Y}; do(\mathbf{X} \leftarrow \mathbf{x}), \mathbf{W} = \mathbf{w}) \tag{13}$$

if $(\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}},\overline{\mathbf{Z(W)}}}}$

where $\mathcal{G}_{\overline{\mathbf{X}}}$ removes all incoming edges to $\mathbf{X}$, $\mathcal{G}_{\underline{\mathbf{Z}}}$ removes all outgoing edges from $\mathbf{Z}$, and $\mathbf{Z(W)}$ are nodes in $\mathbf{Z}$ that are not ancestors of any node in $\mathbf{W}$.

These rules are complete: if a causal query can be computed from observational data given a graph, do-calculus will find the derivation.

A.5 The Causal Hierarchy Theorem

The Causal Hierarchy Theorem establishes fundamental limitations on what can be learned from observational data alone [54].

Theorem 3 (Causal Hierarchy Theorem)**.** *Causal inference problems form a hierarchy:*

1. ***Association*** *($P(Y \mid X = x)$): Statistical dependence observable in passive data*
2. ***Intervention*** *($P(Y; do(X \leftarrow x))$): Effects of deliberate actions*
3. ***Counterfactuals*** *($P(Y_x \mid X = x', Y = y')$): Retrospective reasoning about alternatives*

For almost all SCMs, even with infinite samples from quantities at the level i, one can not be uniquely determine the quantities at level $j > i$.

This theorem directly applies to the spike-wave debate: no amount of observational correlation data can definitively establish whether waves are epiphenomenal without additional causal assumptions or interventional experiments.

B Extended Discussion

B.1 Theoretical Implications

Our causal framework for spike-wave duality represents a methodological shift from correlation-based analysis to principled causal inference. Rather than taking sides in the longstanding debate about whether neural waves are causally relevant or epiphenomenal, we provide formal tools to adjudicate between competing hypotheses. The importance of explicitly stating these competing hypotheses cannot be overstated—only by formalizing different causal structures can we design experiments capable of distinguishing between them. This approach moves neuroscience away from premature commitment to either position without proper causal analysis, toward a more rigorous framework for understanding multi-scale neural phenomena.

B.2 Challenges in Causal Discovery of Spike-Wave Interactions

While our theoretical framework provides a principled foundation, practical implementation in neural data analysis faces several fundamental challenges that must be addressed for successful application.

B.2.1 Unavoidable Hidden Nodes

The contribution of individual neurons to behavior is often limited, and proper causal graph representation would require millions of neurons as nodes. In practice, neural data contains numerous confounding variables that cannot be directly measured or controlled. These include neurons falling outside the recording window, neurons removed due to noise issues, and broader network effects that influence observed signals. Such unobserved confounders can significantly influence the observed neural signals, making it challenging to establish direct causal relationships [1].

Potential solutions include studying causal effects in more controlled environments where confounding factors can be better identified, such as organoids or sliced brain tissue preparations. For in-vivo datasets, methods that explicitly account for existing confounding factors, such as those developed by Gerhardus et al. [64], would be beneficial. However, the fundamental challenge of hidden confounders remains a significant limitation for causal discovery in complex neural systems.

B.2.2 Cyclic Causal Graphs: Recursive Spike-Wave Interactions

Biophysical studies and expert knowledge about spike-field potential interactions at the synaptic level indicate that these interactions are inherently recursive. Spikes influence local field potentials, which in turn can influence subsequent spiking activity through various mechanisms including ephaptic coupling and network-level feedback. This creates cyclic causal relationships that are mathematically much more challenging to analyze than acyclic graphs [65].

Several approaches can address this challenge. Existing causal discovery methods like PCMCI (Peter and Clark Momentary Conditional Independence) are specifically designed to handle cyclic graphs [66]. Alternatively, one can decouple spiking activity from peripheral field potentials by isolating specific components of neural activity for independent analysis, though this approach may sacrifice some biological realism for analytical tractability.

B.2.3 Mixed Linear and Non-Linear Interactions

Neural signals exhibit both linear and non-linear interactions, which are critical for understanding brain function [67, 20]. In the context of spike-wave interactions, the dynamics may be particularly complex: brain wave interactions across regions might be modeled as relatively straightforward linear interactions, whereas spike interactions across regions are often inherently non-linear [68, 69].

The theoretical relationship between presynaptic potentials and spiking activity is fundamentally non-linear due to the sigmoid nature of neuronal activation functions [70, 71]. This complexity poses significant challenges for causal discovery methods. Many traditional algorithms, such as Granger causality, are designed for linear systems and fail to capture the full complexity of neural dynamics [72, 13, 73]. Future applications of our framework will require causal discovery methods that can accommodate both linear and non-linear dynamics simultaneously.

B.2.4 High Autocorrelation of Neural Signals

Autocorrelation—the correlation of a signal with delayed copies of itself—is ubiquitous in neural signals due to intrinsic properties of neuronal firing patterns and synaptic connections [74]. This temporal dependence violates the independence assumptions underlying many causal discovery algorithms, potentially leading to spurious causal inferences.

Addressing this challenge requires time-series analysis methods that explicitly account for autocorrelation. Approaches include autoregressive integrated moving average (ARIMA) models and other methods designed for handling serial dependencies in data. However, incorporating these methods into causal discovery frameworks while maintaining the ability to distinguish genuine causal relationships from autocorrelation-induced spurious associations remains an active area of research.

B.2.5 Violation of Stationarity Assumptions Across Cognitive States

Stationarity assumes that the statistical properties of a signal remain constant over time. Neural signals frequently violate this assumption across different cognitive states, with properties such as mean, variance, and spectral characteristics changing substantially [75]. This non-stationarity can confound causal discovery by making it difficult to distinguish genuine causal relationships from state-dependent correlations.

Simple solutions to uncontrolled nonstationarity often include restricting analysis to intervals with similar cognitive states, though this approach may limit the generalizability of findings. More sophisticated approaches involve non-stationary signal processing techniques such as adaptive filtering or time-varying auto-regressive models. The heterogeneity can in fact aid causal discovery as well; any perturbation to the causal mechanisms reveals information about the structure beyond what is identifiable from mere observations, even if the perturbations are imperfect and due to the nature or unknown sources. By breaking the observational symmetry between the cause and the effect, some existing methods can recover the causal graph beyond what is achievable with perfectly stationary data [76, 77, 78, 79, 80, 81, 82, 83, 77], but our current formulation does not capture such nuanced cases involving multi-domain data.

B.2.6 Spatial Sampling Scale

The historical skepticism toward oscillations stems largely from the coarse spatial sampling of field potential recordings. Conventional electrodes aggregate signals from thousands of neurons, blurring fine-scale structure and making waves appear as diffuse population averages rather than mechanistically grounded dynamics. This resolution gap fostered the view that oscillations cannot exert precise computational influence on single neurons, casting them as epiphenomenal reflections of spiking rather than causal contributors to it.

The fabrication of high-density, high-resolution recording arrays now directly addresses one of the central limitations outlined in the introduction—the coarse spatial sampling that historically relegated oscillations to the status of low-resolution, and thus supposedly epiphenomenal, signals. Platforms such as the Neuropixels probes [84] and the BISC microelectrode arrays [85] allow simultaneous recording of spikes and field potentials across contiguous cortical regions with cellular precision. This convergence eliminates the scale mismatch that once separated the two descriptions, enabling oscillations to be studied as spatially structured, mechanistically grounded phenomena rather than population averages. As these technologies mature, they may reveal whether the apparent coarse structure of waves was a limitation of measurement rather than a property of the brain itself—transforming the spike–wave debate from one constrained by resolution to one grounded in causal interpretation.

B.3 Future Directions

Despite these challenges, our framework opens several promising avenues for future research. First, developing computational tools that implement our theoretical contributions for real neural data, with particular attention to the challenges outlined above. Second, designing controlled experiments that can test specific predictions about when waves are epiphenomenal versus causally sufficient. Third, extending the approach to other multi-scale phenomena in neuroscience beyond spike-wave duality. Finally, investigating how our causal abstraction principles might apply to other complex systems where multiple levels of description compete for explanatory primacy.

The ultimate goal is not to definitively resolve the spike-wave debate, but to provide neuroscience with principled tools for making such determinations based on rigorous causal analysis rather than correlational evidence alone.

C Proofs

C.1 Proof of Proposition 1

Proof. Consider a directed path $\pi : W \to U_1 \to U_2 \to \cdots \to U_n \to B$ where $W \in \mathbf{W}$, $B \in \mathbf{B}$, and $U_i \in \mathbf{U}$ for all $i \in \{1, \ldots, n\}$. For any variable V in a causal graph, we denote by PA_V the parent set of V, defined as the set of all variables that have a direct causal edge pointing into V. In the context of an SCM, these parents appear as arguments in the structural equation that determines V.

The structural equations along the path take the following form. For the first unobserved variable, we have $U_1 \leftarrow f_{U_1}(W, PA_{U_1} \setminus \{W\}, \epsilon_{U_1})$, where $PA_{U_1} \setminus \{W\}$ represents all parents of U_1 except W, and ϵ_{U_1} is an exogenous noise term drawn from the background distribution. For intermediate variables along the path, $U_i \leftarrow f_{U_i}(U_{i-1}, PA_{U_i} \setminus \{U_{i-1}\}, \epsilon_{U_i})$ for $i \in \{2, \ldots, n\}$. Finally, the behavior variable satisfies $B \leftarrow f_B(U_n, PA_B \setminus \{U_n\}, \epsilon_B)$, where again we explicitly separate the contribution from the path variable U_n from other potential parents.

When we perform the intervention $do(W \leftarrow w)$, we fundamentally alter the data-generating process. The structural equation for W is replaced entirely by the constant assignment $W \leftarrow w$, removing all edges into W in the causal graph. This is not mere conditioning—it is a surgical modification of the causal structure. Under this intervention, the distribution of U_1 becomes

$$P(U_1 = u_1; do(W \leftarrow w)) = \int P(\epsilon_{U_1} = e) \cdot 1[f_{U_1}(w, PA_{U_1} \setminus \{W\}, e) = u_1]\, de \tag{14}$$

where $1[\cdot]$ is the indicator function. This distribution differs from the natural distribution $P(U_1)$ whenever the partial derivative $\partial f_{U_1}/\partial W$ is non-zero, which holds for all non-degenerate SCMs where W actually influences U_1.

The change in U_1's distribution propagates forward through the path. Since U_2 depends on U_1 through its structural equation, and the intervention has altered $P(U_1)$, we obtain a modified distribution $P(U_2; do(W \leftarrow w))$ that differs from $P(U_2)$. This cascade continues: each U_i inherits the perturbation from its predecessor U_{i-1}, ultimately reaching B through its dependence on U_n.

Now we consider what happens when we condition on the observed variables $\mathbf{S}$. The critical insight is that conditioning cannot block a causal path unless we condition on at least one variable along that path. Since all intermediate variables $U_1, \ldots, U_n$ belong to the unobserved set $\mathbf{U}$, they are by definition not contained in $\mathbf{S}$. The causal influence from W to B flows through these unobserved intermediaries unimpeded.

To formalize this mathematically, we decompose the conditional distribution using the law of total probability:

$$P(B = b \mid \mathbf{S} = \mathbf{s}; do(W \leftarrow w)) = \int P(B = b \mid \mathbf{S} = \mathbf{s}, U_n = u) \cdot P(U_n = u \mid \mathbf{S} = \mathbf{s}; do(W \leftarrow w))\, du \tag{15}$$

The first term $P(B = b \mid \mathbf{S} = \mathbf{s}, U_n = u)$ is determined by the structural equation for B and does not depend on whether W was intervened upon or merely observed. However, the second term $P(U_n = u \mid \mathbf{S} = \mathbf{s}; do(W \leftarrow w))$ explicitly depends on the intervention. Since the path from

W to U_n passes only through unobserved variables, this distribution differs from the observational $P(U_n = u \mid \mathbf{S} = \mathbf{s})$.

Specifically, in the observational case, U_n 's distribution reflects the natural variation in W and all intermediate variables. Under intervention, W is held fixed at w, and this constraint propagates through to alter U_n's distribution. For any SCM where the composed function from W to U_n has non-zero derivative—that is, where the causal effect actually exists—these two distributions are distinct. Therefore, $P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W})) \neq P(\mathbf{B} \mid \mathbf{S})$. □

C.2 Proof of Proposition 2

Proof. The front-door graph encodes a specific pattern of causal relationships: waves influence behavior exclusively through observed mediators $\mathbf{M} \subseteq \mathbf{S}$, while unobserved confounders $\mathbf{C} \subseteq \mathbf{U}$ create a backdoor path between waves and behavior. Formally, the graph contains edges $\mathbf{W} \to \mathbf{M}$, $\mathbf{M} \to \mathbf{B}$, $\mathbf{C} \to \mathbf{W}$, and $\mathbf{C} \to \mathbf{B}$, with no direct edge from $\mathbf{W}$ to $\mathbf{B}$.

We seek to compare the interventional distribution $P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W}))$ with the observational distribution $P(\mathbf{B} \mid \mathbf{S})$. For clarity, we consider the pure front-door case where $\mathbf{S} = \mathbf{M}$. We begin by deriving the interventional distribution through a systematic application of do-calculus.

Starting with $P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W}))$, we systematically apply the rules of do-calculus to derive an expression in terms of observational quantities. We first apply Rule 2, which allows us to exchange action and observation when appropriate d-separation conditions hold. Since $(\mathbf{S} \perp \mathbf{B} \mid \mathbf{W})$ in $\mathcal{G}_{\overline{\mathbf{WS}}}$ (the graph with incoming edges to both $\mathbf{W}$ and $\mathbf{S}$ removed), we obtain:

$$P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W})) = P(\mathbf{B}; do(\mathbf{W}), do(\mathbf{S})) \tag{16}$$

Applying Rule 3 to remove the intervention on $\mathbf{W}$, we use the fact that $(\mathbf{W} \perp \mathbf{B} \mid \mathbf{S})$ in $\mathcal{G}_{\overline{\mathbf{SW}}}$. Since all paths from $\mathbf{W}$ to $\mathbf{B}$ pass through $\mathbf{S} = \mathbf{M}$ after removing incoming edges, we have:

$$P(\mathbf{B}; do(\mathbf{W}), do(\mathbf{S})) = P(\mathbf{B}; do(\mathbf{S})) \tag{17}$$

We now express this marginal distribution by summing over all possible values of $\mathbf{W}$:

$$P(\mathbf{B}; do(\mathbf{S})) = \sum_{\mathbf{w}'} P(\mathbf{B}, \mathbf{W} = \mathbf{w}'; do(\mathbf{S})) \tag{18}$$

Applying the chain rule to decompose the joint distribution:

$$P(\mathbf{B}, \mathbf{W} = \mathbf{w}'; do(\mathbf{S})) = P(\mathbf{B} \mid \mathbf{W} = \mathbf{w}'; do(\mathbf{S})) \cdot P(\mathbf{W} = \mathbf{w}'; do(\mathbf{S})) \tag{19}$$

For the conditional term, we apply Rule 3. Since $(\mathbf{S} \perp \mathbf{W})$ in $\mathcal{G}_{\overline{\mathbf{S}}}$ (after removing incoming edges to $\mathbf{S}$, there are no common ancestors), we get:

$$P(\mathbf{B} \mid \mathbf{W} = \mathbf{w}'; do(\mathbf{S})) = P(\mathbf{B} \mid \mathbf{S}, \mathbf{W} = \mathbf{w}') \tag{20}$$

Similarly, for the marginal term:

$$P(\mathbf{W} = \mathbf{w}'; do(\mathbf{S})) = P(\mathbf{W} = \mathbf{w}') \tag{21}$$

Finally, applying Rule 2 to convert the intervention on $\mathbf{S}$ back to conditioning (using $(\mathbf{S} \perp \mathbf{B} \mid \mathbf{W})$ in $\mathcal{G}_{\overline{\mathbf{S}}}$), we arrive at:

$$P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W})) = \sum_{\mathbf{w}'} P(\mathbf{B} \mid \mathbf{S}, \mathbf{W} = \mathbf{w}') P(\mathbf{W} = \mathbf{w}') \tag{22}$$

The interventional distribution is a weighted average of the observational conditional $P(\mathbf{B} \mid \mathbf{S}, \mathbf{W} = \mathbf{w}')$ over all possible values $\mathbf{w}'$, weighted by the marginal distribution $P(\mathbf{W} = \mathbf{w}')$. For epiphenomenality, the following must hold,

$$P(\mathbf{B} \mid \mathbf{S}) \stackrel{?}{=} \sum_{\mathbf{w}'} P(\mathbf{B} \mid \mathbf{S}, \mathbf{W} = \mathbf{w}') P(\mathbf{W} = \mathbf{w}'). \tag{23}$$

The only scenarios where this equality could hold observationally are degenerate cases: either the confounding effects are exactly zero (removing the backdoor path entirely), or multiple effects cancel through a precise balance of parameters that has measure zero in the space of all SCMs. For generic SCMs where confounding effects are non-zero and non-canceling, we conclude that $P(\mathbf{B} \mid \mathbf{S}; do(\mathbf{W})) \neq P(\mathbf{B} = \mathbf{S})$. □

C.3 Proof of Theorem 1

Proof. We prove that variables $\mathbf{Z}$ are epiphenomenal to outcomes $\mathbf{Y}$ given predictors $\mathbf{X}$—formally, that $P(\mathbf{Y} \mid \mathbf{X}; do(\mathbf{Z})) = P(\mathbf{Y} \mid \mathbf{X})$—if and only if $\mathbf{Z} \perp\!\!\!\perp_d \mathbf{Y} \mid \mathbf{X}$ in the graph $\mathcal{G}_{\overline{\mathbf{Z}(\mathbf{X})}}$.

Before proceeding, we must carefully define our terms. Let $\mathbf{Z}(\mathbf{X})$ denote the subset of $\mathbf{Z}$ consisting of variables that are not ancestors of any variable in $\mathbf{X}$. Formally, $Z \in \mathbf{Z}(\mathbf{X})$ if and only if there exists no directed path from Z to any $X \in \mathbf{X}$ in the original causal graph $\mathcal{G}$. The modified graph $\mathcal{G}_{\overline{\mathbf{Z}(\mathbf{X})}}$ is constructed from $\mathcal{G}$ by removing all incoming edges to variables in $\mathbf{Z}(\mathbf{X})$. This construction represents the post-intervention graph for a specific subset of $\mathbf{Z}$—those that cannot influence the conditioning set $\mathbf{X}$.

We first establish the forward direction. Assume that $\mathbf{Z} \perp\!\!\!\perp_d \mathbf{Y} \mid \mathbf{X}$ holds in $\mathcal{G}_{\overline{\mathbf{Z}(\mathbf{X})}}$. This d-separation statement means that in the modified graph where we have removed incoming edges to non-ancestors of $\mathbf{X}$ in $\mathbf{Z}$, all paths from $\mathbf{Z}$ to $\mathbf{Y}$ are blocked by $\mathbf{X}$.

Rule 3 of do-calculus provides conditions for deleting interventions from conditional distributions. Specifically, it states that $P(\mathbf{Y} \mid \mathbf{X}; do(\mathbf{Z}), \mathbf{W}) = P(\mathbf{Y} \mid \mathbf{X}, \mathbf{W})$ if $\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z} \mid \mathbf{X}, \mathbf{W}$ in $\mathcal{G}_{\overline{\mathbf{X}}, \overline{\mathbf{Z}(\mathbf{W})}}$. In our case, with empty $\mathbf{W}$, the condition simplifies to requiring $\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z} \mid \mathbf{X}$ in $\mathcal{G}_{\overline{\mathbf{Z}(\mathbf{X})}}$, which is precisely our assumption. Therefore, Rule 3 allows us to conclude:

$$P(\mathbf{Y} \mid \mathbf{X}; do(\mathbf{Z})) = P(\mathbf{Y} \mid \mathbf{X}) \tag{24}$$

establishing that $\mathbf{Z}$ is epiphenomenal to $\mathbf{Y}$ given $\mathbf{X}$.

For the reverse direction, suppose that $P(\mathbf{Y} \mid \mathbf{X}; do(\mathbf{Z})) = P(\mathbf{Y} \mid \mathbf{X})$ holds for all SCMs compatible with causal graph $\mathcal{G}$. This universal validity is crucial—the equality must hold not just for some particular SCM, but for every SCM that could generate the graph structure $\mathcal{G}$.

The completeness theorem for do-calculus, established by Huang and Valtorta (2006) and independently by Shpitser and Pearl (2006), states that if a causal effect is identifiable from a graph, then it can be computed using the three rules of do-calculus. More precisely, if an equality between interventional distributions holds for all SCMs compatible with a graph, then this equality must be derivable using do-calculus rules.

In our case, the equality $P(\mathbf{Y} \mid \mathbf{X}; do(\mathbf{Z})) = P(\mathbf{Y} \mid \mathbf{X})$ represents the deletion of an intervention operator. Among the three rules of do-calculus, only Rule 3 permits such deletion. Rules 1 and 2 deal with insertion/deletion of observations and action/observation exchange, respectively, but cannot eliminate a $do(\cdot)$ operator entirely from a distribution.

Since the equality must be derivable via do-calculus, and only Rule 3 can derive it, the graphical condition required by Rule 3 must hold. This condition is precisely $\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z} \mid \mathbf{X}$ in $\mathcal{G}_{\overline{\mathbf{Z}(\mathbf{X})}}$, completing our proof of the reverse direction.

The theorem provides a complete graphical characterization of epiphenomenality. Variables $\mathbf{Z}$ are epiphenomenal to $\mathbf{Y}$ given $\mathbf{X}$ when their influence on $\mathbf{Y}$ is entirely mediated through $\mathbf{X}$, or when their association with $\mathbf{Y}$ arises solely from common causes that are blocked by conditioning on $\mathbf{X}$. The modification to consider $\mathcal{G}_{\overline{\mathbf{Z}(\mathbf{X})}}$ rather than the original graph $\mathcal{G}$ accounts for the subtlety that we only consider intervening on components of $\mathbf{Z}$ that do not affect our conditioning set $\mathbf{X}$. This distinction is essential for properly capturing the notion of conditional causal irrelevance. □