

Fast Möbius transform: An algebraic approach to information decomposition

Abel Jansma^{1,2,3,*}, Pedro A. M. Mediano^{4,5,†} and Fernando E. Rosas^{6,7,8,9,‡}¹Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany²School of Informatics, University of Edinburgh, Edinburgh, United Kingdom³Dutch Institute for Emergent Phenomena, University of Amsterdam, Amsterdam, the Netherlands⁴Department of Computing, Imperial College London, London, United Kingdom⁵Division of Psychology and Language Sciences, University College London, London, United Kingdom⁶Sussex AI and Sussex Centre for Consciousness Science, Department of Informatics, University of Sussex, Brighton, United Kingdom⁷Principles of Intelligent Behavior in Biological and Social Systems (PIBBS), Prague, Czech Republic⁸Center for Psychedelic Research and Centre for Complexity Science, Department of Brain Science, Imperial College London, London, United Kingdom⁹Center for Eudaimonia and Human Flourishing, University of Oxford, Oxford, United Kingdom

(Received 5 February 2025; accepted 16 June 2025; published 11 July 2025)

The partial information decomposition (PID) and its extension integrated information decomposition (Φ ID) are promising frameworks to investigate information phenomena involving multiple variables. An important limitation of these approaches is the high computational cost involved in their calculation. Here we leverage fundamental algebraic properties of these decompositions to enable a computationally-efficient method to estimate them, which we call the *fast Möbius transform*. Our approach is based on a formula for estimating the Möbius function that circumvents important computational bottlenecks and can in some cases offer a double-exponential speedup. We showcase the capabilities of this approach by presenting two analyses that would be unfeasible without this method: decomposing the information that neural activity at different frequency bands yields about the brain's macroscopic functional organization and identifying distinctive dynamical properties of the interactions between multiple voices in baroque music. Overall, our proposed approach illuminates the value of algebraic facets of information decomposition and opens the way to a wide range of future analyses.

DOI: [10.1103/32sq-z2jf](https://doi.org/10.1103/32sq-z2jf)

I. INTRODUCTION

A. Information decomposition

Understanding how information is distributed and processed across complex systems is a fundamental challenge in various scientific domains, from neuroscience and genetics to machine learning and artificial intelligence [1–3]. The partial information decomposition (PID) framework and its extension integrated information decomposition (Φ ID) have emerged as powerful tools to address this question by decomposing the information provided by different variables into synergistic, redundant, and unique components [4–6]. These frameworks have enabled a range of analyses revealing significant insights into, for example, the inner workings of natural evolution [7], genetic information flow [8], psychometric interactions [9,10], cellular automata [11,12], and polyphonic music [13].

Information decomposition has been particularly useful for studying biological [14–20] and artificial [21–24] neural systems, shedding light on the distinct roles of redundancy and synergy in neural computations [25].

An important limitation of PID and Φ ID is the high computational cost involved in calculating the decomposition when increasing the number of involved variables. One of the most costly steps is the calculation of the information atoms by subtracting values over the so-called redundancy lattice. This operation relies on the Möbius inversion formula [4], which provides a notion of discrete derivative over partially ordered sets [26]. Implementing the Möbius inversion requires knowledge of the Möbius function, but the calculation of this function is, in general, highly nontrivial. In fact, the Möbius function on the redundancy lattice is currently unknown. Because of this, PID atoms are currently calculated by an alternative method that requires solving a system of equations, but the construction of this system quickly becomes unfeasible as the number of variables grows. More broadly, this knowledge gap prevents the information decomposition literature from accessing the rich mathematical toolbox associated with the calculus over lattices.

In this paper, we present an approach that exploits the rich algebraic structure of the redundancy lattice, which leads to a method we call the fast Möbius transform. The core of this method is an efficient way to compute the Möbius function of the redundancy lattice, which relies on algebraic

*Contact author: a.a.a.jansma@uva.nl†Contact author: p.mediano@imperial.ac.uk‡Contact author: f.rosas@sussex.ac.uk

properties of the lattice and the Birkhoff representation theorem [27]. This result allows us to compute information atoms without constructing the underlying lattice or solving a system of equations, enabling the calculation of the PID and Φ ID in larger systems than previously possible. We showcase the capabilities of this approach by presenting two analyses that would have been impossible without it: the decomposition of the information that the five frequency bands of the brain's electromagnetic activity provide about the brain's macroscopic gradient of functional organization and the disentangling of distinctive information dynamics phenomena observed in music scores of J.S. Bach and A. Corelli.

B. Theoretical background

We start by outlining the PID and Φ ID. Throughout the manuscript, we denote partially ordered sets as calligraphic capital letters ($\mathcal{A}$) and antichains of partially ordered sets with bold greek letters (α).

PID [4] builds on the intuition that the information that a set of predictor variables $\mathbf{X} = (X_1, \dots, X_n)$ contains about a target variable Y (typically quantified by Shannon's mutual information $I(\mathbf{X}; Y)$) can be of different types and hence can be decomposed into various subcomponents. For example, for the case of two predictor variables $\mathbf{X} = (X_1, X_2)$, PID states that the information $I(\mathbf{X}; Y)$ can be decomposed as follows:

$$I(\mathbf{X}; Y) = I_{\theta}^{(1)(2)}(\mathbf{X}; Y) + I_{\theta}^{(1)}(\mathbf{X}; Y) + I_{\theta}^{(2)}(\mathbf{X}; Y) + I_{\theta}^{(12)}(\mathbf{X}; Y). \quad (1)$$

The terms on the right-hand side are referred to as information and can be interpreted as follows: $I_{\theta}^{(i)}(\mathbf{X}; Y)$ is interpreted as the unique information carried by X_i , $I_{\theta}^{(1)(2)}(\mathbf{X}; Y)$ corresponds to the redundant information that is carried by both X_1 and X_2 , and $I_{\theta}^{(12)}(\mathbf{X}; Y)$ is the synergistic information carried only by the joint state of both variables together, but not by either of them in isolation.

In general, when decomposing information about a target variable Y given n predictor variables $\mathbf{X}$, the PID decomposes the information into a superexponential number of terms that form a partial order known as the redundancy lattice $\mathcal{R}_n$. The full decomposition is outlined in Methods, Sec. IV A.

Φ ID is an extension of PID that allows one to decompose the information that a set of source variables $\mathbf{X} = (X_1, \dots, X_n)$ provides about a set of target variables $\mathbf{Y} = (Y_1, \dots, Y_m)$ [6]. While it can be applied to decompose the mutual information between any two sets of random variables, it has found natural applications in the analysis of multivariate time series data [15,28,29]. The Φ ID formalism (outlined in Methods, Sec. IV A) can be built following similar steps as the PID, and essentially corresponds to the Cartesian product of two complementary PIDs, one decomposing the information carried by the sources about the targets and the other decomposing the information carried by the targets about the sources.

C. Outline of the paper

The remainder of this paper is organized as follows. Our main results are presented in Sec. II, starting with establishing a relationship between the lattices of redundancies and logical

propositions (Sec. II A). We then present the fast Möbius transform in Sec. II B, which provides a computationally efficient method to calculate partial and integrated information decompositions. These calculations are verified for some example distributions in Sec. II C, after which we present two case studies: information decomposition of brain frequency bands (Sec. II D) and information dynamics in baroque music (Sec. II E). We conclude with a discussion of our results and future work in Sec. III. As our results rely on established theory of information decompositions and algebraic structures, we present the necessary background in Methods, Sec. IV A. Proofs are included in Appendix B.

II. RESULTS

A. A free distributive lattice of redundancies and synergies

The identification of the PID antichains A_n and the construction of their partial ordering via Eq. (23)] is highly nontrivial, and this only becomes worse for Φ ID. Here we present an alternative construction which uses algebraic tools to deliver the same redundancy lattice in a more intuitive manner. Our approach is to build an algebra of redundancies and synergies, extending results reported in Refs. [30,31].

We start from a set of n primitive elements denoted by $(1), (2), \dots, (n)$, corresponding to n sources of information, and two binary operations for building meets $\wedge$ and joins $\vee$. Of these, the join operation ($\vee$) can be thought of as denoting the union between sources and hence we interpret it as synergy, while the meet operation ($\wedge$) can be conceived as the intersection between them and hence we interpret as redundancy. For example, the atom associated with the antichain $(12)(3)$ can be thought of as obtained from the primitives $(1),(2)$ and (3) via the algebraic operations $(1 \vee 2) \wedge (3)$, where $(1 \vee 2)$ relates synergy with the union of (1) and (2) , and $(\cdot) \wedge (3)$ relates redundancy with the intersection of $(\cdot)$ and (3) .

The successive application of these two operations over the set of primitive elements give rise to an infinite number of formulas, which can be seen as a formal language [32]. Then, the second step is to impose properties over the operations to reduce the number of distinct expressions in this language. Specifically, we require the algebraic operations of synergy and redundancy to behave as unions and intersections do — i.e., to be commutative, associative, idempotent, and obey the absorption laws (see Appendix A 2). Additionally, we require redundancy and synergy to distribute over each other as

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c), \quad (2)$$

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c). \quad (3)$$

The set of the resulting distinct expressions is denoted by D_n . Note that the requirement that synergy and redundancy be commutative, associative, and idempotent is compatible with the original intuitions in partial information decomposition. One would expect, for instance, that $(1)(2)$, $(2)(1)$, and $(1)(1)(2)$ refer to the same collection of sources. The absorption laws also have a direct interpretation; for example, $a \wedge (a \vee b) = a$ states that the information provided redundantly by (a) and (ab) is equal to the information given by (a) .

It can be shown that D_n with the ordering implied by the meets and joins forms a lattice $\mathcal{F}_n = (D_n, \wedge, \vee)$ (see Appendix A 2). This type of lattice is known as the free distributed lattice over n generators, as it is freely generated by primitive elements and its meets and joins distribute. Our next result shows that D_n corresponds to the set of antichains A_n , and the lattice $\mathcal{F}_n$ is the same as the redundancy lattice $\mathcal{R}_n$. This result is a natural extension of the findings reported in Ref. [[30], Sec. 5] related to the properties of the union and intersection information, and the results presented in Ref. [[31], Sec. 4] about the relationship between PID antichains and logical propositions.

Proposition 1. Let $\mathcal{R}_n$ be the redundancy lattice on the set of antichains A_n ordered via $\leq$ as in (23), and let $\mathcal{F}_n = (D_n, \wedge, \vee)$ be the free distributive lattice over n elements, ordered as $a \preceq b \iff a \wedge b = a$. Then $A_n = D_n$ and $\mathcal{R}_n = \mathcal{F}_n$.

Proof. See Appendix B 1. ■

Building the redundancy lattice $\mathcal{R}_n$ as the free distributive lattice generated by the set of n source variables can provide conceptual clarity. Indeed, this reveals the redundancy lattice to be a general algebraic construction rather than dependent on a particular interpretation of redundant information, complementing the results reported in Ref. [31]. Additionally, understanding PID as arising from a distributive lattice explains why PID atoms can be visualized as intersections among circles in a Venn diagram—while naive decomposition of entropy cannot. Indeed, this is a direct consequence of the fact that (finite) distributive lattices are isomorphic to lattices built from set intersections and unions, as shown by Birkhoff's representation theorem (see Appendix B 2). See also related discussions presented in Ref. [30].

Naturally, the same algebraic construction can be used to build Φ ID's double redundancy lattice. For this, one needs to (i) construct the normal redundancy lattice $\mathcal{R}_n$, (ii) construct the Cartesian product of nodes, and, finally, (iii) identify the resulting partial ordering by evaluating it coordinatewise. It is worth noting that the product lattice is not in itself a free distributed lattice but just a Cartesian product of two lattices that are. This, nonetheless, is enough to enable an efficient method to construct its Möbius function, as shown in the next subsection.

B. The fast Möbius transform

The computation of PID atoms relies on the Möbius function as a direct application of the Möbius inversion theorem (see Sec. IV B) to the redundancy function $I_\cap^\alpha(X; Y)$ over the redundancy lattice $\mathcal{R}_n$ shows that

$$I_\cap^\alpha(X; Y) = \sum_{\beta: \beta \leq \alpha} I_\cap^\beta(X; Y) \quad (4)$$

$\Downarrow$

$$I_\cap^\beta(X; Y) = \sum_{\alpha: \alpha \leq \beta} \mu_{\mathcal{R}_n}(\alpha, \beta) I_\cap^\alpha(X; Y), \quad (5)$$

where $\mu_{\mathcal{R}_n}$ is the Möbius function on $\mathcal{R}_n$. Below we leverage tools of order theory to enable an efficient method to compute the information decomposition atoms [Eq. (5)], which we call the fast Möbius transform.

Before presenting the main result behind the fast Möbius transform, let us introduce some key ideas. Let us define the complement of an antichain $\alpha = \{a_1, \dots, a_l\} \in A_n$ via the mapping $*$: $\alpha \mapsto \alpha^* = \{\{1, \dots, n\} \setminus a_j \mid a_j \in \alpha\}$, where $A \setminus B$ is the set difference between A and B . For example, if $n = 3$ and $\alpha = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$, then $\alpha^* = \{\{1\}, \{2\}, \{3\}\}$. Let us also introduce the notion of order ideal generated by the antichain α , defined by

$$I_\alpha = \{b \subseteq \{1, \dots, n\} \mid \exists a \in \alpha \text{ such that } b \leq a\}, \quad (6)$$

corresponding to the set of all subsets of elements of α . For example, if $\alpha = \{\{1, 2\}, \{3\}\}$, then $I_\alpha = \{\{1, 2\}, \{1\}, \{2\}, \{3\}, \emptyset\}$ (see Appendixes A and B 2 for more background details).

With these notions at hand, we can now present our next result.

Theorem 1 (Fast Möbius transform for PID). Let P_N be the powerset of the set of indices $\{1, \dots, n\}$, and $\mathcal{B}_n = (P_N, \subseteq)$ the powerset lattice ordered by inclusion. Then, the Möbius function on the redundancy lattice $\mathcal{R}_n$ can be calculated as

$$\mu_{\mathcal{R}_n}(\alpha, \beta) = \begin{cases} (-1)^{|I_{\alpha^*} \setminus I_{\beta^*}|} & \text{if } I_{\alpha^*} \setminus I_{\beta^*} \text{ is an antichain of } \mathcal{B}_n \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Above, I_{α^*} is the ideal in $\mathcal{B}_n$ generated by the complement of α , and $|I|$ is the cardinality of I .

We highlight that being an antichain of $\mathcal{B}_n$ is equivalent to being a PID atom (see Appendix B 2).

Sketch of proof. The proof uses Birkhoff's representation theorem [27], which states that any distributive lattice can be constructed by taking an adequate partially ordered set $\mathcal{T}$ and building the lattice $J(\mathcal{T})$ of all the ideals of $\mathcal{T}$ ordered by set inclusion. This construction is then combined with two facts: (i) that $J(\mathcal{B}_n)$ is isomorphic to the free distributed lattice generated by n primitives and (ii) that the Möbius function on a lattice $J(\mathcal{T})$ can be calculated as [33]

$$\mu_{J(\mathcal{T})}(A, B) = \begin{cases} (-1)^{|B \setminus A|} & \text{if } B \setminus A \text{ is an antichain of } \mathcal{T} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

To use this formula, all that is left is to explicitly construct an isomorphism that maps elements of $\mathcal{R}_n$ to the free distributed lattice generated by n elements, which is done via the map $*$. The full proof is presented in Appendix B 2. ■

Corollary 1 (Fast Möbius transform for Φ ID). The Möbius function on the double redundancy lattice $\mathcal{R}_n \times \mathcal{R}_n$ can be expressed as

$$\mu_{\mathcal{R}_n \times \mathcal{R}_n}(\alpha \rightarrow \beta, \alpha' \rightarrow \beta') = \mu_{\mathcal{R}_n}(\alpha, \beta) \mu_{\mathcal{R}_n}(\alpha', \beta'). \quad (9)$$

Proof. The result follows from the fact that the Möbius function on the product lattice with the Cartesian order is the product of the Möbius function on each of the lattices (see Ref. [[33], Prop. 3.8.2]). ■

This result significantly reduces the computational cost of computing information decomposition atoms. Currently, information atoms are typically calculated by inverting a system of equations [e.g., Eq. (26) for the case of $n = 2$ predictors], which is solved by starting at the lowest redundancy term and

recursively solving for higher terms using

$$I_{\theta}^{\alpha}(X; Y) = I_{\cap}^{\alpha}(X; Y) - \sum_{\beta < \alpha} I_{\theta}^{\beta}(X; Y). \quad (10)$$

This amounts to solving on the order of $|\mathcal{R}_n| = |D_n|$ equations, where $|D_n|$ is the n th Dedekind number. For example, the well-known Python package DIT [34] calculates information atoms using the following steps: (i) construct all antichains (time complexity of $\mathcal{O}(2^{2^n})$), (ii) order the antichains [$\mathcal{O}(|D_n|^2)$], (iii) calculate all redundancies ($\mathcal{O}(c(n)|D_n|)$, where $c(n)$ is a function that depends on the chosen redundancy measure), and, finally, (iv) recursively calculate the atoms [$\mathcal{O}(|D_n|^2)$]. Overall, this approach scales with $\mathcal{O}(2^{2^n})$, and in practice has been found to be feasible only up to $n = 3$ variables.

In contrast, when the Möbius function is known, it can be stored as a $|D_n| \times |D_n|$ matrix, and the calculation of PID atoms reduces to a matrix-vector multiplication:

$$I_{\theta}^{\alpha}(X; Y) = [M^T v]_{\alpha}, \quad (11)$$

where $[M]_{\gamma\delta} = \mu_{\mathcal{R}_n}(\gamma, \delta)$ is the matrix that stores the Möbius function, and $[v]_{\gamma} = I_{\cap}^{\gamma}(X; Y)$. Crucially, while one still needs to know the antichains of $\mathcal{B}_N$, there is no need to order them. In addition, the antichains can be stored together with M at negligible cost. Therefore, this approach effectively bypasses the double exponential scaling when calculating the antichains, as well as the squared Dedekind scaling associated to ordering the lattice. This method thus allows one to exchange time complexity for space complexity, at the cost of calculating the Möbius function once. Moreover, the Möbius function is sufficiently sparse that it can be easily shipped with a package like DIT for up to five variables, as shown in Sec. II C. Calculating any information atom then amounts to the inner product between a row of the Möbius function and the vector of redundancies [$\mathcal{O}(c(n)|D_n|)$]. Therefore, while calculating all atoms, thus still leads to a superexponential scaling $\mathcal{O}(c(n)|D_n|^2)$, we found that this approach is tractable for up to five variables (see Sec. II C).

The approach developed here brings additional benefits: it allows us to compute many atoms even without knowledge about the antichains and the Möbius function. In particular, our next result shows that the topmost synergy atom can be straightforwardly and efficiently calculated by exploiting the structure of the redundancy lattice—but without explicitly building said lattice (thus providing a double-exponential speedup).

Proposition 2. The topmost synergy atom, $I_{\theta}^{(1\dots n)}(X; Y)$, can be directly expressed as a linear combination of the redundancies of all $(n - 1)$ -variable synergies as follows:

$$I_{\theta}^{(1\dots n)}(X; Y) = \sum_{U \subseteq \{1, \dots, n\}} (-1)^{n-|U|} I_{\cap}^{S_n^U}(X; Y), \quad (12)$$

where $S_n^U = \{\{1, \dots, n\} \setminus x \mid x \in U\}$.

Proof. See Appendix B 3. ■

Because the redundancy lattice is isomorphic to its opposite order (by Lemma B 1), Proposition 2 can be dualized to derive a closed-form expression for many of the lower order redundancies as well. Given a set of n variables, let P_n^m , where $0 \leq m \leq n$, denote the set that contains m distinct singleton

sets of variables, and all sets of pairs of variables that do not contain any of the singletons. For $n = 3$, this corresponds to the sets

$$P_3^0 = \{(1, 2), (1, 3), (2, 3)\}, \quad (13)$$

$$P_3^1 = \{(1), (2, 3)\}, \{(2), (1, 3)\}, \{(3), (1, 2)\}, \quad (14)$$

$$P_3^2 = \{(1), (2)\}, \{(1), (3)\}, \{(2), (3)\}, \quad (15)$$

$$P_3^3 = \{(1), (2), (3)\}. \quad (16)$$

Note that elements of P_n^m are clearly antichains of $\mathcal{B}_n$.

Proposition 3. Let P_n^m be the set of low-order redundancies defined above, and let $P_n = \bigcup_{m=0}^n P_n^m$. Then P_n contains 2^n elements and forms a Boolean algebra at the bottom of the redundancy lattice $\mathcal{R}_n$, so for all $m : 0 \leq m \leq n$ any PID atom $\gamma \in P_n^m$ can be expressed as

$$I_{\theta}^{\gamma}(X; Y) = \sum_{k: m \leq k \leq n} (-1)^{k-m} \sum_{\substack{\alpha \leq \gamma \\ \alpha \in P_n^k}} I_{\cap}^{\alpha}(X; Y). \quad (17)$$

Proof. See Appendix B 4. ■

Since constructing set P_n and calculating the Möbius inversion is only exponential in n , this result also provides a double-exponential speedup for the information atoms in P_n , which makes it possible to calculate low-order redundancies for much larger n than was possible before.

C. Computational evaluations of PID and Φ ID

We investigated the performance of the fast Möbius transform [Theorem (1)] in calculating the complete Möbius function for the redundancy lattice $\mathcal{R}_n$ for different numbers of predictors n . The required computations were found to be relatively lightweight for $n = 2, \dots, 5$, taking less than 60 minutes to calculate the $n = 5$ case on a standard laptop. This allow us to efficiently use PID to study scenarios with up to five predictor variables, enabling analyses such as the one presented in Sec. IID.

It is worth mentioning that the full five-variable Möbius function could be stored in under 400 kB, since only around 0.5% of the possible entries are nonzero. This means that these values could be directly hardcoded into software packages featuring functions for information decomposition (e.g., DIT), so it doesn't need to be calculated every run. That said, this approach may not yield useful results for more than $n = 5$ information sources, as the Dedekind number for $n = 6$ is $|D_6| = 7828354$, which means that storing the six-variable Möbius function is likely to take up over 400 GB—assuming similar sparsity. The resulting values of the Möbius function for various values of n are shown in Fig. 1. The matrices with the coefficients of the Möbius function for $n = 2, \dots, 5$ are available at a public repository [35]. The columns are ordered by antichain cardinality, not by the order of the redundancy lattice, as ordering the lattice is not necessary. To order the columns with respect to redundancy ordering, one can invert the Möbius function matrix to obtain the ζ function matrix. This matrix can be interpreted as the adjacency matrix of a directed acyclic graph (DAG). Any topological ordering of this DAG corresponds to a valid ordering of the atoms, under

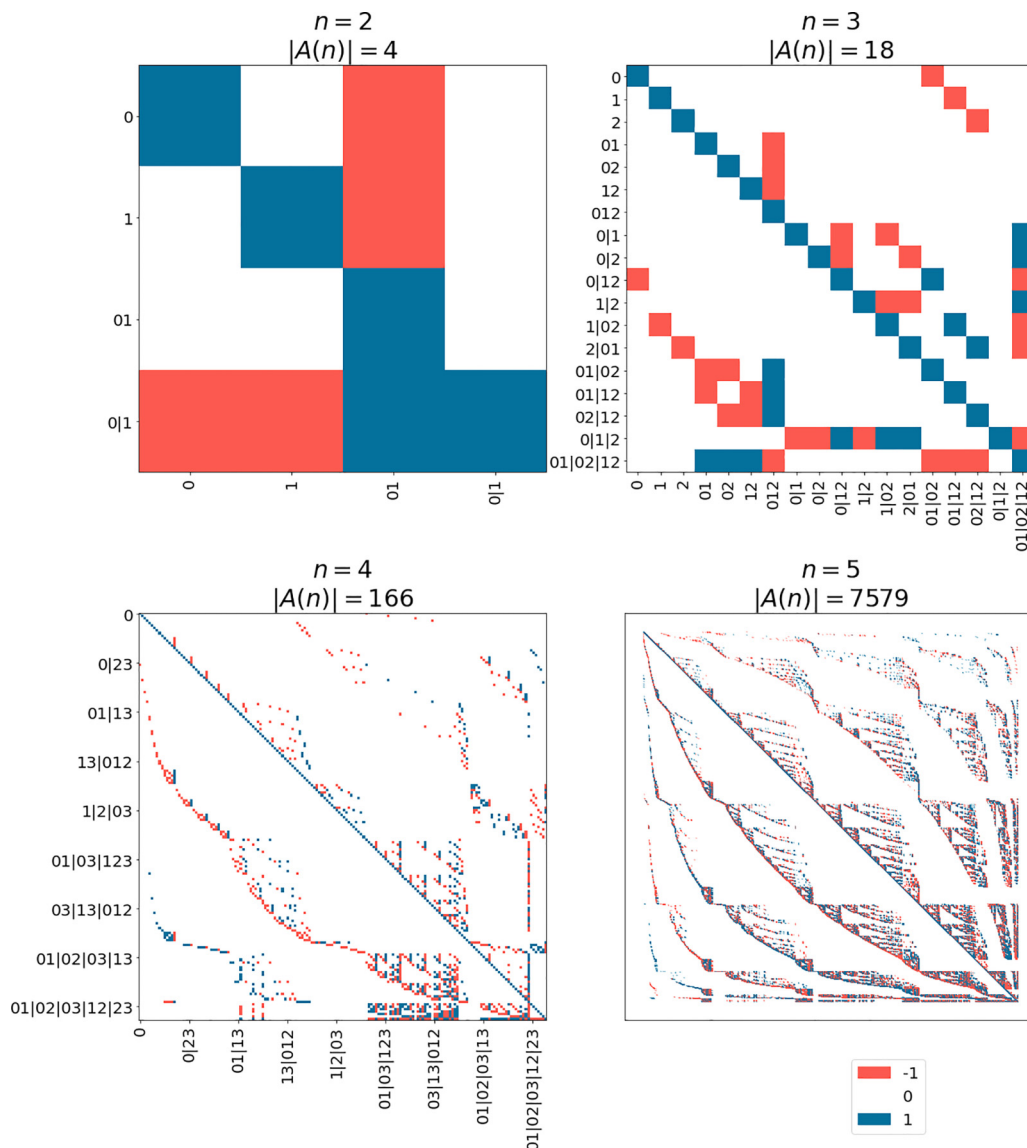


FIG. 1. The Möbius function of the redundancy lattice of up to five variables. The rows and columns are ordered by the cardinality of the antichains. For $n = 4$, only every 20th antichain is indicated along the axes. For $n = 5$, none are shown.

which the ζ and Möbius functions will be lower-diagonal matrices. The fractal-like structure of the matrix that can be observed in Fig. 1 therefore depends on the chosen ordering of the antichains, though it is generally to be expected because the redundancy lattice is a free distributive lattice. A free distributive lattice contains sublattices corresponding to Boolean algebras, which are self-similar. Indeed, the Möbius function of the Boolean algebra shows a similar fractal-like structure (see Appendix C).

With the Möbius function calculated, we calculated the full five-variable PID on four canonical distributions on five binary source variables ($X_1, \dots, X_5$) and a binary target variable Y :

$$p_{\text{uniform}}(X, Y) = \frac{1}{\mathfrak{I}_6}, \quad (18)$$

$$p_{\text{XOR}}(X, Y) = \begin{cases} \frac{1}{2^5} & \text{if } Y = \sum_{i=1}^5 X_i \pmod{2} \\ 0 & \text{otherwise,} \end{cases} \quad (19)$$

$$p_{\text{unq}}(X, Y) = \begin{cases} \frac{1}{2} & \text{if } X_1 = Y \\ 0 & \text{otherwise,} \end{cases} \quad (20)$$

$$p_{\text{red}}(\mathbf{X}, Y) = \begin{cases} \frac{1}{2} & \text{if } \forall i : X_i = Y \\ 0 & \text{otherwise.} \end{cases} \quad (21)$$

We then calculated the PID of these distributions using two well-known measures of redundancy: the measure $I_{\min}$ originally introduced by Williams and Beer [4], and the minimal mutual information (MMI) measure I_{MMI} , introduced by the authors of Ref. [36]. While there are many more measures of redundancy, we chose to focus on the MMI because it is popular, generalises to any number of variables, and is computationally efficient (a comparison of computation time between different measures is presented in Appendix D). However, our method completely decouples the Möbius inversion from the estimation of the redundancy, so the fast Möbius transform can be equally applied to any other measure

TABLE I. Value of specific PID atoms, $n = 5$.

$I_{\min}/\text{MMI}$	(12345)	(1)(2)(3)(4)(5)	(1)
p_{xor}	1	0	0
p_{red}	0	1	0
p_{unq}	0	0	1
p_{uniform}	0	0	0

or estimation method. Our results, shown in Table I, confirmed that I_{MMI} and $I_{\min}$ agreed in all cases and behaved exactly as expected: The information in the XOR distribution is entirely synergistic (indeed, one needs to know all five variables simultaneously to predict the target), while the information about the target variables is carried redundantly by the sources. The unique information carried by X_1 in p_{unq} is also correctly identified, as is the lack of information in the uniform distribution.

Finally, we also evaluated the capabilities of our approach to compute the Möbius inversion for ΦID with various numbers of time source and target variables. The current approach amounts to constructing two PID lattices and recursively solving all $|D_n|^2$ equations. Unfortunately, the largest number of variables this has been possible with thus far has been $n = 2$. With the Möbius function approach, we can write the ΦID equations in terms of $\mu_{\mathcal{R}_n}$ as

$$I_{\delta}^{\alpha \rightarrow \beta}(X; Y) = [(M \otimes M)v]_{\alpha \rightarrow \beta}, \quad (22)$$

where M is the matrix containing the Möbius function of the normal PID lattice [i.e., $[M]_{\gamma\delta} = \mu_{\mathcal{R}_n}(\gamma, \delta)$], v is a vector encoding the redundancy function evaluated in the different ΦID atoms ($[v]_{\alpha \rightarrow \beta} = I_{\cap}^{\alpha \rightarrow \beta}(X, Y)$), and $\otimes$ denotes the Kronecker product. Using the fact that M is sparse, the tensor product $(M \otimes M)$ can be calculated in under 20 ms for the case of four predictors and four target variables, resulting in a matrix with 2 007 889 nonzero values (which is less than 0.3% of all entries). Therefore, using this approach, it is straightforward to calculate the Möbius function of the ΦID lattice for up to four predictors and targets, enabling analyses as the one shown in Sec. II E.

D. Case study: Decomposing the predictive power of the brain's frequency bands

The human brain generates a complex repertoire of activity, commonly measured through techniques such as functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG). In fMRI, it is common to analyze correlations between the time courses of activity across pairs of brain regions, summarized through the so-called functional gradient [37]. In MEG, brain activity is conventionally analyzed through its power spectrum and separated into five canonical frequency bands, referred to as $\delta, \theta, \alpha, \beta, \gamma$ in order of increasing frequency [38]. Comprehensively investigating synergy and redundancy among all five bands was previously an intractable problem. We here present a complete decomposition of the information that these five frequency bands carry about the functional gradient of the human brain.

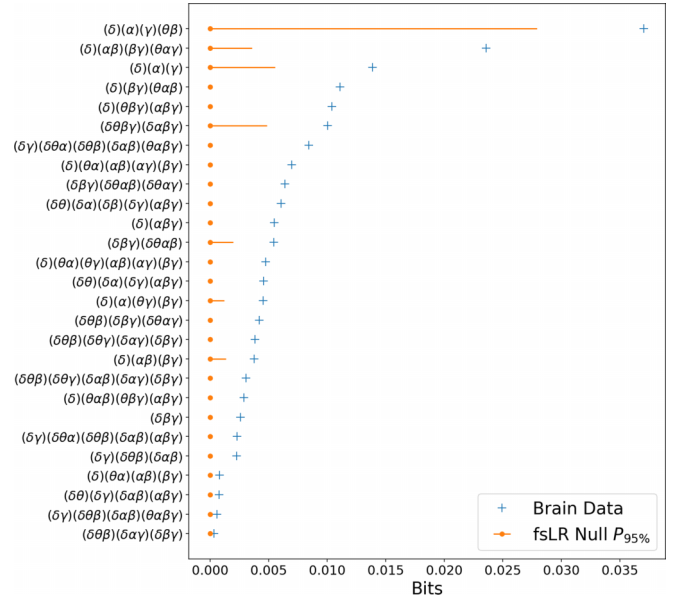


FIG. 2. Information about neural functional connectivity provided by the 27 strongest information atoms of the five canonical brain frequency bands. Blue plus signs denote information atoms in the real brain data, and orange lines denote the 95th percentile of a spatial null model (aka spin test).

Our results reveal that 27 atoms (shown in Fig. 2) were significant beyond the 95th percentile of their null distribution.¹ Interestingly, results suggest there are stark differences in the way different bands carry information about the functional gradient: for example, the δ band appears as an individual source by itself in 12 out of 27 of the significant atoms (more than any other band), whereas the θ and β bands never appear by themselves in any of the 27 atoms. This suggests that the δ band holds much of the redundant information that the brain's electromagnetic activity carries about the functional gradient, whereas the θ and β bands carry this information more synergistically. At the same time, 26 out of 27 atoms contain at least one synergistic source (i.e., one set with two or more bands), emphasising the pervasive role of synergy in linking local and global patterns of brain activity.

The full analysis took around 52 seconds on a standard laptop. Most of this time was spent on calculating the redundancy values, with the actual Möbius inversion over the 7579-node lattice taking only around 0.03 seconds. We tried to calculate the same Möbius inversion with the DIT package, with pre-computed redundancy values, but this did not terminate within the three hours we allowed it to run.

Although the neuroscientific implications of these results require further research, this proof of concept illustrates the possibilities enabled by the fast Möbius transform in empirical data analyses. While the issue of interpretability already arises in this five-variable setting, patterns in the nonzero parts of the decomposition can still reveal informational structure (patterns like “(δ) always appears by itself”). Aggregation

¹Note that, as an exploratory analysis, these results were not corrected for multiple comparisons.

methods can also help characterize the decomposition, like those proposed by Jansma [39], Rosas *et al.* [40], or Mediano *et al.* [6]. In the following section, we propose an extension of the latter approach and group atoms with similar information dynamics.

E. Case study: Cross-scale interdependencies in baroque music

We next turn to the Φ ID framework, and use the Möbius function to investigate the relevance of high-order information dynamics found in music. Our analysis focuses on two sets of repertoire: the well-known chorales for four voices by Johann Sebastian Bach (1685–1750), and the music for string instruments Opus 1 and 3–6 by Arcangelo Corelli (1653–1713). These works are characterized by an elaborate interplay between the melodic lines and rich harmonic progressions, which results in a broad range of chords displayed along the repertoire. Furthermore, as is typical in the Baroque period (approximately 1600–1750), these pieces display a balance in the richness of each of the four voices, which contrasts with the subsequent Classic (1730–1820) and Romantic (1780–1910) periods, where higher voices tended to take the lead while the lower voices provided support. Following the preprocessing pipeline reported in Ref. [41], we studied the chord transitions observed in music scores of Bach and Corelli containing four parts: four voices (soprano, alto, tenor, and bass) in the case of Bach’s chorales and four string instruments (first violin, second violin, viola, and cello) in the case of Corelli’s pieces. Every chord corresponds to a combination of four variables that can take values in an alphabet of 13 possible states (the 12 notes plus silence). Hence, chord transitions can be studied via the probability of observing two different subsequent chords. More details about preprocessing steps are provided in the Appendix.

Results show that Bach’s and Corelli’s scores exhibit significant differences related to their interscale dynamics, as reflected by higher levels of top-down and bottom-up causation in Bach’s chorales (see Fig. 3, categorization of atoms outlined in Methods, Sec. IV D). Other categories of Φ ID atoms did not exhibit significant differences. These results were confirmed by performing additional control analyses, which are reported in Appendix E. Overall, these results suggest that Bach’s chorales excel in how they allow for richer interplay between the whole and the parts, i.e., on how the activity of individual voices affects global chords and vice-versa. Calculating the redundancies in all $166^2 = 27\,556$ antichains took around five minutes, and the Möbius inversion over the corresponding Φ ID lattice took around 0.08 seconds on a standard laptop.

III. DISCUSSION

Here we introduced the fast Möbius transform as an algebraic method to efficiently calculate information decomposition. The fast Möbius transform is based on a closed-form formula for the Möbius function of the redundancy lattice, which turns the calculation of information atoms from the redundancy function into a simple matrix-vector multiplication, opening the way to various numerical optimisation procedures (e.g., parallelization, GPUs). Crucially, this method avoids

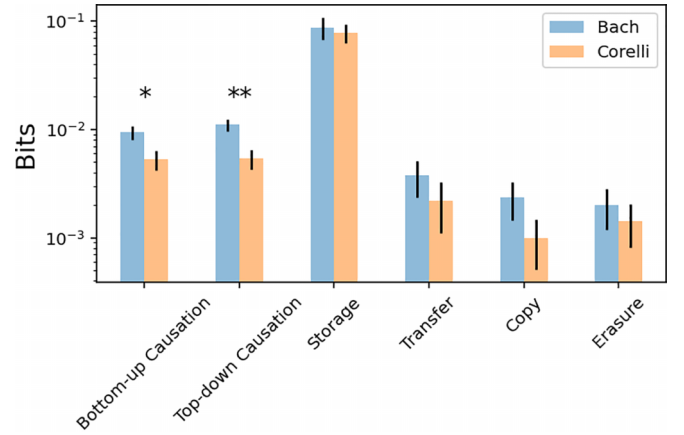


FIG. 3. Relevance of various categories of Φ ID atoms for the Bach’s and Corelli’s music scores. The different Φ ID atoms can be categorized in the above six categories (see Methods, Sec. IV D). Results show that the dynamics within the music of these two composers differ in the interplay between the whole and the parts, which is captured by the top-down and bottom-up causation categories. Significance was calculated via a t-test between the value of the atoms in each category, and significance is indicated as $p \leq 0.05$ with * and $p \leq \frac{0.05}{6}$ with ** (that is, ** indicates significance after Bonferroni correction). Error bars indicate standard error of the mean.

the need to build the lattice itself or to invert a system of equations, which constitutes an important computational bottleneck to compute information decomposition. By doing this, the fast Möbius transform allows one to run information decomposition analyses that were unfeasible before.

The practical utility of the fast Möbius transform was demonstrated in two case studies. First, we decomposed the information about the functional gradient observed in the human brain via fMRI provided by the various frequency bands observed in magnetoencephalographic (MEG) recordings. Results reveal substantial differences in how different bands provide information: the δ band tends to provide information by itself, while θ and β bands tend to carry information synergistically. Second, we disentangled the various information dynamics phenomena observed in the chorales of J.S. Bach and contrasted them against the ones observed in the music of his contemporary Arcangelo Corelli. Results show that Bach’s music possesses a stronger predominance of information modes involving phenomena across scales, where an individual voice affects global chords and vice versa.

The methods behind the fast Möbius transform open promising ways to explore information decomposition from an algebraic perspective. In effect, the results presented here extend the mathematical foundations of information decomposition using the theory of finite distributive lattices and combinatorics. Indeed, showing that the redundancy lattice naturally emerges from a general algebraic construction—that is independent of particular interpretations of the properties of redundant information—supports the validity of this type of decomposition, while presenting their properties from another perspective. We anticipate that this algebraic perspective will open avenues for the theoretical and applied study of multivariate information in complex systems. It should be emphasized that, unfortunately, computing the full PID and Φ ID

beyond $n = 5$ is still difficult, regardless of how efficiently the Möbius function of the corresponding lattices can be calculated. Still, there might be situations in which only some of the atom decompositions are of interest, in which case the Möbius function for specific terms even though a full calculation is infeasible—e.g., using results such as Propositions 2 and 3. Even when one has access to a PID or Φ ID decomposition beyond $n = 5$, interpreting the results can become challenging as the number of atoms grows. To address this, Mediano *et al.* [6] suggested grouping the atoms together into categories like transfer or downward causation. Already in the four-variable case, the results from Sec. II E suggest that this approach can be useful to interpret the results. Alternatively, Jansma [39] suggested grouping atoms together to calculate or bound quantities like aggregate synergy.

More generally, our approach highlights the critical role of the Möbius inversion formula on the construction of information decomposition, following what is believed to be a general theme in the study of complex systems [42,43]. We consequently expect the fast Möbius transform presented here to be useful in every instance where a quantity is decomposed over a free distributive lattice. Examples include the recent decompositions of causal effects [39] and mutual information rate [44], but also the information decompositions that are not based on synergy but on other PID-like measures [45]. We hope this work may motivate further investigations on the algebraic foundations of information decomposition, which may further enhance our computational tools while deepening our fundamental understanding of these frameworks.

IV. METHODS

A. PID and Φ ID

The PID of the information provided by n variables X about Y is built following four steps:

(i) Define as a source of information any subset of such variables, which are indexed as numbers within round brackets; for example, the source constituted by variables X_1 , X_3 , and X_7 is denoted by (137).

(ii) Establish a way to quantify the redundancy between sources, denoted by $I_{\cap}$, which captures the amount of information provided by multiple groups of variables; for example, the redundancy between sources (123) and (45) is captured by the quantity $I_{\cap}^{(123)(45)}(X; Y)$. The redundancy of a source with a single group of variables is required to be equal to the corresponding mutual information, so, for example, $I_{\cap}^{(12)}(X; Y) = I(X_1, X_2; Y)$.

(iii) Assume that the redundancy between sources $S_1, \dots, S_n$ is the same if we remove any sources that contain other sources. For example, $I_{\cap}^{(12)(34)(4)}(X; Y) = I_{\cap}^{(12)(4)}(X; Y)$. This suggests focusing only on the redundancy between collections of sources that form an antichain, where no source is contained into another. Also, it suggests that the redundancy should monotonically follow the following partial relationship:

$$\alpha \leq \beta \iff \forall B \in \beta, \exists A \in \alpha, A \subseteq B. \quad (23)$$

The set of antichains of n target variables is denoted by A_n , and together with the partial order $\leq$ constitutes the *redundancy*

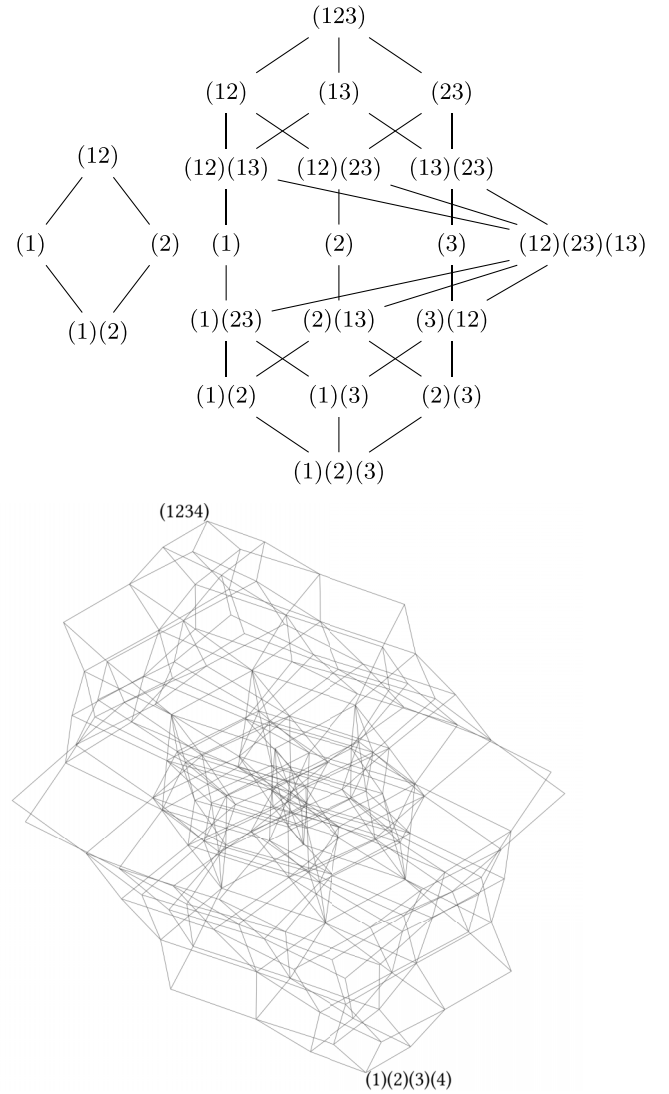


FIG. 4. The redundancy terms of a powerset of n variables can be partially ordered: if A and B are redundancy terms, then $A \leq B$ if for every $b \in B$ there is an $a \in A$ such that $a \leq b$. Shown here are the transitive reduction (Hasse diagrams) of the redundancy lattices $\mathcal{R}_n$ for $n = 2$ (top left), $n = 3$ (top right middle), and $n = 4$ (bottom, labels not shown) variables. Note that elements of $\mathcal{R}_n$ are the antichains in the power set of n variables.

lattice $\mathcal{R}_n = (A_n, \leq)$. The Hasse diagrams of the redundancy lattice are shown in Fig. 4 for $n = 2, 3$, and 4 predictor variables.

(iv) Finally, the PID atoms are defined as the derivative of $I_{\cap}^{\beta}(X; Y)$ on $\mathcal{R}_n$: i.e., via the function $I_{\partial}^{\beta}(X; Y)$ that satisfies the following property for all $\alpha \in A_n$:

$$I_{\cap}^{\alpha}(X; Y) := \sum_{\substack{\beta \in A_n \\ \beta \leq \alpha}} I_{\partial}^{\beta}(X; Y). \quad (24)$$

Using the PID atoms, the total information given by X about Y is decomposed as

$$I(X; Y) = \sum_{\alpha \in A_n} I_{\partial}^{\alpha}(X; Y). \quad (25)$$

PID atoms are usually calculated by inverting a system of linear equations. For example, for the case of $n = 2$ predictors, the PID formalism provides the following system of equations:

$$\begin{aligned} I(X_1, X_2; Y) &= I_{\theta}^{(1)(2)}(X; Y) + I_{\theta}^{(1)}(X; Y) \\ &\quad + I_{\theta}^{(2)}(X; Y) + I_{\theta}^{(12)}(X; Y), \\ I(X_1; Y) &= I_{\theta}^{(1)(2)}(X; Y) + I_{\theta}^{(1)}(X; Y), \\ I(X_2; Y) &= I_{\theta}^{(1)(2)}(X; Y) + I_{\theta}^{(2)}(X; Y). \end{aligned} \quad (26)$$

This provides three equations with four unknowns (the introduction of more predictor variables leads to larger but similarly underdetermined linear systems), which is usually recursively solved by assuming a functional for the redundant information $I_{\theta}^{(1)(2)}(X; Y)$ [4,46]. Alternatively, one can also define either the unique information [47,48] or synergy [40]. Instead of addressing Eq. (26) as a system of equations, in the next section we will investigate the substantial benefits of treating it as a Möbius transform.

The Φ ID framework is a generalization of PID that allows us to decompose the information between two sets of variables X and Y . The decomposition is built in a similar way to PID:

(a) Define a source and target of information to be any subset of the X and Y variables, respectively.

(b) Establish a way to quantify the redundancy between sources and targets, which is denoted as $I_{\cap}^{\alpha \rightarrow \beta}(X; Y)$ with α and β collections of sources and targets, respectively.

(c) Assume that the redundancy between sources and targets is the same if we remove any sources S and targets T that are contained in other sources or targets, respectively, and hence focus on redundancy terms of the form $I_{\cap}^{\alpha \rightarrow \beta}(X; Y)$, where $\alpha \in A_n$ and $\beta \in A_m$ are antichains as in PID. These elements can be naturally ordered as

$$\alpha \rightarrow \beta \leq \alpha' \rightarrow \beta' \quad \text{iff} \quad \alpha \leq \alpha' \text{ and } \beta \leq \beta', \quad (27)$$

where $\leq$ denotes the ordering on $\mathcal{R}_n$. This gives rise to a lattice known as the *double-redundancy lattice* [6].

(d) The Φ ID atoms are obtained via the function $I_{\theta}^{\alpha' \rightarrow \beta'}$ that satisfies

$$I_{\cap}^{\alpha \rightarrow \beta}(X; Y) = \sum_{\alpha' \rightarrow \beta' \leq \alpha \rightarrow \beta} I_{\theta}^{\alpha' \rightarrow \beta'}(X; Y). \quad (28)$$

Then the joint mutual information between X and Y can be decomposed as

$$I(X; Y) = \sum_{\alpha, \beta \in A} I_{\theta}^{\alpha \rightarrow \beta}(X; Y). \quad (29)$$

A Φ ID is the product of two complementary PIDs, one decomposing the information carried by the sources about the targets, and the other decomposing the information carried by the targets about the sources. In fact, the double-redundancy lattice can be shown to correspond to the Cartesian product lattice $\mathcal{R}_n \times \mathcal{R}_n$ with the product order defined above, also known as Cartesian order [[49], Sec. 2.15].

B. Calculus over lattices and the Möbius transform

Sums over lattices, such as the information decompositions stated in Eqs. (24) and (29), can be interpreted as

integrals over an interval of a partially ordered set. In fact, functions defined over partial orders possess a distinct calculus of which we here present some fundamental concepts (more background about partially ordered sets is provided in Appendix A).

An interval between a and b over a partially ordered set $\mathcal{R}$ corresponds to a set $[a, b] := \{x \in \mathcal{R} : a \leq x \leq b\}$. Then, the integral of a function $f(x)$ between a and b over a partially ordered set $\mathcal{R}$ is defined as

$$F(a, b) = \sum_{x \in [a, b]} f(x). \quad (30)$$

To identify the derivative of a function (i.e., the operation that undoes integration), it is useful to first consider the set of all bivariate functions as an algebra under the usual sum and multiplication given by the convolution

$$(F * G)(a, b) = \sum_{x \in [a, b]} F(a, x)G(x, b). \quad (31)$$

Crucially, in this setting, integration can be understood as a particular type of convolution² given by

$$(\zeta * f)(a, b) = \sum_{x \in [a, b]} \zeta(a, x)f(x) = \sum_{x \in [a, b]} f(x), \quad (32)$$

where ζ is the zeta function given by

$$\zeta(a, b) = \begin{cases} 0 & \text{if } [a, b] = \emptyset \\ 1 & \text{otherwise} \end{cases}. \quad (33)$$

When seen from this perspective, the derivative of a function can be defined as its convolution against the multiplicative $*$ -inverse of the zeta function, which is known as the Möbius function. This function was introduced by the mathematician A. F. Möbius in the 19th century in the context of number theory [50], and was extended to partial orders by G-C. Rota in the '60s [26].

Let us now establish the Möbius inversion formula for partially ordered sets, which is analogous to the fundamental theorem of calculus but applied to partial orders. This result states that if f and g are two functions defined over a partially ordered set $\mathcal{T}$, then the two following statements are equivalent [26]:

$$f(a) = \sum_{b \leq a} g(b) \iff g(a) = \sum_{b \leq a} \mu_{\mathcal{T}}(b, a)f(b), \quad (34)$$

where $\mu_{\mathcal{T}}(b, a)$ is the Möbius function of the partial ordering $\mathcal{T}$, which can be recursively constructed via

$$\mu_{\mathcal{T}}(x, y) = \begin{cases} 1 & \text{if } x = y \\ -\sum_{z \in [x, y]} \mu_{\mathcal{T}}(x, z) & \text{if } x < y \\ 0 & \text{otherwise} \end{cases}. \quad (35)$$

If $\mathcal{T}$ is the natural numbers with their usual total order, then the Möbius inversion theorem recovers the discrete fundamental theorem of calculus with $g(x) = (f * \mu_{\mathbb{N}_{\leq}})(0, x) = f(x) - f(x-1)$ being the discrete derivative of f . In general,

²Note that a univariate function $f(x)$ can be seen as bi-variate $f'(x, y) = f(x)$ considering a second variable y with respect to which f' is invariant.

the Möbius function takes a more complex form, which is determined by the chains in the underlying poset [33]. The functional turning f into g is known as the Möbius inversion or *Möbius transform*.³

Overall, the Möbius transform shows that any sum over an interval in a partially ordered set can be inverted, as long as the corresponding Möbius function is known. This result is, indeed, the key that guarantees that PID and Φ ID atoms can be identified from the redundancy function and the corresponding lattice structure. While calculating the Möbius function is often hard, special lattices allow a closed-form expression of their Möbius function: for example, if $\mathcal{T}$ is a powerset ordered by set inclusion, then its Möbius function is known to be $\mu(A, B) = (-1)^{|A|-|B|}$ [33]. Our main result (Theorem 1) extends this finding to the PID and Φ ID lattices.

C. Processing of fMRI and MEG data

We obtained data describing the average functional gradient and spectral power (both in fsLR-4k surface-space) from the neuromaps library [51], resulting in 8000 samples (one per spatial location), each with six real numbers (five frequency bands, to be used as sources, and the functional gradient, to be used as target). The fMRI data is the first principal gradient of functional connectivity reported by Margulies *et al.* [37]. This data was obtained from resting-state fMRI originally collected as part of the Human Connectome Project (HCP) [52] 900 subject release. Resting-state MEG data was originally collected as part of the HCP 1200 subject release and subsequently cleaned and source-reconstructed. Preprocessing steps for both fMRI and MEG were done as per gold-standard HCP pipelines, and full details can be found in the HCP documentation [52].

We quantified redundant information with the MMI measure [53] over binarized data based on its median value. Although binarisation is a common preprocessing step in information-theoretic analysis of neuroimaging data (e.g., in the application of Lempel-Ziv complexity [54]), it has also been criticized on the basis that it removes a large fraction of the information in neural signals [55]. Further work could revisit these analyses using the small but growing toolkit of information decomposition in continuous variables [56].

The calculation of all 7579 information atoms among the frequency bands using the fast Möbius transform took under a minute on a standard laptop. To determine statistical significance, we generated a null distribution for each atom through spatial null models (aka spin tests) that preserve the spatial autocorrelation of the data [57,58].

D. Processing of music scores and Φ ID atoms

Our analyses used electronic scores publicly available from the Kern Scores database at Ref. [60],⁴ focusing on scores

with four melodic lines: four voices (soprano, alto, tenor and bass) in the case of Bach's chorales, and four string instruments (first violin, second violin, viola and cello) in the case of Corelli's pieces. The scores were preprocessed in Python using the MUSIC21 package [59], which allowed us to select only the pieces written in major mode, and to transpose them to C major. The melodic lines were transformed into time series of 13 possible values (one for each note plus one for the silence), using the smallest rhythmic duration as time unit. This generated $\approx 4 \times 10^4$ four-note chords for the chorales and $\approx 8 \times 10^4$ for Corelli's pieces. Chords that persist for multiple measures show up as repeated entries in the time series; for our calculations, we removed any consecutive copies of a chord, keeping only the chord *changes*. We estimated the probability of observing one chord transitioning to another simply by counting the number of such transitions in the repertoire and normalizing this by the total number of chord changes.

We calculated the Φ ID to investigate how chords carry information about the chord that follows them. For this, we consider X^4 and Y^4 to be the four music parts of two subsequent chords, and calculated the corresponding $166^2 = 27\,556$ Φ ID atoms for the considered repertoire of both Bach and Corelli. Following Ref. [6], we grouped these atoms within the following categories:

- (1) Bottom-up causation: Any atom with a single target, where this target is larger than the largest source.
- (2) Top-down causation: Any atom with a single source, where this source is larger than the largest target.
- (3) Storage: Any atom where the source and the target are identical.
- (4) Transfer: Any atom with only one source and one target, and where they are not the same.
- (5) Copy: Any atom with a single source that is a proper subset of the target.
- (6) Erasure: Any atom with a single target that is a proper subset of the source.

Note that, in contrast to the $n = 2$ case, these do not form completely disjoint sets for $n = 4$. For example, the atom $(SAT) \rightarrow (B)$, that describes how the harmony between the soprano, alto, and tenor determines the next melodic movement of the bass is both an instance of top-down causation and transfer. To simplify the analysis, we estimated the relevance of each of these categories of atoms by calculating the mean value of the absolute value of the estimate of each atom involved. Each atom was bias corrected by subtracting the mean value obtained on a null distribution obtained by randomly shuffling the sequences of chords (see Appendix E).

ACKNOWLEDGMENTS

The authors thank A. A. Makkeh for suggesting the name “fast Möbius transform.” A.J. is grateful to J. Jost and E. Olbrich for fruitful discussions on the partial information decomposition and acknowledges support from the Dutch Institute for Emergent Phenomena (DIEP) cluster via the Foundations and Applications of Emergence (FAEME) program, as well as from the Max Planck Institute for Mathematics in the Sciences. F.R. was supported by the PIBBS Affiliate program.

³Not to be confused with *Möbius transformations*, which are rational functions on the complex plane.

⁴At the time of publication, this database is no longer online. The electronic scores as analysed in this research are available at a public repository [35].

DATA AVAILABILITY

The data that support the findings of this article are openly available [37,52,60]. The code to calculate the Möbius function $\mu_{\mathcal{R}_n}$ for up to $n = 5$, as well as the precomputed matrices and the analyses on brain data and music scores, is available at Ref. [35].

APPENDIX A: BACKGROUND

1. Orderings and lattices

An ordering, usually denoted by $\leq$, is a relation linking pairs of objects while observing three properties:

- (i) Reflexivity: $a \leq a$,
- (ii) Antisymmetry: $a \leq b$ and $b \leq a$ if and only if $a = b$,
- (iii) Transitivity: $a \leq b$ and $b \leq c$ imply $a \leq c$.

When $a \leq b$, one says that a is less than or equal to b . Furthermore, when $a \neq b$ and $a \leq b$, then we write $b > a$, and say that a is greater than b .

An example of an ordered set is the natural numbers together with their usual arithmetic ordering. Note that, in this case, any two elements $a, b \in \mathbb{N}$ are comparable (i.e., either $a \leq b$ or $b \leq a$), so $\leq$ is called a total order. However, the properties of orderings presented above allow for more general setups, in which any two elements $a, b \in S$ are either comparable (in which case, either $a \leq b$ or $b \leq a$) or incomparable. Consider, for example, a collection of sets S_i and an ordering $\leq$ under which $S_i \leq S_j$ if and only if $S_i \subseteq S_j$. This is usually not a total order, as there can be two sets where neither contains the other. Such a relationship is called a partial ordering, and a partially ordered set is called a poset, being denoted by calligraphic letters as $(S, \leq) := S$.

While total orderings arrange elements into a relatively simple structure in which all elements form a (potentially infinite) single chain $a \leq b \leq c \leq \dots$, posets can display richer structures. A poset $\mathcal{P}$ can be represented as a graph $G = (\mathcal{P}, E)$, where the set of edges E contains an directed edge $(a, b) \in \mathcal{P} \times \mathcal{P}$ if and only if $a \leq b$. However, a more concise description is the *transitive reduction* obtained by removing all relations that are implied by transitivity. Hence, the graph associated to the transitive reduction contains a directed edge (a, b) if and only if $a \leq b$ and there is no $c \in \mathcal{P}$ such that $a \leq c$ and $c \leq b$. This graph is called the Hasse diagram of the poset, and an example is given in Fig. 5.

A poset $\mathcal{P}$ allows one to define upper and lower bounds. Given a subset $S \subseteq \mathcal{P}$, the element $b \in \mathcal{P}$ is a lower bound for S if $p \leq b$ for all $p \in S$. The greatest lower bound is called the *meet* of S . Similarly, one could define the least upper bound of S , referred to as its *join*. If $S = \{a, b\}$, we denote the join of S as $a \vee b$ and a meet as $a \wedge b$.

A poset where all such meets and joins exist is called a lattice. Please note that not all partial orderings are lattices, as meets and joins are not guaranteed to always exist. Finite lattices have a unique element that is larger than all others, known as the supremum, and a unique element smaller than all others, known as infimum—therefore, their graph representation always looks like a diamond with a single element on top and bottom (for an example, see Fig. 5).

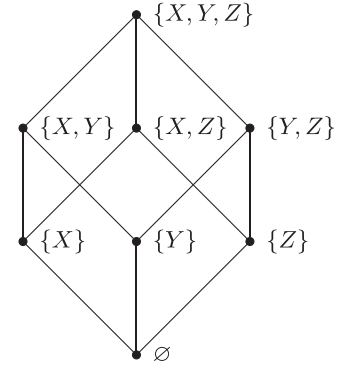


FIG. 5. The power set $\mathcal{P}(\{X, Y, Z\})$ of a set of three variables, ordered by set inclusion, forms a poset known as a Boolean algebra. Shown here is the transitive reduction (Hasse diagram) of the Boolean algebra on three variables. It is common to keep the directionality implicit by drawing the diagram such that all edges are oriented downward. Note that $\{X\} \subseteq \{X, Y, Z\}$, but that the Hasse diagram contains no such edge because $\{X\} \subseteq \{X, Y\} \subseteq \{X, Y, Z\}$. For arbitrary n , the Hasse diagram of a Boolean algebra describes an n cube.

2. Lattices as algebraic structures

The above construction establishes lattices in terms of properties of a partial order. Interestingly, there is an alternative way to build lattices in algebraic terms, which we expose next.

Given a set P , let us introduce two operations $\wedge$ and $\vee$ that take two elements of P as inputs and deliver a third element of P while satisfying three basic properties:

- (1) Commutative: $a \wedge b = b \wedge a$ and $a \vee b = b \vee a$.
- (2) Associative: $(a \wedge b) \wedge c = a \wedge (b \wedge c)$ and $(a \vee b) \vee c = a \vee (b \vee c)$.
- (3) Idempotent: $a \wedge a = a$, $a \vee a = a$.

In addition to this, let us require these operations to interact with each other via the following laws:

- (4) Absorption: $a \wedge (a \vee b) = a$ and $a \vee (a \wedge b) = a$.

These are the same properties that one would naturally require from the operations of union and intersection of sets, and from the logical operations AND and OR.

It can be shown that the algebraic structure that these two operations induce on the underlying set P , written as $\mathcal{L} = (P, \wedge, \vee)$, is precisely a lattice in the order-theoretic sense described above with the order relationship $a \leq b \iff a \wedge b = a$. In other words, an ordering on P is equivalent to an algebraic structure of joins and meets.

The equivalence between lattices and this algebraic construction is important because it opens the door to study ordering relationship with the rich toolkit that can be found in the algebraic literature. One important example is the case of Boolean lattices, which require a third operation equivalent to a negation. Another important example is distributed lattices, where the operations $\wedge$ and $\vee$ are required to distribute over each other.

Another interesting consequence of the link between orderings and algebras is the possibility to build lattices as reflecting a formal language via free algebras. Free algebras are built starting from a number of primitive expressions (say

$a, b, \dots$), and then explore all possible algebraic combinations of them. For example, if $\wedge$ and $\vee$ distribute, then the formal language resulting from a free algebra over two primitives a and b only has four elements: a , b , $a \wedge b$, and $a \vee b$, which has a natural lattice structure considering $\wedge$ and $\vee$ as meets and joins. Such lattices are discussed in more detail in Sec. II A.

APPENDIX B: PROOFS

1. Proof of Proposition 1

Proof. Since the operations $\wedge$ and $\vee$ of $\mathcal{F}_n$ are commutative, associative, and idempotent, each element of this lattice can be reduced to the form $S_1 \wedge S_2 \wedge S_3 \wedge \dots$, where the S_i do not contain $\wedge$ operators and are mutually incomparable (since $A \wedge B = A$ if $A \leq B$). That is, the lattice is an ordering on the set of antichains of $\mathcal{P}(N)$. Left to show is that this ordering is the same.

Given two antichains A and $S = S_1 \wedge S_2 \wedge S_3 \wedge \dots$, assume that $A \leq S$, such that $\forall S_i \in S \exists a \in A : a \subseteq S_i$. Now, note that for any two elements of $\mathcal{F}_n$, if $a \subseteq b$, then $b = a \vee c$ for some $c \in \mathcal{F}_n$. By the absorption laws, one can find that $a \wedge b = a \wedge (a \vee c) = a$. We write

$$A \wedge S = \left(\bigwedge_{a \in A} a \right) \wedge \left(\bigwedge_{S_i \in S} S_i \right). \quad (\text{B1})$$

Because $\wedge$ and $\vee$ are associative and commutative, one is free to rearrange these terms. By assumption, each $a \in A$ can be paired with an $S_a \in S$ such that $a \subseteq S_a$. $A \wedge S$ can thus be written as $a_1 \wedge a_2 \wedge \dots \wedge a_n \wedge S_1 \wedge S_2 \wedge \dots \wedge S_n$, and terms can be freely reordered so until somewhere a term $a_i \wedge S_i$ appears for which $a_i \subseteq S_i$. This reduces to a_i , and this procedure can be done until all S_j are eliminated from the expression. This shows that $A \leq S \Rightarrow A \wedge S = A \iff A \preceq S$.

Left to show is the reverse implication. Assume that $A \wedge S = A$, but that there is an $S_i \in S$ such that $\nexists a \in A$ is $a \subseteq S_i$. This means that the reduction procedure from above can be performed, until one is left with $A \wedge S = A \wedge S_i$. For this to be equal to A for all choices of A and all ordering of elements of A , S_i must be the identity for the $\wedge$ operation. This means that S_i is the maximal element of $\mathcal{F}_n$, which is the set N itself. However, that means that every $a \in A$ is a subset of S_i , which is in contradiction with the assumption that there is no such a . This shows that $A \preceq S \iff A \wedge S = A \Rightarrow A \leq S$ and completes the proof. ■

2. Proof of Theorem 1

This proof requires some preliminary definitions and lemmas. First, we review a general construction that can generate a distributive lattice $\mathcal{L}$ from the ideals of another partially ordered set $\mathcal{P}$, known as Birkhoff's representation theorem. This result is so ubiquitous that it is often referred to as the fundamental theorem of finite distributive lattices.

Before stating this theorem, let us introduce the notion of *ideal* in the context of order theory. For a given poset $\mathcal{P}$, one can define an ideal as a subset $I \subseteq \mathcal{P}$ such that for all $x \in I$

and $y \in \mathcal{P}$, then

$$y \leq x \iff y \in I. \quad (\text{B2})$$

The set of all ideals of $\mathcal{P}$ can be ordered by inclusion and forms a lattice denoted by $J(\mathcal{P})$.

Theorem B1 (Birkhoff [27]). Let $\mathcal{L}$ be a finite distributive lattice. Then there exists a unique (up to isomorphism) partially ordered set $\mathcal{P}$ such that $\mathcal{L}$ is isomorphic to the lattice of ideals $J(\mathcal{P})$.

Given $N = \{1, \dots, n\}$, a set of n indices, let $\mathcal{B}_n$ denote the power set of N ordered by inclusion. Note that, in general, the collection of maximal elements of each ideal within $J(\mathcal{P})$ is equal to the set of antichains of $\mathcal{P}$. Thus, thanks to Birkhoff's result, $J(\mathcal{B}_n)$ is guaranteed to be a distributive lattice over the antichains of $\mathcal{B}_n$. More specifically, it can be shown that $J(\mathcal{B}_n)$ is the free distributive lattice over n generators [33].

One convenient property of distributive lattices constructed as ordered ideals $J(\mathcal{P})$ is that the Möbius function on $J(\mathcal{P})$ can be expressed as [33]

$$\mu_{J(\mathcal{P})}(A, B) = \begin{cases} (-1)^{|B \setminus A|} & \text{if } B \setminus A \text{ is an antichain of } \mathcal{P} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B3})$$

This property would give us a convenient way to calculate the Möbius function of the redundancy lattice, but unfortunately it cannot be used directly as the set of maximal ideals $J(\mathcal{B}_n)$ do not directly give $\mathcal{R}_n$. For example, for $N = \{0, 1, 2\}$ the largest ideal (that is not $\mathcal{B}_n$ itself) has as supremum the element $\{\{0, 1\}, \{0, 2\}, \{1, 2\}\}$, which is not the maximal term on the redundancy lattice (i.e., $\{\{0, 1, 2\}\}$). To instantiate the correspondence, we show that the free distributive lattice is self-dual and explicitly construct the isomorphism between the redundancy lattice and $J(\mathcal{B}_n)$.

Lemma B1. The free distributive lattice $\mathcal{F}_n$ is self-dual, i.e., it is isomorphic to its reverse order, and the isomorphism amounts to exchanging $\wedge$ and $\vee$.

Proof. The lattice $\mathcal{F}_n$ is ordered as $a \leq b \iff a \wedge b = a$. The dual lattice $\mathcal{F}_n^{op}$ is ordered as $a \wedge b = a \iff b \leq a \iff a \leq b$. The absorption law $(a \vee b) \wedge a = a$ also holds in $\mathcal{F}_n^{op}$. If we now assume that $b \leq a$, then $a \wedge b = a$, which can be combined with the absorption law to yield $(a \vee b) \wedge a = a \wedge b$ which is true if and only if $a \vee b = b$. This shows that $\mathcal{F}_n$ is ordered as $a \leq b \iff a \wedge b = a$, and $\mathcal{F}_n^{op}$ is ordered as $a \leq b \iff a \vee b = a$, so exchanging $\wedge$ and $\vee$ in $\mathcal{F}_n$ (or equivalently, redundancy and synergy in $\mathcal{R}_n$) yields the dual lattice $\mathcal{F}_n^{op}$. ■

Lemma B2. Let $J(\mathcal{B}_n)$ be the free distributive lattice over n generators, and let $\mathcal{R}_n$ be the redundancy lattice of n variables. Then $J(\mathcal{B}_n)$ and $\mathcal{R}_n$ are isomorphic. Furthermore, the anti-isomorphism is realized by a map that sends a term in $\mathcal{R}_n$ to the ideal generated by its complement.

Proof. Let A_n be the set of antichains of $\mathcal{B}_n$, and $*$ the map considered in Sect. II B (i.e., $*$: $A_n \rightarrow A_n$ by taking the complement of each element of α as $*$: $\alpha \mapsto \alpha^* = \{N \setminus a \mid a \in \alpha\}$). Note that $\alpha \subseteq \beta \iff \beta^* \subseteq \alpha^*$. Under the map $*$, the ordering of redundancy terms can be rewritten as

$$\alpha \leq \beta \iff \forall b^* \in \beta^* \exists a^* \in \alpha^* \text{ such that } b^* \subseteq a^*. \quad (\text{B4})$$

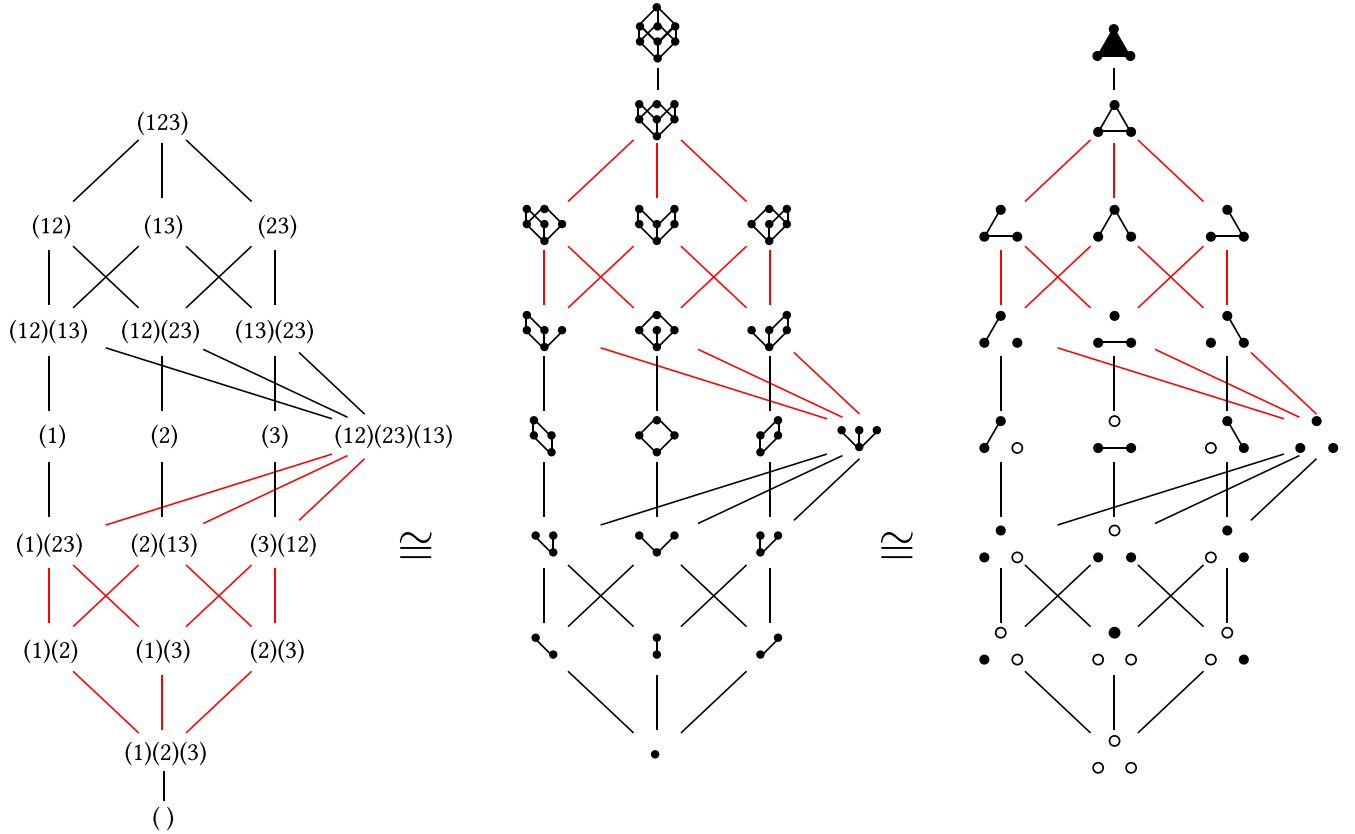


FIG. 6. A visual representation of the isomorphisms used in the proofs, shown here for $n = 3$. Shown are the redundancy lattice $\mathcal{R}_3$ (left), the lattice of ideals of a three-variable Boolean algebra $J(\mathcal{B}_3)$ (middle), and the lattice $\mathcal{C}_2$ of simplicial complexes included in the two-simplex, ordered by inclusion (right). Highlighted in red in all three is the Boolean algebra at the bottom of the redundancy lattice and where it is mapped to under the isomorphisms. The element dangling at the bottom of $\mathcal{R}_n$ and the top of $J(\mathcal{B}_n)$ and $\mathcal{C}_{n-1}$ is usually omitted from the lattice in the PID/ Φ ID literature, but shown here for completeness and clarity.

Let $I_{\mathbf{y}} = \{t \in \mathcal{B}_n \mid \exists x \in \mathbf{y} \text{ s.t. } t \leq x\}$ be the ideal in $J(\mathcal{B}_n)$ generated by $\mathbf{y} \in \mathcal{A}_n$. Equivalently, $I_{\mathbf{y}}$ is the ideal of $\mathcal{B}_n$ with $\mathbf{y}$ as its maximal element. We can then write

$$I_{\alpha^*} = \{t \in \mathcal{B}_n \mid \exists a^* \in \alpha^* \text{ such that } t \subseteq a^*\}, \quad (\text{B5})$$

$$I_{\beta^*} = \{t \in \mathcal{B}_n \mid \exists b^* \in \beta^* \text{ such that } t \subseteq b^*\}. \quad (\text{B6})$$

Now since $J(\mathcal{B}_n)$ is simply the set of all ideals of $\mathcal{B}_n$ ordered by inclusion, i.e., $I \leq I' \iff I \subseteq I'$, we can write the ordering on $J(\mathcal{B}_n)$ as

$$I_{B^*} \leq I_{A^*} \iff \forall x : (x \in I_{B^*} \Rightarrow x \in I_{A^*}). \quad (\text{B7})$$

In other words, both (B4) and (B7) express that any element of B^* is the subset of an element in A^* , so the two orderings are equivalent. This means that $A \leq B \iff I_{B^*} \leq I_{A^*}$, and so the map $A \mapsto I_{A^*}$ is an anti-isomorphism between the redundancy lattice $\mathcal{R}_n$ and $J(\mathcal{B}_n)$. Since $J(\mathcal{B}_n)$ is itself isomorphic to its reverse order (by Lemma B1), we have that $\mathcal{R}_n$ is isomorphic to $J(\mathcal{B}_n)$. ■

With all these elements in place, we can now present the proof of the theorem.

Proof of Theorem 1. The Möbius function of $J(\mathcal{B}_n)$ is given by Eq. (B3). Since the free distributive lattice is self-dual by Lemmas B1 and B2 implies that the Möbius function of $\mathcal{R}_n$ is given by the same expression after each antichain α to its complement-ideal I_{α^*} . ■

This construction can be summarized as follows. Let $i : \mathcal{R}_n \rightarrow J(\mathcal{B}_n)$ map each antichain to the ideal it generates as $i : A \mapsto I_A$. Let $\mu_{\mathcal{R}_n}$ and $\mu_{J(\mathcal{B}_n)}$ be the Möbius functions on $\mathcal{R}_n$ and $J(\mathcal{B}_n)$, respectively. Then the following diagram commutes:

$$\begin{array}{ccc} \mathcal{R}_n \times \mathcal{R}_n & \xrightarrow{(i \circ *) \times (i \circ *)} & J(\mathcal{B}_n) \times J(\mathcal{B}_n) \\ & \searrow \mu_{\mathcal{R}_n} & \downarrow \mu_{J(\mathcal{B}_n)} \\ & & \mathbb{Z} \end{array}$$

3. Proof of Proposition 2

Proof. The proof splits into three parts: (i) We show that the redundancies among $(n - 1)$ -variable synergies form a Boolean algebra at the top of the redundancy lattice. (ii) We show that any other element in the lattice is not part of a Boolean algebra at the top of the lattice. (iii) We show that this leads to the above expression.

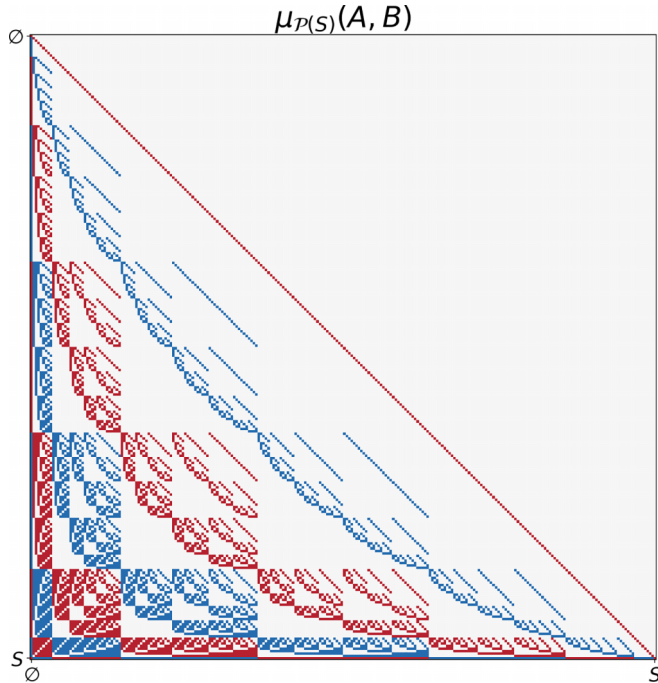


FIG. 7. The Möbius function of the Boolean algebra $\mathcal{B}_n$ of eight variables. The values are shown in a heat map, where the rows and columns are ordered by cardinality, which corresponds to a topological ordering on $\mathcal{B}_n$. The values are given by $\mu_{\mathcal{B}_n}(A, B) = (-1)^{|B \setminus A|}$.

(i) Consider an element of $\mathcal{R}_n$ that corresponds to the redundancy among $(n - 1)$ -variable synergies. It is of the form

$$\{\{X \setminus x\} \mid x \in U\} \quad (\text{B8})$$

for some $U \subseteq X$. Under the bijection $i \circ *$, this is sent to a term

$$I_{\{\{x\} \mid x \in U\}} = \{\{x\} \mid x \in U\} \cup \emptyset. \quad (\text{B9})$$

This bijection is an anti-isomorphism of the redundancy lattice when the subsequent ideals are ordered by inclusion (see

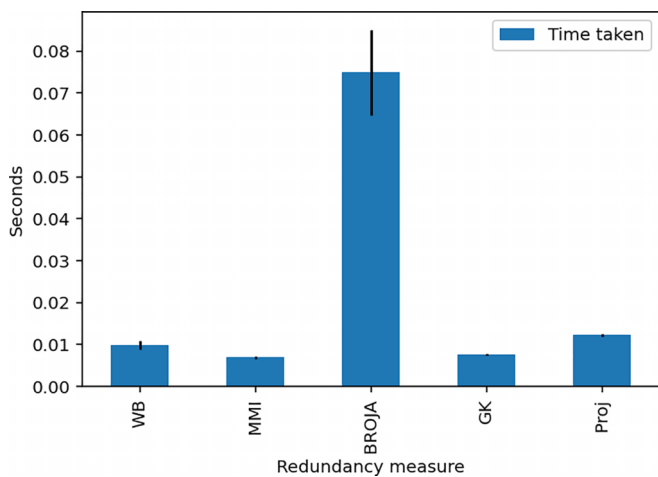


FIG. 8. The time taken to calculate different PID measures on 1000 observations of two variables (the theta and delta band). Error bars indicate the standard deviation across 25 random subsamplings of the full data set.

Fig. 6), which means that the terms of the form (B8) form a Boolean algebra if and only if the terms of the form (B9) do. Note that ordering the latter simply corresponds to ordering by inclusion on all different $U \subseteq X$, so they form a Boolean algebra at the bottom of $J(\mathcal{B}_n)$. Therefore, the redundancies among $(n - 1)$ -variable synergies form a Boolean algebra at the top of $\mathcal{R}_n$. This proves (i).

(ii) We use the fact that the free distributive lattice is isomorphic to the lattice of monotonic Boolean functions (functions of the type $f : \mathcal{B}_n \rightarrow \{0, 1\}$ such that $f(A) \leq f(B) \iff A \leq B$). Note that these functions are fully defined by the set $U \subset \mathcal{B}_n$ that get mapped to zero. We therefore define f_U to be the monotonic Boolean function that maps the set U to zero such that $f_U \leq f_V \iff V \subseteq U$. The topmost function on this lattice is $f_\emptyset$, and below it are functions f_A , where A corresponds to some subset of $\mathcal{B}_n$. When A contains only singletons, then f_A is part of the Boolean algebra between $f_\emptyset$ and $f_{\{\{x\} \mid x \in S\}}$. However, A might also contain nonsingleton elements from $\mathcal{B}_n$. Assume that the interval $[f_A, f_\emptyset]$ is a Boolean algebra and that A contains nonsingleton elements. Let $b \in A$ denote the nonsingleton element. Then construct $f_{A,b} := f_{A \setminus \{\{x\} \mid x \in b\}}$. Since $[f_A, f_\emptyset]$ is a Boolean algebra, every $f_{A,b}$ is guaranteed to exist. However, $f_{A,b}$ maps all elements of b to 1, but b itself to 0. This violates monotonicity, so $f_{A,b}$ is not part of the lattice. This is a contradiction and shows that the only elements that are part of the upper Boolean algebra are the f_A , where A contains only singletons. Since there is thus only one Boolean algebra on n variables at the top, this must be the same as the one from (i). This proves (ii).

(iii) Finally, we show that $\mu_{\mathcal{R}_n}(\alpha, \hat{1}) \neq 0 \iff \alpha$ is part of the upper Boolean algebra. It is known that Eq. (B3) can be written as [33]

$$\mu_{\mathcal{R}_n}(A, B) = \begin{cases} (-1)^{|B \setminus A|} & \text{if } [I_A, I_B] \text{ is a Boolean algebra} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B10})$$

The complete synergy atom can be written as

$$I_\emptyset^{(1 \dots n)}(S, T) = \sum_{\alpha \leq (1 \dots n)} \mu_{\mathcal{R}_n}(\alpha, (1 \dots n)) I_\alpha^\alpha(X; Y), \quad (\text{B11})$$

which reduces to

$$= \sum_{\alpha: [\alpha, (1 \dots n)] \text{ is a Boolean algebra}} (-1)^{|I_\alpha \setminus I_\emptyset|} I_\alpha^\alpha(X; Y). \quad (\text{B12})$$

By (i) and (ii), the only antichains a for which $[a, (1 \dots n)]$ is a Boolean algebra are the redundancies among $(n - 1)$ -variable synergies. This means that

$$I_\emptyset^{(1 \dots n)}(X; Y) = \sum_{U \subseteq \{1, \dots, n\}} (-1)^{n - |U|} I_\emptyset^U(X; Y), \quad (\text{B13})$$

where $S_n^U = \{\{1, \dots, n\} \setminus x \mid x \in U\}$ and completes the proof. ■

4. Proof of Proposition 3

Proof. That there is a unique Boolean algebra at the bottom of the redundancy lattice follows immediately from Proposition 2 and the fact that the redundancy lattice is isomorphic to its dual. It remains to show which antichains are contained in this Boolean algebra. This is most easily seen by

TABLE II. Upon modifying Bach chorales to artificially add redundancy between the soprano and the movement of the other voices, the information atoms change as expected.

	Original	Twofold Red.	Threefold Red.
$S \rightarrow T B$	0.00	1.45	0.00
$S \rightarrow A T B$	0.00	0.00	1.94

noting that $\mathcal{R}_n$ is isomorphic to the lattice $\mathcal{C}_{n-1}$ of simplicial complices included in the $n-1$ -simplex, ordered by inclusion. This isomorphism is essentially the same order-reversing bijection ($i \circ *$) as was established between $\mathcal{R}_n$ and $J(\mathcal{B}_n)$, shown in Fig. 6, so the Boolean algebra at the bottom of $\mathcal{R}_n$ is mapped to the one at the top of $\mathcal{C}_{n-1}$. The top element of $\mathcal{C}_{n-1}$ (not including the full simplex) consists of the simplicial complex on n vertices with all faces on $n-1$ vertices. In Fig. 6, this corresponds to the simplicial complex included in the two-simplex (i.e. the triangle) consisting of all faces on two vertices (i.e., edges). There are $\binom{n}{n-1} = n$ such faces. Removing each of these faces yields another simplicial complex, included in the top one. Since $\mathcal{C}_{n-1}$ is ordered by inclusion, this yields the Boolean algebra that is mapped to the one at the bottom of $\mathcal{R}_n$. The unique smallest element of this Boolean sublattice in $\mathcal{C}_{n-1}$ is the simplicial complex that results from removing all faces on $n-1$ vertices, which corresponds to the complex with all faces on $n-2$ vertices. Under the isomorphism $(i \circ *)^{-1}$ that maps antichains to their complements, this corresponds to the unique redundancy atom among all two-variable synergies. Now consider adding faces on $n-1$ vertices to this smallest element. Each time such a face is added, the corresponding antichain in $\mathcal{R}_n$ obtains the corresponding singleton set (namely, the complement of the $n-1$ vertices), and all pairs that contain that singleton element are dropped from the antichain (this is clear from both perspectives: in $\mathcal{R}_n$ this is necessary to preserve the antichain requirement, and in $\mathcal{C}_{n-1}$ the face on $n-1$ vertices absorbs the faces on $n-2$ vertices that connect to it). This gives the definition of the set P_n of antichains contained in the Boolean sublattice at the bottom of $\mathcal{R}_n$, showing that it is $\mathbb{N}_0$ graded by the number of singletons, and completes the proof. ■

APPENDIX C: MÖBIUS FUNCTION OF THE BOOLEAN ALGEBRA

Shown here in Fig. 7 is the Möbius function of the Boolean algebra $\mathcal{B}_n$ of eight variables. It is given by $\mu_{\mathcal{B}_n}(A, B) = (-1)^{|B \setminus A|}$ and shows the self-similar nature of Boolean algebras.

APPENDIX D: COMPARISON OF COMPUTATIONAL COMPLEXITY FOR PID MEASURES

To illustrate the difference in computational complexity of the different PID measures, we compared a range of different

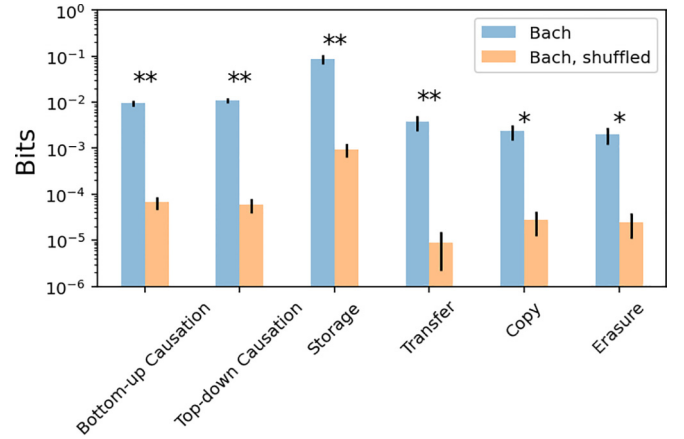


FIG. 9. The mean absolute value of atoms in Bach chorales and a shuffled version in each of the six categories of information dynamics. Only in the first four categories does the original Bach music score significantly differ from a randomly shuffled version after Bonferroni correction. Significance was calculated via a t-test, where * means $p \leq 0.05$ and ** $p \leq \frac{0.05}{6}$ (that is, ** indicates significance after Bonferroni correction). Error bars indicate standard error of the mean.

measures implemented in the DIT package [34]. We compared the measures on the data set from Sec. II D, keeping only the delta and theta bands so bivariate measures like BROJA can be included. We estimated the different measures on 1000 observations on a standard laptop. The results are shown in Fig. 8.

APPENDIX E: ADDITIONAL ANALYSES ON BACH'S CHORALES

While both composers show more downward than upward causation, indicating strong harmonic constraints, Bach scores significantly higher in both categories. To confirm that the atoms indeed capture the musical dependencies, we modified the Bach chorales and verified that the atoms change as expected. The first perturbed version, referred to as twofold redundant Bach, copied the soprano and alto of the original, but replaced the tenor and bass. Let $X_t \in \Sigma$ denote the value of voice X at time step t . We then set $T_{t+1} = (S_t - 4) \bmod 13$ and $B_{t+1} = (S_t - 6) \bmod 13$. The threefold redundant Bach dataset is then formed by, in addition, setting $A_{t+1} = (S_t - 2) \bmod 13$. We then indeed find that the Φ ID atoms change as expected, shown in Table II.

As an additional control, we randomly shuffled the chords from Bach's chorales and compared the mean value of atoms in each of the information dynamics categories with those of the unshuffled Bach. Figure 9 shows that the Bach chorales showed significant information dynamics in all categories, except for copy and erasure. Results for Corelli were identical. This null distribution additionally shows that the effect of bias correction is negligible, as the shuffled atoms are two to three orders of magnitude weaker than the unshuffled atoms.

- [1] M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos* (Simon and Schuster, New York, New York, 1993).
- [2] J. Gleick, *The Information: A History, a Theory, a Flood* (Pantheon Books, New York, New York, 2011).
- [3] Y. Bengio and N. Malkin, Machine learning and information theory concepts towards an AI mathematician, *Bull. Am. Math. Soc.* **61**, 457 (2024).
- [4] P. L. Williams and R. D. Beer, Nonnegative decomposition of multivariate information, [arXiv:1004.2515](https://arxiv.org/abs/1004.2515).
- [5] J. T. Lizier, N. Bertschinger, J. Jost, and M. Wibral, Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work, *Entropy* **20**, 307 (2018).
- [6] P. A. Mediano, F. E. Rosas, A. I. Luppi, R. L. Carhart-Harris, D. Bor, A. K. Seth, and A. B. Barrett, Towards an extended taxonomy of information dynamics via integrated information decomposition, [arXiv:2109.13186](https://arxiv.org/abs/2109.13186).
- [7] H. Rajpal, C. von Stengel, P. A. Mediano, F. E. Rosas, E. Viegas, P. A. Marquet, and H. J. Jensen, Quantifying hierarchical selection, [arXiv:2310.20386](https://arxiv.org/abs/2310.20386).
- [8] Z. Cang and Q. Nie, Inferring spatial and signaling relationships between cells from single cell transcriptomic data, *Nat. Commun.* **11**, 2084 (2020).
- [9] D. Marinazzo, J. Van Roozendaal, F. E. Rosas, M. Stella, R. Comolatti, N. Colenbier, S. Stramaglia, and Y. Rosseel, An information-theoretic approach to hypergraph psychometrics, *Behav. Res.* **56**, 8057 (2024).
- [10] T. F. Varley and P. Kaminski, Untangling synergistic effects of intersecting social identities with partial information decomposition, *Entropy* **24**, 1387 (2022).
- [11] F. Rosas, P. A. Mediano, M. Ugarte, and H. J. Jensen, An information-theoretic approach to self-organization: Emergence of complex interdependencies in coupled dynamical systems, *Entropy* **20**, 793 (2018).
- [12] P. Orio, P. A. Mediano, and F. E. Rosas, Dynamical noise can enhance high-order statistical structure in complex systems, *Chaos* **33**, 123103 (2023).
- [13] T. Scagliarini, D. Marinazzo, Y. Guo, S. Stramaglia, and F. E. Rosas, Quantifying high-order interdependencies on individual patterns via the local O-information: Theory and applications to music analysis, *Phys. Rev. Res.* **4**, 013184 (2022).
- [14] M. Gatica, R. Cofré, P. A. Mediano, F. E. Rosas, P. Orio, I. Diez, S. P. Swinnen, and J. M. Cortes, High-order interdependencies in the aging brain, *Brain Connectivity* **11**, 734 (2021).
- [15] A. I. Luppi, P. A. Mediano, F. E. Rosas, N. Holland, T. D. Fryer, J. T. O'Brien, J. B. Rowe, D. K. Menon, D. Bor, and E. A. Stamatakis, A synergistic core for human brain evolution and cognition, *Nat. Neurosci.* **25**, 771 (2022).
- [16] T. F. Varley, O. Sporns, S. Schaffelhofer, H. Scherberger, and B. Dann, Information-processing dynamics in neural networks of macaque cerebral cortex reflect cognitive state and behavior, *Proc. Natl. Acad. Sci. USA* **120**, e2207677120 (2023).
- [17] T. F. Varley, M. Pope, J. Faskowitz, and O. Sporns, Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex, *Commun. Biol.* **6**, 451 (2023).
- [18] R. Herzog, F. M. Barbey, M. N. Islam, L. Rueda-Delgado, H. Nolan, P. Prado, M. Krylova, I. Izyurov, N. Javaheripour, L. V. Danyeli *et al.*, High-order brain interactions in ketamine during rest and task: A double-blinded cross-over design using portable EEG on male participants, *Transl. Psychiatry* **14**, 310 (2024).
- [19] M. Pope, T. F. Varley, M. G. Puxeddu, J. Faskowitz, and O. Sporns, Time-varying synergy/redundancy dominance in the human cerebral cortex, *J. Phys. Complex.* **6**, 015015 (2025).
- [20] T. F. Varley, O. Sporns, N. J. Stevenson, P. Yrjölä, M. G. Welch, M. M. Myers, S. Vanhatalo, and A. Tokariev, Emergence of a synergistic scaffold in the brains of human infants, *Commun. Biol.* **8**, 743 (2025).
- [21] T. M. Tax, P. A. Mediano, and M. Shanahan, The partial information decomposition of generative neural network models, *Entropy* **19**, 474 (2017).
- [22] A. M. Proca, F. E. Rosas, A. I. Luppi, D. Bor, M. Crosby, and P. A. Mediano, Synergistic information supports modality integration and flexible learning in neural networks solving multiple tasks, *PLoS Comput. Biol.* **20**, e1012178, (2024).
- [23] C. Kaplanis, P. Mediano, and F. Rosas, Learning causally emergent representations, in *NeurIPS 2023 Workshop: Information-Theoretic Principles in Cognitive Systems* (2023).
- [24] X. Kong, O. Liu, H. Li, D. Yogatama, and G. V. Steeg, Interpretable diffusion via information decomposition, [arXiv:2310.07972](https://arxiv.org/abs/2310.07972).
- [25] A. I. Luppi, F. E. Rosas, P. A. Mediano, D. K. Menon, and E. A. Stamatakis, Information decomposition and the informational architecture of the brain, *Trends Cognit. Sci.* **28**, 352 (2024).
- [26] G. C. Rota, On the foundations of combinatorial theory: I. The theory of Möbius functions, in *Classic Papers in Combinatorics* (Birkhäuser Boston, Boston, MA, 1964), pp. 332–360.
- [27] G. Birkhoff, Rings of sets, *Duke Mat. J.* **3**, 443 (1937).
- [28] F. E. Rosas, P. A. Mediano, H. J. Jensen, A. K. Seth, A. B. Barrett, R. L. Carhart-Harris, and D. Bor, Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data, *PLoS Comput. Biol.* **16**, e1008289 (2020).
- [29] P. A. Mediano, F. E. Rosas, J. C. Farah, M. Shanahan, D. Bor, and A. B. Barrett, Integrated information as a common signature of dynamical and information-processing complexity, *Chaos* **32**, 013115 (2022).
- [30] C. Finn and J. T. Lizier, Generalised measures of multivariate information content, *Entropy* **22**, 216 (2020).
- [31] A. J. Gutknecht, M. Wibral, and A. Makkeh, Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic, *Proc. R. Soc. A* **477**, 20210110 (2021).
- [32] J. E. Hopcroft, R. Motwani, and J. D. Ullman, Introduction to automata theory, languages, and computation, *Acm Sigact News* **32**, 60 (2001).
- [33] R. P. Stanley, in *Enumerative Combinatorics*, 2nd ed. Cambridge Studies in Advanced Mathematics Vol. 1 (Cambridge University Press, Cambridge, 2011).
- [34] R. G. James, C. J. Ellison, and J. P. Crutchfield, DIT: A Python package for discrete information theory, *J. Open Source Software* **3**, 738 (2018).
- [35] The Fast Möbius transform: An algebraic approach to decompose information (2024), <https://github.com/AJnsm/algebraicPID>.
- [36] N. Bertschinger, J. Rauh, E. Olbrich, and J. Jost, Shared information—New insights and problems in decomposing information in complex systems, in *Proceedings of the European*

- Conference on Complex Systems 2012* (Springer, Brussels, Belgium, 2013), pp. 251–269.
- [37] D. S. Margulies, S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides *et al.*, Situating the default-mode network along a principal gradient of macroscale cortical organization, *Proc. Natl. Acad. Sci. USA* **113**, 12574 (2016).
 - [38] G. Buzsaki and A. Draguhn, Neuronal oscillations in cortical networks, *Science* **304**, 1926 (2004).
 - [39] A. Jansma, Decomposing interventional causality into synergistic, redundant, and unique components, [arXiv:2501.11447](https://arxiv.org/abs/2501.11447).
 - [40] F. E. Rosas, P. A. Mediano, B. Rassouli, and A. B. Barrett, An operational information decomposition via synergistic disclosure, *J. Phys. A: Math. Theor.* **53**, 485001 (2020).
 - [41] F. E. Rosas, P. A. Mediano, M. Gastpar, and H. J. Jensen, Quantifying high-order interdependencies via multivariate extensions of the mutual information, *Phys. Rev. E* **100**, 032305 (2019).
 - [42] A. Jansma, Higher-order interactions and their duals reveal synergy and logical dependence beyond Shannon-information, *Entropy* **25**, 648 (2023).
 - [43] A. Jansma, Mereological approach to higher-order structure in complex systems: From macro to micro with Möbius, *Phys. Rev. Res.* **7**, 023016 (2025).
 - [44] L. Sparacino, G. Mijatovic, Y. Antonacci, L. Ricci, D. Marinazzo, S. Stramaglia, and L. Faes, Decomposing multivariate information rates in networks of random processes, [arXiv:2502.04555](https://arxiv.org/abs/2502.04555).
 - [45] A. J. Gutknecht, A. Makkeh, and M. Wibral, From Babel to Boole: The logical organization of information decompositions, *Proc. R. Soc. A* **481**, 20240174 (2025).
 - [46] V. Griffith, E. K. Chong, R. G. James, C. J. Ellison, and J. P. Crutchfield, Intersection information based on common randomness, *Entropy* **16**, 1985 (2014).
 - [47] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, Quantifying unique information, *Entropy* **16**, 2161 (2014).
 - [48] R. G. James, J. Emenheiser, and J. P. Crutchfield, Unique information via dependency constraints, *J. Phys. A: Math. Theor.* **52**, 014002 (2018).
 - [49] B. A. Davey and H. A. Priestley, *Introduction to Lattices and Order* (Cambridge University Press, Cambridge, 2002).
 - [50] A. F. Möbius, Über eine besondere art von umkehrung der reihen, *J. Reine Angew. Math.* **9**, 105 (1832).
 - [51] R. D. Markello, J. Y. Hansen, Z.-Q. Liu, V. Bazinet, G. Shafiei, L. E. Suárez, N. Blostein, J. Seidlitz, S. Baillet, T. D. Satterthwaite *et al.*, neuromaps: Structural and functional interpretation of brain maps, *Nat. Meth.* **19**, 1472 (2022).
 - [52] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, WU-Minn HCP Consortium, The WU-Minn human connectome project: An overview, *NeuroImage* **80**, 62 (2013).
 - [53] A. B. Barrett, Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems, *Phys. Rev. E* **91**, 052802 (2015).
 - [54] A. G. Casali, O. Gosseries, M. Rosanova, M. Boly, S. Sarasso, K. R. Casali, S. Casarotto, M.-A. Bruno, S. Laureys, G. Tononi *et al.*, A theoretically based index of consciousness independent of sensory processing and behavior, *Sci. Trans. Med.* **5**, 198ra105 (2013).
 - [55] P. A. Mediano, F. E. Rosas, A. I. Luppi, V. Noreika, A. K. Seth, R. L. Carhart-Harris, L. Barnett, and D. Bor, Spectrally and temporally resolved estimation of neural signal diversity, *eLife* **12**, RP88683 (2023).
 - [56] D. A. Ehrlich, K. Schick-Poland, A. Makkeh, F. Lanfermann, P. Wollstadt, and M. Wibral, Partial information decomposition for continuous variables based on shared exclusions: Analytical formulation and estimation, *Phys. Rev. E* **110**, 014115 (2024).
 - [57] A. F. Alexander-Bloch, H. Shou, S. Liu, T. D. Satterthwaite, D. C. Glahn, R. T. Shinohara, S. N. Vandekar, and A. Raznahan, On testing for spatial correspondence between maps of human brain structure and function, *NeuroImage* **178**, 540 (2018).
 - [58] R. D. Markello and B. Misić, Comparing spatial null models for brain maps, *NeuroImage* **236**, 118052 (2021).
 - [59] <http://web.mit.edu/music21>.
 - [60] C. S. Sapp, Online database of scores in the Humdrum file format, in *Proceedings of the 6th International Conference on Music Information Retrieval, ISMIR* (London, UK, 2005), pp. 664–665.