


Quantifying high-order interdependencies on individual patterns via the local O-information: Theory and applications to music analysis

Tomas Scagliarini,¹ Daniele Marinazzo,² Yike Guo,^{3,4} Sebastiano Stramaglia^{1,5} ,^{1,5} and Fernando E. Rosas^{4,6,7}

¹*Dipartimento Interateneo di Fisica, Università degli Studi Aldo Moro, Bari and INFN, Italy*

²*Department of Data Analysis, Ghent University, Belgium*

³*Department of Computer Science, Hong Kong Baptist University, Hong Kong*

⁴*Data Science Institute, Imperial College London, United Kingdom*

⁵*Center of Innovative Technologies for Signal Detection and Processing (TIRES), Università degli Studi Aldo Moro, Italy*

⁶*Centre for Psychedelic Research, Department of Brain Science, Imperial College London, United Kingdom*

⁷*Centre for Complexity Science, Imperial College London, United Kingdom*



(Received 26 August 2021; accepted 10 February 2022; published 4 March 2022)

High-order, beyond-pairwise interdependencies are at the core of biological, economic, and social complex systems, and their adequate analysis is paramount to understand, engineer, and control such systems. This paper presents a framework to measure high-order interdependence that disentangles their effect on each individual pattern exhibited by a multivariate system. The approach is centered on the *local O-information*, a new measure that assesses the balance between synergistic and redundant interdependencies at each pattern. To illustrate the potential of this framework, we present a detailed analysis of music scores from J. S. Bach, which reveals how high-order interdependence is deeply connected with highly nontrivial aspects of the musical discourse. Our results place the local O-information as a promising tool of wide applicability, which opens other perspectives for analyzing high-order relationships in the patterns exhibited by complex systems.

DOI: [10.1103/PhysRevResearch.4.013184](https://doi.org/10.1103/PhysRevResearch.4.013184)

I. INTRODUCTION

The analysis of interdependence is crucial for understanding the staggering complexity of structures and behaviors manifested in biological, economic, and social systems. The unprecedented amount of data available for scientific scrutiny provides unique opportunities to deepen our understanding of multivariate coevolving complex systems, including the orchestrated activity of multiple brain areas, the interactions between different genes, and the relationship between various econometric indices. Importantly, what allows these systems to be more than the sum of their parts is not to be found in the material nature of their parts, but in the fine structure of their interdependencies [1].

Information theory provides an ideal framework to study interdependencies in multivariate systems, which establishes the notion of *information* as a common currency under which diverse systems can be measured and compared [2]. A particularly promising approach for analyzing the structure of interdependencies is the partial information decomposition (PID), which distinguishes different “modes” of information that multiple predictors convey about a target variable [3–5]. Two paradigmatic examples of such modes are synergy and redundancy [6–10]: redundancy corresponds to information which can be retrieved independently from more than one

source, while synergy corresponds to statistical relationships that exist in the whole but cannot be seen in the parts—this being rooted in the elementary fact that variables can be pairwise independent while being globally correlated.

Despite continuous efforts to develop PID, the precise way in which synergies and redundancies should be calculated is still being revised [11–24]. One way to circumvent this challenge is to avoid computing the full decomposition, and study mixtures of PID atoms that can be captured by linear combinations of Shannon measures. One such measure is the O-information [25], which has been shown to effectively capture the overall balance between redundant and synergistic modes. The effectiveness of the O-information in practical analyses has been verified by recent applications on populations of spiking neurons [26], and the relationship between neural patterns and aging [27].

An important limitation of the O-information is that it characterizes a multivariate system with a single number, which summarized to the aggregated effect of various patterns. Building on the rich literature of pointwise information measures [28–30], in this paper we introduce the *local O-information*, which evaluates each pattern separately—such that its ensemble average recovers the O-information. More specifically, the local O-information constitutes an overall measure that characterizes the high-order interdependencies between the parts of a multivariate system at each possible pattern of activity. Put differently, the local O-information evaluates the “statistical quality” of each realization of collections of random variables, providing a signed scalar that assesses the balance between redundancies and synergies at each individual pattern.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

This paper presents the theory behind the local O-information, and then illustrates its rich capabilities by analyzing the scores of the chorales of J. S. Bach. Our results show how the local O-information is capable of revealing subtle musical relationships, including properties of different intervals, chord dispositions, harmonic depth, and the relationship between music and text. Thanks to its ability to uncover such highly nontrivial relationships, the local O-information is a valuable addition to the toolkit of data analysts interested in the study of complex systems.

The rest of the paper is organized as follows. Section II provides background information about the O-information, introduces the new local O-information, and then illustrates its basic properties on small spin systems. Then, Sec. III presents a detailed analysis of the local O-information on the chorales of J. S. Bach, and finally Sec. IV summarizes our main conclusions.

II. A LOCAL MEASURE OF INFORMATION QUALITY

Let us consider a scientist interested in studying a system of n elements described by the random vector $\mathbf{X}^n = (X_1, \dots, X_n)$, who has enough data to reliably estimate its statistics, which is denoted by $p(\mathbf{X}^n)$. A question of interest is how to leverage the statistics encoded in $p(\mathbf{X}^n)$ in order to deepen our understanding of the structure of interdependencies between the elements of \mathbf{X}^n . Such understanding can lead either to the building of statistical markers to characterize different systems or different states of the same system, or to compare seemingly heterogeneous systems based on the similarity of their relational structure.

Through this section, random variables are denoted by capital letters (e.g., X, Y) and their realizations by lowercase letters (e.g., x, y). Random vectors and their realizations are denoted by capital and lowercase boldface letters, respectively.

A. O-information

Shannon's mutual information is a popular metric of interdependency, which overcomes the limitations of correlation metrics such as Pearson's in that it captures both linear and nonlinear relationships and is applicable to ordinal data. However, the mutual information can only assess the relationships between two (sets of) variables, being unable to fully explore the rich interplay that can take place within triple or higher-order interactions.

Two multivariate extensions of the mutual information are the *total correlation* (TC) [31] and the *dual total correlation* (DTC) [32], which are defined as

$$\text{TC}(\mathbf{X}^n) := \sum_{i=1}^n H(X_i) - H(\mathbf{X}^n),$$

$$\text{DTC}(\mathbf{X}^n) := H(\mathbf{X}^n) - \sum_{i=1}^n H(X_i | \mathbf{X}_{-i}^n).$$

Above, $H(X_i) = -\sum_{x_i} p(x_i) \log p(x_i)$ corresponds to the Shannon entropy, $H(X_i | X_j) = H(X_i, X_j) - H(X_j)$ is the conditional Shannon entropy, and \mathbf{X}_{-i}^n is the vector of all variables except X_i [i.e., $\mathbf{X}_{-i}^n = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$]; hence,

the term $H(X_i | \mathbf{X}_{-i}^n)$ quantifies how X_i is independent from the other $n - 1$ variables. As the mutual information, both TC and DTC are non-negative quantities (see Appendix A), being zero if and only if all variables X_1, \dots, X_n are jointly statistically independent, i.e., if $p(\mathbf{X}^n) = \prod_{i=1}^n p(X_i)$.

Despite the similarities between TC and DTC, these two metrics provide distinct but complementary views on the strength of the interdependencies in a multivariate system. Specifically, the TC accounts for the effect of *collective constraints*, which refer to regions of the phase space that the system explores less [33], while the DTC measures the amount of *shared randomness* between the variables, i.e., the amount of information that can be collected in one variable that also refers to the activity of another [25]. An attractive way to exploit these complementary views is by considering their difference,

$$\Omega_n(\mathbf{X}^n) = \text{TC}(\mathbf{X}^n) - \text{DTC}(\mathbf{X}^n), \quad (1)$$

which is known as the *O-information* [25]. The O-information can be seen as a revision of the measure of neural complexity proposed by Tononi *et al.* in [34], which provides a mathematical construction that is closer to their original desiderata [33]. In effect, the O-information is a signed metric that captures the balance between high- and low-order statistical constraints [35]. By construction, $\Omega(\mathbf{X}^n) < 0$ implies a predominance of high-order constraints within the system \mathbf{X}^n , a condition that is usually referred to as *statistical synergy*. Conversely, $\Omega(\mathbf{X}^n) > 0$ implies that the system \mathbf{X}^n is dominated by low-order constraints, which imply *redundancy* of information. This nomenclature is further supported by the following key properties:

(1) It captures genuine high-order effects, as it is zero for systems with only pairwise interdependencies: if the joint distribution of \mathbf{X}^{n-1} (for n odd) can be factorized as $p_{\mathbf{X}^{n-1}}(\mathbf{x}^{n-1}) = \prod_{k=1}^{n/2} p_{X_{2k}, X_{2k+1}}(x_{2k}, x_{2k+1})$, then $\Omega(\mathbf{X}^n) = 0$.

(2) The O-information is maximized by redundant distributions where the same information is copied in multiple variables, and is minimized by synergistic (“ n -bit paritylike”) distributions: e.g., for binary variables, Ω is maximized by the “ n -bit copy” where X_1 is a Bernoulli random variable with parameter $p = 1/2$ and $X_1 = X_2 = \dots = X_n$, and is minimized when X_1, \dots, X_{n-1} are independent and identically distributed fair coins and $X_n = \sum_{j=1}^{n-1} X_j \pmod{2}$.

(3) The O-information characterizes the dominant tendency, being additive over noninteractive subsystems: if the system can be factorized as $p_{\mathbf{X}^n}(\mathbf{x}^n) = p_{X_1, \dots, X_m}(x_1, \dots, x_m) \times p_{X_{m+1}, \dots, X_n}(x_{m+1}, \dots, x_n)$, then $\Omega(\mathbf{X}^n) = \Omega(X_1, \dots, X_{m-1}) + \Omega(X_m, \dots, X_n)$.

For more details related to the O-information, we refer the reader to Ref. [25].

B. Local O-information

Building on the properties of the O-information reviewed in the previous section, one can design a measure that can capture these effects on a state-by-state basis. In effect, the O-information provides a single scalar that characterizes the interdependencies of a system *on average*. However, in many systems of interest this average represents a midpoint

between highly heterogeneous states, which could potentially be of limited value to understand the role of individual realizations of the corresponding variables. In this section we introduce a *pointwise* measure that allows one to calculate the O-information on individual states.

As a first step, let us introduce a *local total correlation* and *local dual total correlation* which are given by

$$tc(\mathbf{x}^n) := \sum_{j=1}^n h(x_j) - h(\mathbf{x}^n), \quad (2)$$

$$dtc(\mathbf{x}^n) := h(\mathbf{x}^n) - \sum_{j=1}^n h(x_j | \mathbf{x}_{-j}^n), \quad (3)$$

where $h(\mathbf{x}^n) = -\log p(\mathbf{x}^n)$ is the information content of the state \mathbf{x}^n [28]. These quantities capture how the strength of the multivariate interdependencies vary with the state of the system, providing a generalization to the pointwise mutual information introduced in Ref. [36].

Using these definitions, we can define the *local O-information* as follows (see Appendix B):

$$\omega(\mathbf{x}^n) := tc(\mathbf{x}^n) - dtc(\mathbf{x}^n) \quad (4)$$

$$= (n - 2)h(\mathbf{x}^n) + \sum_{j=1}^n [h(x_j) - h(\mathbf{x}_{-j}^n)]. \quad (5)$$

In contrast to $\Omega(\mathbf{X}^n)$, which provides a single value for the random variable, $\omega(\mathbf{x}^n)$ assigns a number for each possible state \mathbf{x}^n . In particular, the local O-information has the following useful properties:

- (1) $\Omega(\mathbf{X}^n) = \mathbb{E}\{\omega(\mathbf{x})\}$.
- (2) $\inf_{\mathbf{x}} \omega(\mathbf{x}) \leq \Omega(\mathbf{X}^n) \leq \sup_{\mathbf{x}} \omega(\mathbf{x})$.

Generally speaking, the local O-information provides more information than its global counterpart as, technically, $\omega(\mathbf{X}^n)$ is a random variable whose mean value is $\Omega(\mathbf{X}^n)$. Therefore, the whole range of values of $\omega(\mathbf{X}^n)$ can naturally provide a more fine-grained description of the system than its mere first moment.

Additionally, note that Lemma 3 of Ref. [25] provides upper and lower bounds for $\Omega(\mathbf{X})$ when the variables take values in a finite alphabet; in particular, if $X_j \in \mathcal{X}$ for all j , then

$$-(n - 2) \log |\mathcal{X}| \leq \Omega(\mathbf{X}^n) \leq (n - 2) \log |\mathcal{X}|. \quad (6)$$

Note that these bounds do not apply to the local O-information; they apply to its average value, but extreme values can be larger than it. Nonetheless, the quantity $(n - 2) \log |\mathcal{X}|$ establishes a natural bound which that, when surpassed, values of ω can be considered to be particularly large. This provides a useful rule of thumb to interpret the ranges of values obtained by evaluations of ω .

C. Proof of concept

A useful way of employing ω is to classify states among different *types*. In particular, building on the properties of the O-information, we say that a state \mathbf{x}^n for which $\omega(\mathbf{x}^n) > 0$ is redundancy dominated, while if $\omega(\mathbf{x}^n) < 0$ we say the state is synergy dominated. In this section we illustrate this capability

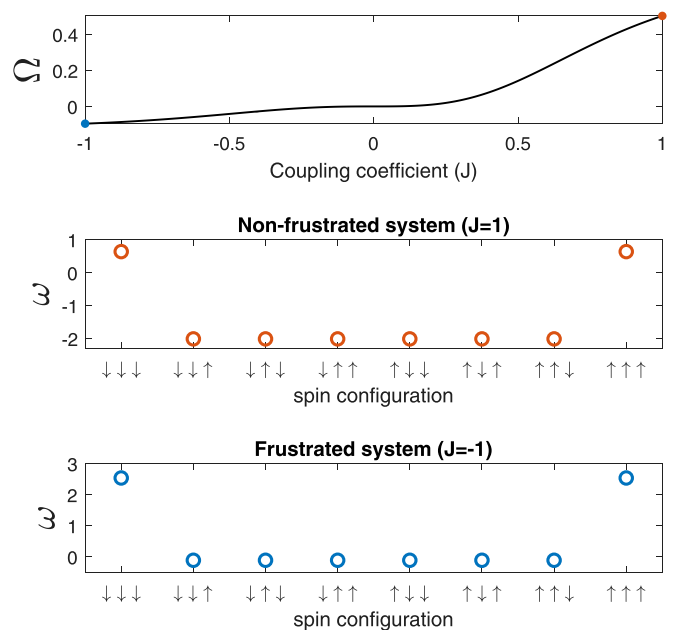


FIG. 1. Concerning the Ising toy model of three spins, Ω is plotted versus the coupling J (top panel). The values of ω for the eight possible configurations of the three spins are shown for the unfrustrated case when $J = 1$ (middle panel) and for the frustrated case, when $J = -1$ (bottom panel).

of the local O-information by using it to analyze a small Ising system.

Let us consider three coupled spins denoted by (S_1, S_2, S_3) , whose joint probability distributions follow a Boltzmann-Gibbs distribution of the form

$$p(s_1, s_2, s_3) = \frac{e^{J(s_1 s_2 + s_1 s_3 + s_2 s_3)}}{Z}, \quad (7)$$

where $Z = \sum_{s_1, s_2, s_3} e^{J(s_1 s_2 + s_1 s_3 + s_2 s_3)}$ is a normalization factor. For positive values of J , the configurations with all the spins in agreement (i.e., $\uparrow\uparrow\uparrow$ and $\downarrow\downarrow\downarrow$) satisfy all bonds, and hence the system may be seen as a small ferromagnet. In contrast, for negative J the system is said to be “frustrated” as there is not a configuration satisfying all bonds simultaneously.

By analyzing Eq. (7) via the O-information, one finds that ferromagnetic behavior is redundancy dominated while frustrated systems are synergy dominated (see Fig. 1). Interestingly, the local O-information shows that configurations of spin agreement are redundancy-dominated states, while the six configurations with disagreement are synergy dominated; this for both positive and negative values of J . This let us conclude that what makes the system redundancy or synergy dominated for different values of J is the different frequency with which either redundancy- or synergy-dominated configurations are visited.

This toy example shows how the local O-information can reveal different qualities of various states—in this case, either states with agreement or disagreement. Furthermore, this example also reveals an intriguing connection between synergy and frustration in spin models, which will be further investigated in a future publication.

III. CASE STUDY: HIGH-ORDER RELATIONSHIPS IN BACH'S CHORALES

To illustrate the usefulness of the local O-information for practical data analysis, this section presents a study of the multivariate statistics of musical scores from the Baroque period. Reference [25] provided an analysis of music scores based on the global O-information; however, such analyses could not provide information about individual chords, and hence could not explore further musical aspects related to harmony and tonality. Here we show how the local O-information can greatly expand this type of analyses, revealing subtle aspects of the musical discourse that are reflected in the high-order interactions.

In the following, Sec. III A describes the procedure to obtain and analyze the data, and Sec. III B discusses our main findings.

A. Processing pipeline

1. Data preprocessing

Our analysis focuses on the chorales for four voices (soprano, alto, tenor, and bass) written by J. S. Bach (1685–1750). These works are characterized by an elaborate counterpoint between the melodic lines that leads to rich harmonic progressions, which in turn results in a broad range of chords displayed along the repertoire. An additional point of interest of these pieces is that, as typical in the Baroque period (approximately 1600–1750), they display a balance in the interest and richness of each of the four voices. This contrasts with the subsequent Classic (1730–1820) and Romantic (1780–1910) periods, where higher voices tend to take the lead while the lower voices provide mere support.

Our analysis is based on the electronic scores publicly available [37], a website that hosts professionally curated digital scores [38]. Our preprocessing pipeline is the same as in Ref. [25], which we describe here for completeness. The scores were preprocessed in Python using the `Music21` package [39], which allowed us to select only the pieces written in Major mode. Each chorale was transposed to *C Major*, and each melodic line was transformed into a time series of 13 possible values (one for each note plus one for the silence), using a small rhythmic duration as common time unit. This resulted in a total of 172 chorales, which gave $\approx 4 \times 10^4$ four-note chords. It is worth remarking that our analyses focus on a specific portion of Bach's corpus—namely, his four-voice chorales—that provides a large and relatively uniform dataset. In particular, the style of Bach's chorales is known to not change significantly throughout his life.

With this data, the joint distribution of the values for the four-note chords was estimated using their empirical frequency [40]. This leads to a probability assigned to each four-note chord, which simply assesses the odds of picking that chord when randomly selecting one out of the whole repertoire. One can express this probability as the multivariate statistic $p(x_1, x_2, x_3, x_4)$, with each variable corresponding to the different voices.

Finally, we used $p(x_1, x_2, x_3, x_4)$ to calculate the local O-information $\omega(x)$ for each chord x using Eq. (5), which determines the dominant statistical behavior (in terms of synergy and redundancy) associated with each chord. The overall

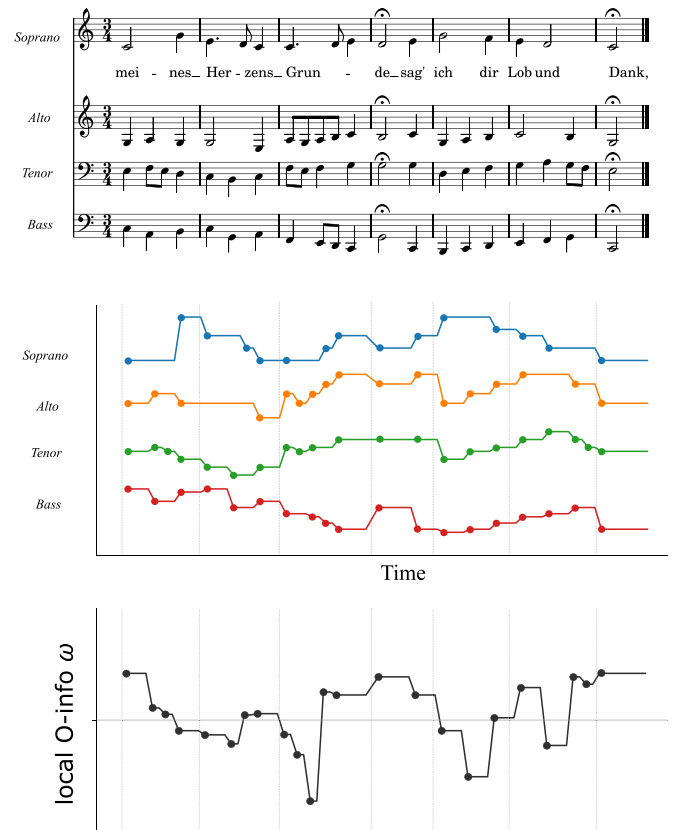


FIG. 2. Processing pipeline. Top: each chorale was selected among those written in major mode and then transposed in C major. Middle: each melodic line was transformed into a time series of 13 possible values. Bottom: for each four-note state the local O-information was calculated using Eq. (5).

pipeline starting from the music score and arriving to the local O-information is illustrated in Fig. 2.

2. Research questions and tools

We studied the multivariate properties of each of the possible four-note chords of Bach's chorales. Our analysis focuses exclusively on harmony and chords, leaving melodic and rhythmic properties to future studies. We focus on the question of what harmonic properties of the music tend to give rise to synergistic or redundant high-order relationships between the four voices [41].

Let us denote by $X = (X_1, X_2, X_3, X_4)$ the random vector that follows the statistics encapsulated by $p(x_1, x_2, x_3, x_4)$. Following standard musical practice, we follow the convention that the variables go from lower to higher range, so that X_1 corresponds to the bass and X_4 is the soprano. Moreover, we use the shorthand notation CEGE when referring to the chord $(x_1, x_2, x_3, x_4) = (C, E, G, E)$.

Note that X can adopt $13^4 = 28\,561$ possible values, and that p is generally not invariant under changes of ordering between the four voices. Since X_1, \dots, X_4 take values among alphabets of cardinality $|\mathcal{X}| = 13$, we do all calculations using logarithms to base 13, so that $H(X_k) \leq 1$ for all $k \in \{1, \dots, 4\}$ —we call this unit a *mut*, for *musical bit*. Equation (6) implies that $-2 < \Omega(X) < 2$, and hence most

values of ω are expected to have absolute value less than 2 muts—which gives a sense of how to interpret the magnitude of local O-information values.

We expected to find a correspondence between tonality and O-information values. In particular, we hypothesize that the principal tonal chords (C, F, and G major) would be associated with redundant behavior, while chords that are farther away from the tonal center (i.e., involve many sharp or flat alterations) would be related to synergistic events. Additionally, we expect dissonance to be associated with less redundancy, as it involves more complex combinations of notes.

B. Results

Here we report analyses that explore various aspects of musical phenomena from the perspective of the local O-information. In the following, Sec. III B 1 first explores properties of the most redundancy- and synergy-dominated chords, providing insights that are then refined in Sec. III B 2 by probing how ω is related with consonant and dissonant intervals. Climbing the ladder of musical complexity, Sec. III B 3 studies the relationship between ω and various aspects of harmony, and finally Sec. III B 4 investigates the role of lyrics.

1. Analysis of the extreme values of the local O-information

Out of the 13^4 possible chords, we found that only 1715 of them were observed at least once in the chorales, corresponding to only 6% of the possibilities, reflecting the specificity of the chord choices used in Bach’s chorales. A weak correlation is observed between frequency and local O-information: more frequent chords tend to have a higher ω , which suggests that more visited chords tend to be made by more redundant parts (see Fig. 3).

Some interesting observations can be made by observing the most positive (redundant) and the most negative (synergistic) states in terms of ω , which are presented in Table I. First, the most redundant states tend to contain few alterations (sharp notes, denoted in the table with the symbol \sharp) and mostly consonant intervals [42]. In contrast, synergistic chords tend to contain more alterations and dissonant intervals, which in the Western culture are typically associated with harshness and unpleasantness. For example, the most synergistic chord contains a major seconds (D-E), while the second most synergistic has one minor second ($F\sharp-G$) and one major second (E- $F\sharp$). The “chord” with highest local O-information is found to be RRRR, where the redundancy can be interpreted as a consequence of the voices doing the same thing—not signing.

2. The role of intervals

The results reported in the previous section imply a link between the musical (i.e., harmonic) properties of a chord and the type of statistical interdependencies among its constituent notes. In particular, results suggest that synergistic interdependencies may be associated with the presence of dissonances, while redundancy may be related to consonance. A consonant interval occurs when the ratio of the frequencies between two notes is very simple, such as (1:2) for the octave, (2:3) for the perfect fifth, or (4:5) for the major third. In the West-

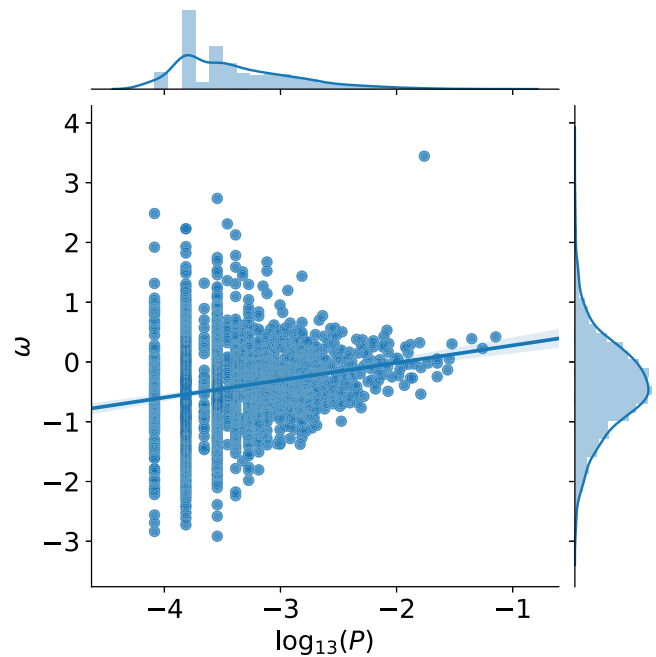


FIG. 3. Scatter plot of the logarithm of the probability of each state versus ω . Almost all the states have ω between -2 and $+2$, i.e., the bounds for Ω . The outlier in redundancy corresponds to the chord RRRR.

ern culture, consonance is typically associated by listeners with pleasantness and acceptability [43]. In Westerner music theory, dissonant intervals typically include the major second (8:9) and minor second (15:16), the major seventh (8:15) and minor seventh (9:16), and the augmented fourth (so-called “tritone” or *diabolus in musica*).

To further explore the relationship between high-order statistics and harmony, we studied how the local O-information depends on the number of dissonant intervals (either seconds/sevenths or augmented fourths) a chord possesses. Results are depicted in Fig. 4. An analysis of variance with Bonferroni correction revealed a significant dependency between the number of dissonances and ω for all number

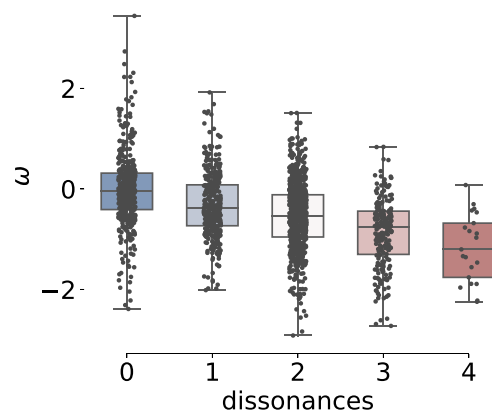


FIG. 4. Dissonance vs local O-information. Each box represents a state with a different number of dissonant intervals. Each category is statistically different from the one with no dissonance (pure consonant intervals).

TABLE I. Chords with the highest (redundance) and lowest (synergy) local O-information. Letters refer to the standard music nomenclature (plus R is for silence), and the ordering of the voices is Bass-Tenor-Alto-Soprano—from left to right.

Redundancy		Synergy	
Chord	ω	Chord	ω
R R R R	3.443	A E D D	-2.916
G D G D	2.736	G B F E	-2.836
F C F C	2.484	B F B B	-2.725
A C A C	2.311	A \sharp E E A	-2.688
C G C C	2.23	G F \sharp F \sharp A	-2.613
E G E G	2.228	G C B A	-2.581
C G C G	2.127	F A \sharp G F	-2.559
A A E A	1.93	G C C A \sharp	-2.522
F D G D	1.921	G E C \sharp A	-2.432
D D A A	1.824	G G G \sharp C	-2.396
G D G G	1.782	R G R E	-2.388
D D A D	1.748	G \sharp F G \sharp C	-2.311
D F C A	1.688	A \sharp F G \sharp C	-2.276
G G D G	1.674	G A F G	-2.245
F F C F	1.594	E G A F	-2.238
E C E C	1.586	E F \sharp C D	-2.221
A C A D	1.544	F \sharp F \sharp C \sharp A	-2.219
F F C D	1.532	G F F A \sharp	-2.185
R R R A	1.522	E A G D	-2.176
G F G D	1.512	C \sharp G G B	-2.173

of chords—except for the contrast between three and four dissonances, arguably due to the small number of chords with four dissonances. The results of all the comparisons are shown in Table II. The chords that contain three or more dissonant intervals are presented in Appendix D (Table VII), most of which exhibits negative values of ω while the only two chords with $\omega > 0$ can be identified as part of G major with added seventh, which is the most frequent dissonant chord in classical harmony.

A question that arises from the results shown in Fig. 4 is how can purely consonant chords be dominantly synergistic—as shown by the variance of the values of ω for zero dissonance. For example, the chord $\omega(\text{CEGC}) = 0.42$ is redundant while $\omega(\text{EGCC}) = -0.45$, being both C major chords

TABLE II. Statistical analysis of Fig. 4. Each row tests the null hypothesis that chords with a different number of dissonances have the same ω .

Dissonances	p-value	Cohen’s d
0-1	<0.0001	0.432
0-2	<0.0001	0.738
0-3	<0.0001	1.187
0-4	<0.0001	1.556
1-2	<0.0001	0.317
1-3	<0.0001	0.821
1-4	<0.0001	1.272
2-3	<0.0001	0.488
2-4	0.0004	0.924
3-4	0.5865	0.447

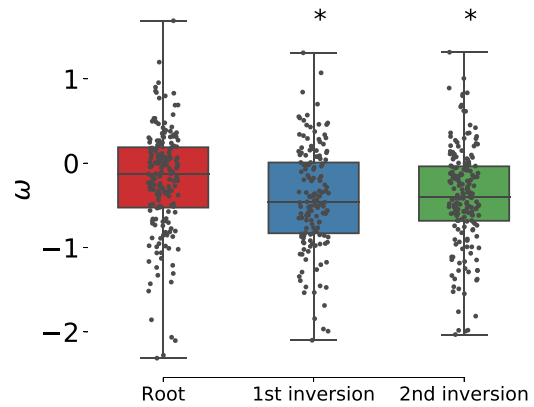


FIG. 5. The effect of chord inversions on the local O-information. Root form occurs when the root note of the triad is in the bottom (e.g., CEG), first inversion when the third note of the triad is the bass note (e.g., EGC), and second inversion when the fifth note is in the bass (e.g., GCE).

but having a different pitch in the bass (the fundamental note in the first case, and the third in the second). Leveraging music theory, a possible explanation from this can be built from the notion of “chord inversion”: a triad chord is in *first inversion* if the third (either major or minor) is in the bass, it is in *second inversion* if the fifth is in the bass, and it is in *root position* if the first/fundamental note is in the bass. In Western classical music each inversion tends to be associated with specific sensations; the first inversion gives a sense of lightness, while the second inversion and root position are typically associated with instability and stability, respectively.

Building on the above considerations, one can conjecture that the realizations of chord inversions may display a trend towards synergy-dominated statistics. By considering the values of ω corresponding to different inversions, a t test shows a tendency towards lower values of ω in chords in first (Cohen’s $d \simeq 0.36$) and second (Cohen’s $d \simeq 0.31$) inversion when compared to chords in root position, as shown in Fig. 5.

Finally, as a complementary way to study the role of intervals on the O-information, we considered the average value of ω for given notes at specific voices—averaging over all possible notes adopted by the other two voices. The results are shown in Fig. 6. It was found that redundancy (i.e., the most positive values of O-information) is “localised” in few intervals, taking place mainly between notes involving the tonic (C major, CEG) or dominant (G major, GBD) chords, or between silences. Also, most redundancy in the bass is associated to the fundamental note of each chord—either C or G. In contrast, synergy (i.e., the most negative values of O-information) are much more widespread. Interestingly, the redundancies between the two extreme voices (soprano and bass) are relatively weak (except between their silence), while synergies between them are not. Please note that the extreme voices tend to carry an important role in Bach chorales—the soprano carrying out the main melody, and the bass leading the harmony.

3. Harmonic depth

While the previous section focused on the role of single intervals, now our analyses focus on harmonic considerations.

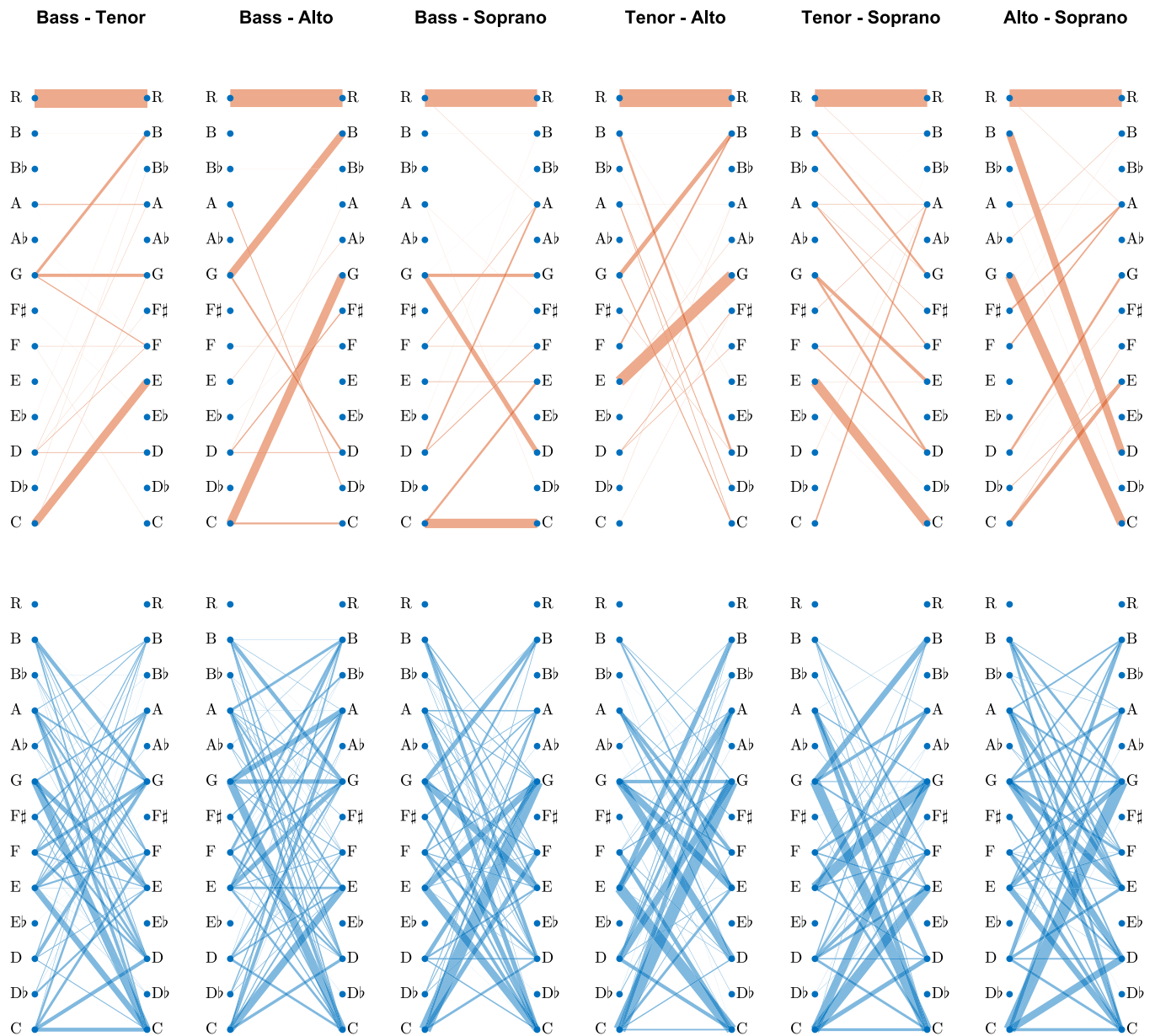


FIG. 6. Networks of redundant (orange) and synergy-dominated (blue) relationships in Bach’s chorales. For each pair of voices the link between two notes was created by averaging the local O-information of all chords containing those two notes.

Harmony is organized around a *tonality* (also called “key”), which plays the role of center of gravity around which music discourse revolves. The axial pitch of a tonality is called the *root*, which in turn gives name to the tonality, e.g., C is the root of the tonality of C major. In classical Western music there are 12 different major tonalities, one for each of each pitch. Also, each major tonality has an associated minor tonality, which is located a minor third below (e.g., C major is associated with A minor). Each of these 12 major tonalities is made of seven distinct pitches, and are naturally ordered by a notion of proximity depending on how many pitches they have in common. This gives rise to the *circle of fifths*: major keys separated by a fifth have only one note different. For example, C and G major are only distinguished by the note F, which is sharp for the latter but natural for the former.

A simple way to explore the impact of harmony on the high-order statistics is by analyzing the dependency between ω and the number of alterations (sharps or flats) that a chord has. In effect, please recall that all the chorales analyzed are in major mode, and have been shifted to C (see Sec. III A). Moreover, chords belonging to C major have no alterations, while chords for more distant tonalities have progressively more alterations—either sharps if going up the cycle of fifths, or flats otherwise. Therefore, we ran statistical analyses (*t* test corrected for multiple comparisons) on the effect of the number of alterations on ω , whose results are shown in Fig. 7. Results revealed significant decreases of ω ($p < 0.01$) for states with one (Cohen’s $d \simeq 0.41$) or two (Cohen’s $d \simeq 0.44$) alterations with respect to chords without any alterations.

In order to deepen our understanding of how harmony and O-information are related we introduced the notion of

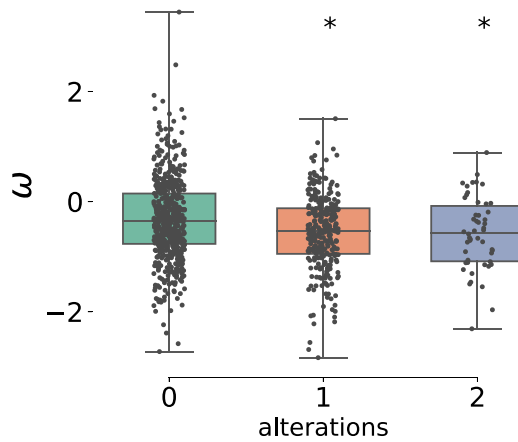


FIG. 7. The local O-information is related to the number of alterations (sharps and flat notes) inside each state. States with one and two alterations have ω significantly lower ($p < 0.05$) than states with no alterations.

harmonic depth. Harmonic depth corresponds to the smallest number of steps (going clockwise or anticlockwise) in the circle of fifths that are required to go from C major to a given tonality, or from A minor in the case of minor tonality. For example, the chord D major has an harmonic depth of +2, while the chord F minor (denoted as Fm) has a harmonic depth of +4.

We are interested to study the relationship between ω and harmonic depth. For this purpose, we consider the values of ω for triads that belong to a specific tonality, regardless of the arrangement of notes between the voices. For example, determining a chord is C just accounts for triads containing only the pitches C, E, G, regardless of how they are arranged among the voices. Results are shown in Fig. 8, and show that the value of ω decreases as soon as the tonality moves away from C major, being this difference is more pronounced in minor chords. This suggests that synergy may also be associated with more complex harmonic explorations involving more harmonically distant chords.

4. Lyric analysis

As a final step in our analysis, we investigated the relationship between the values of ω and the corresponding chords during which a given word is sung as part of the lyrics.

For this purpose, we consider the different values of ω that correspond to each time a given word is sung. In cases of melismas (i.e., when many notes are sung under the same syllable), the values of the whole progression were averaged and counted as one realization of the word.

As a first analysis, we calculated a word cloud where words associated with negative or positive values of ω are represented in red and blue, respectively. As shown in Fig. 9, many of the most common words (like *Gott*, *Herr*, *Sohn*) are redundant, with the exception of *Jesu*, which is synergistic. As the majority of the chords explored by Bach are synergistic (the average O-information is negative; see Fig. 3), this prevalence of redundant words is highly nontrivial.

Another insight that can be drawn from the word cloud is that words that are not the subject of the phrase seem to be

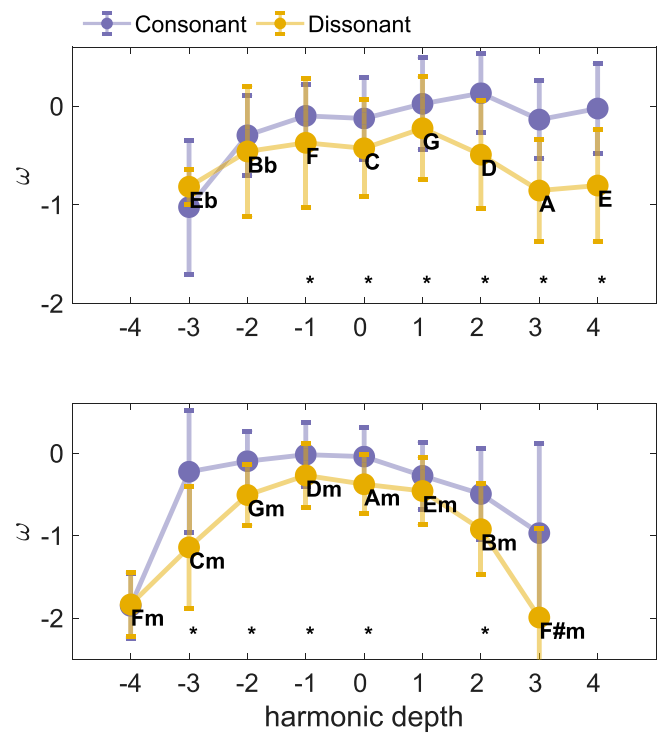


FIG. 8. The effect of harmonic distance. The figure shows the mean values (and confidence intervals) of ω for musical states in every different major (top) and minor (bottom) chord. The purple line indicates purely consonant triads, while the yellow line corresponds to chords with one or more dissonances.

more synergistic. To verify this, we evaluated the effect on ω of words being in root form (nominative case) with respect to all others [44]. The results confirmed our conjecture, showing that words in root form have a tendency towards higher values of ω (see Fig. 10). As a speculation, this may be interpreted by noting that root form words correspond to the most important part of the sentence, and hence a redundant underlying harmony might contribute to an easier comprehension. The most frequent cases of words for which we found both the root and the nonroot word are shown in Table III.

Further results are provided in Appendix D, where the most frequently encountered chords are shown in Table IV, while the ω extreme values and the most common tonal chords



FIG. 9. Word cloud of Bach's chorales lyrics. Most common words found in text, with their size representing their frequency and their color their local O-information sign (plus or minus).

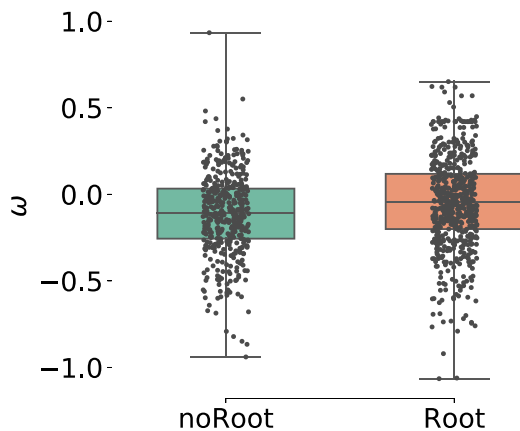


FIG. 10. Root vs nonroot words. Each box shows the distribution of local O-information ω for each category. A statistical comparison between the two populations has been carried out with a two sample t -test, which rejected the null hypothesis with $p < 0.01$; the effect size can be expressed as Cohen’s $d \simeq 0.26$.

TABLE III. Most common words and their average local O-information ω .

Root	ω	No Root	ω
Gott	0.008	Gottes	-0.233
		Gotte	-0.002
		Gotts	-0.256
Herr	0.096	Herre	-0.007
		Herren	-0.152
Christ	-0.127	Christen	-0.442
		Christe	-0.092
		Christenheit	-0.342
ewig	-0.159	Ewigkeit	-0.024
		ewiglich	-0.264
Geist	0.098	Geistern	-0.866
		soll	-0.111
		sollst	-0.444
Himmel	-0.188	solls	-0.009
		Himmels	-0.314
		Seele	-0.044
Gnad	0.084	Gnaden	0.052
		Gnade	0.066
Seel	-0.02	Seelen	0.083
		voll	-0.458
Mensch	-0.07	voller	-0.16
		Menschen	-0.224
		hoch	0.059
Ehr	-0.277	höchsten	-0.158
		Lichtes	-0.288
		Ehre	-0.107
Licht	0.137	ohne	-0.336
		ohne	-0.336
Ehr	-0.277	Tröster	-0.236
		ohne	-0.336
ohn	-0.214	Ende	0.074
		Tröster	-0.236
Trost	-0.286	Herzen	0.132
		Ende	0.074
End	0.106	Herzens	-0.149
		Herze	0.32
Herz	0.175		

are listed in Tables V and VI, respectively. Please note that an analysis done on time-shuffled surrogate data shows that bias correction is not necessary on this data (see Appendix C).

5. Music analysis based on frequencies

As a control, in Appendix E we investigated the possibility of retrieving similar results using just the frequencies of occurrences. For this purpose, we redo each analysis using p instead of ω versus the musical quantities of interest. The results of such analyses can be seen in Fig. 11, and Table VIII compares the statistical results obtained when studying the various musical properties in terms of the frequencies or ω . Overall, there are some associations that also replicate with the frequencies, but in general the significance is weaker and the effect sizes are smaller. This let us conclude that ω provides information about these aspects of music that cannot be retrieved from the frequency of occurrence.

IV. CONCLUSIONS

This paper introduces a framework to study the high-order interdependencies observed in complex multivariate systems, which is capable of disentangling their effects on individual patterns of activity. The approach is centered on the local O-information, a measure that quantifies the balance between redundancy and synergy at each pattern. Because of its information-theoretic nature, this measure is widely applicable, being suitable to assess systems with categorical, discrete, and continuous variables.

The capabilities of the proposed framework were showcased in an analysis of the scores of the chorales of J.S. Bach, which illuminated the high-order relationships that exist between the different voices. In particular, our results showed that synergy-dominated interdependencies tend to be associated with complex musical elements, including dissonances, chord inversions, and harmonic distance from the tonal center [45]. Taken together, our findings provide converging evidence about the relationship between statistical synergy and the complexity of the musical discourse.

These findings have interesting parallels with recent studies on the human brain, which are revealing a close relationship between synergistic interdependencies in neural activity and high cognitive functions. Historically, the notion of synergistic information was originated in theoretical neuroscience as an effort to characterize aspects of complex neural activity [6,7,34,46–48]. Moreover, recent empirical work has shown that synergy characterizes the interactions between areas of the human brain that have undergone a more pronounced evolutionary expansion [49], and is also associated with changes in brain function due to anaesthesia and disorders of consciousness [50]. Under the light of these findings, the results presented in this work let us speculate that elaborated musical discourse may have key similarities with the type of neural activity that underpins high brain functions. One possible commonality could be the presence of emergent phenomena, which have been recently characterized formally in terms of statistical synergy [51]. This would not be the first time music is shown to share some of the hallmark properties of complex systems; in effect, properties of musical discourse

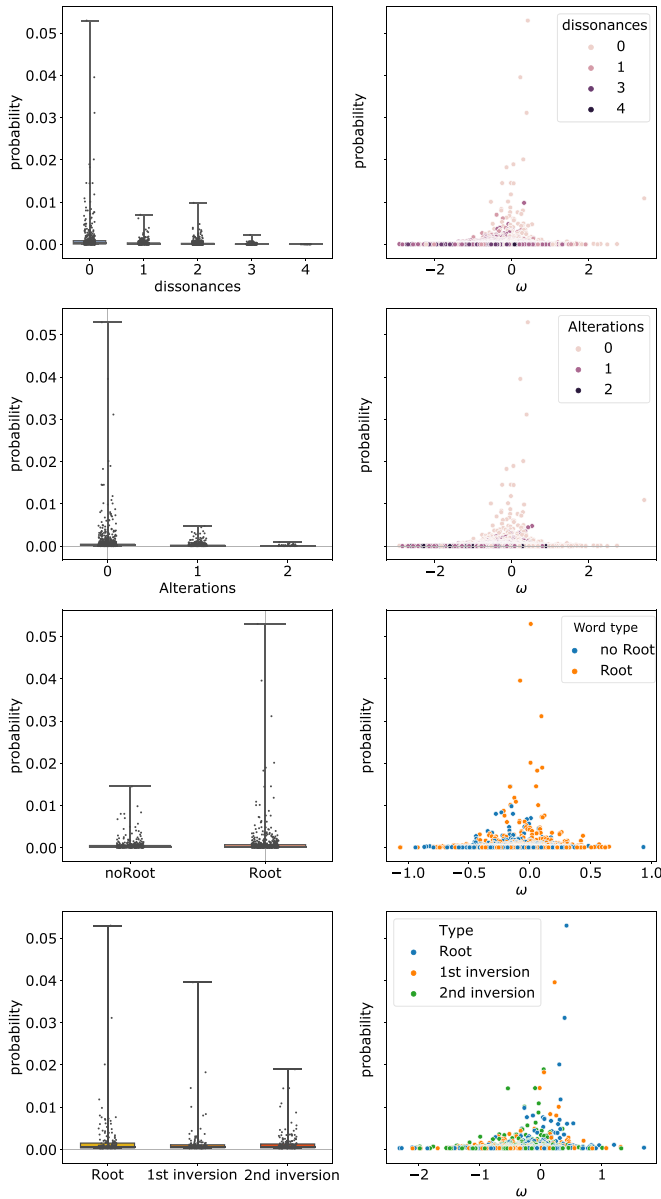


FIG. 11. Musical analysis using probabilities instead of local O-information ω .

have been shown to be related to nonlinear fluctuations and self-organized criticality [52,53], and also to entropy production and irreversibility [54].

Pointwise information measures, initially proposed in Ref. [29] with respect to the local transfer entropy, are a promising set of techniques that can provide a detailed description of information transfer mechanisms in complex systems. Recently this paradigm has been applied to implement the local Granger causality [30], which has shown interesting results on physiological and neural data. The fine descriptions allowed by the formalism introduced in this paper bring another perspective over high-order interdependencies, which complement existent pointwise information decomposition approaches (e.g., Ref. [16]) by being applicable to larger systems, hence greatly extending their domain of practical applicability. The extension of these ideas to dynamical in-

formation decomposition, such as the integrated information decomposition framework [55], constitutes a promising direction for future research.

ACKNOWLEDGMENTS

The authors thank Pablo Padilla and Alejandro Reyes for insightful discussions. F.E.R. was supported by the Ad Astra Chandaria Foundation. S.S. was supported by MIUR project PRIN 2017WZFTZP, “Stochastic forecasting in complex systems.”

Code for the analysis reported here is publicly available [56].

APPENDIX A: PROOF OF NON-NEGATIVITY OF TC AND DTC

The TC and DTC leverage relationships between three building blocks: the joint entropy [$H(X^n)$] and the sum of—conditioned or not—marginal entropies [$\sum_{i=1}^n H(X_i|X_{-i}^n)$ and $\sum_i H(X_i)$, respectively]. This Appendix provides proofs for

TABLE IV. Most common chords.

Chord	ω	Frequency
C E G C	0.4201	1881
C G C E	0.226	1405
G G B D	0.3887	1106
G B D G	0.3019	715
G B G D	0.0464	674
C C G E	-0.1328	649
G D B G	-0.0182	517
C G E C	-0.2457	516
A E A C	0.0517	514
G D G B	-0.0189	421
R R R R	3.4431	388
E G C G	-0.0938	361
C E C G	-0.5384	359
G F B D	0.3229	350
B D G D	-0.0342	311
E C G C	-0.2682	298
E G C E	0.2924	286
C C E G	-0.2694	286
B G D G	-0.0387	272
D A D F	0.1538	271
A A C F	0.2383	260
F C F A	0.2522	257
G G C D	-0.3909	250
F F C A	-0.058	241
F F A C	-0.1293	226
F A C D	-0.2281	220
D F A D	0.0358	217
A A C E	0.0617	214
A C E C	0.0578	203
A A E C	-0.2498	190
C E G E	0.307	177
F A C F	0.1666	175
G G D B	-0.491	174
B G D F	-0.0853	172
D D F# A	0.5305	170

TABLE V. First states with the highest (redundancy) and lowest (synergy) ω values. The letters refer to the standard nomenclature of notes in music (R is for rest) and the order of the voices is Bass-Tenor-Alto-Soprano, from left to right.

Redundancy		Synergy	
Chord	ω	Chord	ω
R R R R	3.443	A E D D	-2.916
G D G D	2.736	G B F# E	-2.836
F C F C	2.484	B F B B	-2.725
A C A C	2.311	A# E E A	-2.688
C G C C	2.23	G F# F# A	-2.613
E G E G	2.228	G C B A	-2.581
C G C G	2.127	F A# G F	-2.559
A A E A	1.93	G C C A#	-2.522
F D G D	1.921	G E C# A	-2.432
D D A A	1.824	G G G# C	-2.396
G D G G	1.782	R G R E	-2.388
D D A D	1.748	G# F G# C	-2.311
D F C A	1.688	A# F G# C	-2.276
G G D G	1.674	G A F G	-2.245
F F C F	1.594	E G A F	-2.238
E C E C	1.586	E F# C D	-2.221
A C A D	1.544	F# F# C# A	-2.219
F F C D	1.532	G F F A#	-2.185
R R R A	1.522	E A G D	-2.176
G F G D	1.512	C# G G B	-2.173
F G C D	1.508	G A C# A	-2.173
F# A C A	1.506	A D# A D	-2.144
B G B G	1.5	E E F# A	-2.143
C C G A	1.482	F# B E E	-2.142
F A A F	1.46	A G A# F	-2.115
G B G B	1.446	F A C# G	-2.114
C C G C	1.435	F B G# C	-2.103
G G D D	1.361	F# E B D	-2.099
G F A D	1.318	G E D A#	-2.09
C A D E	1.315	C# E F# B	-2.086

basic inequalities between these terms, which guarantee the non-negativity of the TC and DTC.

Let us first show that $H(\mathbf{X}^n) \leq \sum_i H(X_i)$. This inequality follows directly from the chain rule applied to the joint entropy, as follows:

$$H(\mathbf{X}^n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1) \leq \sum_{i=1}^n H(X_i),$$

where the inequality follows from the fact that conditioning cannot increase the entropy.

Let us now show that $H(\mathbf{X}^n) \geq \sum_{i=1}^n H(X_i | \mathbf{X}_{-i}^n)$. This can be proven using again the chain rule, but in a different manner:

$$\begin{aligned} H(\mathbf{X}^n) &= \sum_{i=1}^n H(X_i | X_1, \dots, X_{i-1}) \\ &\geq \sum_{i=1}^n H(X_i | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n). \end{aligned}$$

TABLE VI. Most common tonal chords.

Chord	ω	Occurrence
C E G C	0.42	1881
C G C E	0.226	1405
C C G E	-0.133	649
C G E C	-0.246	516
E G C G	-0.094	361
C E C G	-0.538	359
E C G C	-0.268	298
E G C E	0.292	286
C C E G	-0.269	286
C E G E	0.307	177
E G C C	-0.452	132
E C G G	-0.302	122
E C G E	0.015	113
E G G C	-0.577	102
E E G C	0.37	99
G G C E	0.149	90
E E C G	0.002	78
E G E C	-0.045	74
E C E G	0.093	71
E C C G	-0.467	52
C C E C	0.442	52
C C E E	0.033	44
C E C E	0.774	41
C G E G	-0.106	37
G G E C	-0.812	35
G E C E	-0.977	26
C C G C	1.435	26
E E G G	-0.16	25
C G E E	0.538	24
G E C C	-1.132	22

APPENDIX B: PROOF OF EQ. (5)

Here we provide an explicit derivation of the local O-information formula. By recalling the definitions of local total correlation and dual total correlation, and that $h(x_i) = -\log p(x_i)$, one can find that

$$\begin{aligned} \omega(\mathbf{x}^n) &:= \text{tc}(\mathbf{x}^n) - \text{dtc}(\mathbf{x}^n) \\ &= \left[\sum_{j=1}^n h(x_j) - h(\mathbf{x}^n) \right] \\ &\quad - \left[h(\mathbf{x}^n) - \sum_{j=1}^n h(x_j | \mathbf{x}_{-j}^n) \right] \\ &= \sum_{j=1}^n h(x_j) - 2h(\mathbf{x}^n) + \sum_{j=1}^n h(x_j | \mathbf{x}_{-j}^n) \\ &= -2h(\mathbf{x}^n) + \sum_{j=1}^n [h(x_j) + h(\mathbf{x}^n) - h(\mathbf{x}_{-j}^n)] \\ &= (n-2)h(\mathbf{x}^n) + \sum_{j=1}^n [h(x_j) - h(\mathbf{x}_{-j}^n)]. \end{aligned}$$

TABLE VII. Most common dissonant chords (only chords with three or more dissonant intervals are shown).

Chord	ω	Occurrence
G G B F	-0.2391	81
D D F# C	-0.4532	28
G F G B	-0.8748	27
C A# C E	-0.1148	27
D D C F#	-0.5139	26
G G F B	-0.8308	25
G F B G	0.226	25
G B G F	-0.7421	24
B A C F	-0.455	22
E F B D	-0.3088	22
D F# D C	-0.713	18
D C D F#	-0.7643	17
E D G# E	-0.262	17
B C A F	-0.7095	16
G B F G	0.2502	16
E F B G	-0.4072	14
C C E A#	-0.2633	14
C E D E	-0.7414	13
G F B C	-0.8746	12
D F C E	-0.5583	12
C D F# D	-1.0235	10
E E D G#	-0.7605	10
D C F# D	-0.1526	10
F B F A	-0.1821	9
B C F C	-1.8124	8
G F A# A	-1.5336	8
E D E G#	-0.7108	8
A E A# C	-0.6605	8
F G C# E	-0.5848	8
B C F A	-0.5059	8

TABLE VIII. Comparison between each musical analysis done using probability and O-information.

	Probability			O-information		
	p-value	Cohen's d	Significant	p-value	Cohen's d	Significant
Dissonances						
0-1	<0.0001	0.331	1	<0.0001	0.432	1
0-2	<0.0001	0.446	1	<0.0001	0.738	1
0-3	<0.0001	0.387	1	<0.0001	1.187	1
0-4	0.0395	0.338	0	<0.0001	1.556	1
1-2	1	0.198	0	<0.0001	0.318	1
1-3	1	0.482	0	<0.0001	0.821	1
1-4	1	0.503	0	<0.0001	1.272	1
2-3	1	0.274	0	<0.0001	0.489	1
2-4	1	0.372	0	0.0004	0.924	1
3-4	1	0.38	0	0.5865	0.447	0
Root inversion						
root-1st	0.7603	0.1069	0	0.0032	0.3562	1
root-2nd	0.4917	0.1443	0	0.0085	0.3139	1
1st-2nd	1	0.0267	0	1	0.047	0
Diesis						
0-1	<0.0001	0.2242	1	<0.0001	0.4061	1
0-2	0.0332	0.2463	0	<0.0001	0.4434	1
1-2	1	0.3109	0	1	0.0568	0
Words						
root - noRoot	0.0178	0.1493	0	<0.0001	0.2553	1

APPENDIX C: BIAS CORRECTION

In this work, we estimate all the information-theoretic quantities by first estimating the underlying joint probability of occurrences of four-note chords. As mentioned in Sec. III A 1, we estimate these probabilities directly from the frequency of occurrences, without employing regularization methods, as some chords are not representative of the Baroque repertoire and hence will never appear, irrespective of the sample size.

Estimators of information-theoretic quantities can suffer from bias if the sample size is small. To estimate the bias in our scenario, we constructed a null model by time shuffling the original series, which effectively destroys the interdependencies while preserving the marginal statistics of each of the four voices. By calculating the average of ω for multiple realizations of these shuffled series, results show that the values are usually of the order of 10^{-1} , being one order of magnitude smaller than the typical values of ω observed in our analyses (see Fig. 12). For this reason, adding a bias correction would not alter our results.

APPENDIX D: TABLES

In Tables IV–VIII, we provide a series of further results that complete the musical analysis presented in the text.

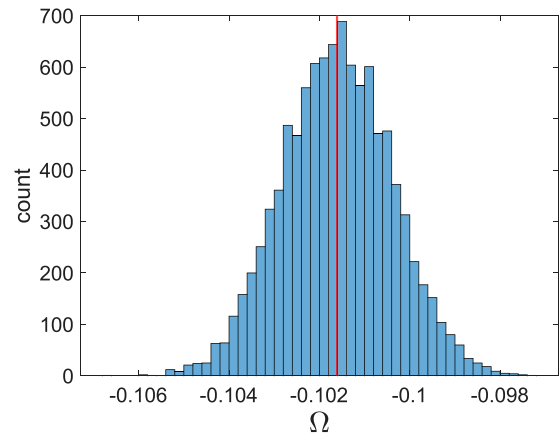


FIG. 12. Bias estimation for the musical analysis. Each realization of the null model was built by shuffling the original musical scores and then by calculating the values of ω as done in the text. The histogram shows the distribution of the average of ω for each realization, showing a bias of $\approx -10^{-1}$ —significantly smaller than the typical values of ω reported in our results.

APPENDIX E: COMPARISON WITH PROBABILITY

In Fig. 11, we present the results of each analysis performed in the text using the frequencies of occurrence instead of the local O-information.

[1] J. P. Crutchfield, The calculi of emergence: computation, dynamics and induction, *Phys. D (Amsterdam, Neth.)* **75**, 11 (1994).

[2] J. P. Crutchfield and D. P. Feldman, Regularities unseen, randomness observed: Levels of entropy convergence, *Chaos* **13**, 25 (2003).

[3] P. L. Williams and R. D. Beer, Nonnegative decomposition of multivariate information, [arXiv:1004.2515](https://arxiv.org/abs/1004.2515).

[4] V. Griffith, *Quantifying Synergistic Information*, Ph.D. thesis, California Institute of Technology, 2014.

[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] I. Gat and N. Tishby, Synergy and redundancy among brain cells of behaving monkeys, in *Advances in Neural Information Processing Systems 11, Proceedings of the 1998 Conference*, edited by Michael J. Kearns, Sara A. Solla, and David A. Cohn (MIT Press, Boston MA, 1999).

[7] E. Schneidman, W. Bialek, and M. J. Berry, Synergy, redundancy, and independence in population codes, *J. Neurosci.* **23**, 11539 (2003).

[8] N. Timme, W. Alford, B. Flecker, and J. M. Beggs, Synergy, redundancy, and multivariate information measures: an experimentalist’s perspective, *J. Comput. Neurosci.* **36**, 119 (2014).

[9] F. Rosas, V. Ntranos, C. J. Ellison, S. Pollin, and M. Verhelst, Understanding interdependency through complex information sharing, *Entropy* **18**, 38 (2016).

[10] M. Wibral, V. Priesemann, J. W. Kay, J. T. Lizier, and W. A. Phillips, Partial information decomposition as a unified approach to the specification of neural goal functions, *Brain Cogn.* **112**, 25 (2017).

[11] M. Harder, C. Salge, and D. Polani, Bivariate measure of redundant information, *Phys. Rev. E* **87**, 012130 (2013).

[12] Virgil Griffith, Edwin K. P. Chong, R. G. James, C. J. Ellison, and J. P. Crutchfield, Intersection information based on common randomness, *Entropy* **16**, 1985 (2014).

[13] A. B. Barrett, Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems, *Phys. Rev. E* **91**, 052802 (2015).

[14] E. Olbrich, N. Bertschinger, and J. Rauh, Information decomposition and synergy, *Entropy* **17**, 3501 (2015).

[15] R. A. A. Ince, The partial entropy decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal, [arXiv:1702.01591](https://arxiv.org/abs/1702.01591).

[16] C. Finn and J. T. Lizier, Pointwise partial information decomposition using the specificity and ambiguity lattices, *Entropy* **20**, 297 (2018).

[17] R. G. James, J. Emenheiser, and J. P. Crutchfield, Unique information via dependency constraints, *J. Phys. A: Math. Theor.* **52**, 014002 (2019).

[18] R. G. James, J. Emenheiser, and J. P. Crutchfield, Unique information and secret key agreement, *Entropy* **21**, 12 (2019).

[19] N. Ay, D. Polani, and N. Virgo, Information decomposition based on cooperative game theory, [arXiv:1910.05979](https://arxiv.org/abs/1910.05979).

[20] F. E. Rosas, P. A. M. Mediano, B. Rassouli, and A. B. Barrett, An operational information decomposition via synergistic disclosure, *J. Phys. A: Math. Theor.* **53**, 485001 (2020).

[21] C. Finn and J. T. Lizier, Generalised measures of multivariate information content, *Entropy* **22**, 216 (2020).

- [22] A. J. Gutknecht, M. Wibrál, and A. Makkeh, Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic, *Proc. R. Soc. A* **477**, 20210110 (2021).
- [23] K. Schick-Poland, A. Makkeh, A. J. Gutknecht, P. Wollstadt, A. Sturm, and M. Wibrál, A partial information decomposition for discrete and continuous variables, [arXiv:2106.12393](https://arxiv.org/abs/2106.12393).
- [24] A. Makkeh, A. J. Gutknecht, and M. Wibrál, Introducing a differentiable measure of pointwise shared information, *Phys. Rev. E* **103**, 032149 (2021).
- [25] F. E. Rosas, P. A. M. Mediano, M. Gastpar, and H. J. Jensen, Quantifying high-order interdependencies via multivariate extensions of the mutual information, *Phys. Rev. E* **100**, 032305 (2019).
- [26] S. Stramaglia, T. Scagliarini, B. C. Daniels, and D. Marinazzo, Quantifying dynamical high-order interdependencies from the O-information: An application to neural spiking dynamics, *Front. Physiol.* **11**, 1784 (2021).
- [27] M. Gatica, R. Cofré, P. A. M. Mediano, F. E. Rosas, P. Orió, I. Diez, S. P. Swinnen, and J. M. Cortes, High-order interdependencies in the aging brain, *Brain Connectivity* (Mary Ann Liebert, Inc., New York, 2021).
- [28] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, Transfer entropy, *An Introduction to Transfer Entropy* (Springer, New York, 2016), pp. 65–95.
- [29] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Local information transfer as a spatiotemporal filter for complex systems, *Phys. Rev. E* **77**, 026110 (2008).
- [30] S. Stramaglia, T. Scagliarini, Y. Antonacci, and L. Faes, Local Granger causality, *Phys. Rev. E* **103**, L020102 (2021).
- [31] S. Watanabe, Information theoretical analysis of multivariate correlation, *IBM J. Res. Dev.* **4**, 66 (1960).
- [32] H. Te Sun, Nonnegative entropy measures of multivariate symmetric correlations, *Inform. Control* **36**, 133 (1978).
- [33] In the context of this study, “statistical constraints” refer to conditions encoded in the joint probability distribution related to some states being not frequented as often as others. If one particular state has probability zero, that implies a hard constraint (i.e., the system cannot adopt that state), while if the probability is nonzero but low, this implies a soft constraint (i.e., the system tends to not adopt that state often).
- [34] G. Tononi, O. Sporns, and G. M. Edelman, A measure for brain complexity: relating functional segregation and integration in the nervous system, *Proc. Natl. Acad. Sci. USA* **91**, 5033 (1994).
- [35] For a comparison between the original measure proposed in Ref. [34] and the O-information, please see Ref. [25].
- [36] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications* (MIT, Cambridge, MA, 1968).
- [37] <http://kern.ccarh.org>.
- [38] C. S. Sapp, Online database of scores in the humdrum file format, in ISMIR, 2005 (unpublished), pp. 664–665.
- [39] <http://web.mit.edu/music21>.
- [40] While low-order constraints impose strong restrictions on the system and allow little amount of shared information between variables, high-order constraints impose collective restrictions that enable large amounts of shared randomness.
- [41] Please note that our results establish associations but not causal relationships, in neither direction, between musical and statistical phenomena. The possibility of such causal relationships is of great interest, but their proper assessment may require substantial further work.
- [42] Please note that regularization methods, such as Laplace smoothing, can have significant effects on the results. We decided not to use such methods, as some chords (e.g., C-C♯-D-D♯) are not representative of the Baroque repertoire.
- [43] I. Lahdelma and T. Eerola, Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension, *Sci. Rep.* **10**, 8693 (2020).
- [44] For a description of consonant and dissonant intervals, please see the next section.
- [45] German language has four cases: nominative (subject), accusative (direct object), dative (indirect object), and genitive (possessive).
- [46] G. Tononi, G. M. Edelman, and O. Sporns, Complexity and coherency: integrating information in the brain, *Trends Cognit. Sci.* **2**, 474 (1998).
- [47] P. E. Latham and S. Nirenberg, Synergy, redundancy, and independence in population codes, revisited, *J. Neurosci.* **25**, 5195 (2005).
- [48] E. Ganmor, R. Segev, and E. Schneidman, Sparse low-order interaction network underlies a highly correlated and learnable neural population code, *Proc. Natl. Acad. Sci. USA* **108**, 9679 (2011).
- [49] A. I. Luppi, P. A. M. 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, *BioRxiv*, 2020, Cold Spring Harbor Laboratory, <https://doi.org/10.1101/2020.09.22.308981>.
- [50] A. I. Luppi, P. A. M. Mediano, F. E. Rosas, J. Allanson, J. D. Pickard, R. L. Carhart-Harris, G. B. Williams, M. M. Craig, P. Finoia, A. M. Owen, L. Naci, D. K. Menon, D. Bor, and E. A. Stamatakis, A synergistic workspace for human consciousness revealed by integrated information decomposition, *BioRxiv*, 2020, Cold Spring Harbor Laboratory, <https://doi.org/10.1101/2020.11.25.398081>.
- [51] F. E. Rosas, P. A. M. 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).
- [52] L. Telesca and M. Lovallo, Revealing competitive behaviours in music by means of the multifractal detrended fluctuation analysis: application to Bach’s Sinfonias, *Proc. R. Soc. London, Ser. A* **467**, 3022 (2011).
- [53] A. González-Espinoza, H. Larralde, G. Martínez-Mekler, and M. Müller, Multiple scaling behaviour and nonlinear traits in music scores, *R. Soc. Open Sci.* **4**, 171282 (2017).
- [54] A. González-Espinoza, G. Martínez-Mekler, and L. Lacasa, Arrow of time across five centuries of classical music, *Phys. Rev. Res.* **2**, 033166 (2020).
- [55] P. A. M. Mediano, F. Rosas, R. L. Carhart-Harris, A. K. Seth, and A. B. Barrett, Beyond integrated information: A taxonomy of information dynamics phenomena, [arXiv:1909.02297](https://arxiv.org/abs/1909.02297).
- [56] https://github.com/tomscag/local_O_information.