


Article

# A Synergistic Perspective on Multivariate Computation and Causality in Complex Systems

Thomas F. Varley 

Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405, USA; tfvarley@uvm.edu

**Abstract:** What does it mean for a complex system to “compute” or perform “computations”? Intuitively, we can understand complex “computation” as occurring when a system’s state is a function of multiple inputs (potentially including its own past state). Here, we discuss how computational processes in complex systems can be generally studied using the concept of statistical synergy, which is information about an output that can *only* be learned when the joint state of all inputs is known. Building on prior work, we show that this approach naturally leads to a link between multivariate information theory and topics in causal inference, specifically, the phenomenon of causal colliders. We begin by showing how Berkson’s paradox implies a higher-order, synergistic interaction between multidimensional inputs and outputs. We then discuss how causal structure learning can refine and orient analyses of synergies in empirical data, and when empirical synergies meaningfully reflect computation versus when they may be spurious. We end by proposing that this conceptual link between synergy, causal colliders, and computation can serve as a foundation on which to build a mathematically rich general theory of computation in complex systems.

**Keywords:** Berkson’s paradox; synergy; multivariate information theory; higher-order interactions; partial information decomposition



**Citation:** Varley, T.F. A Synergistic Perspective on Multivariate Computation and Causality in Complex Systems. *Entropy* **2024**, *26*, 883. <https://doi.org/10.3390/e26100883>

Academic Editors: Jiang Zhang, Peng Cui and Hector Zenil

Received: 15 August 2024

Revised: 17 October 2024

Accepted: 18 October 2024

Published: 21 October 2024



**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

What does it mean to say that a complex system, such as the brain, “computes”? The metaphor of computation is a canonical one throughout the field of complex systems [1,2] although unlike digital and analog computers designed by engineers, complex systems are self-organizing, noisy, and often continuous. While early cyberneticians made use of concepts like Turing machines and Boolean logic gates as analogies [3], as the field has progressed, it has become increasingly clear that complex systems are not merely elaborate Turing machines. Formal theories that rely on mathematical notions of well-defined, computable functions are often too rigid for the noisy, multifaceted, ill-defined, and non-deterministic world of complex systems. As a result, like many other topics in the field of complex systems, applying the concept of computation often relies on the intuition of researchers rather than objective quantification.

Despite this, it is also clear that many systems *do* seem to do something very much like the most general theories of computation. Consider a neuron: it is well known that neurons receive inputs from pre-synaptic parent neurons (in the form of excitatory and inhibitory neurotransmitters) and account for those inputs when “deciding” (computing) their behavior in the immediate future (whether to fire an action potential or not). Here, we suggest a relaxed and pragmatic definition of computation, in the spirit of the simple mapping account [4,5], that incorporates these basic features. In the context of complex systems, we operationalize computation as occurring when a system, or subset of a system, accounts for the state of multiple exogenous or endogenous inputs when deciding its next state. While this definition is very general, this may ultimately be a virtue; it does not require that the process of accounting for inputs be mathematically well defined, nor does it assume a discrete, finite state machine, and it works just as well for continuous systems

as discrete ones. For instance, an economist might say that a business computes the optimal price of its products as some nebulous function of inputs (costs of material, competitors' pricing, etc).

The specific mechanisms and physical implementations of these computations are all different, but they share a common high-level feature: information about the states of many elements, possibly including the past state of the computer itself, is integrated to "decide" on a given next state, out of some space of possibilities. We might call this the "integrated information" account of computation (note that "integrated information" in this context is not necessarily the same as integrated information theory, which is a proposed theory of consciousness).

Here, we show that this definition of computation is particularly useful in the context of multivariate information theory [6–8], which concerns itself with the interactions between three or more variables (higher-order interactions). For such an account to be useful, particularly one as general as this, it would be useful if it pointed to some analytic tools that scientists could use to try and understand the computational structure of whatever system they are interested in. Based on the integrated information account of computation, we would expect that an observer attempting to model a putative computational process would find that predicting the output requires knowing the joint state of all of the inputs simultaneously. This notion is formalized by the concept of *synergy*: information about a target that is only disclosed by the joint state of all of the inputs and no simpler combination of sources [9].

Much of this work is inspired by the developments of Lizier et al. [10], who proposed that "non-trivial" computation could be recognized by multivariate dependencies that were "greater than the sum of their parts". This approach was further refined by Lizier, Flecker, and Williams in 2013 [11], who combined their method with the recently proposed approach using the partial information decomposition (explained in Section 2.2.1) as a principled measure of synergy. Subsequently, Timme et al. proposed using synergy as a measure of computation in neocortical circuits [12], an approach that has led to a series of studies in the area of computational neuroscience [13–16]. While highly productive, many instances of this approach have relied on intuitive arguments relating synergy to computation, and the formal connections have been largely unexplored. In this paper, we propose to refine the mathematical links between synergy and the integrated information account of computation, showing that the two can be linked to another set of fundamental concepts in complex systems science: collider bias and causal inference.

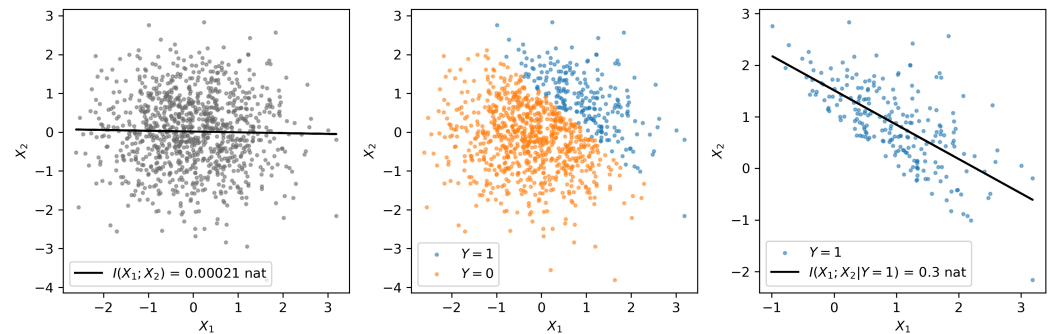
## 2. Berkson's Paradox, Colliders, and "Computation"

Berkson's paradox is one of the most well-known veridical paradoxes in statistics and is often used to demonstrate the effects of non-random sampling of data [17]. The usual presentation is as follows:

Alice is a social psychologist who is interested in how desirable traits (intelligence, attractiveness, etc.) are distributed over the U.S. population. Her data include information from a large number of Hollywood actors (a convenience sample; talent scores can be inferred from movie reviews and attractiveness from tabloids). Using movie reviews and tabloids, every actor is assigned a score for both talent and attractiveness. When Alice plots acting talent against attractiveness, she finds a statistically significant negative correlation. With a  $p$ -value  $\leq 0.05$  in hand, Alice rushes off to publish a paper about how attractiveness and acting talent are anticorrelated and then writes a grant proposing a research program into how the genes that code for acting talent and facial attractiveness are antagonistic.

When phrased in this somewhat cartoonish way, Alice's mistake is fairly obvious: she has an extremely biased sample (indeed, Berkson's paradox is sometimes called selection bias [18]). Hollywood celebrity is fickle; however, it is generally true that to reach the status of celebrity, one must be either unusually attractive or unusually talented. Ugly and untalented people do not (typically) make it to the silver screen, and so Alice's data are

fatally compromised from the outset. Thinking in terms of computation, we might say that “the media system” is “computing” whether a person reaches some threshold baseline of joint talent and attractiveness to be allowed onto the red carpet (and, by extension, onto Alice’s list). For a geometric visualization of Berkson’s paradox, see Figure 1. In the next section, we will discuss the usual interpretation of Berkson’s paradox in the context of collider bias and causal inference, before moving on to an alternative interpretation that relates collider bias and Berkson’s paradox to higher-order information sharing and synergy.



**Figure 1.** **Left:** Two random, normally distributed variables  $X_1$  and  $X_2$  with zero mean, unit variance, and  $X_1 \perp X_2$ . There is no mutual information between them. In the example of movie stars,  $X_1$  and  $X_2$  might be analogous to attractiveness and talent. **Center:** we can define a third random variable as a threshold function:  $Y = (X_1 + X_2) > 1$ , analogous to the “selection” criteria that define who will be a movie star as a function of whether they reach some minimal threshold of attractiveness and/or talent. **Right:** If we then compute the mutual information between  $X_1$  and  $X_2$  conditional on  $Y = 1$ , we see a spurious correlation, suggesting that  $X_1$  and  $X_2$  are anticorrelated, when they are actually independent. This is Berkson’s paradox.

### 2.1. The Usual Interpretation: Collider Bias

Berkson’s paradox appears to be about two variables (such as attractiveness and acting talent) but is, in fact, actually about a higher-order interaction between three variables (one of which is “hidden”). This third variable is whether a given member of the population is selected to be in Alice’s dataset or not. Importantly, the state of this hidden variable,  $Y$ , is itself a function of the two other variables we are interested in. Both talent and attractiveness causally influence an actor’s success. In the case of Alice’s study, the inclusion criteria could be a threshold function that computes whether someone is talented enough or attractive enough to make it to Hollywood stardom.

Written out in the language of causal diagrams (directed, acyclic graphs, or DAGs), we can represent our system of three variables as a *collider*:

$$X_1 \rightarrow Y \leftarrow X_2 \quad (1)$$

Colliders are something of a bugbear in modern causal inference and multivariate statistics [19]. We will discuss the problem of causal structure inference in more detail below, but briefly, when attempting to understand the structure of a system, the presence of colliders makes it difficult to understand when there is real dependence between two variables, or if they are actually (“causally”) independent and the empirical dependence is illusory. Much has been written on the problem of collider bias, but from an intuitive perspective, the source of the bias is fairly simple. When we sample data from a system with a collider, we are unwittingly creating an unbiased sample. If we then try and learn correlations between features, we are actually computing the conditional mutual information  $I(X_1; X_2|Y)$  when we might believe that we are computing the bivariate mutual information  $I(X_1; X_2)$ . The subtlety emerges when we realize that conditioning the mutual information is not a purely subtractive operation (which it is commonly believed to be). It is entirely possible for the conditional mutual information to be greater than

the unconditioned mutual information. In the language of causal diagrams, Pearl might say that conditioning on the collider opens a path between  $X_1$  and  $X_2$  [20]: an illusion of dependency has been created, one that may not be there in the original dataset.

## 2.2. The Synergistic Interpretation

In the example of Alice and the movie stars, the two input variables (talent and attractiveness) are a priori independent (i.e.,  $I(X_1; X_2) = 0$  bit), but when conditioned on the third variable (whether they are attractive enough, or talented enough, to make it in Hollywood), a significant dependency appears (i.e.,  $I(X_1; X_2|Y) \gg 0$  bit). A very natural question to ask might be how different is the true dependency between  $X_1$  and  $X_2$  from the conditional dependency? We can compute the simple difference as follows:  $I(X_1; X_2) - I(X_1; X_2|Y)$ .

For readers already familiar with information theory, this relationship should be recognizable. It defines the *co-information* [21], one of three proposed multivariate generalizations of the classic, bivariate, Shannon mutual information:

$$\begin{aligned} Co(X_1; X_2; Y) &= I(X_1; X_2) - I(X_1; X_2|Y) \\ &= I(X_1; Y) - I(X_1; Y|X_2) \\ &= I(X_2; Y) - I(X_2; Y|X_1) \end{aligned} \quad (2)$$

As a generalized mutual information, co-information is interesting because it does not obey the non-negativity of either the bivariate mutual information or the two other generalizations (both the total correlation [22] and the dual total correlation [23] are strictly non-negative). Let us first imagine the case where  $Co(X_1; X_2; Y) > 0$  bit. If the co-information is positive, then  $I(X_1; X_2) > I(X_1; X_2|Y)$ . In this case, learning the state of  $Y$  “explains away” the dependency between  $X_1$  and  $X_2$ . There must be some uncertainty that is “triply redundant” between all three variables, such that learning the state of any individual variable accounts for correlations between the other two (consider the symmetry of the three equations in Equation (2)). The case of the negative co-information is a little bit harder. Clearly, this occurs when  $I(X_1; X_2) < I(X_1; X_2|Y)$ , which implies that learning  $Y$  somehow “illuminates” (or perhaps “hallucinates”) a dependency between  $X_1$  and  $X_2$  that is not visible when the two are considered on their own. Interpreting this phenomenon (which is widespread in empirical data [24]) has been a standing puzzle in information theory since co-information was introduced in the 1950s. Various information theorists have proposed a variety of interpretations and decompositions, including in terms of joint and conditional mutual information, sums and differences of entropies [25], or lattice structures [26]. However, it was not until the development of the partial information decomposition (PID) by Williams and Beer in 2010 that the negativity of the co-information was finally interpreted in terms of purely higher-order interactions [9].

The PID (explained below) reveals that the co-information is negative (and by extension, Berkson’s paradox occurs) when there is information about  $Y$  that is only learnable when the state of the rest of the system (i.e.,  $X_1$  and  $X_2$ ) is known jointly. We will argue below that, in the context of a collider where  $Y$  is causally influenced by both  $X_1$  and  $X_2$ , this is indicative of a “computational process” happening in  $Y$ , where multiple streams of information are integrated as part of  $Y$ ’s “choice” of what state to adopt.

### 2.2.1. Explaining Partial Information Decomposition

The partial information decomposition (PID) is a framework for assessing how the interaction of many “input” variables reveals information about the state of a target variable [9,27]. As previously mentioned, the PID was not the first attempt to study different orders of information sharing in the case of multiple sources and a single target (see [25] for an early example), but the PID is unique in that it revealed the entire structure of multivariate information and formalized the notion of multiple “kinds” of higher-order

interaction. For two inputs and the target, it is straightforward to compute the joint mutual information that both sources provide about the target as follows:

$$I(X_1, X_2; Y) = H(Y) - H(Y|X_1, X_2) \tag{3}$$

However, this scalar value does not reveal the structure of the information: What information about  $Y$  could only be learned by observing  $X_1$ ? Or only by observing  $X_2$ ? What information could be learned by observing either  $X_1$  alone or  $X_2$  alone? What “synergistic” information can only be learned by observing  $X_1$  and  $X_2$  jointly (and no simpler combination of elements)? The PID was proposed in a seminal paper by Williams and Beer [9] and decomposes the joint mutual information into a set of non-overlapping “partial information atoms”:

$$I(X_1, X_2; Y) = Red(X_1, X_2; Y) + Syn(X_1, X_2; Y) + \tag{4}$$

$$Unq(X_1; Y/X_2) + Unq(X_2; Y/X_1) \tag{5}$$

In the two-variable case, these four terms describe every possible dependency between inputs and the target. The redundant information  $Red(X_1, X_2; Y)$  is the information that is duplicated over  $X_1$  and  $X_2$  such that it could be learned by observing either  $X_1$  alone or  $X_2$  alone. The unique information term  $Unq(X_i; Y/X_j)$  is the information about  $Y$  that could only be learned by observing  $X_i$ . Finally, the synergistic information  $Syn(X_1, X_2; Y)$  is the information about  $Y$  that can only be learned when the states of  $X_1$  and  $X_2$  are observed together and, crucially, cannot be learned by any simpler set of sources.

The marginals’ mutual information can be similarly decomposed as follows:

$$I(X_1; Y) = Red(X_1, X_2; Y) + Unq(X_1; Y/X_2) \tag{6}$$

$$I(X_2; Y) = Red(X_1, X_2; Y) + Unq(X_2; Y/X_1) \tag{7}$$

Note here that context is important: since we are analyzing the bivariate relationship  $X_i - Y$  in the context of the collider  $\{X_1, X_2\} \rightarrow Y$ , the redundant information  $Red(X_1, X_2; Y)$  still counts toward  $I(X_i; Y)$ . For a visualization of the PID using Venn diagrams, see Figure 2.

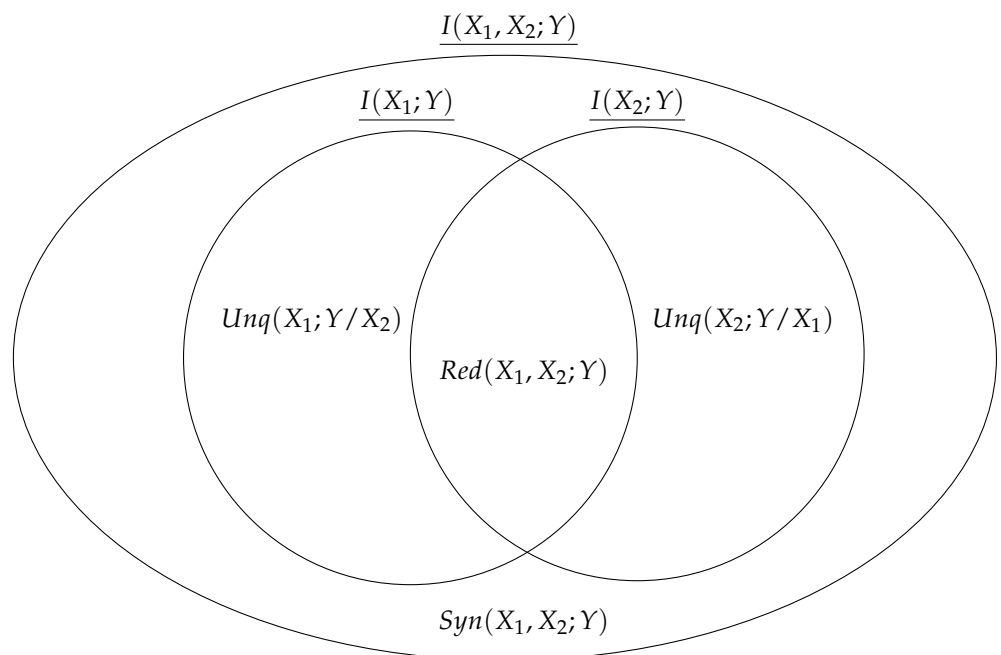


Figure 2. Visual intuition behind the bivariate PID.

There remains an active debate about the best way to define redundancy. Unlike the Shannon entropy, which is uniquely specified by Shannon's original axioms, many different possible redundancy functions can be shown to satisfy the initial Williams and Beer axioms. These functions can return meaningfully different results (for further discussion, see [28,29]); thus, in any given analysis of empirical data, care should be taken when choosing one or another. By and large, however, the arguments made in this paper do not require operationalizing a specific measure. In cases where we want to compute a PID for didactic purposes (see Section 2.3), we use the minimum mutual information [30]:

$$Red(X_1, X_2; Y) := \min_i I(X_i; Y) \quad (8)$$

Intuitively, we can understand the minimum mutual information and estimate how much uncertainty, on average, about  $Y$  would be resolved regardless of whether we learn  $X_1$  alone or  $X_2$  alone. Once again, this is not the only viable measure and has a number of limitations; however, it is also conceptually the simplest, which suits our purposes here. One significant benefit of this redundancy function is that it is known to be analytically correct for multivariate Gaussian variables [31] and has the nice property that it only depends on the pairwise, marginal relationship between  $X_i$  and  $Y$  (for more, see [32,33]).

### 2.2.2. Decomposing Conditional Mutual Information and Co-Information

With the PID established, it is possible to gain new insights into measures like co-information and conditional mutual information. For example, consider the following identity:

$$I(X_1; X_2|Y) = I(X_1; X_2, Y) - I(X_1; Y) \quad (9)$$

If we substitute in Equations (4) and (6), we find

$$I(X_1; X_2|Y) = Unq(X_1; X_2/Y) + Syn(X_1, Y; X_2) \quad (10)$$

$$= Unq(X_1; X_2/Y) + Syn(X_2, Y; X_1) \quad (11)$$

where  $Unq(X_1; X_2/Y)$  refers to the unique information about  $X_2$  disclosed by  $X_1$  in the context of  $Y$ .

We can see that the effect of conditioning on the collider is to make visible the higher-order interaction between all three variables:  $Syn(X_1, Y; X_2)$  is the synergistic information about  $X_2$  that is only learnable when  $X_1$  and  $Y$  are known together (and likewise for  $Syn(X_2, Y; X_1)$ ). This provides the crux of why conditioning on a collider can cause dependencies to appear: the observer suddenly gains access to higher-order information (in the form of the synergistic dependency) that they can only "see" when only considering  $X_1$ ,  $X_2$ , and  $Y$  together.

We can further refine this intuition by considering the following co-information:

$$Co(X_1; X_2; Y) = I(X_1; X_2) - I(X_1; X_2|Y) \quad (12)$$

Decomposing the co-information with respect to the target,  $Y$  reveals that

$$Co(X_1; X_2; Y) = Red(X_1, X_2; Y) - Syn(X_1, X_2; Y) \quad (13)$$

We can see that, whenever Berkson's paradox occurs and  $Co(X_1; X_2; Y) < 0$ , synergy outweighs redundancy. Said differently, information about the state of  $Y$  is dominated by an "integrated" interaction between multiple sources.  $Y$  is "choosing" its state as a function of multiple inputs. Accordingly, we say that synergy is reflecting this high-order computation. This link between Berkson's paradox and synergy was first noted by Rosas et al. [34] in an excellent study of higher-order information sharing. Here, we elaborate on this crucial insight by extending it using the PID, as well as discussing how it may be applied in empirical contexts.

### 2.3. Case Study I: Summing Dice

To demonstrate the proposed link between synergy and computation, it helps to go through a simple example. Here, we have (arguably) the simplest “computation” possible: addition. To model it, consider two fair, six-sided dice,  $D_1$  and  $D_2$ , with  $D_1 \perp D_2$ . We will create a third variable  $S$ , such that  $S = D_1 + D_2$ . We hope it is not controversial to say that  $S$  implements a computation (in this case, addition). While the distributions of  $D_1$  and  $D_2$  are maximally entropic, this is not the case for  $S$ : the probability that  $S = 2$  is  $1/36$  (since there is only one combination of inputs that leads to that outcome). In contrast, the probability that  $S = 6$  is  $1/6$  since there are six possible combinations of inputs that sum to 6.

By construction, we know that  $I(D_1; D_2) = 0$  bit. However,  $I(D_1; D_2 | S) \approx 1.896$  bit, and so  $Co(X_1; X_2; Y) \approx -1.896$  bit. The synergistic component of the information dominates and so the co-information is negative and Berkson’s paradox has occurred!

If we compute the actual PID using the minimum mutual information function for redundancy, we find that  $Red(D_1, D_2; S) \approx 0.689$  bit and  $Syn(D_1, D_2; Y) \approx 2.585$  bit (and both unique terms are 0 bit exactly since the dice are identical). Where do these numbers come from? Intuitively, we can understand the non-zero redundancy by noticing that, even though  $D_1 \perp D_2$ , learning the state of either die alone is enough to “rule out” some possible states of  $S$ . For example, learning that  $D_1 = 1$  or  $D_2 = 1$  is sufficient to exclude all possible values of  $S > 7$ , since to achieve such a value requires that at least one die be greater than or equal to two. This is what Harder refers to as “mechanistic redundancy”, in contrast to “source redundancy”, which is attributable to correlations between the inputs [35].

The remaining information about  $S$  is synergistic. This is because completely resolving all the uncertainty about the state of  $S$  requires knowing both inputs. Learning that  $D_1 = 1$  is sufficient to rule out any  $S > 7$ ; however, it is not enough to specify the value of  $S$  any further. Doing so requires knowing both inputs, and this is the nature of computation. We should note that we are not saying that synergistic information *is* computation in any fundamental sense. Whatever physical process that reads the value of the individual dice and computes the sum is what is performing the computation. The synergistic information is merely a kind of statistical fingerprint, indicating that information has been integrated in some higher-order fashion.

### 2.4. Case Study II: Transfer Entropy

The perspective that synergistic information represents some putative notion of “computation” can help us make sense of difficult-to-interpret phenomena in complex systems research. As a second case study, let us consider the transfer entropy [36,37]. The transfer entropy from some variable  $X$  to another variable  $Y$  is defined as follows:

$$TE(X \rightarrow Y) = I(X_{past}; Y_t | Y_{past}) \tag{14}$$

where  $X_{past}$  refers to the (potentially multivariate) embedded past of  $X$ ,  $Y_{past}$  refers to the (potentially multivariate) embedded past of  $Y$ , and  $Y_t$  refers to the state of  $Y$  at time  $t$ . The logic behind the transfer entropy is to capture that information about  $Y$ ’s next state disclosed by  $X$ ’s past that is above and beyond the information disclosed by  $Y$ ’s own past (hence the conditioning).

As we have already established, conditioning mutual information is not a purely subtractive manipulation. Generally, the tacit assumption implicit in applications of the transfer entropy is that conditioning on  $Y_{past}$  will reduce the apparent correlation between  $X_{past}$  and  $Y_t$ . However, this is not always the case: how do we interpret the case of  $I(X_{past}; Y_t) < I(X_{past}; Y_t | Y_{past})$ ?

Williams and Beer [38] showed that the transfer entropy can be decomposed into

$$TE(X \rightarrow Y) = Unq(X_{past}; Y_t / Y_{past}) + Syn(X_{past}, Y_{past}; Y_t) \tag{15}$$

The first term,  $Unq(X_{past}; Y_t/Y_{past})$ , is generally understood to refer to the true bivariate information transfer and was termed the “state-independent transfer entropy” (SITE) by Williams and Beer, as it does not depend on the past state of the target. The second term,  $Syn(X_{past}, Y_{past}; Y_t)$ , is less clear: it refers to the information about  $Y_t$  that can only be learned when both  $X$ ’s past and  $Y$ ’s past are known together. Williams and Beer termed it the “state-dependent information transfer” (SDTE) since it does depend on the past state of the target.

This decomposition shows that the putatively bivariate transfer entropy actually conflates pairwise and higher-order dependencies (this is also generally true of any conditional mutual information). This discovery prompted intense criticism of the transfer entropy as a technique for complex network inference by James and Crutchfield [39], as it obviously challenges the interpretation that  $TE(X \rightarrow Y)$  measures unidirectional information “flow”. Consequently, they argue that the transfer entropy cannot be properly localized to a single source. Since the critique, however, this facet of the bivariate transfer entropy has received little attention, despite its significance. Recently, Daube et al. [40] showed that the SDTE contributes to the estimation of transfer entropy from simulated data. Furthermore, contra Barrett [31], who suggested stripping the SDTE and using only the “true” bivariate SITE transfer, Daube et al. argued that the SITE actually does worse than the SDTE.

Based on both empirical findings and theoretical arguments presented here, it seems implausible to continue to treat the transfer entropy as a measure of conveniently bivariate information flow. However, none of the existing treatments meditate on how to best intuitively understand what kinds of dynamics are represented by the SDTE. Formally, the synergy is well understood and easy to derive. However, a unifying interpretative framework has largely been lacking. If we take that synergy refers to integrative computation, then the SDTE starts to become a little bit less mysterious: any process that “decides” its next state as a function of some combinations of memory and external perturbation can be modeled as performing a “computation” on two inputs: one is the state of its own past, and one is the value of the input it receives. State-dependent transfer entropy could be thought of as reporting that part of the transfer entropy reflects the computation that the target element performs in deciding its next state as a function of what is stored in its memory and what input it receives from another part of the system.

### 3. Computation and Causality in Complex Systems

So far, we have assumed that the causal model of the collider is known, and notions such as “inputs” and “outputs” are well specified. In the case of summing the values of the dice, the inputs ( $D_i$ s) are obviously the individual dice, and the output (or target,  $S$ ) is the sum of the input. In this context, it is obvious that the relevant information decomposition is of  $I(D_1, D_2; S)$ , and the notion of synergy makes intuitive sense. Now suppose that we have three new variables,  $A_1$ ,  $A_2$  and  $B$ , and that rather than being related by a collider motif, they are related by a broadcasting motif as follows:

$$A_1 \leftarrow B \rightarrow A_2 \quad (16)$$

where  $B$  broadcasts its state to each  $A_i$ . In this case, Berkson’s paradox will not hold: if  $A_1$  and  $A_2$  are otherwise independent, when conditioning on  $B$ , we would expect  $I(A_1; A_2|B) = 0$  bit. Given the definition of the co-information given in Equation (12), it is easy to see that  $Co(A_1; A_2; B) = I(A_1; A_2)$ . Since  $Co(A_1; A_2; B) \gg 0$  bit, we can infer that the information structure of the triad is, overall, dominated by redundancy. This is consistent with the intuition that a broadcaster is duplicating information over both  $A_1$  and  $A_2$  (as well as representing it in itself).

There is, however, an apparent problem. From Equation (13), we know that

$$Co(A_1; A_2; B) = Red(A_1, A_2; B) - Syn(A_1, A_2; B) \quad (17)$$



If  $Co(A_1; A_2; B) \gg 0$ , then how might we interpret the possibility of non-zero synergy in this context? From a mathematical perspective, it is straightforward: there is uncertainty about the state of  $B$  that can only be resolved when  $A_1$  and  $A_2$  are both known. Unfortunately, this seems to complicate our argument that synergy is a statistical fingerprint of computation; the broadcaster motif does not seem to be performing any meaningful “computation”. It is merely duplicating information over downstream elements.

To resolve this uncertainty, it is important to remember that the PID is an inherently directed analysis, dividing the system into inputs and targets. As previously mentioned in the case of the collider, the inputs and targets map neatly onto the causal model of the collider itself. In the case of the broadcaster motif, this is not the case. While it is numerically possible to compute  $Syn(A_1, A_2; B)$ , doing so is running “against the causal flow” and, consequently, appears to be meaningless from a directed perspective. It could be said that the map and territory are in conflict: the orientation of the PID is misaligned with the causal structure of the system. This is important to highlight, as it shows that not all computable synergies are necessarily meaningful: the structure of the underlying causal model is a key consideration, and researchers interested in applying these analyses should be mindful of the different contextual factors that might change the interpretation of the numerical value.

What about in cases where the causal ground truth is not known? It is easy to construct tractable toy models in theoretical contexts, but in natural sciences, rarely is a blueprint handed to us by God. Instead, it must be inferred. Suppose we have some triad,  $\{X, Y, Z\}$ , which could be either a collider or a broadcaster and we do not know which. Furthermore, we lack any domain expertise that can be leveraged to assume a structure. In the material we have covered so far, disambiguating these two cases is difficult. We could look for Berkson’s paradox, assuming that if  $X \rightarrow Z \leftarrow Y$  then  $Co(X; Y; Z) < 0$ , but even if the structure is a collider, if  $X$  and  $Y$  are correlated, we may still find that redundancy outweighs synergy. In systems that have a large synchronous component to their dynamics (for example, the nervous system), this is a real concern.

If a scientist has a dataset from which synergies can be computed, how could they disambiguate between those synergies that represent some kind of computational process (as in the collider) and those that do not (as in the broadcaster)? We argue that, ultimately, this is a question of causal inference [20] and leveraging both domain expertise, as well as more refined techniques for structure learning in complex systems [41,42] must be incorporated into the analysis. If, given some dataset, it is possible to construct a causal (or merely effective) scaffold of the system, then colliders can be identified as putative sites of computation. This approach has been leveraged in computational neuroscience: given finely resolved, temporally ordered spiking data, it is possible to infer effective connectivity networks using measures like the bivariate or multivariate transfer entropy [15,16], from which colliders can be sampled.

Alternately, if one has strong priors on the structure of the system under study, then those may be leveraged. For example, in a quantitative study of intersectionality in social systems, Varley and Kaminski studied how demographic factors such as race and sex synergistically disclose information about outcomes such as incomes [43]. In this context, there is ample prior evidence to assume a collider-like structure, while a broadcaster-like structure can be ruled out because there is no plausible way that income could causally impact features like race or sex.

Recently, considerable work has involved the problem of inferring directed models that account for higher-order interactions. It has long been known that strictly pairwise approaches to structure learning fail in the presence of higher-order interactions [8], and finding optimal ways to account for higher-order redundancies and synergies remains an open problem. One popular approach is the multivariate transfer entropy [44,45], which takes a serial conditioning approach to account for higher-order information in transfer entropy, although other approaches include the framework proposed by Mijatovic et al., which introduced a balance index (B-index) that quantified whether the interaction between

two elements was redundancy-dominated or synergy-dominated in the context of the rest of the system. Beyond the pairwise network approach, representing higher-order information with hypergraphs has been recently explored [46,47] as well. Similarly, generalizations of the partial information decomposition to stochastic processes [48] can reveal time-directed higher-order information in dynamical systems.

Assuming that one is able to construct a plausible causal (or effective) model of the system under study, there is still the issue that the joint mutual information itself is not a causal measure. Previous work has established that even time-directed measures of information flow, such as temporal mutual information or transfer entropy, poorly reflect causal dependencies [49,50]. The link between information flow, causality, and structure inference is a rich field that we cannot completely survey (for further discussion, see [51–53]), and the problem of true causal information decomposition remains an outstanding issue. Ultimately, the question of attributing higher-order, synergistic causation may not have a unique solution but, like the question of redundancy, may depend on the specific way one operationalizes the notion of “causality” (interventional, counterfactual, etc). In the short term, however, we propose that a small modification of the joint mutual information may be useful here. In 2003, Tononi and Sporns proposed a measure they termed “effective information”, which is the mutual information from a source to a target when the entropy of the sources is maximal [54]:

$$EI(\mathbf{X} \rightarrow Y) = I(\mathbf{X}^{H_{\max}}; Y) \quad (18)$$

By forcing a maximum entropy distribution on the inputs, the effective information attempts to simulate random assignment in interventional approaches to causality. Later, Hoel et al. [52] argued that the effective information could be understood as an instance of Pearl’s do-calculus applied to mutual information:  $EI(\mathbf{X} \rightarrow Y) = I(Y; do(\mathbf{X}))$ . Being modified mutual information, effective information can be decomposed using the PID, lending a more causal flavor to the synergy than is usually the case (as well as removing any possible source redundancy [35]). For an example of this approach to discrete data, see [43]. We should be very clear, however, that while effective information and do-calculus approximate a causal inference (and by extension, causal computation), they are still ultimately based on correlational analyses, which can break down in certain contexts. It is known that the Shannon entropy can misrepresent structures in complex systems [55]. For researchers interested in applying these methods, considering the appropriateness of an information-theoretic model should be seen as a necessary first step in experimental design. Future work considering other notions of information structure (such as algorithmic information theory and algorithmic information dynamics [56]) will complement this work.

#### 4. Discussion

Since the notion of synergy was formalized in the context of the PID, it has been widely applied to a number of different fields. Each of these applications brings with it distinct interpretations, and there is not much cross-talk between disciplines as to the general meaning of synergistic information. For example, when considering sociological identities, synergy is equated with the sociological concept of “intersectionality” [43]. In the context of biological neural networks, synergy has been associated with information processing in the nervous system [12,13,15,16,57,58], while in artificial neural networks, synergy has been associated with the complexity of representations [59] or multi-task learning [60]. Synergy has been identified in the cardiovascular system as well [48]. When considering econometric data, synergy has been associated with economic sophistication [61], and synergy has been found to correlate with the level of consciousness in patients who have been anesthetized or are comatose [58,62]. In climatological data, synergy has been identified, although its significance remains unclear [57,63,64], and in theoretical biology, synergy has been associated with notions of “individuality” or organismic autonomy [65].

This list is non-exhaustive, but it is sufficient to indicate that synergy is ubiquitous in the natural world; far from being some kind of strange or exotic dependency that only

occurs in constructed models, synergy has been found in complex systems at almost every scale. Despite this ubiquity, the variety of different, context-dependent interpretations shows that science in general lacks a unified framework for understanding the significance of synergy generally. Here, we propose that synergy can be understood as a kind of fingerprint of computation in complex systems: when an element of a system “chooses” its next state based on some multi-argument function that accounts for multiple inputs, synergistic information will almost certainly exist. By showing that synergistic dynamics are intimately related to the notion of colliders in causal inference, we hope to ground future studies of synergistic dynamics in complex systems in a common framework.

Since the groundbreaking work of Williams and Beer first formalized the notion of synergy in the context of the PID over a decade ago [9], the majority of the focus of the field has been on solving the problem of how to recognize it in data. This has included work on a plethora of redundancy functions (for review, see [28,29]), extensions of the PID (such as the multi-target integrated information decomposition [66,67] and the partial entropy decomposition [68,69]), and the generalized information decomposition [70], as well as non-PID-based heuristics such as the O-information [71], the O-information rate O-information rate [48] (a derivative of the O-information specifically designed for dynamical system), or the  $\alpha$ -synergy decomposition [72].

At this point, the problem of recognizing synergistic dependencies in data is, if not totally solved, well studied enough that the field has a rich toolkit of measures to apply to any given dataset. This puts us in the position of being able to move beyond asking merely “is there synergy?” or “where is the synergy?” Instead, we must begin asking “what is synergy for?”

### Conclusions

In this paper, we explored the idea that modeling “computation” in complex systems is related to the information-theoretic notion of statistical synergy. Synergy occurs when there is information about the state of an output variable that can only be learned when the joint state of all of its input variables is known and proposed that this is a statistical “fingerprint” reflecting some physical process by which an output element “chooses” its next state as a function of multiple inputs (rather than merely propagating information forward). To further formalize this notion, we related this process to the collider motif studied in the field of causal inference, where a single variable is causally influenced by multiple inputs simultaneously. We showed that the well-known collider bias (also known as Berkson’s paradox) occurs because of a synergistic interaction between the inputs and output, providing a mathematical link between the world of multivariate information theory and causal inference or structure learning. We further proposed that this link deepens our understanding of both statistical synergy as a concept and the phenomenon of computation in complex systems.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created for this article.

**Acknowledgments:** I would like to thank Maria Pope for extensive discussions around the links between synergy and Berkson’s paradox and Olaf Sporns for support.

**Conflicts of Interest:** The author declares no conflicts of interest.

### References

1. Flake, G.W. *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation*; MIT Press: Cambridge, MA, USA, 2000.
2. Mitchell, M. *Complexity: A Guided Tour*; Oxford University Press: Oxford, UK, 2009.

3. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [[CrossRef](#)]
4. Piccinini, G. Mapping Accounts. In *Physical Computation: A Mechanistic Account*; Piccinini, G., Ed.; Oxford University Press: Oxford, UK, 2015. [[CrossRef](#)]
5. Schweizer, P. Computation in Physical Systems: A Normative Mapping Account. In *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*; Berkich, D., d'Alfonso, M.V., Eds.; Series Title: Philosophical Studies Series; Springer International Publishing: Cham, Switzerland, 2019; Volume 134, pp. 27–47. [[CrossRef](#)]
6. Lizier, J.T. *The Local Information Dynamics of Distributed Computation in Complex Systems*; Springer Theses; Springer: Berlin/Heidelberg, Germany, 2013. [[CrossRef](#)]
7. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
8. Varley, T.F. Information Theory for Complex Systems Scientists. *arXiv* **2023**, arXiv:2304.12482. [[CrossRef](#)]
9. Williams, P.L.; Beer, R.D. Nonnegative Decomposition of Multivariate Information. *arXiv* **2010**, arXiv:1004.2515.
10. Lizier, J.T.; Prokopenko, M.; Zomaya, A.Y. Detecting Non-trivial Computation in Complex Dynamics. In *Proceedings of the Advances in Artificial Life*; Almeida e Costa, F., Rocha, L.M., Costa, E., Harvey, I., Coutinho, A., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2007; pp. 895–904. [[CrossRef](#)]
11. Lizier, J.T.; Flecker, B.; Williams, P.L. Towards a Synergy-based Approach to Measuring Information Modification. *arXiv* **2013**, arXiv:1303.3440. [[CrossRef](#)]
12. Timme, N.M.; Ito, S.; Myroshnychenko, M.; Nigam, S.; Shimono, M.; Yeh, F.C.; Hottowy, P.; Litke, A.M.; Beggs, J.M. High-Degree Neurons Feed Cortical Computations. *PLoS Comput. Biol.* **2016**, *12*, e1004858. [[CrossRef](#)]
13. Faber, S.P.; Timme, N.M.; Beggs, J.M.; Newman, E.L. Computation is concentrated in rich clubs of local cortical networks. *Netw. Neurosci.* **2018**, *3*, 1–21. [[CrossRef](#)]
14. Sherrill, S.P.; Timme, N.M.; Beggs, J.M.; Newman, E.L. Correlated activity favors synergistic processing in local cortical networks in vitro at synaptically relevant timescales. *Netw. Neurosci.* **2020**, *4*, 678–697. [[CrossRef](#)]
15. Newman, E.L.; Varley, T.F.; Parakkattu, V.K.; Sherrill, S.P.; Beggs, J.M. Revealing the Dynamics of Neural Information Processing with Multivariate Information Decomposition. *Entropy* **2022**, *24*, 930. [[CrossRef](#)]
16. Varley, T.F.; Sporns, O.; Schaffelhofer, S.; Scherberger, H.; Dann, B. Information-processing dynamics in neural networks of macaque cerebral cortex reflect cognitive state and behavior. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2207677120. [[CrossRef](#)]
17. Berkson, J. Limitations of the Application of Fourfold Table Analysis to Hospital Data. *Biom. Bull.* **1946**, *2*, 47–53. [[CrossRef](#)]
18. Westreich, D. Berkson's bias, selection bias, and missing data. *Epidemiology* **2012**, *23*, 159–164. [[CrossRef](#)] [[PubMed](#)]
19. Holmberg, M.J.; Andersen, L.W. Collider Bias. *JAMA* **2022**, *327*, 1282–1283. [[CrossRef](#)] [[PubMed](#)]
20. Pearl, J.; Glymour, M.; Jewell, N.P. *Causal Inference in Statistics: A Primer*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
21. Matsuda, H. Physical nature of higher-order mutual information: Intrinsic correlations and frustration. *Phys. Rev. E* **2000**, *62*, 3096–3102. [[CrossRef](#)] [[PubMed](#)]
22. Watanabe, S. Information Theoretical Analysis of Multivariate Correlation. *IBM J. Res. Dev.* **1960**, *4*, 66–82. [[CrossRef](#)]
23. Abdallah, S.A.; Plumbley, M.D. A measure of statistical complexity based on predictive information with application to finite spin systems. *Phys. Lett. A* **2012**, *376*, 275–281. [[CrossRef](#)]
24. Varley, T.F.; Pope, M.; Faskowitz, J.; Sporns, O. Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex. *Commun. Biol.* **2023**, *6*, 1–12. [[CrossRef](#)]
25. McGill, W.J. Multivariate information transmission. *Psychometrika* **1954**, *19*, 97–116. [[CrossRef](#)]
26. Bell, A.J. The co-information lattice, Nara, Japan, 2003. In Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003), Nara, Japan, 1–4 April 2003.
27. Gutknecht, A.J.; Wibral, M.; Makkeh, A. Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2021**, *477*, 20210110. [[CrossRef](#)]
28. Kolchinsky, A. A Novel Approach to the Partial Information Decomposition. *Entropy* **2022**, *24*, 403. [[CrossRef](#)]
29. Kay, J.W.; Schulz, J.M.; Phillips, W.A. A Comparison of Partial Information Decompositions Using Data from Real and Simulated Layer 5b Pyramidal Cells. *Entropy* **2022**, *24*, 1021. [[CrossRef](#)]
30. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared Information—New Insights and Problems in Decomposing Information in Complex Systems. *arXiv* **2013**, arXiv:1210.5902. [[CrossRef](#)]
31. Barrett, A.B. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E* **2015**, *91*, 052802. [[CrossRef](#)] [[PubMed](#)]
32. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying Unique Information. *Entropy* **2014**, *16*, 2161–2183. [[CrossRef](#)]
33. Rauh, J.; Bertschinger, N.; Olbrich, E.; Jost, J. Reconsidering unique information: Towards a multivariate information decomposition. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 2232–2236. [[CrossRef](#)]
34. Rosas, F.; Ntranos, V.; Ellison, C.J.; Pollin, S.; Verhelst, M. Understanding Interdependency Through Complex Information Sharing. *Entropy* **2016**, *18*, 38. [[CrossRef](#)]
35. Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2013**, *87*, 012130. [[CrossRef](#)]
36. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [[CrossRef](#)]

37. Bossomaier, T.; Barnett, L.; Harré, M.; Lizier, J.T. *An Introduction to Transfer Entropy: Information Flow in Complex Systems*; Springer: Berlin/Heidelberg, Germany, 2016.
38. Williams, P.L.; Beer, R.D. Generalized Measures of Information Transfer. *arXiv* **2011**, arXiv:1102.1507.
39. James, R.G.; Barnett, N.; Crutchfield, J.P. Information Flows? A Critique of Transfer Entropies. *Phys. Rev. Lett.* **2016**, *116*, 238701. [[CrossRef](#)]
40. Daube, C.; Gross, J.; Ince, R.A.A. A whitening approach for Transfer Entropy permits the application to narrow-band signals. *arXiv* **2022**, arXiv:2201.02461.
41. Scanagatta, M.; Salmerón, A.; Stella, F. A survey on Bayesian network structure learning from data. *Prog. Artif. Intell.* **2019**, *8*, 425–439. [[CrossRef](#)]
42. Kitson, N.K.; Constantinou, A.C.; Guo, Z.; Liu, Y.; Chobtham, K. A survey of Bayesian Network structure learning. *Artif. Intell. Rev.* **2023**, *56*, 8721–8814. [[CrossRef](#)]
43. Varley, T.F.; Kaminski, P. Untangling Synergistic Effects of Intersecting Social Identities with Partial Information Decomposition. *Entropy* **2022**, *24*, 1387. [[CrossRef](#)] [[PubMed](#)]
44. Novelli, L.; Lizier, J.T. Inferring network properties from time series using transfer entropy and mutual information: Validation of multivariate versus bivariate approaches. *Netw. Neurosci.* **2021**, *5*, 373–404. [[CrossRef](#)] [[PubMed](#)]
45. Wollstadt, P.; Lizier, J.T.; Vicente, R.; Finn, C.; Martinez-Zarzuela, M.; Mediano, P.; Novelli, L.; Wibral, M. IDTxL: The Information Dynamics Toolkit xl: A Python package for the efficient analysis of multivariate information dynamics in networks. *J. Open Source Softw.* **2019**, *4*, 1081. [[CrossRef](#)]
46. Varley, T.F.; Pope, M.; Puxeddu, M.G.; Faskowitz, J.; Sporns, O. Partial entropy decomposition reveals higher-order structures in human brain activity. *arXiv* **2023**, arXiv:2301.05307. [[CrossRef](#)]
47. Marinazzo, D.; Van Roozendaal, J.; Rosas, F.E.; Stella, M.; Comolatti, R.; Colenbier, N.; Stramaglia, S.; Rosseel, Y. An information-theoretic approach to build hypergraphs in psychometrics. *Behav. Res. Methods* **2024**, *56*, 8057–8079. [[CrossRef](#)]
48. Faes, L.; Mijatovic, G.; Antonacci, Y.; Pernice, R.; Barà, C.; Sparacino, L.; Sammartino, M.; Porta, A.; Marinazzo, D.; Stramaglia, S. A New Framework for the Time- and Frequency-Domain Assessment of High-Order Interactions in Networks of Random Processes. *IEEE Trans. Signal Process.* **2022**, *70*, 5766–5777. [[CrossRef](#)]
49. Lizier, J.T.; Prokopenko, M. Differentiating information transfer and causal effect. *Eur. Phys. J. B* **2010**, *73*, 605–615. [[CrossRef](#)]
50. Eldhose, E.; Chauhan, T.; Chandel, V.; Ghosh, S.; Ganguly, A.R. Robust Causality and False Attribution in Data-Driven Earth Science Discoveries. *arXiv* **2022**, arXiv:2209.12580. [[CrossRef](#)]
51. Ay, N.; Polani, D. Information flows in causal networks. *Adv. Complex Syst.* **2008**, *11*, 17–41. [[CrossRef](#)]
52. Hoel, E.P.; Albantakis, L.; Tononi, G. Quantifying causal emergence shows that macro can beat micro. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19790–19795. [[CrossRef](#)] [[PubMed](#)]
53. Goodwell, A.E.; Jiang, P.; Ruddell, B.L.; Kumar, P. Debates—Does Information Theory Provide a New Paradigm for Earth Science? Causality, Interaction, and Feedback. *Water Resour. Res.* **2020**, *56*, e2019WR024940. [[CrossRef](#)]
54. Tononi, G.; Sporns, O. Measuring information integration. *BMC Neurosci.* **2003**, *4*, 31. [[CrossRef](#)] [[PubMed](#)]
55. Zenil, H.; Kiani, N.A.; Tegnér, J. Low-algorithmic-complexity entropy-deceiving graphs. *Phys. Rev. E* **2017**, *96*, 012308. [[CrossRef](#)]
56. Zenil, H.; Kiani, N.A.; Tegnér, J. *Algorithmic Information Dynamics: A Computational Approach to Causality with Applications to Living Systems*; Cambridge University Press: Cambridge, UK, 2023. [[CrossRef](#)]
57. Antonacci, Y.; Minati, L.; Nuzzi, D.; Mijatovic, G.; Pernice, R.; Marinazzo, D.; Stramaglia, S.; Faes, L. Measuring High-Order Interactions in Rhythmic Processes Through Multivariate Spectral Information Decomposition. *IEEE Access* **2021**, *9*, 149486–149505. [[CrossRef](#)]
58. Luppi, A.I.; Mediano, P.A.M.; Rosas, F.E.; Allanson, J.; Pickard, J.D.; Carhart-Harris, R.L.; Williams, G.B.; Craig, M.M.; Finoia, P.; Owen, A.M.; et al. A Synergistic Workspace for Human Consciousness Revealed by Integrated Information Decomposition. *eLife* **2023**, *12*, RP88173. [[CrossRef](#)]
59. Ehrlich, D.A.; Schneider, A.C.; Priesemann, V.; Wibral, M.; Makkeh, A. A Measure of the Complexity of Neural Representations based on Partial Information Decomposition. *arXiv* **2023**, arXiv:2209.10438.
60. Proca, A.M.; Rosas, F.E.; Luppi, A.I.; Bor, D.; Crosby, M.; Mediano, P.A.M. Synergistic information supports modality integration and flexible learning in neural networks solving multiple tasks. *arXiv* **2022**, arXiv:2210.02996. [[CrossRef](#)]
61. Rajpal, H.; Guerrero, O.A. Quantifying the Technological Foundations of Economic Complexity. *arXiv* **2023**, arXiv:2301.04579. [[CrossRef](#)]
62. Luppi, A.I.; Mediano, P.A.M.; Rosas, F.E.; Allanson, J.; Pickard, J.D.; Williams, G.B.; Craig, M.M.; Finoia, P.; Peattie, A.R.D.; Coppola, P.; et al. Reduced emergent character of neural dynamics in patients with a disrupted connectome. *NeuroImage* **2023**, *269*, 119926. [[CrossRef](#)]
63. Goodwell, A.E.; Kumar, P. Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resour. Res.* **2017**, *53*, 5920–5942. [[CrossRef](#)]
64. Goodwell, A.E.; Kumar, P. Temporal Information Partitioning Networks (TIPNets): A process network approach to infer ecohydrologic shifts. *Water Resour. Res.* **2017**, *53*, 5899–5919. [[CrossRef](#)]
65. Krakauer, D.; Bertschinger, N.; Olbrich, E.; Flack, J.C.; Ay, N. The information theory of individuality. *Theory Biosci.* **2020**, *139*, 209–223. [[CrossRef](#)]

66. Mediano, P.A.M.; Rosas, F.E.; Luppi, A.I.; Carhart-Harris, R.L.; Bor, D.; Seth, A.K.; Barrett, A.B. Towards an extended taxonomy of information dynamics via Integrated Information Decomposition. *arXiv* **2021**, arXiv:2109.13186.
67. Varley, T.F. Decomposing past and future: Integrated information decomposition based on shared probability mass exclusions. *PLoS ONE* **2023**, *18*, e0282950. [[CrossRef](#)]
68. Ince, R.A.A. The Partial Entropy Decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal. *arXiv* **2017**, arXiv:1702.01591.
69. Finn, C.; Lizier, J.T. Generalised Measures of Multivariate Information Content. *Entropy* **2020**, *22*, 216. [[CrossRef](#)]
70. Varley, T.F. Generalized decomposition of multivariate information. *PLoS ONE* **2024**, *19*, e0297128. [[CrossRef](#)]
71. Rosas, F.; Mediano, P.A.M.; Gastpar, M.; Jensen, H.J. Quantifying High-order Interdependencies via Multivariate Extensions of the Mutual Information. *Phys. Rev. E* **2019**, *100*, 032305. [[CrossRef](#)]
72. Varley, T.F. A scalable synergy-first backbone decomposition of higher-order structures in complex systems. *Npj Complex.* **2024**, *1*, 1–11. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.