

Modes of Information Flow

Ryan G. James,^{1,*} Blanca Daniella Mansante Ayala,^{1,†} Bahti Zakirov,^{2,‡} and James P. Crutchfield^{1,§}

¹*Complexity Sciences Center and Physics Department,
University of California at Davis, One Shields Avenue, Davis, CA 95616*

²*Department of Engineering Science and Physics, College of Staten Island,
The City University of New York, 2800 Victory Blvd., Staten Island, NY 10314*

(Dated: August 20, 2018)

Information flow between components of a system takes many forms and is key to understanding the organization and functioning of large-scale, complex systems. We demonstrate three modalities of information flow from time series X to time series Y . *Intrinsic* information flow exists when the past of X is individually predictive of the present of Y , independent of Y 's past; this is most commonly considered information flow. *Shared* information flow exists when X 's past is predictive of Y 's present in the same manner as Y 's past; this occurs due to synchronization or common driving, for example. Finally, *synergistic* information flow occurs when neither X 's nor Y 's pasts are predictive of Y 's present on their own, but taken together they are. The two most broadly-employed information-theoretic methods of quantifying information flow—time-delayed mutual information and transfer entropy—are both sensitive to a pair of these modalities: time-delayed mutual information to both intrinsic and shared flow, and transfer entropy to both intrinsic and synergistic flow. To quantify each mode individually we introduce our *cryptographic flow ansatz*, positing that intrinsic flow is synonymous with secret key agreement between X and Y . Based on this, we employ an easily-computed secret-key-agreement bound—intrinsic mutual information—to quantify the three flow modalities in a variety of systems including asymmetric flows and financial markets.

Keywords: stochastic process, transfer entropy, intrinsic mutual information, mutual information, information flow.

PACS numbers: 05.45.-a 89.70.+c 05.45.Tp 02.50.Ey 02.50.-r 02.50.Ga

I. INTRODUCTION

Information flow is an important signature of truly complex systems—an incisive proxy for their structure and behavior. Truly complex systems are becoming increasingly familiar to most all the sciences and engineering. Certain expressed genes modulate the expression of others, leading to structured biological processes in metabolism, development, and evolution. The price of a good signals scarcity and demand to consumers and producers, respectively. During specific cognitive tasks, only a small portion of the brain's connectome is activated and so determining the underlying subnetwork is critical to identifying neural function. To these ends, methods of tracking and quantifying information flow form a core toolset for analyzing large-scale complex systems.

Methodologically, it is difficult to find an alternative toolset with the many features that recommend using information theory [1, 2] to analyze complex systems. First, information accounts for any type of co-relation [1]. Whereas, statistical correlation requires an under-

lying model and typically captures only linear statistical dependencies [3]. Second, information is broadly applicable. Many systems across the sciences simply do not have an “energy” and this precludes appealing to physics-based modeling that starts from a system Hamiltonian. In contrast, information can be defined for mechanical, chemical, biological, social, and engineered systems. Third, information provides directly comparable quantitative units across qualitatively different systems. Pairwise statistical correlation comes in fixed, domain-specific units—meters squared, concentration squared, volts squared, dollars squared, and so on. Information is universally measured in bits. Fourth, there is a rough equivalence between probability theory and information theory [4] and a substantial foundational overlap between mathematical statistics and information theory; e.g., see Refs. [5] and [1, Ch. 11]. Finally, and perhaps most importantly, large-scale complex systems generate emergent patterns—patterns that an analyst does not know a priori. Information does not require prior knowledge of an appropriate representational basis, which is essential when attempting to discover new patterns not seen before [6]. Despite a number of technical challenges, to date it appears that the tools of information are the most general, workable alternative available in the pursuit of complex systems. The following provides new results that remove several of the remaining impediments.

* rgjames@ucdavis.edu

† masante@ucdavis.edu

‡ bahtizakirov@gmail.com

§ chaos@ucdavis.edu

In particular, despite its long-lived intuitive appeal [7–11], information flow is not yet a concretely defined concept. One consequence is that many methods of calculating it fail surprisingly when deployed in unfamiliar contexts [12]. Here, we posit that these failures are not directly due to shortcomings of the quantitative measures themselves, but rather arise due to analysts adhering to the concept of a unitary information flow. In contrast, we demonstrate that the flow between two time-series, X and Y , takes on three qualitatively distinct modes: *intrinsic*, *shared*, and *synergistic* information flows. Unfortunately, to date methods of computing information flow do not quantify these distinct modes. Here, we solve this problem via a novel adaptation of cryptography.

Our development proceeds as follows. Section II discusses and exemplifies the three modes of information flow. As an aid in this, App. A briefly reviews the necessary notation and concepts from elementary probability (random variables X and Y), time series (random variable sequences $X_{t:t'}$), and information measures ($H[X]$, $H[Y|X]$, $I[X : Y]$, $I[X : Y|Z]$). Section III then surveys extant measures of information flow within multivariate time series. Section IV introduces our cryptographic flow ansatz and shows how it isolates intrinsic information flow and so yields the full three-way decomposition of information flow. Section V explores components of the decomposition in a variety of settings, including asymmetric flows and financial indices.

II. MODES OF INFORMATION FLOW

Colloquially, information flow is the movement of information from one agent or system to another. Unlike many flows considered in physics, such as electric current or fluid flow, there is no single conservation law for information. This makes quantifying information flow vastly more challenging. In light of this, definitions of information flow have been somewhat ad-hoc, though they typically employ either some form of mutual information [13] or quantify the influence one agent has on another [14]. They are often backed with examples where the proposed measure performs admirably, though performance in other settings can be mixed or misleading. We propose that this inconsistency is due to conflating distinct modes of information flow.

Specifically, information flow from time series X to time series Y can take three qualitatively distinct forms. The first, *intrinsic flow*, is when the past behavior of the X time series is directly predictive of the present behavior of the Y time series in a fashion that the past behavior of Y is not. For example, this occurs when an “upstream”

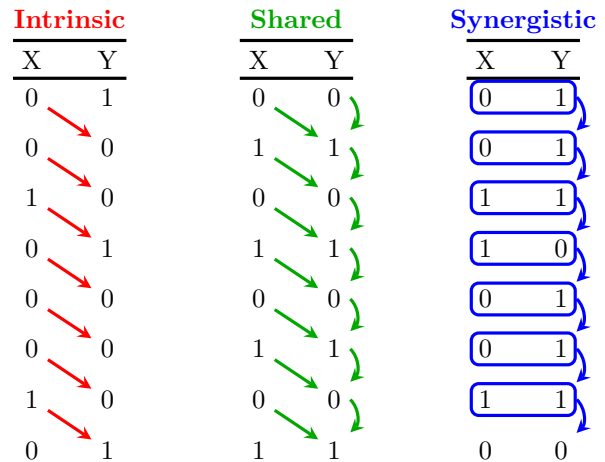


FIG. 1. Modes of information flow: Intrinsic flow is exemplified by $Y_0 = X_{-1}$. Shared flow by $Y_0 = \neg X_{-1} = \neg Y_{-1}$. Synergistic flow by $Y_0 = X_{-1} \oplus Y_{-1}$.

X drives a “downstream” Y . The second mode, *shared flow*, is when the present behavior of the Y time series can be inferred from the prior behavior of either the X time series or the Y time series. This occurs due to, say, a common driver or synchronization within a system. The third mode, *synergistic flow*, occurs when both the past of the X time series and the past of the Y time series are each independent of the present of the Y time series, but when combined the two pasts become predictive of it. This occurs in systems where the behavior of a component strongly depends upon its context within the system. Appendix B gives a more detailed rationale for the three modalities of flow.

Markovian examples of these three types of flow are illustrated in Fig. 1. Exemplifying intrinsic flow is the case where X_t is random and Y simply follows it: $Y_0 = X_{-1}$. Shared flow is demonstrated with synchronization, where $Y_0 = \neg X_{-1} = \neg Y_{-1}$, where \neg is the “not” operation. Finally, synergistic flow can be seen when X_t is random and $Y_0 = X_{-1} \oplus Y_{-1}$, where \oplus is the exclusive-OR operation. Both shared and synergistic flows are symmetric, in that they cannot be said to originate from either X ’s past or Y ’s past. Whereas, intrinsic flow is uniquely attributed to X ’s past.

Now, the question is: How do we detect and quantify these three forms of dependence from time series observations?

III. EXTANT MEASURES OF INFORMATION FLOW

Historically, information flow has been measured via the *time-delayed mutual information* [15]:

$$I[X_{-1}:Y_0] . \quad (1)$$

It posits that information flow from X to Y is the information shared between X 's past observations X_{-1} and Y 's present observation Y_0 . As such, it is sensitive to both intrinsic and shared dependence, as seen in Fig. 2. While time-delayed mutual information captures a restricted notion of causality, it "... fails to distinguish information that is actually exchanged from shared information due to common history and input signals " [16]. That is, it conflates intrinsic and shared dependence.

To cleave away the shared dependence from the time-delayed mutual information Ref. [16] proposed the *transfer entropy*:

$$I[X_{-1}:Y_0 | Y_{-1}] , \quad (2)$$

—the information shared by X 's past and Y 's present, given Y 's past—and correctly intuited that the influences of common history and input signals on shared information "... are excluded by appropriate conditioning of transition probabilities." Unfortunately, this ignores the possibility of conditional dependence. And, the transfer entropy suffers as a result, failing to distinguish intrinsic flow from synergistic flow; again, see Fig. 2. (That transfer entropy separates into two components is not new, see App. C.) Taken at face value, this presents the unfortunate situation of being unable to quantify any specific mode of information flow.

A short aside will highlight the issue here. Previous efforts to measure flow rest on the misunderstanding that conditioning is *subtractive*: "in our new approach, these influences are excluded by appropriate conditioning of transition probabilities" [16]. The erroneous assumption here being that conditioning only *excludes* dependency. In point of fact, information-theoretic conditioning and probabilistic conditioning, for that matter, are generically *not* subtractive operations, as demonstrated in Fig. 2 and Problem 10 in Chapter 2 of Ref. [17]. Ignoring this has led to the belief that conditioning on more and more time series results in a more incisive analysis of information flow within a system [16, 18, 19]. It need not.

Intrinsic				Shared				Synergistic			
X_{-1}	Y_0	Y_{-1}	Pr	X_{-1}	Y_0	Y_{-1}	Pr	X_{-1}	Y_0	Y_{-1}	Pr
0	0	0	1/4	0	1	0	1/2	0	0	0	1/4
0	0	1	1/4	1	0	1	1/2	0	1	1	1/4
1	1	0	1/4					1	0	1	1/4
1	1	1	1/4					1	1	0	1/4

$$\begin{array}{lll}
 I[X_{-1}:Y_0] = 1 & I[X_{-1}:Y_0] = 1 & I[X_{-1}:Y_0] = 0 \\
 I[X_{-1}:Y_0 | Y_{-1}] = 1 & I[X_{-1}:Y_0 | Y_{-1}] = 0 & I[X_{-1}:Y_0 | Y_{-1}] = 1
 \end{array}$$

FIG. 2. Three canonical types of dependence between two variables X_{-1} and Y_0 in the context of a third Y_{-1} : (Left) Intrinsic dependence exists between the first two variables in spite of the third. (Middle) Shared dependence exists synchronously with the third. (Right) Synergistic dependence exists only when also observing the third.

IV. CRYPTOGRAPHIC COMMON INFORMATION

Overcoming the challenge of information flow requires adopting a different viewpoint—one that directly addresses when two system components, and only two system components, possess common information. Solving this would circumvent the open-ended issue of conditioning on all of a system's other, possibly unspecified, variables besides the two of interest; a strategy that, on its own, fatally ignores conditional dependence, as we just argued. Thus, we must simultaneously solve a definitional problem and a technical problem: respectively (i) acknowledging distinct modes of flow and (ii) accounting for both conditional independence and dependence. Our solution appeals to cryptography and the information theory of two parties sharing secret keys. We introduce an ansatz that directly quantifies intrinsic flow and, thereby, completes the decomposition of information flow into its three modes.

Consider again the flow of intrinsic information from X_{-1} to Y_0 . The flow implicates some sort of dependency or correlation between X and Y that can unambiguously be attributed to X . Were this dependency able to be reproduced from other aspects of the system, it could not be said to have originated from X_{-1} . This observation evokes the cryptographic idea of *secret key agreement* and leads to our ansatz:

Cryptographic Flow Ansatz

Intrinsic information flow exists exactly when X_{-1} and Y_0 can agree on a secret key, while the past of the rest of the system eavesdrops.

Quantitatively, intrinsic information flow is the rate at which information secret to X_{-1} and Y_0 can be extracted from observations of the system.

A. Secret Key Agreement Rate

In this way, we identify the intrinsic information flow as the secret-key agreement rate [20], which is defined as follows. Consider a joint random variable (X, Y, Z) , where Alice has access to the X realizations, Bob the Y realizations, and Eve the Z realizations. Given N IID realizations of the joint variable, let X^N denote Alice's observations, Y^N Bob's, and Z^N Eve's. Random variable V represents the public communication that all three observe. Let S denote the secret key that Alice and Bob wish to have in common.

Now, let S_X and S_Y represent the secret keys that Alice and Bob, respectively, distill from their private observations as well as from the public communication V : $S_X = f(X^N, V)$ and $S_Y = g(Y^N, V)$. In this, functions f and g represent the mechanism by which Alice and Bob construct their copy of secret key S . A *secret key agreement scheme* defines the allowed public communications V and mechanisms f and g . If the scheme is any good, S_X and S_Y will be identical and equal to S with high probability: $\Pr(S_X = S_Y = S) \geq 1 - \epsilon$. Moreover, being secret, the key S should have arbitrarily small correlation with Eve's private observations Z and the public communication V : $I[S : V, Z^N] \leq \epsilon$. The *secret-key agreement rate* $S(X : Y \parallel Z)$ then is the maximum rate R such that:

$$\lim_{N \rightarrow \infty} \frac{1}{N} H[S] \geq R - \epsilon,$$

for $N > 0$ and $\epsilon > 0$. In other words, the secret key agreement rate is the largest rate at which a secret key S can be successfully produced.

Effectively, given realizations of the three-way joint random variable, there exists a scheme by which Alice and Bob can publicly exchange information and then distill their public and private information into a secret key upon which they both agree with arbitrarily high probability, but which has arbitrarily little information shared with all information available to Eve. The challenge now is to determine the secret-key extraction functions f and g , as well as what public communication V is necessary. See Ref. [21] for concrete examples.

B. An Easily Computed Upper Bound

Though we identified the secret key agreement rate with intrinsic information flow, the nonconstructive nature of its definition mandates we appeal to some proxy if we wish to practically estimate it. While several lower and upper bounds exist for the secret key agreement rate, here we use the *intrinsic mutual information* [22]. We

recommend this upper bound due to its nontrivial behavior (exemplified shortly) and relatively straightforward estimation.

An eavesdropper not only has access to her observations z , but also to any (local) modification $\Pr(\bar{z}|z)$ of them. Therefore, the ability of Alice and Bob to agree upon a key cannot be reduced if Z is replaced by any "corruption" \bar{Z} . This observation simplifies the secret key optimization, leading to a constructive bound on the secret key agreement rate:

$$\begin{aligned} S(X : Y \parallel Z) &\leq \min_{\Pr(\bar{z}|z)} I[X : Y \mid \bar{Z}] \\ &= I[X : Y \downarrow Z]. \end{aligned} \quad (3)$$

The last quantity—*intrinsic mutual information*—is therefore an upper bound on the secret key agreement rate. It can be easily verified that $I[X : Y \downarrow Z]$ is bounded from above by both $I[X : Y]$, when $\Pr(\bar{z}|z)$ is constant, and $I[X : Y \mid Z]$, when $\Pr(\bar{z}|z)$ is the identity. Fortunately, this optimization is not difficult due to the boundedness of \bar{Z} . (In fact, $|\bar{Z}| \leq |Z|$ [23]). Appendix D explains how to calculate the intrinsic mutual information.

To illustrate its behavior, consider the distribution in Fig. 3 [22]. This distribution has two qualitatively distinct sets of events. The first, encoded using 0s and 1s, exhibits conditional dependence: any pair of X , Y , or Z are independent, but given the third they are perfectly correlated. The second, encoded using 2s and 3s, exhibits conditional independence: any pair is perfectly correlated and is also correlated with the third. The mutual information $I[X : Y] = H[1/2, 1/4, 1/4] = 3/2$ bit, reflecting that they share 01-, 2-, and 3-ness. However, the conditional mutual information $I[X : Y \mid Z] = 1/2$ bit, reflecting the conditional dependence that occurs half the time. The mapping of Z to \bar{Z} given in the table demonstrates that the conditional dependence can be destroyed while preserving the conditional independence: $I[X : Y \mid \bar{Z}] = 0$ bit $= I[X : Y \downarrow Z]$. This indicates that both the mutual information and conditional mutual information misleadingly identify dependencies that do not belong to X and Y alone, but rather are shared by, or induced by, Z .

C. Flow Decomposition

Recall that we showed the time-delayed mutual information $I[X_{-1} : Y_0]$ captures both intrinsic and shared flows, while transfer entropy $I[X_{-1} : Y_0 \mid Y_{-1}]$ captures intrinsic and synergistic flows. Together with our cryptographic flow ansatz that quantifies intrinsic flow, simple

X	Y	Z	\bar{Z}	Pr	
0	0	0	0	1/8	$I[X : Y] = 3/2$ bit
0	1	1	0	1/8	
1	0	1	0	1/8	
1	1	0	0	1/8	$I[X : Y Z] = 1/2$ bit
2	2	2	2	1/4	$I[X : Y \downarrow Z] = 0$ bit
3	3	3	3	1/4	

FIG. 3. The intrinsic mutual information $I[X : Y \downarrow Z]$ can be less than both $I[X : Y]$ and $I[X : Y | Z]$: The mutual information $I[X : Y]$ captures the 01-, 2-, and 3-ness that is shared by X and Y , while the conditional mutual information $I[X : Y | Z]$ captures the fact that knowledge of Z indicates whether the 0s and 1s of X and Y are the same or different. Neither of these dependencies are held by X and Y alone, and so the intrinsic mutual information $I[X : Y \downarrow Z]$ vanishes.

Intrinsic				Shared				Synergistic			
X_{-1}	Y_0	Y_{-1}	Pr	X_{-1}	Y_0	Y_{-1}	Pr	X_{-1}	Y_0	Y_{-1}	Pr
0	0	0	1/4	0	1	0	1/2	0	0	0	1/4
0	0	1	1/4	1	0	1	1/2	0	1	1	1/4
1	1	0	1/4					1	0	1	1/4
1	1	1	1/4					1	1	0	1/4

intrinsic = 1	intrinsic = 0	intrinsic = 0
shared = 0	shared = 1	shared = 0
synergistic = 0	synergistic = 0	synergistic = 1

FIG. 4. Flow decomposition of the canonical dependency types of Fig. 2: Each dependency type is associated with a unique flow mode: intrinsic with intrinsic, shared with shared, and synergistic with synergistic.

algebra now gives a full and constructive decomposition of the distinct flow modes:

Intrinsic Flow: $I[X_{-1} : Y_0 \downarrow Y_{-1}]$

Shared Flow: $I[X_{-1} : Y_0] - I[X_{-1} : Y_0 \downarrow Y_{-1}]$

Synergistic Flow: $I[X_{-1} : Y_0 | Y_{-1}] - I[X_{-1} : Y_0 \downarrow Y_{-1}]$

To illustrate how this works, Fig. 4 returns to the canonical dependency types of Section II and applies the decomposition to each. In the case of the Intrinsic pair of time series, we find that intrinsic flow is 1.0 bit while shared and synergistic are both 0.0 bit. The synchronized or Shared pair of time series decomposes with 1.0 bit of shared flow, with 0.0 bit of both intrinsic and synergistic flows. Finally, the Synergistic pair has no intrinsic or shared information flow, but 1.0 bit of synergistic flow. These observations justify the names given to the three modes.

V. RESULTS

With their definitions and estimation methods laid out, we now turn to demonstrate the diversity and advantage of quantifying separate modes of information flow in settings ranging from asymmetric flow to financial indices.

A. Asymmetric Information Flow?

At first blush, it seems difficult for there to be shared information flow from X to Y , but not from Y to X . Here, we provide a relatively simple example and an intuitive explanation of this phenomenon. Consider the following (jointly) Markovian transition matrix defining a pair of time series X and Y :

$$T_{X_{-1}, Y_{-1} \rightarrow X_0, Y_0} = \begin{matrix} & 0,0 & 0,1 & 1,0 \\ 0,0 & \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ 0,1 & & & \\ 1,0 & 4/9 & 5/9 & 0 \end{matrix}, \quad (5)$$

with stationary distribution $\pi = (2/11, 9/22, 9/22)$. The time series' temporal joint distribution is:

X_{-1}	Y_{-1}	X_0	Y_0	Pr
0	0	0	1	2/11
0	1	1	0	9/22
1	0	0	0	2/11
1	0	0	1	5/22

The joint time series has information flows from Y to X of 0.526 200 bit intrinsic, 0.449 821 bit shared, and 0.0 bit synergistic; and from X to Y of 0.0 bit intrinsic, 0.044 381 bit shared, and 0.120 759 bit synergistic. Note that from Y to X , the intrinsic information flow is equal to the transfer entropy, while from Y to X the intrinsic information flow is equal to the time-delayed mutual information.

To gain a more intuitive understanding of the asymmetry in this, let us isolate the transition matrix corresponding to the generation of the X time series:

$$T[X_{-1}, Y_{-1} \rightarrow X_0] = \begin{matrix} & 0 & 1 \\ 0,0 & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ 0,1 & & \\ 1,0 & 1 & 0 \end{matrix}. \quad (6)$$

In this instance, $I[X_{-1} : Y_{-1}] = 0.449 821$ bit while $I[X_{-1} : Y_{-1} | X_0] = 0$ bit. This indicates that the variables form a Markov chain $X_{-1} - X_0 - Y_{-1}$ and so any

information X_{-1} and Y_{-1} share is contained within X_0 . Therefore, the shared information flow is 0.449 821 bit. From this, the intrinsic information flow from Y to X is $I[Y_{-1} : X_0] - 0.449 821 \text{ bit} = 0.526 200 \text{ bit}$ and synergistic is $I[Y_{-1} : X_0 | X_{-1}] - 0.449 821 \text{ bit} = 0.0 \text{ bit}$.

Looking toward the Y time series, we find the following transition matrix:

$$T[X_{-1}, Y_{-1} \rightarrow Y_0] = \begin{matrix} & & 0 & 1 \\ & 0,0 & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ 0,1 & & & \\ 1,0 & \begin{pmatrix} 4/9 & 5/9 \end{pmatrix} & & \end{matrix} \quad (7)$$

Here, consider a locally-modified Y'_{-1} constructed by passing Y_{-1} through a channel that preserves the value 0, but maps a 1 to a 0 with probability $16/45$. This results in the modified transition matrix:

$$T[X_{-1}, Y'_{-1} \rightarrow Y_0] = \begin{matrix} & & 0 & 1 \\ & 0,0 & \begin{pmatrix} 4/9 & 5/9 \\ 1 & 0 \end{pmatrix} \\ 0,1 & & & \\ 1,0 & \begin{pmatrix} 4/9 & 5/9 \end{pmatrix} & & \end{matrix} \quad (8)$$

Given Y'_{-1} , X_{-1} , and Y_0 are independent. This implies that no information about Y_0 can be uniquely attributed to X_{-1} , since there is a method of reconstructing any influence X_{-1} has on Y_0 using Y_{-1} alone. We then conclude that the intrinsic information flow from X to Y is 0.0 bit. Shared information flow is then $I[X_{-1} : Y_0] = 0.044 381 \text{ bit}$, while synergistic flow is $I[X_{-1} : Y_0 | Y_{-1}] = 0.120 759 \text{ bit}$.

These jointly Markovian time series exemplify the degree of asymmetry that can exist in information flow. Specifically, it is not immediately obvious that shared information flow—due to common driving or synchronization, for example—can be large in one direction while small or nonexistent in the other. These sorts of asymmetries appear in a variety of data sets, and so its demonstration in a relatively simple Markovian setting is pedagogically helpful. These relationships are summarized in Fig. 5.

B. Financial Information Flows

We next analyze information flows between a financial index and its constituent stocks. The value of a financial index is the weighted average of the value of its constituent stocks. Here, we highlight our analysis of the Standard & Poor’s 500 (S&P 500), while App. E compares information flows in the S&P 400 and S&P 600 indices. The S&P 500 consists of 500 “large cap” stocks, whose total value is approximately \$23.9 trillion dollars or 80% of the

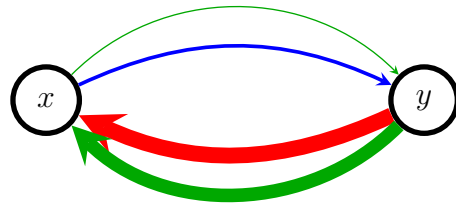


FIG. 5. A representation of the information flows between x and y . The color of an arrow corresponds to its mode: red for intrinsic, green for shared, and blue for synergistic. The width of an arrow corresponds to its strength: the wider the arrow, the larger that flow mode. We see that there is relatively little flow from x to y and its modes are shared and synergistic, while there is significantly more information flow from y to x and its modes are intrinsic and shared.

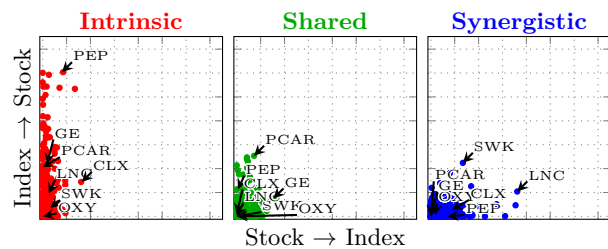


FIG. 6. Distinct information flows within the “large cap” S&P 500 financial index. Axis scales are identical with minimum 0.0000 bit and maximum 0.0126 bit.

US market. The other indices are smaller, engineered to reflect financial dynamics at other economic scales. The time series consist of the sign of the change in the daily closing price of each stock and the index between January 1st 2000 and December 31st 2008. We only include stocks whose symbol was in the index for the entirety of the date range. We estimate each information flow measure utilizing a past of length 1. These methods match those of Ref. [24], where the transfer entropy between the S&P 500 and its constituents was analyzed. As noted, interpreting the transfer entropy as an information flow is unclear.

To probe market behavior with our more refined scheme, we evaluate the intrinsic, shared, and synergistic information flows between each stock and its index. The S&P 500’s information flows are given in Fig. 6. The analysis immediately reveals that intrinsic information flow is heavily skewed: the index value drives many stock values, but individual stock values are not directly predictive of the index. Shared information flow also skews, but only slightly. Thus, there is common behavior to both stock values and the index value. This common behavior is predictive of a stock’s value, but less predictive of the index. Synergistic flow is also skewed, but in a different way. Here, there is some joint feature of the prior

values of the stock and index that is predictive of the index value, but not so of the stock. In Ref. [24], the asymmetry of the transfer entropy—the sum of intrinsic and synergistic flows—in the S&P 500 was noted. Our flow decomposition refines this asymmetry: stocks whose transfer entropy skewed more heavily in index-to-stock directly do so due to intrinsic information flow, while those that skew more in the direction of stock-to-index do so synergistically.

We can further analyze the behavior of specific stocks. PEP (PepsiCo), for example, is strongly driven—via intrinsic flow—by the behavior of the S&P 500. LNC (Lincoln Financial Group) is most strongly influenced synergistically by both its own past behavior combined with the past behavior of the S&P 500. PCAR (PACCAR Inc) is influenced in approximately equal measure by the S&P 500 intrinsically and in a shared fashion with its own behavior. OXY (Occidental Petroleum) operates virtually independently from the S&P 500, with exceedingly little information flow of any mode either to the stock from the index or *vice versa*.

The S&P 400 and S&P 600 have also been analyzed, and the results are in Appendix E. These mid- and small-cap stocks generally have less total information flow in both the stock-to-index and index-to-stock directions. Intrinsic flow is broader in these indices, without the significant asymmetry seen in the S&P 500. Shared flow, however, shows a strong asymmetry in the S&P 400 while is nonexistent in the S&P 600. Synergistic flow, again, shows none of the asymmetry that is seen in the S&P 500.

VI. CONCLUSION

Detecting and quantifying information flow is both important and ill-defined—proposals to date have led to ambiguous, misleading, or inconsistent interpretations of behavior and structure. Conceptually, information flow is the medium through which causality propagates. Here, we proposed that one of the primary impediments to successfully diagnosing information flow is that it is not a singular concept. Rather, information flow can take on several qualitatively distinct modes. The intrinsic information flow is the mode most closely aligned with prior intuitions, such as that motivating the transfer entropy [16].

To quantify the intrinsic information flow, we proposed the cryptographic flow ansatz that posits intrinsic flow is synonymous with the ability to construct a secret key. This obviated the infinite regress of conditioning on all of a system’s other, possibly unspecified, components and the effects arising from conflating conditional indepen-

dence and dependence. This enabled us to quantify intrinsic information flow using the intrinsic mutual information, an easily computed upper bound on the secret key agreement rate. With this in hand, the remaining flow modes are quantified with the aid of the time delayed mutual information and the transfer entropy.

When appealing to cryptographic secret key agreement rate, we made a choice to approximate it with the intrinsic mutual information. Though, tighter upper bounds on the secret key agreement rate exist, they are generally much more difficult to estimate [25]. This makes them generally impractical in all but the smallest and simplest of cases. There also exist lower bounds on the secret key agreement rate [25], though these too are computationally prohibitive for general practice. Presumably, using improved bounds would be justified by an application’s need for more accuracy.

Refinements aside, the distinct quantification of each mode of information flow is broadly applicable. Demonstrating its consistency and discriminating power, we computed the intrinsic, shared, and synergistic information flows for key base cases and between several financial indices and their constituent stocks. As a new lens into stock market dynamics, these led to a significantly more nuanced view of the interactions between individual companies and the market. For example, we discovered that those stocks whose transfer entropy from the S&P 500 is large are that way due to intrinsic flow; further there is no stock that intrinsically drives the S&P 500.

Additionally, shared information flow is often entirely neglected in analyses due to the prevailing opinion that transfer entropy supplants time-delayed mutual information whereas when considering information flow as multimodal the latter plays a first-class role. Without observations such as these it is impossible to paint a complete picture of how information is shuttled throughout a complex system. Looking forward, we believe that quantifying the distinct modes of information flow in an even broader variety of settings will lead to substantial improvements in our understanding how a system and its components interact to generate truly complex behavior.

ACKNOWLEDGMENTS

We thank D. Feldspar for helpful discussions. As an External Faculty member, JPC thanks the Santa Fe Institute and the Telluride Science Research Center for their hospitality during visits. This material is based upon work supported by, or in part by, John Templeton Foundation grant 52095, Foundational Questions Institute grant FQXi-RFP-1609, the U.S. Army Research Laboratory and the U. S. Army Research Office under contract

W911NF-13-1-0390 and grant W911NF-18-1-0028, and via Intel Corporation support of CSC as an Intel Parallel Computing Center. BZ was supported by the 2017 NSF Research Experience for Undergraduates Program and BDMA by a UC MEXUS-CONACYT Doctoral Fellowship.

Appendix A: Background: Times Series and Information

Let us summarize notation. We denote random variables using capital letters (X), realizations of random variables using lower case (x), and the event space of a random variable using calligraphics (\mathcal{X}). We denote a sequence of temporally-ordered random variables (a time series) using a Python-like slice notation: $X_t, X_{t+1}, X_{t+2}, \dots, X_{t+\tau-1} = X_{t:t+\tau}$. We suppress the starting or ending index of a slice if it is infinite; to wit, a bi-infinite time series is simply $X_{:}$. Throughout we assume stationarity— $X_{t:t+\tau} = X_{0:\tau}$ —and that time series are ergodic. A time series is IID when its random variables are independent and identically distributed.

We next review several fundamental information-theoretic measures; for a more detailed introduction please refer to any standard text; e.g., Refs. [1, 26, 27]. The central measure of information theory is a random variable's *entropy*:

$$H[X] = - \sum_{x \in \mathcal{X}} p(x) \log_2 p(x) . \quad (\text{A1})$$

The entropy of a joint variable is defined similarly:

$$H[X, Y] = - \sum_{x, y \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_2 p(x, y) . \quad (\text{A2})$$

These quantify the total amount of uncertainty that exists within a set of random variables. Given two random variables, the *conditional entropy* quantifies uncertainty of one given knowledge of the other:

$$H[X|Y] = H[X, Y] - H[Y] . \quad (\text{A3})$$

That is, it is the uncertainty “left over” after the uncertainty of Y is removed from the joint uncertainty of X and Y .

Entropy is, generally, subadditive. Its degree of subadditivity is known as the *mutual information* and quantifies

the dependence between two variables:

$$I[X : Y] = H[Y] + H[X] - H[X, Y] \quad (\text{A4})$$

$$= H[X, Y] - H[X|Y] - H[Y|X] \quad (\text{A5})$$

$$= \sum_{x, y \in \mathcal{X}, \mathcal{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} . \quad (\text{A6})$$

The *conditional mutual information* measures the additional change in uncertainty about Y given X , when given a third variable Z :

$$I[X : Y|Z] = H[Y|Z] - H[Y|X, Z] \quad (\text{A7})$$

$$= \sum_{x, y, z \in \mathcal{X}, \mathcal{Y}, \mathcal{Z}} p(x, y|z) \log_2 \frac{p(x, y|z)}{p(x|z)p(y|z)} . \quad (\text{A8})$$

Note that conditioning can increase statistical dependence; that is, $I[X : Y|Z] > I[X : Y]$. This reflects the fact that conditional mutual information is sensitive to both intrinsic dependencies between X and Y , as well as dependencies induced by Z . In other words, dependence between X and Y may be revealed through their relationship with Z . Such dependencies can occur even when X and Y are independent: $I[X : Y] = 0$.

References [28, 29] review how these elementary information quantities extend to measure randomness and correlation in time series.

Appendix B: Joint Analysis of Information Flow

To gain a greater understanding of how joint interactions among time series lead to the three modes of information flow, consider all variables of interest simultaneously. Figure 7 represents all interactions of X_{-1} , X_0 , Y_{-1} , and $\overline{Y_{-1}}$ in the form of an *I-diagram* [30]. Three of the regions are identically zero due to the constraints placed on $\overline{Y_{-1}}$; namely, that the variables form a Markov chain $X_{-1}Y_0 - Y_{-1} - \overline{Y_{-1}}$. The time delayed mutual information $I[X_{-1}:Y_0] = a + b + c$; the transfer entropy $I[X_{-1}:Y_0 | Y_{-1}] = a$; and the intrinsic flow $I[X_{-1}:Y_0 \downarrow Y_{-1}] = a + b$. Since the intrinsic mutual information is bound from above by the conditional mutual information, we conclude that information atom b is necessarily nonpositive.

As noted, intrinsic flow is given by $a + b$. Shared flow is then given by the time delayed mutual information minus the intrinsic flow, which is c . Synergistic flow is given the transfer entropy minus the intrinsic flow, that is, $-b$. Together, these measure the total flow or $a + c$.

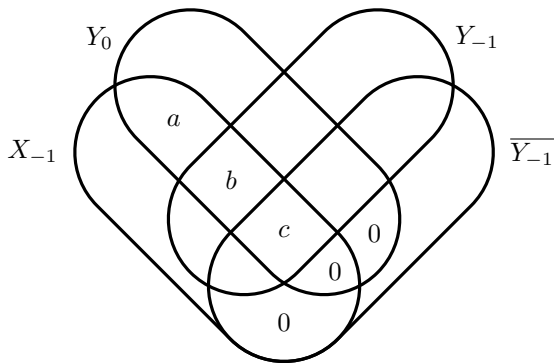


FIG. 7. Analysis of how the three measures of information flow are related. From the definition of intrinsic mutual information, $I[X_{-1}Y_0 : \bar{Y}_{-1}|Y_{-1}] = I[X_{-1} : \bar{Y}_{-1}|Y_{-1}] = I[Y_0 : \bar{Y}_{-1}|Y_{-1}] = 0$. Furthermore, $b \leq 0$. $I[X_{-1}:Y_0] = a + b + c$, $I[X_{-1}:Y_0|Y_{-1}] = a$, and $I[X_{-1}:Y_0 \downarrow Y_{-1}] = a + b$. From this, we can determine that intrinsic information flow is $a + b$, shared information flow is c , and synergistic flow is $-b$. Together, the total information flow is $a + c = I[X_{-1} : Y_0] - I[X_{-1} : Y_0 : Y_{-1}|\bar{Y}_{-1}]$.

Appendix C: Relationship With the Partial Information Decomposition

The separation of transfer entropy into two components is not a new idea. Utilizing the *partial information decomposition* [31] Williams *et al.* [32] decompose the transfer entropy from X to Y into two components: that *uniquely* provided by X_{-1} (“state independent transfer”), and that *synergistically* provided by both X_{-1} and Y_{-1} (“state dependent transfer”):

$$I[X_{-1}:Y_0|Y_{-1}] = I_{\cap}[X_{-1} \rightarrow Y_0 \setminus Y_{-1}] + \quad (C1)$$

$$I_{\cap}[X_{-1}Y_{-1} \rightarrow Y_0] . \quad (C2)$$

While our intrinsic and synergistic information flows are qualitatively very similar to this idea, there are some important distinctions. First, it is well-known that the intrinsic mutual information, used here to quantify intrinsic flow, is incompatible as a measure of unique information within the partial information decomposition [33]. Second, the goal in the partial information decomposition is to decompose (for example) $I[X_{-1}, Y_{-1} : Y_0]$. Here, we do not constrain our flows to any particular sum. This is in line with the lack of any conservation law governing information [34]. Furthermore, with the lack of widely accepted method of computing the partial information decomposition and the hints of incompleteness when the number of inputs exceeds two [35], there is always the unfortunate possibility that such an endeavor simply can not be realized.

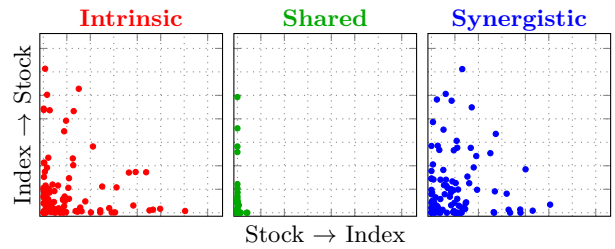


FIG. 8. Information flows within the “mid cap” S&P 400. Axis scales are identical with minimum 0.0000 bit and maximum 0.0025 bit.

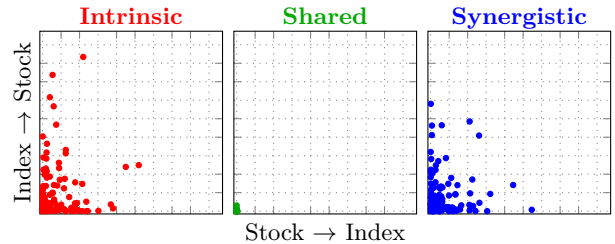


FIG. 9. Information flows within the “small cap” S&P 600. Axis scales identical with minimum 0.0000 bit and maximum 0.0032 bit.

Appendix D: Computing the Intrinsic Mutual Information

All computations were performed with `dit` [36]. Its implementation of the intrinsic mutual information utilizes the basin hopping method from `scipy.optimize`. While this calculation is straightforward for discrete probability distributions, there are significant difficulties in computing the intrinsic mutual information for continuous random variables. In essence, and barring a particularly clever and unforeseen method, one must establish a repertoire of potential transformations to apply to the variable Z . The conditional mutual information then must be minimized over the space of these transformations. Its evaluation can be performed using estimation methods such as the standard KSG estimator [37].

Appendix E: Financial Indices Detailed Analysis

Here, we compare our analysis of the S&P 500 financial index in the main text to flows detected in the S&P 400 and S&P 600 indices. The S&P 400 consists of 400 “mid cap” stocks, whose total value is approximately \$2.1 trillion dollars or 7% of the US market. The S&P 600 consists of 600 “small cap” stocks, whose total value is approximately \$896 billion dollars or 3% of the US market.

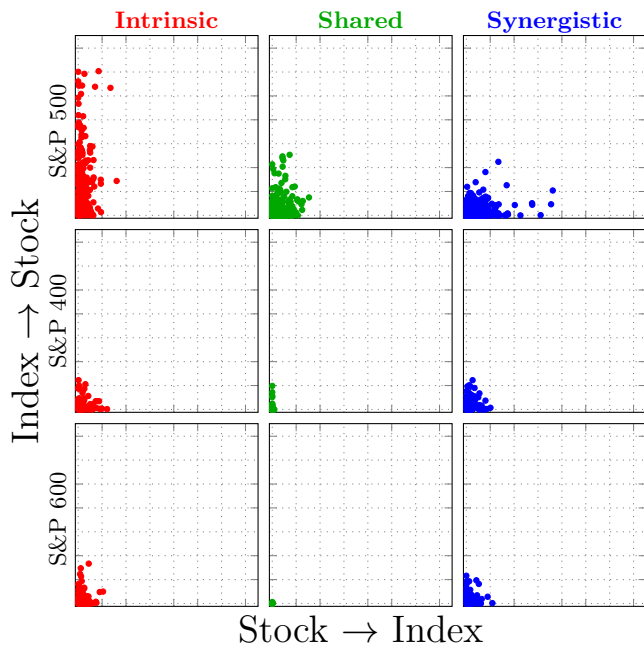


FIG. 10. Information flows in the S&P 500, 400, and 600 stock indices compared. Axis scales identical with minimum 0.0000 bit and maximum 0.0126 bit.

Again, the time series consist of the sign of the change in the daily closing price of each stock and the index between January 1st 2000 and December 31st 2008 for stocks whose symbol was in the index for the entirety of the date range. And, as for the S&P 500, we estimate

each information flow measure utilizing a past of length 1.

a. *S&P 400*

The information flows within the S&P 400 are illustrated in Fig. 8. These mid cap stocks display very different dynamics than the large cap S&P 500. Neither the intrinsic nor the synergistic flows display any marked asymmetry in directionality. The shared information flow, however, demonstrates a strong asymmetry where the stock and index are both predictive of the stock. Overall, the amount of information flow is significantly smaller than in the S&P 500.

b. *S&P 600*

The S&P 600 dynamics are again different from the large and mid cap stocks. The intrinsic flows are skewed in a similar manner as the S&P 500, though not as strongly. Shared information flow, however, is effectively nonexistent. Synergistic flows are largely symmetric, similar to the mid cap stocks.

The dynamics of all three indices are displayed on equal scales in Fig. 10.

-
- [1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, second edition, 2006. 1, 8
- [2] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge, United Kingdom, 2003. 1
- [3] W. Feller. *An Introduction to Probability Theory and its Applications*. Wiley, New York, third, revised edition, 1970. 1
- [4] R. S. Ingarden and K. Urbanik. Information without probability. *Colloq. Math.*, IX:131, 1962. 1
- [5] S. Kullback. *Information Theory and Statistics*. Dover, New York, 1968. 1
- [6] J. P. Crutchfield. Between order and chaos. *Nature Physics*, 8(January):17–24, 2012. 1
- [7] C. E. Shannon. The lattice theory of information. *Trans. IRE Prof. Group Info. Th.*, 1(1):105–107, 1953. 2
- [8] W. Ross Ashby. *An Introduction to Cybernetics*. John Wiley and Sons, New York, second edition, 1960.
- [9] R. Shaw. Strange attractors, chaotic behavior, and information flow. *Z. Naturforsch.*, 36a:80, 1981.
- [10] K. Kaneko. Lyapunov analysis and information flow in coupled map lattices. *Physica*, 23D:436, 1986.
- [11] J. M. Horowitz and H. Sandberg. Second-law-like inequalities with information and their interpretations. *New J. Phys.*, 16:1–29, 2014. 2
- [12] R. G. James, N. Barnett, and J. P. Crutchfield. Information flows? A critique of transfer entropies. *Phys. Rev. Lett.*, 116(23):238701, 2016. 2
- [13] J. Hlinka and J. Kořenek. Causal network discovery by iterative conditioning: comparison of algorithms. *arXiv:1804.08173*, 2018. 2
- [14] X. S. Liang. The Liang-Kleeman information flow: Theory and applications. *Entropy*, 15(1):327–360, 2013. 2
- [15] H. Kantz and T. Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge University Press, Cambridge, United Kingdom, 2004. 3
- [16] T. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85(2):461, 2000. 3, 7
- [17] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, first edition, 1991. 3
- [18] J. Sun and E. M. Bollt. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory

- couplings. *Physica D: Nonlinear Phenomena*, 267:49–57, 2014. 3
- [19] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier. *An Introduction to Transfer Entropy: Information Flow in Complex Systems*. Springer Intl. Publishing, New York, 2016. 3
- [20] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. Info. Th.*, 39(3):733–742, 1993. 4
- [21] D. Jost, U. Maurer, and J. L. Ribeiro. Information-theoretic secret-key agreement: The asymptotically tight relation between the secret-key rate and the channel quality ratio. *IACR Cryptology ePrint Archive*, 1130, 2017. 4
- [22] U. M. Maurer and S. Wolf. Unconditionally secure key agreement and the intrinsic conditional information. *IEEE Trans. Info. Th.*, 45(2):499–514, 1999. 4
- [23] M. Christandl, R. Renner, and S. Wolf. A property of the intrinsic mutual information. In *IEEE Intl. Symp. Info. Th.*, pages 258–258, 2003. 4
- [24] O. Kwon and G. Oh. Asymmetric information flow between market index and individual stocks in several stock markets. *EPL (Europhysics Letters)*, 97(2):28007, 2012. 6, 7
- [25] A. Gohari, O. Günlü, and G. Kramer. On achieving a positive rate in the source model key agreement problem. *arXiv:1709.05174*. 7
- [26] D. J. C. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press, Cambridge, United Kingdom, 2003. 8
- [27] R. W. Yeung. *Information theory and network coding*. Springer, New York, 2008. 8
- [28] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS*, 13(1):25–54, 2003. 8
- [29] R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *CHAOS*, 21(3):037109, 2011. 8
- [30] R. W. Yeung. A new outlook on Shannon’s information measures. *IEEE Trans. Info. Th.*, 37(3):466–474, 1991. 8
- [31] P. L. Williams and R. D. Beer. Nonnegative decomposition of multivariate information. *arXiv:1004.2515*. 9
- [32] P. L. Williams and R. D. Beer. Generalized measures of information transfer. *arXiv:1102.1507*. 9
- [33] N. Bertschinger, J. Rauh, E. Olbrich, and J. Jost. Shared information—new insights and problems in decomposing information in complex systems. In *Proceedings of the European Conference on Complex Systems 2012*, pages 251–269. Springer, 2013. 9
- [34] R. G. James, N. Barnett, and J. P. Crutchfield. Information flows? A critique of transfer entropies. *Phys. Rev. Lett.*, 116(23):238701, 2016. 9
- [35] J. Rauh. Secret sharing and shared information. *Entropy*, 19(11):601, 2017. 9
- [36] R. G. James, C. J. Ellison, and J. P. Crutchfield. dit: a Python package for discrete information theory. *The Journal of Open Source Software*, 3(25):738, 2018. 9
- [37] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys. Rev. E*, 69(6):066138, 2004. 9