Modes of Information Flow

Bahti Zakirov[1 2]

[1] *City University of New York, College of Staten Island, Department of Engineering Science and Physics*
[2] *City University of New York, Macaulay Honors College*

Abstract

Information flow is a highly useful concept for understanding the behavior of systems. There have been numerous attempts to quantify information flow, but there exists confusion about the meaning of these measures. We consider two common, though flawed measures of information flow, time delayed mutual information and transfer entropy, and demonstrate that it is erroneous to conflate the results given by these tools with what one is to intuitively believe constitutes information flow. We separate information flow into three modalities of shared, intrinsic, and conditional. In this context, we demonstrate that time delayed mutual information and transfer entropy actually turn out to provide combinations of *shared*, *conditional*, and *intrinsic* information flow, and that a third measure is needed to fully be able to disaggregate the types of information flow that exist within a system. We then propose a new measure, *intrinsic transfer entropy*, which utilizes intrinsic conditional mutual information from information theoretic cryptography. This provides the first concrete method of separating information flow into its *intrinsic*, *conditional*, and *shared* components. We apply intrinsic transfer entropy to a variety of systems to demonstrate its usefulness.

This work was done at the UC Davis NSF REU in the summer of 2017. My advisor was Professor Jim Crutchfield. I was supervised directly by Dr. Ryan James. Pretty much all of Crutchfield's complexity sciences center group were very helpful in some way.

1 Overview

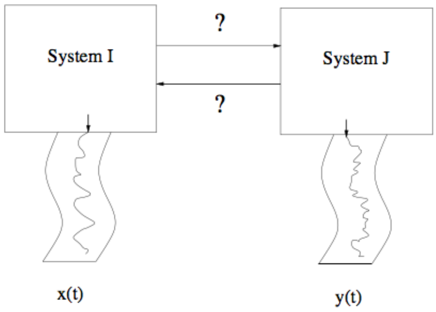

Figure 1: Image credit: Thomas Schreiber [1]

Information transfer is a cool concept. Suppose you are given two spooky black boxes both of which output time series data (as shown in figure 1). If one can ascertain that there is information transfer between the two time series, it follows that whatever mysterious processes are going on in the boxes one in some way is exhibits Granger causality towards the other [2]. By kicking the right box in this scenario, you can be absolutely certain that the time series of the other one will in some way be influenced. All of that can be obtained with no requirement of a model, and no undertanding of what is actually happening inside those boxes. The tricky part lies in quantifying this notion.

Currently, the state of the art is to view information transfer as a single entity that is measured in its entirety by *transfer entropy*. Recent results [3], however, have shown that it is erroneous to conflate the results of any of the current measures of information flow with the thing itself.

We assert that there is not one singular entity that is 'information transfer', but rather three separate modalities- the combination of which can be used to obtain the original desired definition. One purpose of this work is in part to build on the growing body of evidence that the most correct interpretion of information transfer is the one presented here.

2 Information

2.1 Information Theory

Shannon Entropy

$$H(X) = -\Sigma_p p \log_2 p \tag{1}$$

In Information Theory, probability distributions can be described in terms of Shannon Entropy. Shannon Entropy is roughly the amount of information that is required to fully describe the event. It is a quantification of the uncertainty involved in the value of a random variable. [4]

Outcome	Probability
Heads	$\frac{1}{2}$
Tails	$\frac{1}{2}$

The Shannon Entropy of the coin flip is given by

$$-(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}) = -1\ bit$$

So, given one bit of information about a coin flip, the outcome would be absolutely certain. A visual representation of Shannon entropy can be seen in figure 2.

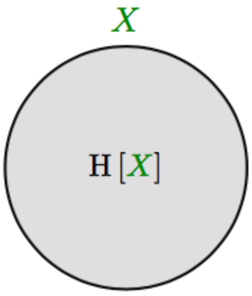

Figure 2: Since Shannon entropy corresponds to a cardinality of some set, it is natural to visualize information entropy quantities as Venn diagrams. Here is the Shannon entropy of one random variable X. [5]

Mutual Information Mutual Information is defined as follows

$$I(X:Y) = \Sigma_{x\in X}\Sigma_{y\in Y}p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)} \tag{2}$$

Mutual information is another fundamental concept in Information Theory. It is an aggregate quantity of two distributions that measures the ‘information overlap’ of the two distributions. Intuitively, it measures how much knowing one of the variables reduces uncertainty about the other.A visual representation of mutual information can be seen in figure 3.

Outcome	Probability
Heads Heads	$\frac{1}{2}$
Tails Tails	$\frac{1}{2}$

For example, if one flipped two coins that were taped together, knowing the outcome of one coin flip completely reveals the outcome of the second one. The mutual information of that coin system would be 1 bit, which is the entire uncertainty of the coin system. All of the information in this case is shared. Alternatively, two coins that are not taped together do not inform on each other's outcomes at all. Therefore, their mutual information would be 0.

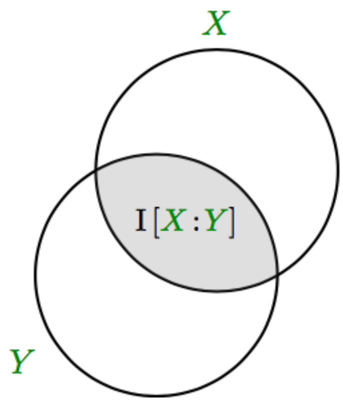

Figure 3: This figure demonstrates mutual information [5]

Joint Entropy

$$H(X,Y) = H(X) + H(Y) - I(X,Y) \tag{3}$$

The joint entropy is the total Shannon entropy of two distributions. The formula can be explained as follows: $H(X) = H(X|Y)+I(X,Y)$ and $H(Y) = H(Y|X)+I(X,Y)$ where the conditioning can be interpreted as 'excluding', or information contained in X that is unique to X and unknown to Y. So, $H(X,Y)$ can be read off as $H(XExcludingY) + H(YExcludingX) + 1 * I(X,Y)$.

Conditional Mutual Information

$$I(X:Y|Z) = \Sigma_{z\in Z}p(z)\Sigma_{x\in X}\Sigma_{y\in Y}p(x,y|z)\log_2\frac{p(x,y|z)}{p(x|z)p(y|z)} \tag{4}$$

Conditional Mutual Information is the Mutual Information between two distributions conditioned on a third distribution. It is the mutual information between X and Y 'given' Z. The meaning of this conditioning is oft misinterpreted. In many cases, conditioning on a third variable reduces the entropy. However, conditioning is not a purely a reductive operation. Conditioning can actually *induce* dependence between formerly independent time series. As an example, consider X and Y to be fair coin flips and Z to be their exclusive OR. In this case X and Y share no information ($I[X:Y]=0$), yet given knowledge of Z we know whether X and Y are equal or not, thus making X and Y have shared information ($I[X:Y|Z]=1$). This phenomenon of dependence induced by conditioning is dubbed 'Conditional/Synergistic Dependence' and is the source of synergistic information flow in the case where we begin using conditional mutual information to calculate transfer entropy. People's unawareness about the phenomenon of conditional dependence is part of the reason why it is not yet well known that information flow has multiple different modalities [3]. A few of the information measures can be seen in figure 4.

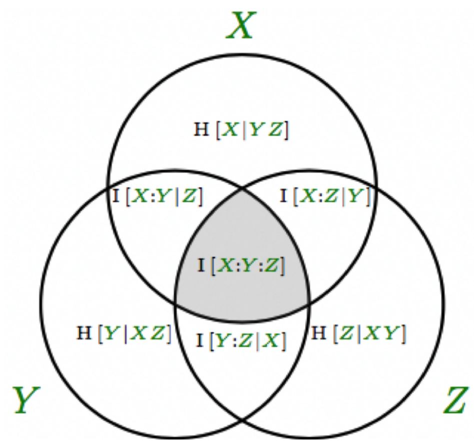

Figure 4: This figure demonstrates all of the above information measures using three random variables X,Y, and Z. [5]

Entropy Rate In a stochastic process:

$$h(\chi) = lim_{n->\infty}\frac{1}{n}H(\chi_1, \chi_2, ..., \chi_n) \tag{5}$$

In a time series it can be thought of as: [1]

$$h(\chi) = H(x_t : x_{t-1}x_{t-2}...x_{t-\infty}) \tag{6}$$

If the markov order, n, is finite, $h(\chi)$ only goes off up to n rather than infinity. It is important to note that we are switching from looking at random variables to talking about time series. Each x in this example corresponds to a different random variable that is generated by counting methods from a time series also called X. The entropy rate of a process is the mutual information between its past and its present. It is the answer to the question 'How does the entropy of this time series increase with each new step'. Although entropy rate is mathematically well defined for stochastic processes [4], what happens algorithmically here is that the time series is broken up into up to n-length chunks which all get turned into a distribution through counting methods, and the basic question that gets answered is how does the entropy of the sequence grow with n. Note that this is a function of the history length. In applying this to a time series, we assume that it is stationary Markov process of order n. For the purposes of this work, I have assumed Markov order 1 for the stocks and iterated prisoners dilemma, and Markov order 7 for the random Boolean networks..

2.2 Information Transfer

The story of information transfer starts with Time Delayed Mutual Information.

$$TDMI(X,Y) = I(X_t : Y_{t-1}) \tag{7}$$

Time Delayed Mutual Information is one of the first proposed measures for information transfer, and is sadly still used today. It was very quickly replaced by transfer entropy, and the reason for that is because it conflated synchronized, periodic processes with causation. For instance, if time series Y was an independent sine wave, and time series X was an independent cosine wave, time delayed mutual information would output that X is causing Y, or vice versa. However, this is clearly nonsensical, as we have just defined the two processes to be completely independent of one another. So, the way to interpret the result of TDMI is as a combination of *shared* flow and *intrinsic* flow- where shared flow is the portion that has been contaminated by the potentially periodic behavior.

TDMI's failure vis a vis synchronized, periodic processes prompted Thomas Schreiber to create a new measure [1] which he called transfer entropy.

$$TE(X,Y) = I(X_{t-1} : Y_t | Y_{t-1}) \tag{8}$$

Transfer entropy accounts for potential periodicity by conditioning on the past- which essentially removes everything we would have known about Y just from the past of Y alone. However, in conditioning on the Y_{t-1}, transfer entropy induces a new kind of dependence that corresponds to a synergistic [6] multivariate information interaction. Hence, Transfer Entropy consists of a combination of *synergistic* information flow plus *intrinsic* information flow.

Now, it is clear that if we are to disaggregate the three types of information flow, we require a third measure. Intrinsic Transfer Entropy is that third measure, based on intrinsic conditional mutual information from information theoretic cryptography, it finds the minimum transfer entropy across all possible 'fuzzings' of the variable being conditioned on, Z. ITE detects exclusively *intrinsic* information flow.

$$ICMI = I[X : Y \downarrow Z] = \min_{p(\bar{Z}|Z)} I[X : Y | \bar{Z}] \tag{9}$$

Given these three tools, we can filter out any form of information transfer that we wish.

- Shared Information Transfer = $TDMI_{X \to Y} - ITE_{X \to Y}$
- Synergistic Information Transfer = $TE_{X \to Y} - ITE_{X \to Y}$
- Intrinsic Information Transfer = $ITE_{X \to Y}$

3 Example Cases

In the examples below, we demonstrate some scenarios in which the distinction between the modes of information flow may be of importance. All information measures are contained in the dit python library. [7]

3.1 Information Transfer in Stocks and Indices

An index is a financial instrument that consists of stocks. It is a single entity that is entirely determined by the combination of its constituents. Standard and Poor's 500, for example, is an index containing about 500 stocks. Its market price is a weighted sum of the market prices of its constituents.

In a previous work [8], Kwon et al. measured the transfer entropy of the time series data of discretized closing prices of stocks, and discretized closing prices of the indices that they belong to. Instead of floating point numerical values, the numbers 1,0, and -1 were used for when the price increased, stayed the same, or decreased. These data spanned eight years, starting in January 2000 and ending in December 2008. Their results (shown below in figure 5) indicate that there is an asymmetry in the transfer entropies- with the index value driving the value of the stock more than the stock driving the value of the index.

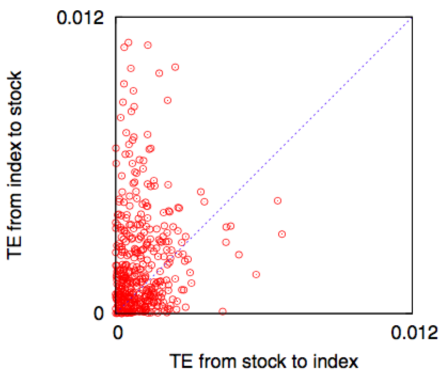

Figure 5: This image is credit to Kwon et al. Each data point is a particular stock in S&P500, where the x-coordinate is the transfer entropy from stock to index and the y-coordinate is that from index to stock. The prominent feature of this graph is the asymmetry of transfer entropy. One possible explanation of this phenomenon is that people looking to invest their money tend to watch the value of the indices, rather than that of each individual stock. In this way, it's possible that the value of an index influences investor behavior, and as a consequence, stock value.

Kwon and Oh concluded that there exists 'downward causation' in the stock-index system. The emergent index is more significant as a driver to the system than its constituents.

They operated under the assumption that Transfer Entropy was equivalent to information transfer. However, the conclusions they drew indicate that what they were really trying to get at was what we now call intrinsic information transfer. If, for example, the only asymmetry was in the synergistic flow- it would be impossible to attribute these results to downward causation. So, to test their conclusion under our interpretation of what information transfer really is, we repeated their analysis (seen in figure 6).

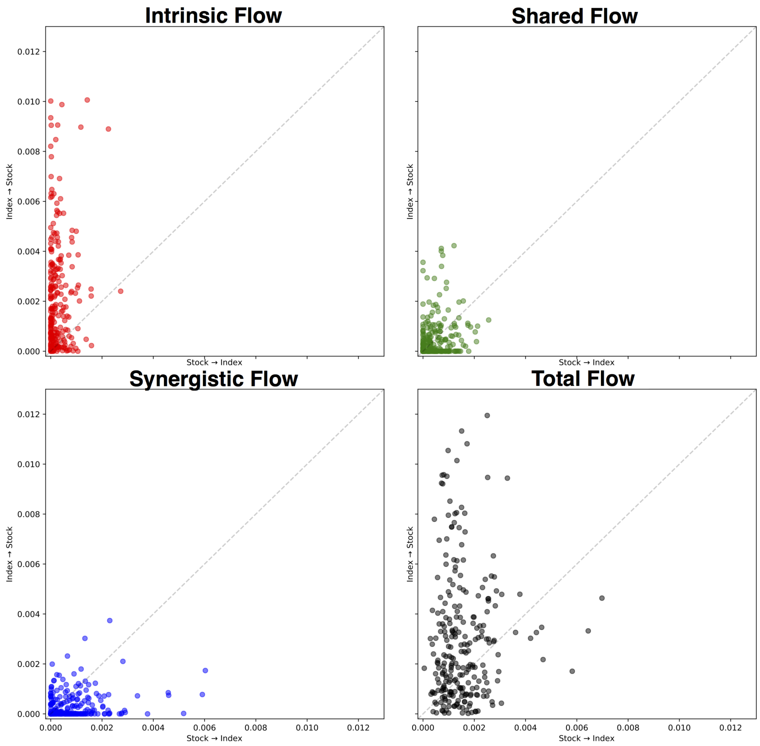

Figure 6: This is the decomposition of Kwon and Oh's analysis into the different modalities of information flow. It is clear that the intrinsic information flow, the one that is conceptually closest to what Kwon and Oh seemed to be trying to measure, is in fact even more strongly asymmetrical than presented in the original analysis. Although these results indicate that there is in fact a strong asymmetry in the information dynamics of stocks and indices, this asymmetry is specifically in the intrinsic information flow. Notice, however, that the shared flow is not asymmetrical, and that the synergistic flow appears to be asymmetrical in the opposite direction. In this particular scenario, the intrinsic flow is probably the correct mode to consider- but this figure clearly shows that there is more to information transfer than just transfer entropy. [5]

3.2 Information Transfer in Random Boolean Networks

A random Boolean network (RBN) is a collection of discrete Boolean states with discrete time evolution. Each node in the network has a state, and a certain number of connections drawn from a Poisson distribution for each node. The time evolution of these nodes is determined by a random lookup table, which is unique to every node. At every time step, each node is updated based on the state values of the other nodes it is randomly connected to. Since the connections remain constant for all time steps of a particular network, the system often enters attracting fixed points, limit cycles, and sometimes chaotic attractors that characterize the network dynamics. An interesting feature of RBN's is that they undergo a phase transition from an ordered phase to a chaotic phase as one varies the mean of the Poisson distribution (aka the connectivity) that determines the number of connections each node has. As the network gains connectivity, at around the 2.7 point, the dynamics of the RBN turn chaotic. These effects can be seen in figure 7.

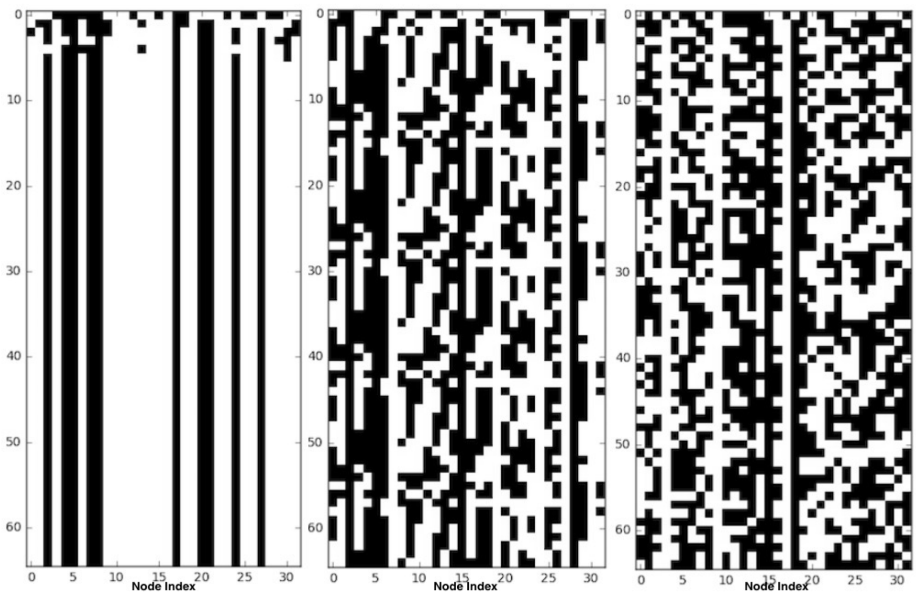

Figure 7: This image depicts some archetypal RBN state diagrams. The left RBN is of connectivity value K=1. The middle network is on 'the edge of chaos' [9], with a connectivity value exactly at 2.7. The rightmost network has a connectivity value of 3, and is well into the chaotic phase. One can see, looking from up to down, the fixed behavior of the node in a stable phase, the somewhat periodic behavior of the critical node, and the almost entirely unpredictable behavior of the one on the right.

Information Dynamics Lizier et al propose that global information transfer in the RBN's peaks at their critical connectivity values [9]. In their work, they set up thousands of RBN's with length 50 at various connectivity values and evolved them for 250 time steps. They then calculated a global aggregate transfer entropy for each network- where the state histories of 50 pairs of nodes were randomly sampled from each network, and the sum of their transfer entropy values was considered as a 'Global Transfer Entropy'. It turned out that in fact Transfer Entropy does peak at the critical value of connectivity that marks the phase transition from ordered to chaotic. However, once again the Transfer

Entropy tool has been conflated with the concept of information transfer. So, we constructed some RBN's (results seen in figure 8).

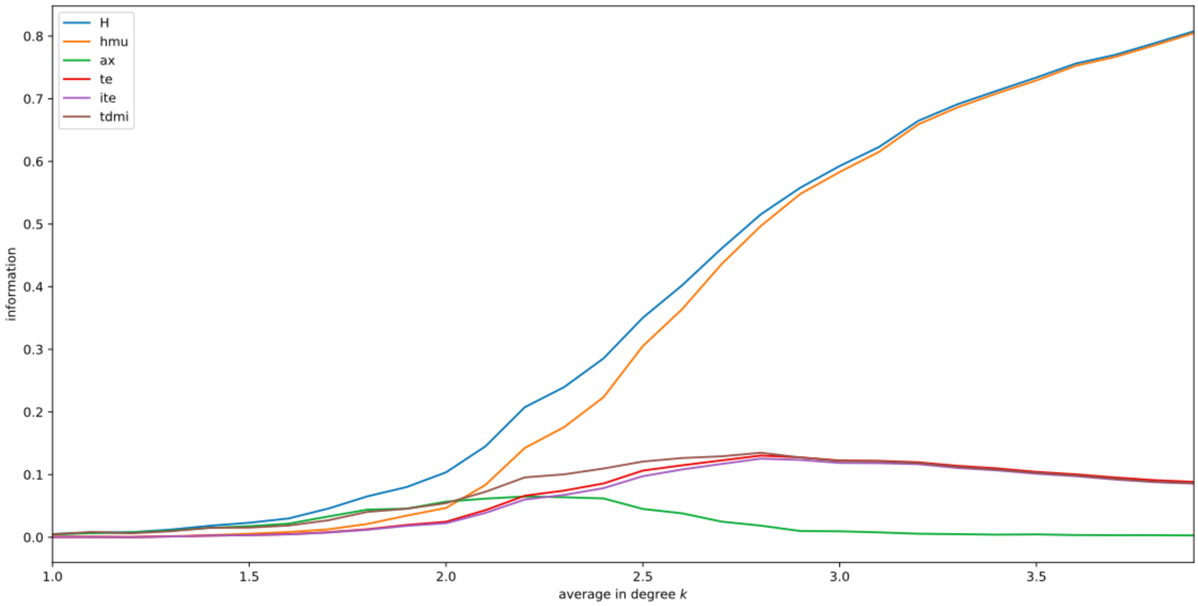

Figure 8: In the scenario of random Boolean networks, it appears that the intrinsic information flow peaks at approximately the critical value of connectivity. Therefore, the transfer entropy values consisted mostly of their intrinsic components. So, a different interpretation of Lizier et al.'s conclusion would be that intrinsic information flow peaks at the 'edge of chaos'.

3.3 Future Work: Information Transfer in the Iterated Prisoners Dilemma

The Game Iterated Prisoners Dilemma is a game where N agents are forced to interact with each other in a specific way. The premise of the game is based on the one turn prisoners dilemma, where the optimal strategy is famously to defect. In IPD, what happens is that each player is entered into a tournament where one by one they are forced to play a prisoners dilemma game with each other player for a certain number of rounds. Just like in one shot prisoners dilemma, their options in the game are to cooperate or defect. For each round in a single game, the action of the first player and the action of their opponent are input to a payoff matrix, which then assigns them points. Ultimately, the goal is to acquire as many points as possible during each game with the other players in order to win the tournament. Avoiding the less important details, the payoff matrix most commonly used with this game is

	Player 2 Cooperates	Player 2 Defects
Player 1 Cooperates	2,2	3,-1
Player 1 Defects	-1,3	0,0

The payoff matrix rewards the defector, just like the one shot prisoners dilemma. However, the important feature of this game is that repeated interactions incentivize long term cooperation. Therefore, there is a unique requirement placed on would-be strategies to both attempt to cooperate while at the same time not falling prey to other defecting players.

This game originally gained popularity when political scientist Robert Axelrod hosted a public contest to submit the best strategy for this game to be pitted against the rest of them. The winning submission in Axelrod's original tournament was a strategy called Tit for Tat- which very simply copied the opponents last move. The phenomenon of cooperative strategies being successful in IPD came to be known as the emergence of cooperation after Axelrod published his book about it titled 'The Evolution of Cooperation'. [10]

The Dynamic In this version of IPD, we create an evolutionary component by having the lowest scoring player copy the strategy of the highest scoring player at the end of each tournament. This way, a time series is generated of the past strategies for each player. This is not the only type of evolutionary dynamic that is possible within an IPD game. One popular such dynamic is called a Moran Process, where an individual is selected stochastically based on their score (to promote fitness of winners) every round to copy the strategy of another stochastically selected individual. We chose not to use this process because the non stochastic version is more intuitive, runs significantly faster, and is not likely to change the final outcome for our purposes.

The Strategies In this experiment, we drew multiple random samples of size 10 from a pool of approximately 200 strategies in the Axelrod library. [11] The pool spanned all of the popular strategies, including Press and Dyson's Zero Determinant extortionate strategies [12], as well as more obscure ones that came to being as a result of experimentation with genetic algorithms [13]. One instance of a typical game can be seen below in figure 9. A complete list of strategies can be found in the documentation for the Axelrod python library. [11].

Analysis What we did here was run the different information measures between every pair of players- where the players correspond to the columns in figure 9. Each time series entry is a strategy that the player has at a given time step. The results of the analysis can be seen in figure 10.

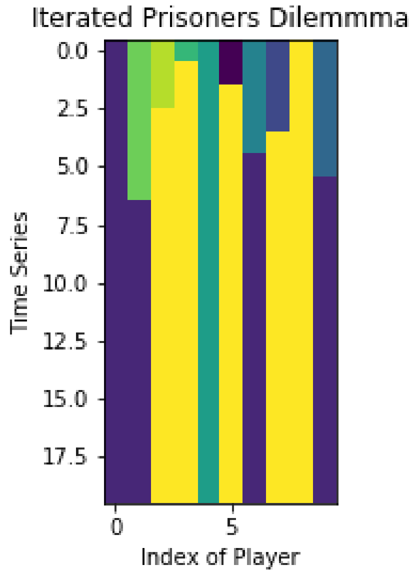

Figure 9: This figure displays the time series of all players in a typical game. Each color in this image corresponds to a different strategy, and each player's entire history of strategies becomes their time series. It appears as though in this one, a couple of dominant strategies have taken over the game and crowded out the others.

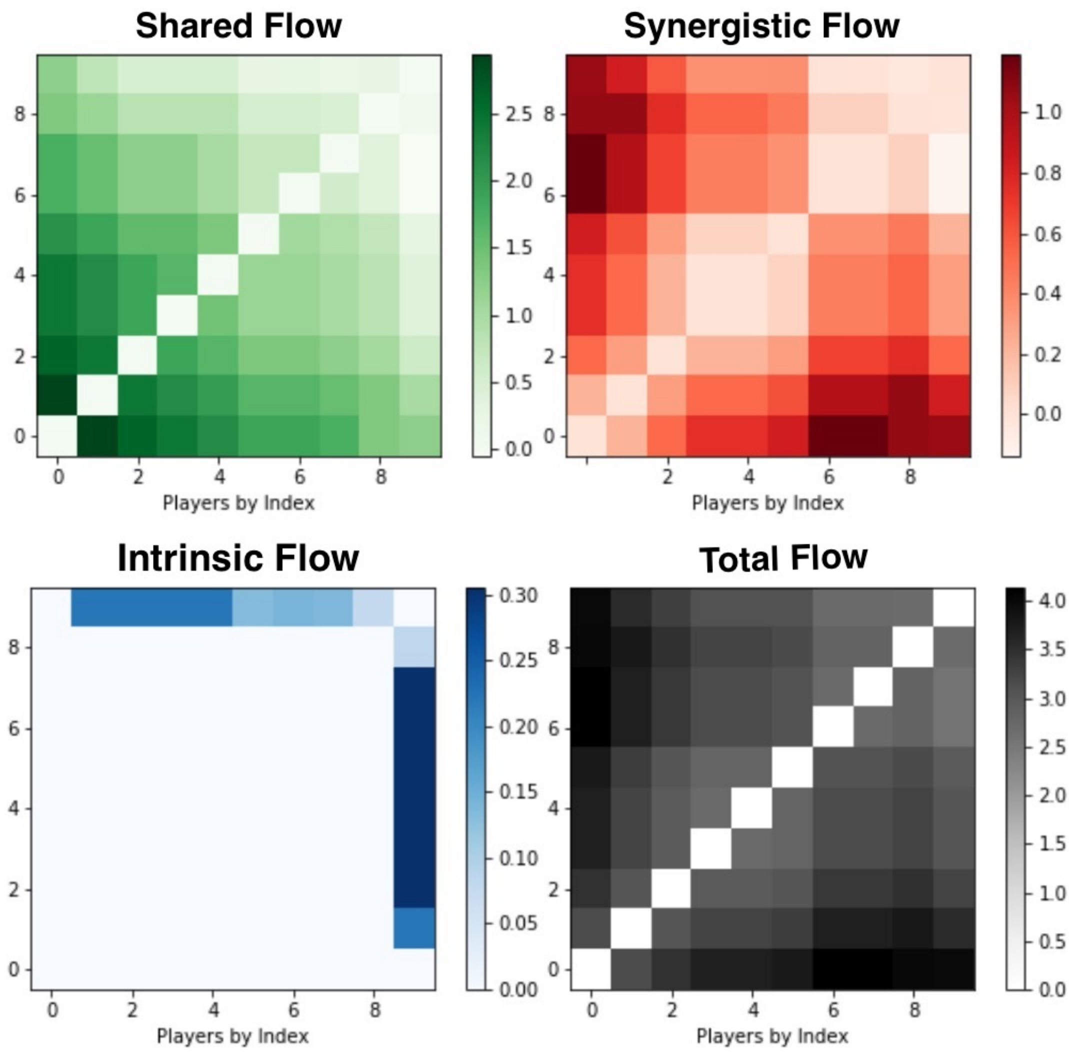

Figure 10: This figure displays all of the different modalities of information flow on a pairwise basis. First, notice how the shared and synergistic flows are symmetric across the diagonal. This symmetry is not present in the intrinsic flow. It is also clear that the magnitude of the shared flow is far greater than that of the other two. This is expected, since shared flow tends to overshoot in the presence of synchronized, periodic behavior. Another significant feature is that only one of the players has any sort of intrinsic flow at all- the player of index 9 that spent most of its time in the 'Gradual' strategy [13]. Perhaps this indicates that the Gradual strategy is in some way more influential over the dynamics of the game than the other players.

4 Conclusion

In this work, we have unpacked the concept of information flow into three distinct modalities. We have demonstrated that these modalities have different forms (such as the stocks), and may have distinct non overlapping utilities. Ultimately, we have substantiated James et al.'s [3] suspicion that there is neither one unified notion of information transfer, nor a single tool that can measure it.

Acknowledgements This work was supported by NSF REU PHY-1560482 I thank Dr. Ryan James, intrepid leader, apex postdoc, world's okayest dad
I thank Professor Jim Crutchfield for the opportunity to work in his lovely cohort, Professor Rena Zieve for engineering a perfect summer for seventeen nerds, Seventeen nerds for their friendship, and Zuni for his love and support.

References

[1] Schreiber, T. (2000). Measuring information transfer. Physical Review Letters, 85(2):461464

[2] arXiv:0910.4514v2

[3] James, R.G.; Barnett, N.; Crutchfield, J.P. Information flows? A critique of transfer entropies. Phys. Rev. Lett. 2016, 116, 238701.

[4] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 1991

[5] Image credit to Ryan James

[6] arXiv:1004.2515v1

[7] Dit python library docs.dit.io

[8] 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.

[9] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. The information dynamics of phase transitions in random Boolean networks. In S. Bullock, J. Noble, R. Watson, and M. A. Bedau, editors, Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems (ALife XI), Winchester, UK, pages 374381. MIT Press, Cambridge, MA, 2008.

[10] https://en.wikipedia.org/wiki/The_Evolution_of_Cooperation

[11] Axelrod Python Library Axelrod.readthedocs.io

[12] Press W. & Dyson F. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. Proc. Natl. Acad. Sci. USA 109, 1040910413 (2012).

[13] http://www.lifl.fr/IPD/references/from_lifl/alife5/html/