# An Introduction to Transfer Entropy

Terry Bossomaier · Lionel Barnett
Michael Harré · Joseph T. Lizier

# An Introduction to Transfer Entropy

## Information Flow in Complex Systems

Springer

Terry Bossomaier
School of Computing and Mathematics
Charles Sturt University
Bathurst, NSW
Australia

Lionel Barnett
Department of Informatics
University of Sussex
Brighton
UK

Michael Harré
Department of Civil Engineering
University of Sydney
Darlington, NSW
Australia

Joseph T. Lizier
Department of Civil Engineering
University of Sydney
Darlington, NSW
Australia

ISBN 978-3-319-43221-2 ISBN 978-3-319-43222-9 (eBook)
DOI 10.1007/978-3-319-43222-9

Library of Congress Control Number: 2016954697



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

This book is aimed at advanced undergraduate and graduate students across a wide range of fields, from computer science and physics to the many current and potential application areas of transfer entropy. Other researchers interested in this new and fast-growing topic will also find it useful, we hope.

It sits at the nexus of information theory and complex systems. The science of complex systems has been steadily growing over the last few decades, with a range of landmark events, such as the formation of the Santa Fe Institute in 1984, and the fundamental work of physics Nobel Laureates Murray Gell-Mann and Phillip Anderson. But precisely defining complex systems proved illusive. There are many examples, properties, ways of simulating and a diversity of theoretical suggestions. But it is only after 30 years that the pieces are finally falling into place.

Information theory, dominated by Claude Shannon's mathematical theory of communication, was one of the great theoretical ideas of the 20th century. It proved a valuable tool in analysing some complex systems, but it was only much later, with Schreiber's transfer entropy, that the relationship between information flow and complexity became apparent.

This book, like any complex system, emerged in parallel, with the synchronisation of ideas and thinking of the four authors. Terry's involvement in information theory goes back a very long way to its use in understanding images and animal vision. But he became interested in complex systems two and a half decades ago and the possibility that information theory would be a key tool was always in the background.

It was through the neuroscience dimension that Terry met Mike, while he was a PhD student at the Centre for the Mind at the University of Sydney. While working there Mike collaborated with David Wolpert of NASA Ames and it was David who introduced Mike to maximum entropy techniques and their application to economic game theory. This collaboration lead to several key findings regarding tipping points in microeconomics, 'persona choice' in behavioural game theory, and contributed significantly to Mike's PhD. During this time Mike also developed the idea of using mutual information as a tool to study financial market crashes in the same way that mutual information had been used to characterise phase transitions in physics.

Terry's collaboration with the University of Sussex began in the mid-1990s, but he and Lionel did not actually engage in any detailed discussions until the Artificial Life Conference in Lisbon in 2007. Lionel, along with Anil Seth, had been working on causality measures, particularly with applications to neuroscience and consciousness, for some while before getting interested in transfer entropy. Lionel then began a series of annual month-long visits to the Centre for Research in Complex Systems at Charles Sturt University, where some of the research in this book had its genesis.

Joe, meanwhile, had been working on transfer entropy during his PhD, finding some extraordinary results for simple systems, such as cellular automata. Although Terry and Joe met in Lisbon, it was not until the IEEE ALife conference in Paris that any sort of real dialogue began. In many ways, that conference was instrumental in formulating the ideas which led to this book.

The structure of the book is a bit like stone fruit, with a soft wrapping of a hard core, although the non-mathematical reader might find it something like climbing a mountain. After a qualitative introduction, Chap. 2 introduces ideas of statistics, which will be familiar to many readers. The going then gets tougher, or at least more mathematical, reaching its zenith in Chap. 4 where the main ideas of transfer entropy are worked out. We adopt Knuth's dangerous bend symbol, ☡ and ☡☡. The reader already familiar with information theory could perhaps go straight to Chap. 4, but other readers would need the background in Chap. 3. The later chapters of the book introduce a variety of applications, from simple, canonical systems to finance and neuroscience. The full details of Chap. 4 are not necessary to get an idea of the kind of applications covered. Transfer entropy is hard to calculate from real data. Some robust software is now available and new applications are appearing at an increasing rate.

Many people have been influential over the years in the development of this book, and we thank them all. Alan Kragh and John Lewis at Ilford Ltd. gave much encouragement to Terry in the pursuit of theoretical metrics for imaging science. The seminal work by Linfoot and Fellgett was pivotal at that time, although Terry never had the opportunity to meet either. But his real work in information theory began at the Australian National University with Allan Snyder FRS, Mike's PhD supervisor years later. His interest in complexity was stimulated by collaboration with David Green in the 1990s.

Lionel has been supported by the Sackler Centre at the University of Sussex, led by Anil Seth, with whom he has published extensively.

Joe was introduced to complex systems by Terry Dawson, while at Telstra Research Laboratories. This interest was fused with information theory under the guidance of Mikhail Prokopenko, then at CSIRO, now at the University of Sydney. Mikhail played a pivotal role in supervising Joe's PhD, also under Albert Zomaya at Sydney. Joe's work on information theory continued in his postdoc years at the Max Planck Institute for Mathematics in the Sciences in Leipzig, Germany, with Juergen Jost.

With regards to this book, Joe thanks in particular Michael Wibral, Juergen Pahle, Greg Ver Steeg and Mikhail Prokopenko for valuable discussions, comments and feedback on draft material.

The authors thank Carolyn Leeder for administrative assistance.

Some of the original research by the authors described in the book was funded by the Australian Research Council.

This book would have taken ten times as long to produce had it not been for Donald Knuth's TEX mathematical typesetting package and Leslie Lamport's LATEX. We use GNUPlot frequently, and Terry uses Emacs extensively almost every day. So thanks, also, to Richard Stallman.

# Contents


**1 Introduction** ... 1
1.1 Information Theory ... 2
1.2 Complex Systems ... 2
1.2.1 Cellular Automata ... 3
1.2.2 Spin Models ... 4
1.2.3 Oscillators ... 5
1.2.4 Complex Networks ... 5
1.2.5 Random Boolean Networks ... 7
1.2.6 Flocking Behaviour ... 7
1.3 Information Flow and Causality ... 9
1.4 Applications ... 10
1.5 Overview ... 10

**2 Statistical Preliminaries** ... 11
2.1 Set Theory ... 12
2.2 Discrete Probabilities ... 13
2.3 Conditional, Independent and Joint Probabilities ... 14
2.3.1 Conditional Probabilities ... 14
2.3.2 Independent Probabilities ... 14
2.3.3 Joint Probabilities ... 15
2.3.4 Conditional Independence ... 16
2.3.5 Time-Series Data and Embedding Dimensions ... 17
2.3.6 Conditional Independence and Markov Processes ... 18
2.3.7 Vector Autoregression ... 20
2.4 Statistical Expectations, Moments and Correlations ... 20
2.5 Probability Distributions ... 22
2.5.1 Binomial Distribution ... 22
2.5.2 Poisson Distribution ... 23
2.5.3 Continuous Probabilities ... 24
2.5.4 Gaussian Distribution ... 25
2.5.5 Multivariate Gaussian Distribution ... 25

2.6 Symmetry and Symmetry Breaking . . . 28

**3 Information Theory** . . . 33
3.1 Introduction . . . 33
3.2 Basic Ideas . . . 35
3.2.1 Entropy and Information . . . 35
3.2.2 Mutual Information . . . 38
3.2.3 Conditional Mutual Information . . . 42
3.2.4 Kullback–Leibler Divergence . . . 43
3.2.5 Entropy of Continuous Processes . . . 45
3.2.6 Entropy and Kolmogorov Complexity . . . 50
3.2.7 Historical Note: Mutual Information and Communication . . . 50
3.3 Mutual Information and Phase Transitions . . . 51
3.4 Numerical Challenges . . . 52
3.4.1 Calculating Entropy . . . 53
3.4.2 Calculating Mutual Information . . . 59
3.4.3 The Non-stationary Case . . . 63

**4 Transfer Entropy** . . . 65
4.1 Introduction . . . 65
4.2 Definition of Transfer Entropy . . . 66
4.2.1 Determination of History Lengths . . . 69
4.2.2 Computational Interpretation as Information Transfer . . . 72
4.2.3 Conditional Transfer Entropy . . . 74
4.2.4 Source–Target Lag . . . 77
4.2.5 Local Transfer Entropy . . . 77
4.3 Transfer Entropy Estimators . . . 78
4.3.1 KSG Estimation for Transfer Entropy . . . 79
4.3.2 Symbolic Transfer Entropy . . . 80
4.3.3 Open-Source Transfer Entropy Software . . . 81
4.4 Relationship with Wiener–Granger Causality . . . 82
4.4.1 Granger Causality Captures Causality as Predictive of Effect 83
4.4.2 Definition of Granger Causality . . . 83
4.4.3 Maximum-Likelihood Estimation of Granger Causality . . . 86
4.4.4 Granger Causality Versus Transfer Entropy . . . 88
4.5 Comparing Transfer Entropy Values . . . 90
4.5.1 Statistical Significance . . . 90
4.5.2 Normalising Transfer Entropy . . . 91
4.6 Information Transfer Density and Phase Transitions . . . 92
4.7 Continuous-Time Processes . . . 93

**5 Information Transfer in Canonical Systems** . . . 97
5.1 Cellular Automata . . . 98
5.2 Spin Models . . . 104
5.3 Random Boolean Networks . . . 106

5.4 Small-World Networks . . . . . 111
5.5 Swarming Models . . . . . 115
5.6 Synchronisation Processes . . . . . 119
5.7 Summary . . . . . 122

**6 Information Transfer in Financial Markets** . . . . . 125
6.1 Introduction to Financial Markets . . . . . 126
6.2 Information Theory Applied to Financial Markets . . . . . 128
6.2.1 Entropy and Economic Diversity: an Early Ecology of Economics . . . . . 128
6.2.2 Maximum Entropy: Maximum Diversity? . . . . . 129
6.2.3 Mutual Information: Phase Transitions and Market Crashes . 129
6.3 Information Transferred from One Market Index to Another . . . . . 130
6.4 From Indices to Equities and from Equities to Indices . . . . . 133
6.4.1 Economics of Beauty Pageants . . . . . 134
6.5 The Internal Economy and Its Place in the Global Economy . . . . . 135

**7 Miscellaneous Applications of Transfer Entropy** . . . . . 139
7.1 Information Transfer in Physiological Data . . . . . 139
7.2 Effective Network Inference . . . . . 143
7.2.1 Standard Pairwise TE Approach for Effective Network Inference . . . . . 144
7.2.2 Addressing Redundancy and Synergy in the Data . . . . . 145
7.2.3 Applications of Effective Network Inference . . . . . 148
7.3 Applications in Neuroscience . . . . . 149
7.3.1 TE for Pulse Sequences . . . . . 149
7.3.2 Direct TE Estimation Between Spiking Neurons . . . . . 151
7.3.3 TE in Brain Imaging . . . . . 152
7.4 Information Transfer in Biochemical Networks . . . . . 153
7.5 Information Transfer in Embodied Cognitive Systems . . . . . 157
7.6 Information Transfer in Social Media . . . . . 162
7.7 Summary . . . . . 164

**8 Concluding Remarks** . . . . . 167
8.1 Estimation . . . . . 167
8.1.1 Non-parametric Estimation . . . . . 167
8.1.2 Parametric Estimation . . . . . 168
8.1.3 Non-stationary Systems . . . . . 169
8.2 Systems with Many Variables . . . . . 169
8.3 Touching the Void: the Link to Thermodynamics . . . . . 170

**References** . . . . . 171

**Index** . . . . . 187

# List of Key Ideas


**Key Idea 1** We can accurately reconstruct the *state* of a $d$-dimensional, non-linear dynamical system $y_t = f(\mathbf{x}_t)$ by observing the $\mathbf{m} : d \leq \mathbf{m} \leq 2d + 1$ past data points of the one-dimensional time series $y_t$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

**Key Idea 2** The information of information theory has nothing to do with meaning. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

**Key Idea 3** Shannon information is a property of sets of objects, not the objects themselves. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

**Key Idea 4** All the system-level information-theoretic quantities may be expressed as expectation values over the pointwise (local) quantities. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37

**Key Idea 5** Mutual information is the total marginal entropy minus the joint entropy, or the Kullback–Leibler divergence of the product of marginal distributions from the joint distribution. . . . . . . . . . . . . . . . 39

**Key Idea 6** The properties of the differential entropy can be counter-intuitive in comparison with those of the Shannon entropy (of discrete variables); e.g. it can be negative. . . . . . . . . . . . . . . . . . . . . 45

**Key Idea 7** Other information-theoretic terms (e.g. conditional entropies, MI and conditional MI) applied to multivariate distributions may be formed as the sums and differences of the underlying entropy terms (with each evaluated as per Eqn. 3.25). . . . . 46

**Key Idea 8** Crucially, the differential MI (and conditional MI) has certain properties matching those for discrete variables (i.e. being non-negative), and does not change with scaling of the variables. . . . . 46

**Key Idea 9** The MI between two Gaussian variables is completely determined by their correlation coefficient $\rho$ in Eqn. 3.34, increasing with the magnitude of $\rho$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47

**Key Idea 10** Mutual information peaks at a second-order phase transition, across very many systems. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

**Key Idea 11** Naively calculating information from frequency estimates is just that, naive! . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

**Key Idea 12** There is a trade-off between bias and variance in the calculation of entropy. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

**Key Idea 13** Calculating mutual information is tricky and needs to be validated case by case. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60

**Key Idea 14** The key innovation of the KSG algorithm is getting the numerical errors to partially cancel in the marginal and joint entropy estimates. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62

**Key Idea 15** Schreiber and Paluš' insight was that, to assess the influence of the past of $Y$ on current $X$, the shared information between $X$ and its own past must be accounted for. . . . . . . . . . . . . . . . 67

**Key Idea 16** $\mathbf{T}_{Y \to X}(t)$ with lag 1 may be interpreted intuitively as the degree of uncertainty about current $X$ resolved by past $Y$ and $X$, over and above the degree of uncertainty about current $X$ already resolved by its *own* past alone. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68

**Key Idea 17** Transfer entropy measures how much information the source process provides about state transitions in the target. . . . . . . . . 70

**Key Idea 18** $\mathbf{T}^{(k,\ell)}_{Y \to X}(t)$ may be interpreted intuitively as the degree of uncertainty about current $X$ resolved by the past *states* $Y$ and $X$, over and above the degree of uncertainty about current $X$ already resolved by its *own* past *state* alone. . . . . . . . . . . . . . . . . . . . . . . . . . . . 72

**Key Idea 19** Information transfer and causality are related but distinct concepts. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74

**Key Idea 20** $\mathbf{T}_{Y \to X|Z}(t)$ may be interpreted intuitively as the degree of uncertainty about current $X$ resolved by the past state of $Y$, $X$ and $Z$ together, over and above the degree of uncertainty about current $X$ already resolved by its own past state *and the past state of* $Z$. . . . . 75

**Key Idea 21** TE terms of various orders are all complementary, and *all* of these orders of TE terms are required to properly account for the information in the target $X_t$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76

**Key Idea 22** The term *information dynamics* [195, 196, 198, 182, 199] is used to refer to investigations of the decomposition of information storage and transfer components in Eqn. 4.21–Eqn. 4.23, and also their local dynamics in space and time (see e.g. local transfer entropy in Sect. 4.2.5). . . . . . 77

**Key Idea 23** The local transfer entropy tells us about the *dynamics* of information transfer in time. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78

**Key Idea 24** Granger causality is based on the premise that cause precedes effect, and a cause contains information about the effect that is unique, and is in no other variable. . . . . . . . . . . . . . . . . . . . . . . . 83

**Key Idea 25** $Y$ Granger-causes $X$ iff $X$, conditional on its own history, is *not* independent of the history of $Y$. . . . . . . . . . . . . . . . . . . . . 84

**Key Idea 26** Theorem 4.2 blurs the boundaries between Granger causality and transfer entropy; thus we might consider the ML estimator (4.43) as defining a generalised (non-linear) Granger causality or, alternatively, a parametric transfer entropy statistic. . . . . 88

**Key Idea 27** Finally, we should stress that, for non-Gaussian processes, transfer entropy and Granger causality are simply *not measuring the same thing!* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

**Key Idea 28** Using transfer entropy, even in these simple systems, requires some subtlety and thought about which information channels to measure and how to approach such measurement. . . . . . . 98

**Key Idea 29** The notion of ecological diversity, as measured by entropy and its generalisations, can help us understand the interconnectedness, stability and sustainability of our modern financial systems. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128

**Key Idea 30** Jaynes' MaxEnt principle can be used to model the decisions of economic agents in micro-economics. . . . . . . . . . . . . . . . . . 129

**Key Idea 31** Information theory can be used to analyse the critical phenomena of financial markets, such as market crashes, just as it can be used in other complex systems. . . . . . . . . . . . . . . . . . . . . . . . . . 130

**Key Idea 32** In which direction does the net information in markets flow, from the equity to the index or from the index to the equity? . . . 134

**Key Idea 33** Western countries are globally the most influential, and Japan has become less influential following the Asian financial crisis in 1997. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 137

**Key Idea 34** Understanding both the strength and the direction of macro-economic indicators provides an important insight into the knock-on effects that other countries feel as a result of a country's internal economic distress. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 137

**Key Idea 35** TE can give quite complex answers, even for apparently simple questions, and remind us of the care required in selection of estimators and parameters in order to achieve robust and reliable results. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139

**Key Idea 36** Effective network analysis examines *directed* (time-lagged) relationships between nodes from their time-series data, and seeks to infer the "minimal neuronal circuit model" which can replicate and indeed *explain* the time series of the nodes [311, 95]. 144

**Key Idea 37** Transfer entropy has been recognised by the research community as a natural fit for effective connectivity inference, since it measures the directed relationship between nodes in terms of the predictivity (or explanation) added by the source node about the target. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 144

**Key Idea 38** *Iterative* or *greedy* approaches with conditional transfer entropy can both capture synergies *and* eliminate (only non-required) redundancies [200, 85, 315, 213]. . . . . . . . . . . . . . . . . . . 147

**Key Idea 39** *Iterative* or *greedy* approaches with conditional transfer entropy infer an effective network in which a directed link indicates that the source is *a* parent of the target, in conjunction with the other parent nodes. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 148

**Key Idea 40** There is significant potential for transfer entropy to produce key insights regarding the time-series dynamics on biochemical networks—measuring predictive effects of one gene on another, modulation of such effects over time, and indeed inferring effective networks. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154

**Key Idea 41** Sensorimotor interaction and morphological structure induce information structure in the sensory input and neural system, promoting information processing and flow between sensory input and motor output [206] *which can be quantified by transfer entropy*. 158

**Key Idea 42** Such coherent wave structures may emerge as a resonant mode in evolution for information flow. . . . . . . . . . . . . . . . . . 159

# List of Open Research Questions

**1** How should synergy and redundancy components of mutual information from a set of sources to a target be properly measured? Indeed, is this possible in general, or only in limited circumstances? . 43
**2** Can wavelet methods be used to get better mutual information for non-stationary systems? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
**3** What are the best estimators for different probability distributions and for large dimensionality? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
**4** Are there better methods for calculating TE, suitable for real data, for non-stationary systems without ensemble data? . . . . . . . . . . . . . . . 80
**5** Is transfer entropy invariant under arbitrary *non-linear* invertible causal filtering? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 86
**6** Can more sophisticated estimators (kernel-based, adaptive partitioning, $k$-nearest neighbour, etc., see Sect. 3.4.2) be expressed as predictive parametric models, to which Theorem 4.2 applies? . . . . 88
**7** Can local (or another variant of) transfer entropy be used to formally separate complex from ordered or chaotic dynamics? . . . . . . 103
**8** Which of the above techniques, a mix of them, or additions to them will prove most convincing for inferring effective connections, whilst eliminating redundancies, capturing synergies, and adapting to the size of available data sets? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 148
**9** How can transfer entropy be formulated for irregular pulse sequences or spike trains? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 150
**10** What happens to EEG transfer entropy after conditioning out other electrodes for each electrode pair? . . . . . . . . . . . . . . . . . . . . . . . . . 153
**11** How can transfer entropy be computed for irregularly sampled time series? For example, using kernel methods and resampling techniques to pre-process the data [38]. . . . . . . . . . . . . . . . . . . . . . . . . . 155
**12** Can we determine direct relationships between transfer entropies in biochemical networks and metabolic costs in the system? . . . . . . . . 156
**13** What informational features "distinguish biological networks from other classes of complex physical systems"? . . . . . . . . . . . . . . . . 157

**14** What are the more important information channels to focus on regarding information flow in embodied cognitive systems—between nodes in an agent's neural network, from actuators to sensors through the environment, or between distributed agents in the system? . . . . . 161

**15** Are there characteristics in the dynamics of transfer entropy that can be linked to key evolutionary or adaptive steps in an embodied agent's development? . . . . . 161

**16** Can transfer entropy or other measures of information dynamics be utilised as an application-independent, intrinsic goal to drive the guided self-organisation of embodied cognitive systems, via adaptation or evolution? For which types of behaviour would this provide a useful template (e.g. top-down causation [342])? How could the intrinsic capability conferred by guiding for high transfer entropy then be built on to produce application-specific utility? . . . . . 161

**17** On which information channels in social media networks will transfer entropy prove to be most revealing of underlying structure? . 164

**18** Given high dimensionality, and limited samples per user, how should one pre-process social media data in order to best capture the relevant information and yield to transfer entropy analysis? . . . . . 164

**19** How do the entropy and mutual information estimators perform on different known statistical distributions, especially in cases where the theoretical distribution is known [124, 144]? . . . . . 168

**20** Are there additional good non-parametric estimators for transfer entropy which avoid summation of entropic quantities, following the extension of [93, 110, 337, 350] for KSG-style TE estimation? . . 168

**21** How can non-parametric estimators for global TE and pairwise conditioning be improved, in terms of efficiency as well as robustness to small data sets? . . . . . 169

**22** Can we relate the energy of communication, in neurons or other systems, to the transfer entropy required of the communication? . . . . 170

# List of Key Results

**1** Local transfer entropy provides the first *quantitative* evidence that particles are the dominant information transfer agents in cellular automata. This result holds for related moving coherent spatiotemporal structures in other systems—see Sect. 5.5. . . . . . . . . . . 101
**2** Neither a perspective of information transfer in computation nor causality in mechanics is more correct than the other—they both provide useful insights and are complementary. . . . . . . . . . . . . . . . . . . 103
**3** High average TE does not imply the presence of coherent particle structures; only the local TE can reveal this. . . . . . . . . . . . . . . . . . . . . . 103
**4** The ordered phase in RBNs is dominated by information storage (information already in nodes dominates their next states; the chaotic phase is dominated by information transfer (information from incoming links, in the context of the nodes' past, dominates their next states); there appears to be a balance between these operations near the critical phase. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 109
**5** Conditional and pairwise transfer entropies reveal different aspects of the dynamics of a system—neither is more correct than the other; they are both useful and complementary. . . . . . . . . . . . . . . . . . . . 110
**6** Networks with low levels of rewiring $\gamma$ (more regular structure) and small activity $r$ exhibit more ordered dynamics which is dominated by information storage, while networks with higher levels of rewiring $\gamma$ (more random structure) and higher activity $r$ exhibit more chaotic dynamics which is dominated by information transfer. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
**7** Small-world networks hold computational advantages over regular or random network structures, in supporting both intrinsic information storage and transfer operations. . . . . . . . . . . . . . . . . . . . . . 115
**8** Wang et al. provided the first quantification of coherent information cascades in the swarm as waves of large, coherent information transfer. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 118

**9** The transfer entropy dropped to zero significantly earlier than the order parameter indicated that synchronisation had been achieved. . . 122
**10** Strong correlations were observed between node degree and outgoing transfer entropy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 122
**11** TE analysis is difficult to get right, and is best performed using estimators which are stable with respect to parameter changes (in particular the KSG estimator). One should take care with such parameters, as well as ensuring that data is embedded correctly. . . . . . . 142
**12** This approach using transfer entropy revealed how information was distributed *spatially* and *temporally* in the system, allowing a precise description of how the embodied computation took place in the agent. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 161
**13** Inner TE activity in Twitter becomes suppressed when transfer from Google is high, then increases as such incoming flow reduces (suggesting activation of default mode activity following reaction to stimulus). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 163
**14** If the time series of edits of a source editor on Wikipedia is predictive of edits by a target editor (as measured by TE), then this is a useful implication of whether the two actually interact [32]. . . . . . 164

# Symbols

| | |
|---|---|
| $\beta$ | Inverse temperature |
| $\mathbf{I}(X : Y \mid Z)$ | conditional mutual information |
| $\mathbf{T}_{Y \to X \mid Z}(t)$ | conditional transfer entropy |
| $\mathbf{H}$ | entropy, arbitrary number of dimensions |
| $F$ | Granger causality |
| $\xi$ | information discrimination |
| $\mathbf{t}_{x \to y \mid z}$ | conditional local transfer entropy |
| $\mathbf{i}$ | local mutual information |
| $\mathbf{t}_{x \to y}$ | local transfer entropy |
| $\mathbf{I}$ | mutual information between two probability distributions $X$,$Y$ |
| $\mathbf{I}(X_1 : X_2 : X_3)$ | multi-information among $X_i$ (mutual information for 3 variables) |
| $N_G$ | number of nodes in a graph or network |
| $\Omega$ | sample space |
| $\psi(x)$ | digamma function |
| $\rho(x, y)$ | Pearson correlation coefficient between $x$ and $y$ |
| $\eta(x)$ | information or surprise. Could also be called local entropy using the definitions of this book |
| $\tau$ | time delay or lag |
| $\mathbf{T}$ | transfer entropy |
| $\mathbf{A}$ | coupling matrix for VAR process |
| $\mathbf{S}$ | VAR process |
| $\mathbf{G}(p : q)$ | cross entropy between $p$ and $q$ |
| $d$ | embedding dimension |
| $L$ | path length in a graph |

# Acronyms

| | |
|---|---|
| AO | Australian Share Market |
| CPI | Consumer Price Index |
| DAX | Frankfurt Stock Index |
| DDLab | Discrete Dyamics Lab (Andy Wuensche) |
| DJIA | Dow Jones Share Market |
| ECA | Elementary Cellular Automata |
| EEG | Electroencephalography |
| ET | Effective Transfer Entropy |
| FTSE | London Stock Exchange (Financial Times Stock Exchange) |
| G | Cross Entropy, **G** |
| GDP | Gross Domestic Product |
| GLM | Generalised Linear Model |
| GTE | Global Transfer Entropy |
| JDIT | Java Information Dynamics Toolkit |
| KLD | $\mathscr{K}(X\|\|Y)$ Kullback–Leibler Divergence between $X$ and $Y$ |
| kT | Product of Boltzmann's constant $k$ and absolute temperature $T$ |
| LTCM | Long Term Capital Management |
| MI | Mutual Information |
| ML | Maximum Likelihood |
| QRE | Quantal Response Equilibrium |
| REA | Relative Explanation Added |
| ROC | Receiver Operating Characteristic |
| S&P | Standard and Poor's Stock Index |
| TB | Trade Balance |
| TE | Transfer Entropy |
| TSE | Tononi–Sporns–Edelman (complexity) |
| XOR | Exclusive OR |
| XR | Exchange Rate |

# List of Tables

3.1 Fruit and vegetable occurrence table . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
3.2 Exclusive OR (XOR) Boolean operation $X = Y \texttt{ XOR } Z$. Resulting values for $X$ are listed in the logic table for each $Y, Z$ pair . . . . . . . . . . 43
3.3 Cairns climate data: mean daily maximum temperature and mean monthly rainfall retrieved from the Australian Bureau of Meteorology (http://www.bom.gov.au, 18 August 2013). Mean daily maximum temperature and monthly rainfall are 29.0°C and 168 mm. $p(\texttt{wet})$ and $p(\texttt{hot})$ are illustrative constructions for the rainfall or temperature being above some threshold each day . . . . . . . 45

5.1 Rule table for ECA rule 110. The Wolfram rule number for this rule table is composed by taking the next cell value for each configuration, concatenating them into a binary code starting from the bottom of the rule table as the most significant bit (e.g. b01101110 = 110 here), and then forming the decimal rule number from that binary encoding. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 99

# List of Figures


1.1 Starlings swarming over Brighton West Pier . . . . . . . . . . . . . . . . . . . . . 8

2.1 Taking the areas in this Venn diagram as representing the relative occurrence of the events in sets $A$, $B$ and $A \cap B$, then $p(A \cap B) = \frac{area(A \cap B)}{area(A \cup B)} = \frac{area(A \cap B)}{area(A) + area(B) - area(A \cap B)}$, $p(A|B) = \frac{area(A \cap B)}{area(B)}$ and $p(B|A) = \frac{area(A \cap B)}{area(A)}$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
2.2 A network of statistical dependencies between the stochastic variables $a_i$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
2.3 A continuous Gaussian distribution (red) and one possible discretisation (bars) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
2.4 Two coupled Gaussians with a correlation coefficient $\rho = 0.75$; the marginal probability distributions and the discretised and normalised histograms are projected onto their respective "rear walls" of the plot . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
2.5 A potential function described by $\phi(Q) = Q^4 + \mu Q^2$ . . . . . . . . . . . . . . 29
2.6 A bifurcation plot of the equilibrium solutions to the equation $Q = \tanh(\beta Q/2)$ showing that, as $\beta$ varies, the number of solutions changes from one ($\beta < 2$) to three ($\beta > 2$). The blue lines represent the expected activity (mean magnetisation) of the system; around each point of the blue lines there will be some minor thermal fluctuations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

3.1 Low- and high-entropy fur! How would you interpret the entropy of feline fur? There is no one answer to this. Is it meaningful to talk about the fur entropy of the calico cat? Where does the giant statue in Barcelona fit in? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

3.2 Two Gaussian distributions with different mean and variance. The KLD depends on how much the distributions overlap, shown here as a yellow area in the left-hand figure. As the yellow area increases, as the two curves move closer, the KLD decreases, reaching zero when the curves overlap completely. To see the asymmetry in the KLD, the right-hand figure shows the integrand of Eqn. 3.20: the red curve (plus signs) is $\mathscr{K}(a||b)$ and the blue curve (circles) is $\mathscr{K}(b||a)$, where $a$ is the curve with the maximum to the left of $b$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48

4.1 Transfer entropy $\mathbf{T}^{(k,1)}_{Y \to X} = 0$ plotted against coupling parameter $c$ for increasing target history length $k$ for Example 4.1 . . . . . . . . . . . . . 69

5.1 Local transfer entropy dynamics in ECA rule 54 . . . . . . . . . . . . . . . . . . 100
5.2 Local transfer entropy dynamics in ECA rule 18 . . . . . . . . . . . . . . . . . . 101
5.3 Mutual information and transfer entropy for the Ising model. The red vertical line denotes the phase transition (Curie temperature). The green line shows the position of the peak for the global transfer entropy (after [24]) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 107
5.4 Dynamics of RBNs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 108
5.5 Average information dynamics versus connectivity in RBNs . . . . . . . . 110
5.6 Information measures versus $\gamma$, for networks with $\bar{K} = 4$ and $r = 0.36$ (after [190]). Information measures are in bits and plotted against the left $y$-axis: entropy, $\mathbf{H}(X)$; active information storage, $\mathbf{A}^{(k=14)}_X$; entropy rate, $\mathbf{H}'_X$; pairwise TE, $\mathbf{T}^{(k=14)}_{Y \to X}$; complete TE, ${}^c\mathbf{T}^{(k=14)}_{Y \to X}$. Note that the entropy rate here represents the sum of all orders of transfer entropy terms $H_{\mu X}$ (see Sect. 4.2.2). A measure of complexity in dynamics, $\sigma_\delta$ (a standard deviation of perturbation avalanche sizes; see [190] for full definition), is plotted against the right $y$-axis, with its peak indicating the critical regime of dynamics here—we have a subcritical regime to the left of this peak, and supercritical to the right. Error bars indicate the *standard deviation* of the values across the 250 sampled networks. (The standard error of the mean is too small to be visible) . . . . . . . . . . . . . . . . . . . . . . . . 113
5.7 Motifs implicated in calculation of information storage at node $i$ include *directed feedback cycles* and *feedforward loop motifs* (loops of length 3 shown for both types). This figure first appeared in [185] and is © American Physical Society, and is reprinted with permission . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
5.8 Schooling groups of predator and prey fish. Schooling in fish produces apparent information cascades [67, 39], e.g. in handling predator avoidance by the school. This figure “Moofushi Kandu fish.jpg” is copyright by Bruno de Giusti, used under Creative Commons CC-BY-SA-2.5-IT [75] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116

5.9 Local transfer entropy at each agent in a swarm at several time steps as three separate swarms merge. The *x–y* coordinates of each agent in the swarm are indicated by the axes; the colour of each agent represents its local TE (averaged over TE contributions from each source to that agent)—red represents positive local TE, while blue is negative. These figures were first published in [345], and are copyright to the authors of that paper; the figures are re-used under the Creative Commons attribution licence. A video showing the local TE during this merge in more fine-grained detail is available on YouTube at http://youtu.be/vwfhijoq4cs, with further videos available in the playlist http://goo.gl/3QbQE8 .................... 118
5.10 Snapshots during a synchronisation process ........................ 119

6.1 Each country is connected to a number of other countries through a global network of economic relationships. Internally a country is governed by social, political, economic and geological constraints and relationships such as transport networks, natural resources, manufacturing centres as well as less obvious networks of social and political influence. These in turn are reciprocally coupled to the internal dynamics of other countries through trade, foreign exchange markets, political relationships and geographical considerations. Understanding how these factors influence one another, in particular the strength and direction of the connections, is of key importance for our understanding of how stable and sustainable our socio-economic systems are ..................... 136

7.1 Transfer entropy in heart–breath data ........................... 141
7.2 Sample effective network diagrams ............................. 146
7.3 Local transfer entropy shown on snakebot modules ............... 160