MuST 10: Causation and Complexity

10th Munich-Sydney-Tilburg Conference in the Philosophy of Science

1-3 March 2017
Charles Perkins Centre – D17

ABSTRACTS

Munich Centre for Mathematical Philosophy (MCMP), Ludwig Maximilian University, Munich
Sydney Centre for the Foundations of Science (SCFS), University of Sydney
Tilburg Centre for Logic, Ethics and Philosophy of Science (TiLPS), University of Tilburg

In collaboration with the Centre for Complex Systems (CCS), University of Sydney

D17 - Charles Perkins Centre
The University of Sydney NSW 2006
Australia

T +61 2 8627 1940
E Info.perkins@sydney.edu.au

ABN 15 211 513 464
CRICOS 00026A

Statistical Laws in Complex Systems*

Eduardo G. Altmann
School of Mathematics and Statistics, University of Sydney, 2006 NSW, Australia
ega@maths.usyd.edu.au

January 19, 2017

Abstract

Power-law distributions are a cornerstone in the field of complex systems. They appear in Zipf's pioneering observations (1930s-1940s), in Simon-Mandelbrot disputes on the mechanism generating these distributions (1950s-1960s), in self-organized criticality (1980s-1990s), and in the observations of networks with scale-free degree distribution (1990s-2000s). Power-law distributions are the best known example of *statistical laws* which are claimed to be *universally* valid in different complex systems. The goal of this contribution is to critically discuss the concept of statistical laws, their validity and role in complex-systems research (results based on Refs. [1,2,3]).

A renewed interest in statistical laws is happening because of the increasingly important role played by data in different scientific fields. On the one hand, analysis of larger datasets typically corroborate the observations motivating the proposal of statistical laws. Examples are shown in Ref. [1], e.g., for the famous Zipf's law for word-frequency distribution. On the other hand, larger dataset tend to fail statistical analysis testing the validity of statistical laws. This happens because even small (systematic) fluctuations not captured by the (often symplistic) statistical laws become incompatible with the statistical fluctuations expected at large sample sizes. These seemingly contradictory consequences of larger datasets demand for a more careful discussion of statistical laws.

The unsatisfactory nature of the interpretation of statistical laws is a source of great misunderstanding and conflicts in the complex-systems community. While until the mid 2000's it was common to see publications having as main claim the observation of power laws in particular systems (often the degree distribution of networks), more recently the community shifted to a critical view in which power-law distributions are claimed to be almost nowhere valid (see Refs. [4,5]). As we argue below, both claims are based on misunderstanding of the role of statistical laws and of the statistical methods used to test them.

There are three main results that will be presented in this talk:

(i) We argue that statistical laws should be assessed in comparison to alternative models and not in terms of testing its validity [1]. In statistical terms this means that the analysis should focus on model comparison and not on hypothesis testing. One reason for this is that complex-system data often show (long-range) correlations while the usual maximum-likelihood fitting and statistical tests assume independence of observations. We argue that many previous claims of rejection of power-law distributions [4,5] should be revisited because they fail to account for known correlations in the data.

(ii) Complex-system research often treats the statistical analysis of the law (both fitting and test) as a step that precedes and is independent of the discussion about the generative models behind the data. In this approach, statistical laws act as *stylized facts* that summarize a property of the system which generative models are required to reproduce. We will argue against the separation of the two steps because the different fluctuations predicted by the different generative models are essential in the statistical analysis of the data [3].

*Abstract submitted to the conference *MuST10: Causation and Complexity*

(iii) We argue that one of the main advantages of simplistic statistical laws is that they allow us to understand general statistical properties of the system. This is often independent of the specific parametric choice of the statistical law. This points to an interpretation of statistical laws as a convenient symplification for understanding purposes and not as a signature of some underlying common mechanism. For instance, we show how statistical properties of fat-tailed datasets are conveniently modeled by power-law distributions and allow us to estimate the accuracy of the estimation of entropic measures in texts [2].

The general points discussed above are based on Refs. [1,2,3] and will be illustrated mainly in the analysis of word frequencies.

References

[1] E. G. Altmann and M. Gerlach, "Statistical laws in linguistics", chapter in the book Creativity and Universality in Language , M. Degli Esposti, E. G. Altmann, F. Pachet (Eds)., Lecture Notes in Morphogenesis (Springer), 2016.

[2] E. G. Altmann, L. Dias, M. Gerlach, "Generalized Entropies and the Similarity of Texts", Pre-print: [arXiv:1611.03596]

[3] J. C. Leitao, J.M. Miotto, M. Gerlach, and E. G. Altmann, "Is this scaling nonlinear?", Royal Society Open Science 3, 150649 (2016)

[4] A. Clauset, C. R.Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. SIAM Review. 51, 661703 (2009).

[5] M.P.H. Stumpf and M. A. Porter, "Critical Truths About Power Laws", Science, 335, 665666 (2012).

A new physical interpretation of drift in a deterministic setup

Pierrick Bourrat
Research Fellow
University of Sydney/Macquarie University

Genetic Drift is often characterized in statistical terms as resulting from deviations from expected reproductive outputs or expected fitness. Yet a purely statistical notion of drift will be unsatisfactory if one wants to distinguish two or more distinct types of causal processes that can lead to the very same observed phenomenon. It seems therefore desirable to have a concept of drift in which deviations from expected values are explained physically rather than assumed as mathematical/statistical truths.

What is known as the propensity interpretation of fitness permits one physical interpretation of drift. An expected reproductive output or fitness, under this interpretation, is viewed as a tendency or disposition comparable to familiar examples like fragility. Entities have a disposition to produce a certain number of offspring in the same way glass has a disposition to break. Evolutionary change due to drift increases proportionally with the deviations from expected reproductive outputs resulting from the dispositional properties of entities to produce offspring.

However, propensity interpretations of probability in general or when applied to fitness are controversial. In fact, propensity interpretations have been claimed to be empty accounts of probability because it is not clear what propensities represent. Furthermore, it is classically assumed that the least problematic propensity interpretation of probabilities, namely single-case propensities, requires indeterminism. If fitnesses are (single case) propensities and they play a causal role in evolution, they are simply brute indeterministic facts about the physical reality. Yet, it is conceivable that in many cases (maybe the majority of them) what is referred as drift is occurring in a fully deterministic set-up, or sufficiently close to one. At any rate, an increasing number of philosophers believe that indeterminism is eliminable from evolutionary theory which questions the adequacy of propensity interpretations of fitness and the physical interpretation of drift that derives from it. It is perfectly reasonable to have a concept of fitness that relies on propensities and ultimately indeterminism, but it is more controversial to apply it when there is no reason to suppose indeterminism and strong reasons to suppose determinism. In such cases the distribution of reproductive outputs of entities of the same type and the deviations from expected values must necessarily be accounted by invoking other factors than propensities. Small differences in initial conditions of the environment seem, *prima facie*, to be a good candidate and it is the type of factor I will explore in this talk.

In light of the above remarks it seems reasonable to demand a physical interpretation of drift satisfying two related desiderata. First, it is desirable to have a physical account of drift that does not rely exclusively on one kind of interpretation of fitness, namely the propensity interpretation, especially because it is grounded, as we have seen, in one of the most controversial interpretation of probability. Second, and related to the previous point, our account should be compatible both with a deterministic and an indeterministic world. This is because even if in some cases indeterministic processes seem to have consequences on reproductive output, it is plausible that many cases of drift are straightforward cases of determinism, or at least that they can be considered as such. The main aim of this talk will be to develop an alternative or at least

complementary physical interpretation of drift to the interpretation borne out of the propensity interpretation of fitness and that satisfies the two desiderata.

My talk will run as follows. I will start by presenting Godfrey-Smith's (2009) framework for drift, which happens to be compatible with the two desiderata. More particularly I present his view that drift results on the one hand from differences in reproductive output due to differences in extrinsic properties as opposed to difference in intrinsic properties (which should be attributed to natural selection), and on the other hand from the population exhibiting what Godfrey-Smith calls a 'low continuity'. In the second part, I will demonstrate that although it is on the right track, this framework is problematic for a number of reasons. Starting from Godfrey-Smith's framework, I expose my physical account of drift. This account is fundamentally independent from probabilities but makes it possible to explain in physical terms the probabilities of the classical models of evolutionary change. In the last part, I will respond to some objections one might have with my framework and show that under some particular conditions I briefly describe, it is compatible with a new objective interpretation of probability in deterministic setups. I will call this interpretation "natural-range interpretation of probability" following Rosenthal (2010).

Work Cited

Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford ; New York: Oxford University Press.

Rosenthal, J. (2010). The natural-range conception of probability. In *Time, chance, and reduction: Philosophical aspects of statistical mechanics* (pp. 71–90). Cambridge: Cambridge University Press.

Application to MUST10: Causation and Complexity

Affiliation

Brett Calcott
Department of Philosophy
University of Sydney
brett.calcott@gmail.com

Title

The role of causal influence in the evolution of the genetic toolkit.

Abstract

Changes in gene regulation have played a key role in the production of the vast diversity of life on our planet. One explanation for this is that the regulatory genes provide a *genetic toolkit*, a set of resources that can be re-used and recombined to produce phenotypic diversity and novelty. The search for principles that underly this capacity for combinatorial change has been dominated by a focus on the modular topology of gene networks. It is assumed that networks with clusters of genes that interact strongly with one another and weakly with other such clusters will be easier to rewire and recombine.

Using simulations of gene regulatory networks, I demonstrate a different approach to thinking about why gene networks can be easily rewired. I focus instead on the evolution of key genes that acquire a stable causal relationship to the upstream environments in which they are expressed, and to the downstream adaptive responses they control. Such genes are typically called selector genes, input-output genes, or

master regulatory genes, and are often represented using a bow-tie topology, where the gene forms a narrow funnel, integrating multiple upstream inputs, and controlling multiple downstream outputs.

I detect and track these genes as they evolve using information theoretic measures that are computed by manipulating the networks in ways that resemble experimental interventions. I show that these measures capture the key causal properties of the genes even in architectures where a bow-tie architecture does not exist. It is these causal properties, I argue, that enable the network to recombine existing functionality.

I demonstrate this with three different phases of simulations. First, I outline the plausible conditions under which such key genes can evolve. Then, I show how a gene network can evolve multiple subsystems containing such genes. Finally, I look at how parts of these subsystems can then be rapidly recombined using these key genes, and show that the resulting network modifications closely match a prominent example of regulatory evolution: the rapid appearance of wing-spots in flies.

Lastly, I argue that—like the appeals to modularity—my approach has similarities to well-known design features that engineers use to build systems that are easily reconfigured. I outline the notion of a messaging system, a fundamental architecture that exists in almost all software we now interact with, and suggest why it may provide a better framework for thinking about evolvability in gene regulatory networks than current appeals to modularity.

Emanuele Crosato
Centre for Complex Systems
Faculty of Engineering & IT
The University of Sydney

emanuele.crosato@sydney.edu.au

Information Dynamics in Swarms

Aggregation in groups is a common phenomenon in many animal species. Living in a group offers a range of benefits, such as shelter, anti-predator vigilance, predator satiation and confusion, more efficient foraging strategies, easier access to mates and division of labor [1]. Animal groups build complex structures, e.g., termite nests, and efficiently perform various collective tasks: birds are known to reduce their energy costs during migration, and ants stigmergically mark out trails for recruitment in foraging. Remarkably, complex collective formations and behaviours often self-organise from very simple interactions among individuals. Such self-organisation has motivated research into collective dynamics within different areas of science.

Originally explored by biologists and animal behaviourists, animal collective behaviour is now a well-established cross-disciplinary topic, studied by physicists, mathematicians and engineers. The collective problem-solving capabilities of animal groups have imbued the computational sciences, and inspired new approaches to artificial intelligence. The life-like features of animals aggregations, such as self-assembly, scalability, adaptation to the environment and self-reconfiguration after external intrusions are particularly useful in engineering autonomous systems. The intricate mechanics of collective motion have also attracted physicists searching for a fundamental theory of active matter. Among all these disciplines, there is a common understanding emerging that information plays a dynamic role in coordination of adaptive behaviours, and that distributed information processing is a specific mechanism that endows the group with collective computational capabilities.

Collective motion is one of the most striking examples of aggregated coherent behaviour in animal groups, dynamically self-organising out of local interactions between individuals. It is observed in different animal species, such as schools of fish, flocks of birds, colonies of insects and herds of ungulates. For example, in response to a predator, many schools of fish display complex collective motion patterns, including compression, "hourglass", "vacuole", "flash expansion", or form highly parallel translating groups [2].

While it is generally accepted that, when moving in groups, animals process information to coordinate their motion, quantifying such information processing has remained elusive. It has been observed that small perturbations cascade through an entire swarm in a wave-like manner, with these cascades conjectured to embody information transfer. For example, it has been shown that fish increase velocity in response to a predator, and this change of motion spreads across the group faster than the speed of an approaching predator — a phenomenon labelled the "Trafalgar effect" [3]. The "wave collective behaviour in flocks" was further quantified, confirming that the velocity of the perturbation's propagation within the group is greater than the velocity of the predator. This phenomena of coherent wave formation is observed widely, across insect swarms, bird flocks, fish schools and mammalian herds, as well as elsewhere in biology (e.g. perturbation waves in protein networks and dynamic opening and

closing of stomatal apertures in plants), and in all cases can be (at least qualitatively) interpreted as serving to transfer information rapidly within a group.

In these studies information cascades are associated to some specific behavioural propagation, while principled methods to quantify information transfer have not being used. Part of our research focuses on developing an information-theoretical framework, called *information dynamics*, and applying it to study complex systems. Distributed information processing is typically dissected into its three primitive functions: the transmission, storage and modification of information. Information dynamics characterises and measures each of these primitives information-theoretically.

Using these measures it is possible to quantify the swarm's dynamics with respect to specific operations on information: collective memory, communication, synergy and so on. One of our recent studies uses information dynamics to characterise coordination during group motion by quantifying dynamic information flows in space and time across the collective, using (variants of) transfer entropy and mutual information. We used these measures to provide the first quantification of information dynamics in a school of fish, identifying information-processing patterns within it during directional changes around a circular tank. This analysis reveals peaks in information flows during collective U-turns, with transfer entropy shown to anticipate turns. We also reveal how the information flows depend on relative position and alignment between fish pairs, and how the effective neighbourhood size fluctuates in terms of information flows.

Information Theory also allows to study critical phenomena. Swarming behaviours may be interpreted in terms of collective system-level patterns which emerge and exist only within a narrow range of system parameters, often in a vicinity of critical dynamics. In general, a critical regime is observed during a transition between ordered and chaotic phases of system dynamics, which may be triggered by varying a control parameter, e.g., the strength of interactions within the system. Following Ginzburg-Landau theory of phase transitions developed in physics, Haken introduced order parameters in explaining structures that spontaneously self-organise in nature [4]. When energy or matter flows into a system typically describable by many variables, it may approach a threshold and undergo a phase transition. At this stage, the behaviour of the overall system can be described by only a few order parameters (degrees of freedom) that characterise newly formed patterns.

Another example of collective behaviour emerging near criticality is distributed computation at "the edge of chaos": it has been observed that these primitive functions have maximum capacity near order-chaos phase transitions. This phenomenon has been studied in various complex systems, such as Cellular Automata, Ising spin models, Genetic Regulatory Networks and other biological networks, neural information processing, and many others.

Our findings may provide a step forward towards a general "information dynamics theory of collective motion", and bring insights in swarm engineering, especially in the area of Guided Self- Organisation (GSO), which aims to exploit the advantages of self-organisation while still being able to indirectly affect the outcome of the self-organising process.

References

[1] Jens Krause and Graeme D Ruxton. *Living in groups*. Oxford University Press, 2002.

[2] Julia K Parrish, Steven V Viscido, and Daniel Grünbaum. Self-organized fish schools: an examination of emergent properties. *The biological bulletin*, 202(3):296–305, 2002.

[3] J. E. Treherne and W. A. Foster. Group transmission of predator avoidance behaviour in a marine insect: The trafalgar effect. *Anim. Behav.*, 29(3):911 – 917, 1981.

[4] Herman Haken. Synergetics. *Physics Bulletin*, 28(9):412, 1977.

Authors:

Mirko Di Bernardo, Philosophy of Science, University of Rome, Tor Vergata, Italy and Theological Faculty of Florence, Italy

Francesca Tramontano, Philosophy of Law, University of Turin, Italy and Theological Faculty of Florence, Italy

Allosteric chemistry, biological autonomy and causality. An informational approach

Abstract

Combining Monod's notions of gratuity of cellular processes and stereospecific recognition with Kauffman's last results on thermodynamic work[1], propagating organization of process and embodied or biological information - deeply connected within an epistemological multidimensional and hierarchical framework - this study aims to discuss, demonstrate and assess that there is an intimate connection between biochemical information, biological autonomy and causation. It is, in fact, the general arbitrariness of allosteric chemistry that allows arbitrary molecules to cause events. Within this framework, information is actually related to the biological and "cognitive" conditions that allow cells to be maximally efficient in the execution of a high variety of roles within a given range of available energetic sources. Our aim is to provide evidence of the need of new informational measures of biological complexity - new axiomatic systems - not just concerning the statistical rarity or computational necessity, but addressing the *temporality* of the living beings *as informational causation*. This study aims to lay the foundations for a *bio-semantic approach*, that may explain the generative and constructive dimension of the internal temporalization and space processes of living systems.

Keywords: The living time as construction, informational causation, the gratuity of cellular processes, thermodynamic work, biological complexity

[1] We compared Longo, G., Montevil, M., *Perspectives on Organisms. Biological time, Symmetries and Singularities*, 2014, Springer; Kauffman, S., Logan, R.K., Este, R. et al. Biol Philos, *Propagating Organisations: an enquiry*, 2008, Springer Kauffman, S., *Humanity in a Creative Universe*, 2016, Oxford University Press.

The Impossibility of Causal Claims in Psychiatry: An Analysis of the Russo-Williamson Thesis and Its Implications

Sydney Katherine Green
Doctoral Candidate
University of Antwerp
sydneykatherine.green@uantwerpen.be

The Russo-Williamson Thesis [hereafter RWT] maintains that, in order for a causal claim to hold water, evidence of both a statistical correlation and a mechanism is required (Russo and Williamson [2007, 2011]; Illari [2011]; Clarke *et al.* [2014]). For instance, in order to prove that Drug X cures Disease Y, researchers must provide both (1) evidence that there is a strong statistical correlation between the use of X and the elimination or amelioration of Y, and (2) evidence of a mechanism by which X eliminates or ameliorates Y. This requirement is intended to prevent the acceptance of causal claims which later turn out to be spurious. Given that the human body is a complex system, the two types of evidence are needed to guarantee that we have an adequate understanding of the individual components within a causal relationship, and how they work together to bring about a cure. However, this two-part evidentiary standard runs counter to the current hierarchy of evidence as set forth in Evidence-Based Medicine [hereafter EBM]. EBM's hierarchy maintains that evidence of statistical correlation, as gathered in randomized controlled trials and other statistically-focused studies, is the strongest type of evidence for a causal claim, and should be considered sufficient (OCEBM [2011]). Furthermore, evidence of mechanisms is far less reliable, and thus unnecessary.

I argue that, while the RWT is correct in its epistemic elevation of mechanistic evidence, it nevertheless cannot be defended in its current formulation. While its implications have been discussed extensively in connection to medicine and the social sciences, its implications for the mental health sciences (i.e., psychiatry and psychotherapy) have not received the same level of attention. It is my contention that these implications are of great importance, and have the power to reveal key shortcomings in the RWT's current formulation. Namely, as it stands, the RWT eliminates the possibility of causal claims within the mental health sciences. This is because causal claims made in these fields do not rely on two separate types of evidence, one of statistical correlations and one of mechanisms. Instead, such claims *derive* mechanistic explanations from statistical correlations, developing likely causal stories which are not independently verified. In order to illustrate this point, I discuss a number of historical examples from psychiatry, including the dopamine hypothesis for schizophrenia and the monoamine hypothesis for depression. Following this historical examination, I discuss the RWT's implications for scientific practices and decision-making in the mental health sciences, arguing that a more relaxed standard of evidence is preferred.

References:

Clarke, Brendan, Donald Gillies, Phyllis Illari, Federica Russo and Jon Williamson [2014]: 'Mechanisms and the Evidence Hierarchy', *Topoi*, **33**, 339–360.

Illari, Phyllis [2011]: 'Mechanistic Evidence: Disambiguating the Russo-Williamson Thesis', *International Studies in the Philosophy of Science*, **25**, 139–157.

OCEBM Levels of Evidence Working Group [2011]: 'The Oxford 2011 levels of evidence', http://www.cebm.net/index.aspx?o=5653.

Russo, Federica and Jon Williamson [2007]: 'Interpreting Causality in the Health Sciences', *International Studies in the Philosophy of Science*, **21**, 157–170.

— [2011]: 'Epistemic Causality and Evidence-Based Medicine', *History and Philosophy of the Life Sciences*, **33**, 563–582.

Crick information: giving substance to biological information

Arnaud Pocheville (Sydney)
Paul Griffiths (Sydney)

The idea that biological information is created by evolution, passed on in heredity, and expressed during development is an attractive gloss on what has been revealed by the last century of advances in biology. But on closer examination it is hard to see what scientific substance corresponds to this vision. Several biologists and philosophers of biology have suggested that 'biological information' is no more than a collection of loose metaphors. Others have offered their own theories of biological information, but most of these have been oddly unrelated to actual biological practice. Here we argue that the conception of information used by Francis Crick in his 'sequence hypothesis' and 'central dogma', a conception closely related to the older idea of 'biological specificity', is adequate to state a substantial, general theory of biological information. There are two aspects to this account, corresponding to a fundamental duality in information theory between Shannon and Kolmogorov measures.

Anne-Marie Grisogono

Title: The science of making sense of complex systems

Abstract:
What we mean by causation in everyday life seems a long way removed from the roles played by causality in fundamental physics. Yet there are links, and understanding the emergence of the processes that generate, modify and maintain order in complex systems is key to making sense of them. It is a notoriously difficult challenge to understand how the consequences of changes and events propagate through the dense networks of interactions in complex situations to produce outcomes – often unexpected, and often undesirable. We also want to learn from such experiences and understand how people have contributed to causing the outcomes, so that we can hold them responsible for their actions, and perhaps enact changes to avoid a repetition. Taking both reductionist and collective systems approaches to complexity and causation, together with some insights from cognitive psychology, we will explore the science of making sense of complex systems. A perhaps surprising conclusion is that causation may be not a fundamental but an emergent process.

Discrete Fractal Structures in Layered Erdős–Rényi Networks

Michael S. Harré, michael.harre@sydney.edu.au
Complex Systems Research Group,
Faculty of Engineering and IT,
The University of Sydney.

Complex system that are made up of many interacting components can exhibit several well known macroscopic patterns that indicate particular types of microscopic relationships that are driving the observed non-trivial or 'self-organising' patterns of global behaviour. Network theory [1], the mathematical study of the graphs that describe these systems using *nodes* and *links*, is an important approach to understanding complex systems as it can be applied to a large number of different systems and the mathematical results of the theory can then be applied directly to the system of interest. In the Erdős–Rényi (E-R) model [2] of network formation there is a critical probability p_c of any two nodes being connected, below this probability the network as a whole is largely disconnected, above this probability the network spontaneously forms a giant connected component, the formation of this component is a macroscopic indicator of the change in a key microscopic interaction. In the Barabási-Albert (B-A) [3] preferential attachment model each new node prefers to attach to a node that already has a large number of connections. This local decision rule used by each new node connecting to the network results in a scale-free distribution of links across the nodes and it is this global scale-free distribution of links that is used as an indicator of the local self-organising rules that are driving the global topology of the network. These local interactions between nodes are key to understanding the emergent macroscopic properties of a dynamical system and so in this talk I will present and expand on our recent theoretical results [4] in modeling the emergence of discrete scale-invariance, or fractals, in network theory using constrained layers of E-R networks. The local constraints between nodes relate to the average number of links each node can form and the emergence of fractal structures is a new macroscopic indicator of complex microscopic interactions in complex systems. It is a fascinating theoretical discovery that Fractals, one of the most important signals of complexity, has been largely absent from the simplest network topologies but it emerges naturally as a result of a discrete constraint optimization problem using Erdős–Rényi networks as a base model. These new results suggest a range of interesting theoretical and applied directions for research that will be outlined in this talk.

[1] "Complex networks: Structure and dynamics", S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, *Physics Reports*, 2006, Vol. 424, Issues 4–5, Pages 175–308

[2] "Emergence of Scaling in Random Networks", Albert-László Barabási, Réka Albert, *Science,* Vol. 286, Issue 5439, 1999: pp. 509-512

[3] "On random graphs, I", P. Erdös, A. Rényi, *Publicationes Mathematicae*, Vol. 6, 1959, pp. 290-297

[4] "The social brain: scale-invariant layering of Erdős–Rényi networks in small-scale human societies", Michael S. Harré, Mikhail Prokopenko, *Journal of the Royal Society Interface*, Vol. 13, Issue 118, 2016

Presenter: Stephan Hartmann

Title: Conditionals, Testimony, and Causal Structure

To be advised.

Is there a continuing role for dynamical mathematical models to understand complexity in human organisational systems?

Alexander Kalloniatis

Defence Science & Technology Organisation

24 Scherger Drive, Canberra Airport, ACT

Alexander.Kalloniatis@dsto.defence.gov.au

For several decades, the scientific means of grappling with complexity in human organisations has seen a strong disparity between broad qualitative models, at one extreme, or most recently heavily data-driven workflow or agent-based models. In this talk, I explore the space for a model based on somewhat elementary coupled differential equations that also invokes more contemporary mathematics from network science. The model is a stochastic adaptation of the famous Kuramoto model of network coupled oscillators but represents both the social dimension of human organisational work, through the network variables, but also individual human cognitive aspects through representation of non-Gaussian stochasticity and the Perception-Action cycle. The model also allows for representation of artificial/technological agents such as information objects, artefacts or decision-support tools. I show an example of application to an existent human organisation – a military headquarters – illustrating an initial validation. I argue that there remains indeed space for such models to enable cross-validation across a suite of models of different levels of fidelity. Most importantly, such predictive models allow for an efficient exploration of intervention strategies to improve human organisational performance while also allowing for sufficient solvability that genuine insight into the factors for – or causes of - success may be gained. The talk is based on work published by the presenter at the 21st International Command and Control Symposium, Physica A and Physica D in 2016.

Stuart Kauffman
Australia Talk Abstract
Jan 17, 2017

The Emergence and Evolution of Life Beyond Physics

The emergence and evolution of life is based on physics but is beyond physics. Evolution is an historical process arising from the non-ergodicity of the universe above the level of atoms. Most complex things will never exist. Human hearts exist. Prebiotic chemistry saw the evolution of many organic molecules in complex reaction networks, and the formation of low energy structures such as membranes. Theory and experiments suggest that from this, the spontaneous emergence of self reproducing molecular systems could arise and evolve. Such "collectively autocatalytic systems" cyclically link non-equilibrium processes whose constrained release of energy constitutes "work" to construct the same constraints on those non-equilibrium processes. Cells yoke a set of non-equilibrium processes and constraints on the energy released as work to build their own constraints and reproduce.

Such systems are living, and can propagate their organization with heritable variations, so can be subject to natural selection. In this evolution, these proto-organisms emerge unprestatably, and afford novel niches enabling, not causing, further types of proto-organisms to emerge. With this, unprestatable new functions arise. The ever-changing phase space of evolution includes these functionalities. Since we cannot prestate these ever new functionalities, we can write no laws of motion for this evolution, which is therefor entailed by no laws at all, and thus not reducible to physics. Beyond entailing law, the evolving biosphere literally constructs itself and is the most complex system we know in the universe.

Choosing a Level of Causal Description: A Pragmatic Approach

London School of Economics (ANU Visitor February - September 2017)
d.b.kinney@lse.ac.uk

November 30, 2016

Several recent authors in philosophy of science—including Weslake (2010), Woodward (2010, 2016), Weatherson (2012), and Franklin-Hall (2016)—argue that the most appropriate description of a particular causal relationship in nature is not necessarily the most detailed or fine-grained description of that relationship. This claim is held to be true in natural sciences such as biology and neuroscience, social sciences such as economics, and sciences that lie somewhere in between, such as epidemiology. For example, consider the following two descriptions of the causal relationship between smoking and lung cancer in a population:

i. Smoking causes lung cancer.

ii. Smoking Marlboros causes lung cancer.

There is an intuitive sense in which (i) is preferable to (ii) as a description of the etiology of rates of lung cancer in most populations. However, the relevant question for my purposes here is this: can we provide a general framework for determining when a causal description is given at the correct or optimal level?

I argue that the notion of the "correct" level of causal description is essentially pragmatic, in the sense that it ought to be defined in relation to a particular decision problem. I then show that for some decision problems, an agent would not pay any more to learn the value of a more fine-grained causal variable than she would pay to learn the value of a more coarse-grained causal variable. In these cases, the more fine-grained description is not worth anything to the agent, and therefore the more coarse-grained description can be preferred. Alternatively, if the agent *would* pay more to learn the value of the more fine-grained causal variable, then she should use the more fine-grained variable in her description.

	Lung Cancer	No Lung Cancer
Approve	-15	100
Deny	15	-110

Table 1: Payoff Matrix For Life Insurance Decision

This criterion is pragmatic in that it defines the choiceworthiness of a level of causal description in terms of a particular decision problem faced by an agent. In this respect, my approach here is novel. More typically, authors such as Craver (2007) and Woodward (2010) argue that the level of *proportionality* between cause and effect provides a sufficient standard for determining the correct level of causal description. I argue against this view by showing that in probabilistic cases a proportionality-based criterion over levels of causal description leads to undesirable or unrealistic conclusions. This critique extends to a recent account from Pocheville et al. (forthcoming) that defines proportionality in probabilistic contexts using an information-theoretic mathematical approach. I conclude that my more pragmatic approach to choosing the appropriate level of causal description is preferable to a proportionality-based approach.

I can provide a glimpse into the details of my account via an example. Suppose that an analyst has to decide whether to deny or approve an individual for life insurance. In an obvious oversimplification, suppose that only one fact—whether or not the potential customer develops lung cancer—will determine whether offering the policy will lead to a positive or negative outcome. How much should the analyst pay to learn whether or not the potential customer smokes?

We can begin answering this question by explicating the decision problem that the analyst faces. Table 1 specifies the payoffs of each of her feasible actions in each state of the world (the payoffs here are arbitrary in the sense that any set of real numbers could have been used to explicate FPCI as a concept). If the analyst knew for certain that the potential customer would develop lung cancer, then she would choose Deny, and receive a payoff of 15. If she knew for certain that the agent would not develop lung cancer, then she would choose Approve, and receive a payoff of 100. However, suppose that the analyst cannot learn whether the customer will develop lung cancer. Instead, she can only learn whether the customer smokes. Suppose further that if a person smokes, then there is a 90% probability that she will develop lung cancer. If she does not smoke, then there is a 20% probability that she will not develop lung cancer. Now

suppose that the analyst learns that the potential customer does in fact smoke. Under these conditions, the expected payoff of Approve is $.9(-15) + .1(100) = -3.5$ and the expected payoff of Deny is $.9(15) + .1(-110) = 2.5$. So the agent would choose to Deny, and expect a payoff of 2.5. Next, suppose that the analyst learns that the person does not smoke. Under this supposition, the expected payoff of Approve would be $.2(-15) + .8(100) = 77$, and the expected payoff of Deny would be $.2(15) + .8(-110) = -85$. So the agent would choose to Approve, and expect a payoff of 77. Finally, let us suppose that the analyst has the prior belief that a person is equally likely to smoke or not smoke. Under this supposition, the expected payoff from learning the potential customer's smoker status is $.5(2.5) + .5(77) = 39.75$.

However, one also has to consider what the agent would have done had she not learned the potential customer's smoker status. Assuming that the analyst's credences obey Bayes' Theorem, her prior beliefs about whether the potential customer smokes and her posterior beliefs about whether the potential customer develops lung cancer (given the customer's smoker status) jointly imply that the potential customer has a 55% probability of developing lung cancer, and therefore a 45% probability of not developing lung cancer. Thus, if the analyst does not learn whether the patient smokes, then the expected payoff of Approve is $.55(-15) + .45(100) = 36.75$ and the expected payoff of Deny is $.55(15) + .45(-110) = -41.25$. Since the analyst always acts to maximize expected payoff, in the absence of information as to the potential customer's smoker status, she would choose Approve. If we subtract the expected payoff of Approve when smoker-status is unknown from the expected payoff of learning whether the potential customer smokes, we get $39.75 - 36.75 = 3$. This is the fair price of information regarding whether or not the potential customer smokes; the analyst should only pay up to three units of whatever currency the payoffs are expressed in to learn whether or not the potential customer smokes.

My positive proposal, which I motivate over the course of this paper, is that we should use a more fine-grained causal variable in our representations of the world if and only if the more fine-grained variable exhibits a greater fair price of information than its coarse-grained counterpart. I conclude by defending my approach against the objection that it is overly pragmatic.

References

C. F. Craver. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press, Clarendon Press, 2007.

L. R. Franklin-Hall. High-level explanation and the interventionist's 'variables problem'. *British Journal for the Philosophy of Science*, 67(2):553–577, 2016.

A. Pocheville, P. Griffths, and K. Stotz. In *Proceedings of the 15th Congress of Logic, Methodology and Philosophy of Science*, pages 93–102. London: College Publications, forthcoming.

B. Weatherson. Explanation, idealisation and the goldilocks problem. *Philosophy and Phenomenological Research*, 84(2):461–473, 2012.

B. Weslake. Explanatory depth. *Philosophy of Science*, 77(2):273–294, 2010.

J. Woodward. Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy*, 25(3):287–318, 2010.

J. Woodward. The problem of variable choice. *Synthese*, 193(4):1047–1072, 2016.

Kevin Korb

Title: Causal explanation with Bayesian networks

Abstract:

How to generate explanations from Bayesian networks is a long-standing problem that has attracted many different answers, for example, using mutual information (Suermondt, 1992). We have recently developed a causal information theory that can assist with this, combining mutual information with an interventionist theory of causality (e.g., Korb, Nyberg & Hope, 2011). While causal information theory is a promising tool, the theory itself is not fully developed and its precise application in making sense of causal explanations is not clear. Example of the former: multiple causes often interact with one another in their joint effects, but thus far accounts of the types of interactions and how to measure them are deficient. Example of the latter: causal explanation depends upon context (e.g., what conditions are assumed as a part of a causal query), but how to translate explanatory context into conditions or settings for a causal Bayesian network is not well understood.

Differentiating information transfer and causal effect, or: Why you should be interested in more than causality

Dr. Joseph Lizier
Centre for Complex Systems, and Complex Systems Research Group
Faculty of Engineering and IT
The University of Sydney
joseph.lizier@sydney.edu.edu

The concepts of information transfer and causal effect have received much recent attention, yet often the two are not appropriately distinguished and certain measures have been suggested to be suitable for both. We discuss two existing measures, transfer entropy and information flow, which can be used separately to quantify information transfer and causal information flow respectively. We apply these measures to cellular automata on a local scale in space and time, in order to explicitly contrast them and emphasize the differences between information transfer and causality. We also describe the manner in which the measures are complementary, including the conditions under which they in fact converge. We show that causal information flow is a primary tool to describe the causal structure of a system, while information transfer can then be used to describe the emergent computation on that causal structure.

J.T. Lizier and M. Prokopenko, "Differentiating information transfer and causal effect", *European Physical Journal B*, vol. 73, no. 4, pp. 605-615, 2010. doi:10.1140/epjb/e2010-00034-5

Structures, dynamics and mechanisms – an integrative account

Holger Lyre
Philosophy Department & Center for Behavioral Brain Sciences
University of Magdeburg

December 2016

Abstract

I develop and defend the following theses about the relationship between dynamical and mechanistic explanations: (1) that dynamical explanations are essentially structural, (2) that they are multiply realizable, (3) possess realizing mechanisms, and (4) provide a powerful top-down heuristic. Four examples shall support my points: the harmonic oscillator, the Haken-Kelso-Bunz model of bimanual coordination, the Watt governor and the Gierer-Meinhardt model of biological pattern formation. My comparative analysis leads to the picture of "horizontal" and "vertical" explanations to illustrate the orthogonal perspectives of the dynamical and the mechanistic approach as well as their integration at suitable intersection points.

Extended Abstract

Strong Mechanism versus Strong Dynamicism

Dynamical systems play a major role in science. It remains, however, particularly controversial whether and in which sense dynamical models and explanations fit into the overall doctrine of mechanism. According to this doctrine, genuine explanations especially in the life sciences are mechanistic. „Strong dynamicists" radically disagree with the mechanist doctrine. Chemero and Silberstein (2008) and Stepp, Chemero and Turvey (2011) argue in favor of dynamicism and dynamic explanations as self-contained, i.e. they

understand dynamical models as genuinely explanatory whether or not the variables in the explanation refer to mechanisms.

Proponents of „strong mechanism“, on the other hand, consider the mechanistic approach as the exclusive account of explanation in the neurosciences. They have postulated two rigorous constraints: the “3M-constraint” according to which in *successful* explanatory models a model-mechanism-mapping must exist (Craver and Kaplan 2011), and the “details-constraint”, according to which completeness and specificity about the details of a mechanism is an explanatory virtue (Kaplan 2011). Dynamical explanations are either mechanistic in the sense that they conform to the 3M- and details-constraint or are no explanations at all.

Dynamical explanations as structural: the harmonic oscillator

The model of the harmonic oscillator explains the occurrence of periodic motion behavior in a wide variety of natural and artificial systems (at least to 1st order approximation). The key physical idea is that harmonic oscillation occurs whenever the restoring force is directly proportional to the displacement, hence, $F \sim - x(t)$. This leads to the well-known oscillator equation as a dynamical law: $d^2 / dt^2 \, x(t) = -k \, x(t)$.

The oscillator model depends on one dynamical variable, x(t), only. In broadest terms, x(t) represents some quantity of change. It is a relational rather than an intrinsic property. Hence, the dynamical equation doesn't capture its target system(s) intrinsically, but rather picks out the decisive temporal variation of a relational property that is causally relevant to bring about the observed behavior. Moreover, the harmonic oscillator possesses a wide and impressive variety of (seemingly heterogeneous) realizations. This is also grounded in the fact that the oscillator equation picks out a relational dynamical variable without specifying the relata any further. And it is a general feature of many cases of alleged multiple realizability: the instantiations or realizers may only superficially look drastically different, but do in fact share a particular set of relevant properties. It is then perfectly understandable why it is possible to have, on the one hand, a 'higher-level' generalization, while having, on the other hand, the 'lower-level' realizers appear to be different. The 'higher-level' law simply quantifies over the commonalities only: the shared properties of

the realizers. The situation becomes even more tricky in cases where the realizers share all and only relational properties (Lyre 2009).

We may summarize the above analysis in the following two theses:

- (1) Dynamical explanations are essentially structural: they individuate their entities only relationally by focusing on the relevant spatiotemporal-cum-causal structure of their target systems.
- (2) Dynamical laws are multiply realizable. Their multiple realizability is grounded in shared structure of the realizers.

Furthermore, mechanisms are said to be multilevel. One particular mechanism comprises at least two levels: the higher level on which the phenomenon occurs and the lower level on which the mechanism's components, their operations and their mechanistic organization “live”. This picture is highly misleading if not flatly wrong once applied to the class of dynamical explanations. The dynamical, 'higher-level' explanation picks out properties that are already existent on the 'lower level'. The 'higher-level' dynamical equation picks out relational properties of the 'lower-level' realizers, i.e. the components and operations of the realizing mechanisms. In other words: the relevant 'higher-level' and 'lower-level' relations coincide. The dynamical explanation turns out to be a reductionist explanation in the full Nagelian sense. We may think of dynamical equations as providing us with reducing Nagelian identities: they provide us with an insight into the identity of the structure of the dynamical equation(s) and the structure of the realizing mechanisms. The levels of the realizing mechanisms that underlie dynamical equations collapse. This can be captured in the following thesis:

- (3) The realizers of dynamical laws are realizing mechanisms.

Two more examples, the Haken-Kelso-Bunz model of bimanual coordination and the Watt governor, shall be used in the paper to support the above three theses in more detail. The final thesis will be that (4) dynamical explanations very often provide a powerful top-down heuristic to predict and identify realizing mechanisms. It shall be illustrated by discussing the Gierer-Meinhardt model of biological pattern formation. This is also connects to the overall picture of explanation that is developed in the paper and according to which dynamical and mechanistic explanations provide “horizontal” and “vertical” perspectives of explanation.

“Horizontal” and “vertical” explanations

Mechanistic explanations are short-range and local, whereas dynamical explanations are, typically, long-range and global. Moreover, mechanisms occur, typically, in nested multi-level hierarchies. We can think of such hierarchies as vertical towers of nested mechanisms. In this sense, mechanical explanations are “vertical”. By way of contrast, dynamical explanations are “horizontal”, they focus on long-ranging, generalizable laws with a wide (and potentially even universal) scope. Theses (1) and (3) allow for a natural intersection between the two, orthogonal approaches of explanation: the realizing mechanisms of dynamical explanations can be seen as "intersection points" of the horizontal and vertical perspectives of dynamical and mechanistic explanations. This provides us with an integrated account of explanation according to which, in many cases, a complete explanation and full understanding of a target phenomenon demands the identification of intersections of horizontal and vertical explanations.

Mesoscopic Modeling as a Cognitive Strategy for Handing Complex Biological Systems

Miles MacLeod

University of Twente

m.a.j.macleod@utwente.nl

One on the principal goals of modern computational systems biology is the production of reliable large scale models of complex gene regulatory and metabolic systems suitable for predicting system behavior in response to perturbation, and thus for testing new medical procedures or drug inventions. For many systems biologists this goal requires the production of complex high-fidelity models. While it might have been thought that modern computational and mathematical resources would help overcome the challenges complex nonlinearity throws the way of mathematical modelers and enable the construction of such models, the reality is proving otherwise. Various constraints modelers face combined with this complexity render the production of such models particularly difficult. As a result the field is falling behind on one of its signature goals, producing models capable of guiding medical intervention. The purpose of this paper is thus to help build a more sound *cognitive* rationalization of how modelers do in fact approach these systems, and what any individual modeling project should aim to achieve. While models may fall short of their target of capturing biological systems with high fidelity, they can provide meaningful starting points for cognitively tractable development of larger scale models but over a more extended time frame. This strategic approach relies on a pattern of modeling called *mesoscopic modeling*.

In this paper then we will explore some of the constraints modelers are acting under, and ways in mesoscopic modeling create manageable pathways for handling these. Part of the information from which these insights are drawn derives from a 5 year ethnographic study of model building and cognitive practices in two systems biology labs. The most significant constraint relates to limited data and data uncertainty. Uncertain or incomplete data sets create mathematical problems which are very difficult to resolve, namely that parameter fitting processes are incapable of finding unique solutions. Different solutions may agree well over a certain domain but behave quite differently over others, making it very hard to rely on such models predictively. Data uncertainty plus the complexity of the models creates complex patterns of parameter compensation (or model sloppiness). This is a real hindrance to model production which becomes more acute the more complex a model of a nonlinear system that is attempted. Likewise the larger the scale models the more difficult it is cognitively to debug errors, which is very important given the often inaccurate knowledge of any given biological network modelers must work with. Obstacles like these cannot necessarily be resolved through applying greater computational power alone. As such the project of building large scale models upfront by relying on brute computational force faces numerous challenges.

However a look around the field in general shows that most modelers are not really attempting large-scale predictive models upfront, but models of rather smaller scale to more heuristic or investigative purposes. These purposes make use of the power of computation to explore and evaluate new hypotheses about biological networks, and thus contribute to the process of building up better information about such systems. These uses do not require the same standards of accuracy and reliability that predictive goals do. An important part of this work is mesoscopic modeling. While mesoscopic modeling can have both epistemological and ontological interpretations, in this case we

focus mainly on the epistemological interpretations, which are connected in biology to Noble's (2008) notion of middle-out modeling. In systems biology mesoscopic models may be said to be "mesoscopic" in a couple of senses. Firstly representations are middle-sized. Systems are identified or abstracted to keep the number of interacting elements to a moderate scale. Secondly representations occupy a middle-ground by being relying on simplified representation of both the system but also system-level phenomena. They are approximations and abstractions at both ends. For instance sets of interactions are black-boxed and included in such a model in the form of a single dependent variables which represent the output of these interactions. The interaction relationships between biochemicals are approximated and simplified in place of using the more detailed representations that molecular biologists prefer such as the Michaelis-Menten model of enzyme kinetics. At the systems-level only certain relationships, such as dominant or equilibrium relationships, might be thought important to capture initially and these may themselves be simplified or approximated. This makes such models unsuited as "disease simulators" (Voit et al., 2012; p23). They simply cannot capture the required ranges of system behavior accurately enough.

However mesoscopic models can be rationalized cognitively, as a starting point for building more complex representations in a controlled cognitively manageable way without necessarily running into large parameter uncertainty problems. These models afford modelers a relatively simple representation of a system, which are easier to fit, encounter problems of parameter uncertainty at a lower level, and from which modelers can "locally increase granularity" and in turn increase model scale. Dependent variables in the abstract model can be swapped for more detailed mechanisms, which generate better accuracy and ability to account for other system behaviors. Mesoscopic models are simplified enough to facilitate computational solutions and mathematical analysis. At the same time they are simplified enough that the mechanisms by which the model operates can be understood and tracked by the modeler.

Analysis of these model building practices finds that modelers can use such models to generate an understanding of how lower levels and higher levels in a model are related in the neighborhood of overarching solutions to a given problem. This understanding is extremely important for inferring where and how to augment the model to improve model scale and thereby accuracy in respect to specific tasks. While multilevel problems are usually hugely complex since lower level perturbations to a model do not track linearly to higher-level perturbations, having a good mesoscopic model allows the causes and effects of changes and additions to the model to be more easily discerned and tested computationally. Cautious modification to the model can be simulated, and the effects tracked through the model and understood. At the same time the model itself provides certain expectations about patterns of cause and effect between levels researchers can rely upon for their next moves. These models in other words bring levels into partial coordination which facilitates their further coordination to be more easily explored and identified in smaller manageable steps. In cognitive terms building such models promotes hierarchical learning of system relationships.

Such practices provide a first-hand case of what cognitive strategy for managing complexity looks like, which in turn provides a rationalization for building models which fall short of predictive reliability. Mesoscopic modeling replaces any notions of relying on computational power alone to break through complexity, with a more traditional step-by-step approach which prioritizes understanding and using tractable models to scaffold the development of more complex representational ones.

References

Noble, D. (2008). *The music of life: biology beyond genes*. Oxford: Oxford University Press.

Voit, E. O., Qi, Z., and Kikuchi, S. (2012). "Mesoscopic Models of Neurotransmission as Intermediates between Disease Simulators and Tools for Discovering Design Principles." *Pharmacopsychiatry* 45(1): 22.

Degrees of Causal Contribution

Cei Maslen, Senior Lecturer at Victoria University of Wellington, New Zealand

Cei.Maslen@vuw.ac.nz

In this paper I discuss prospects for using Shapley Value (a concept from cooperative game theory) as a way of defining degrees of causal contribution and dividing legal and moral responsibility for different types of cases. I also explore a problem Sara Bernstein has called the Moral Difference Puzzle and how it bears on these issues.

The basic idea of Shapley Values is that members of a group receive a share of the total payoff of the game that is proportional to a kind of average of their marginal contributions over all possible groupings of players. For some limited cases, Shapley's four axioms (symmetry, zero player, aggregation, efficiency) justify applying this measure for measuring degrees of causal contribution. However, obstacles arise when attempting to apply this to other cases.

One problem is that Shapley Value only gives us a quantitative measure of the relative degrees of causal contribution of different causes, whereas intuitively we need a measure of the absolute degrees of causal contribution of different causes for many applications. Reflection on Bernstein's Moral Difference Puzzle helps to illustrate this problem.

Partial References:

(1) Bernstein (forthcoming). "Causal Proportions and Moral Responsibility".

(2) Braham & van Hees (2009). "Degrees of Causation".

(3) Ferey & Dehez (2016). "Overdetermined Causation Cases, Contribution and the Shapley Value".

(4) Kaiserman (forthcoming). "Causal Contribution"

(5) Maslen (2004), "Degrees of influence and the problem of pre-emption".

(6) Roth & Verecchia (1979). "The Shapley Value as applied to Cost Allocation: a reinterpretation".

(7) Shapley (1953). "A Value for n-person Games".

VARIABLE DEFINITION AND CAUSAL INFERENCE: THE ROLE OF INDEPENDENT COMPONENT ANALYSIS

Marco CAPASSO[1] Lorenzo CASINI[2] Alessio MONETA[3]

January 20, 2017

Causal discovery methods based on graphical models are widespread in the social and biomedical sciences and are proved to be very powerful under a set of conditions. Much literature has discussed violations of assumptions such as the causal Markov condition and the faithfulness condition (cf. Cartwright, 2002; Dawid, 2010). Less importance, however, has been given to the problem of variable definition.

It has been recently pointed out that a problem arises when the variables object of investigation are "ill-defined", in the sense that they have easily identifiable causal roles, either because they are "aggregates", that is, sums or averages of variables with heterogeneous causal roles (Woodward, 2016; Eberhardt, 2016), or when, for a variety of reasons, they do not completely "transmit" causal information (Neapolitan, 2003). Moreover, Spirtes and Scheines (2004) have pointed out the case of "ambiguous manipulation", in which an observed variable such as, for example, total cholesterol is "ill-defined", in the sense that it is the sum of two non-observed variables, viz. HDL (high density lipoproteins) plus LDL (low density lipoproteins), each of which has a different causal role: LDL cholesterol causes heart disease, while HDL cholesterol prevents heart disease. Thus, an intervention on total cholesterol is ambiguous and has unpredictable effects, since it is not specified whether it acts on the causal path linking LDL cholesterol to heart disease or rather on the path linking HDL cholesterol to heart disease (or on both paths).

Notwithstanding the attention to ill-defined variables, a rigorous and comprehensive study of the cases where the problem of variable definition arises and makes troubles for causal inference has not been yet produced. This study seems particularly desirable in special sciences like economics, in which the applications of methods for causal inference must often cope with complex or "non-standard" settings. In this paper we attempt to precisely fill this gap. Moreover, we propose to address the problem of variable definition by means of a modeling approach grounded on "independent component analysis" (ICA), viz. a statistical technique that under the assumption of non-Gaussianity and linearity transforms a set of data in a mixture of independent shocks. We show that in several cases

[1]Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway. Email: *marco.capasso@gmail.com*.

[2]Department of Philosophy, University of Geneva, Switzerland. Email: *lorenzodotcasini@gmail.com*.

[3]Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy. Email: *a.moneta@santannapisa.it*.

our proposed approach can solve the problem of variable definition. Furthermore, we show that in other cases, where it cannot reliably infer causal structures in the presence of ill-defined variables, our approach is at least able to flag their presence. Our argument shall proceed as follows.

In the first part of the paper, we show that simple structures over three-variable sets including one ill-defined variable violate the causal Markov and the faithfulness condition. Moreover, we show which (possibly "unfaithful" and "non-Markovian") relations of conditional independence are implied by different structures including such a variable. The aim is not just to highlight violations of the aforementioned assumptions, but to find some regularities in the relationship between structures with aggregates and properties of conditional independence. For example, in a structure $X \longleftarrow Z \longrightarrow Y$ in which Z is the sum of two variables, each of which causes separately X and Y, X is independent of Y, but, conditional on Z, X is dependent on Y. Thus, in this case many causal search algorithms would detect a collider whereas the original structure is instead a graphical fork. Next, we show which relations of (conditional or not) dependence and independence are implied when the original structure is a graphical chain where the ill-defined variable Z is in the middle. We also highlight the analogous pitfalls that emerge in the application of causal search algorithms to this case.

In the second part of the paper, we present the ICA approach and demonstrate how its application to the cases of ambiguous manipulations not only helps detecting the problem of variable definition but also uncovers the true causal structure under assumptions as weak as those made by typical causal search algorithms (e.g. causal sufficiency, non-strictly deterministic causes, non-Gaussianity, linearity). Many studies (cf. Shimizu et al., 2006; Moneta et al., 2013; Capasso and Moneta, 2016) have already shown the potential of ICA for causal inference. However, ICA's potential for detecting which variables transmit the causal information, or—in the language of Spirtes et al. (2000)—which graphical causal paths are "active", has not yet been studied. To this end, we distinguish between "common" shocks—each of which affects two or more variables on the same active path—and "idiosyncratic" shocks—each of which affects a single variable. We argue that a structure with an ill-defined variable displays patterns of common and idiosyncratic shocks that are different from those displayed by a structure over variables with unambiguous causal roles. To illustrate this point, we start from the simplest cases of two-variable sets and then move on to the cases of three-variable sets analysed in the previous section. For example, in the case of a structure comprising just an ill-defined variable like "total cholesterol" and an effect of only one part of it (e.g. a specific disease caused by HDL only), the structure of the shocks is constituted by a common shock (affecting both variables) and two idiosyncratic shocks (one for each variable). If one were to observe HDL instead of total cholesterol, one would detect just one idiosyncratic shocks since the shock to HDL is completely transmitted to the disease and is thus a common shock.

In sum, we propose a search procedure that allows one to detect cases of ill-defined variables via the identification of latent sources of variation by means of ICA. In some cases, building on the regularities found in the first part of the paper with regard to the relationships between structures with aggregate variables and properties of conditional independence, our proposed search procedure also exploits conditional independence tests. As a consequence, the search procedure naturally lends itself to correcting typical mis-

takes of causal search algorithms in the presence of ill-defined variables.

References

Capasso, M. and A. Moneta (2016). Macroeconomic responses to an independent monetary policy shock: a (more) agnostic identification procedure. *LEM Working Paper Series*.

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: comments on Hausman and Woodward. *British Journal for the Philosophy of Science 53*(3), 411–453.

Dawid, A. P. (2010). Beware of the dag! *Journal of Machine Learning Research 6*, 59–86.

Eberhardt, F. (2016). Green and grue causal variables. *Synthese 193*(4), 1029–1046.

Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics 75*(5), 705–730.

Neapolitan, R. (2003). Stochastic causality. paper presented at the International Conference on Cognitive Science, Sydney, Australia, July 2003.

Shimizu, S., P. O. Hoyer, A. Hyvärinen, and A. Kerminen (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research 7*, 2003–2030.

Spirtes, P., C. N. Glymour, and R. Scheines (2000). *Causation, prediction, and search*. MIT press.

Spirtes, P. and R. Scheines (2004). Causal inference of ambiguous manipulations. *Philosophy of Science 71*(5), 833–845.

Woodward, J. (2016). The problem of variable choice. *Synthese 193*(4), 1047–1072.

Causal Patterns as Intrinsic Emergence

George Nguyen
University of Queensland
✉ g.nguyen@uq.edu.au

ABSTRACT

Barbour *et al.* (2014) introduce a measure of configurational complexity based on dynamical quantities. Their simulations of the dynamics of self-gravitating systems show a monotonic increase of this complexity defining an arrow of time. They demonstrate further, based on the results of Marchal & Saari (1976), that there is a growth of information storage in local subsystems as the global complexity increases. In this paper, we argue that both the general agreement of causal directionality with time's arrow and the stability of causal mechanisms are emergent patterns of the dynamical information growth in the subsystem to which agents are internal—a phenomenon Crutchfield (1994) calls "intrinsic" emergence. In light of this, we discuss the relation between causation and dynamics with respect to the levels of resolution at which agents model the world.

References

- Barbour, J., Koslowski, T. & Mercati, F. Identification of a gravitational arrow of time. *Phys. Rev. Lett.* **113**, 181101 (2014).
- Crutchfield, J.P. The calculi of emergence: computation, dynamics and induction. *Physica* D **75**, 11 (1994).
- Marchal, C. & Saari, D.G. On the final evolution of the n-body problem. *J. Diff. Equ.* **20**, 150 (1976).

Title

X-Separation: Minimising Bayesian network parameterisation for informal reasoning.

Author, Affiliation, Email Address

Erik P Nyberg, Monash University, erik.nyberg@monash.edu

Extended Abstract

Intelligence analysis is a complex business. Analysts need to combine many pieces of interrelated information to assess the probabilities of competing hypotheses and give good reasons for their conclusions. History shows that they are prone to the same reasoning errors as everyone else (groupthink, confirmation bias, etc), but when analysts make these errors the consequences can be disastrous. For example, all sides agree that the US intelligence analysis of Iraq's weapons program was extremely poor, and provided the flawed justification for invading Iraq. Consequently, the US government is funding a major research program, CREATE, to develop software, training and procedures to help analysts produce more well-reasoned reports. These tools will embody cognitive strategies to assist with informal reasoning and decision making in complex situations.

Constructing a Causal Bayesian Network (CBN) is one such approach, which has already demonstrated its value in analysing complex legal evidence and arguments. A reasoner can explicitly represent relevant beliefs about the world in a probabilistic causal map, e.g. *Breast Cancer* shown below. Software performs the complex calculations required to combine all the information, and can also identify which evidence, e.g. a biopsy result, would greatly affect the probability of a conclusion, e.g. having breast cancer. This lowers the reasoner's cognitive load and clarifies why and how much their beliefs support their conclusion. More detail, including a simplified legal example showing how two frequent probabilistic fallacies can easily be overcome, is provided in Korb & Nyberg (2016). Pioneers of this approach include Korb (2004), Hahn & Oaksford (2007), Fenton, Neil & Lagnado (2013), and Nicholson et al (in preparation). All these researchers, combined in one interdisciplinary team, have won a CREATE grant to continue this work.

There are several competing structured methods for clarifying informal reasoning, including Wigmore charts (Wigmore, 1913), formal proof tableaux, argument schemes (e.g. Walton, 2008), argument scenarios (e.g. Verheij et al, 2016), and argument diagrams (e.g. Toulmin, 1958; van Gelder et al, 2009). The most important difference is that CBNs quantify the strength of individual connections. They use conditional probabilities: the probability of each effect given each possible combination of causes, e.g. the probability of an abnormal imaging result given that the test is performed, the clinical examination result was abnormal, and the patient has breast cancer. Utilities can also be quantified, e.g. the cost of performing the biopsy. Quantifying individual connections and utilities can be an enormous advantage for assessing the overall strength of an argument. However, it is also a well-known weakness, because the workload required to estimate these parameters can be prohibitive.

We fully parameterised *Breast Cancer* using publicly available Australian statistics. Other networks must use more subjective estimates. In either case, providing each parameter requires some time and effort, and as the size of a CBN increases, the number of parameters generally

increases super-exponentially. Even *Breast Cancer*, where many of the variables are binary and few variables are directly connected, still required 240 estimates. Full parameterisation is a necessary evil for other applications of CBNs, so it is often presumed to be necessary for assessing specific arguments. Not so: our aim here is to outline a set of widely applicable techniques that vastly reduce the number of parameters required.

One key is that the conclusion of an argument relates to only a small part of the CBN. Often this is the probability distribution over a single target variable, such as probability of breast cancer, or the existence of a single causal arrow. The remainder of the CBN is only relevant insofar as it has an impact on this part. Another key is that some pieces of evidence make other aspects of the CBN irrelevant in formally identifiable ways. So, given the graph structure G, the conclusion H, and the evidence E, software can automatically identify the unnecessary parameters.

d-separated: The paradigmatic case is *d*-separation, which describes when specific conditions have 'blocked' all paths from some variable to the conclusion, making the variable probabilistically irrelevant. We define several weaker analogues that apply to parameters, and show that given each separation the parameter is independent of the conclusion.

0-separated: If the probability of some value is zero, then probabilities conditional on this value become irrelevant. E.g. if we know the age of the patient, then probabilities for breast cancer given other ages become irrelevant.

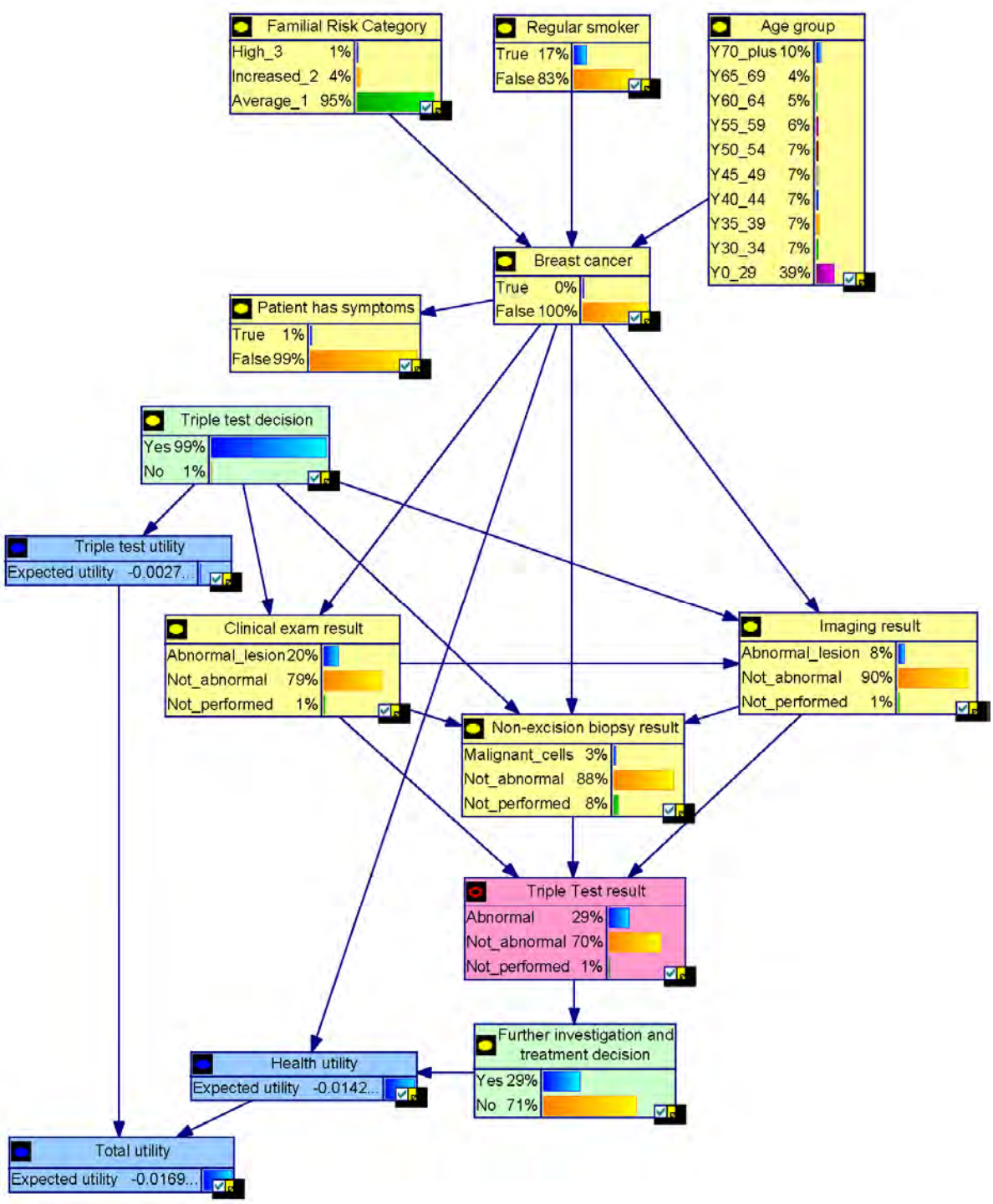
Λ-separated: If we know an effect value, then all we need to update the distribution over a cause variable is the ratio of the conditional probabilities for the effect, not their absolute values. E.g. if there is an abnormal biopsy, then it is sufficient to know that this is 19 times more likely with breast cancer than without.

E-separated: If we are arguing about what probability an event had before it occurred, then effects of the event shouldn't be used to directly update the probability of that event.

J-separated: If V_2 is *d*-connected by some path via V_1 to the conclusion H, and we know the probability distribution of V_1 (by Jeffrey conditionalization), then the parameters connecting V_2 to V_1 are irrelevant to H unless we acquire new information about V_2. E.g. we know the probability distribution for familial risk, so it is unnecessary to specify how familial risk is computed from more specific information unless we acquire it.

X-separated: A parameter is '*X*-separated' iff it is *d*- or 0- or Λ- or *E*- or *J*-separated. This disjunction is the broader notion of separation we need.

Some arguments are intended to establish the absolute probability of their conclusion, whereas other arguments are only intended to raise its relative probability. For absolute arguments, *X*-separated parameters are unnecessary. For relative arguments, we need to distinguish background knowledge from the additional evidence. We then define '*r*-separation', which identifies the parameters structurally irrelevant to the additional evidence. For relative arguments, both *r*-separated and *X*-separated parameters are unnecessary. Using these definitions, our CREATE software will automatically minimise parameterisation, which further simplifies an effective cognitive strategy to deal with a risky and complex world.



Breast Cancer CBN

References

Fenton N., Neil M., & Lagnado D.A. (2013). A General Structure for Legal Arguments About Evidence Using Bayesian Networks. *Cognitive Science, 37*(1), 61–102.

van Gelder T., ter Berg T., Patterson F., & Teppema S. (2009). *Critical Thinking: Reasoning and Communicating with Rationale.* Pearson Education Benelux, Amsterdam, NL.

Hahn U., & Oaksford M. (2007). The Rationality of Informal Argumentation: A Bayesian Approach to Reasoning Fallacies. *Psych. Rev., 114*(3), 704–732.

Korb K.B. (2004). Bayesian informal logic and fallacy. *Informal Logic*, *24*(1), 41–70.

Korb K.B. & Nyberg E.P. (2016). Analysing Arguments Using Causal Bayesian Networks. *The Reasoner, 10*(4), 26–27.

Nicholson A.E., Mascaro S., Thakur S., Korb K.B., & Ashman R. (in preparation). Delphi Elicitation for Strategic Risk Assessment.

Toulmin S.E. (1958). *The Uses of Argument*. Cambridge University Press, Cambridge, UK.

Verheij B., Bex F., Timmer S.T., Vlek C.S., Meyer J.C., Renooij S. & Prakken H. (2016). Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, *15*(1), 35–70.

Walton D., Reed C. & Macagno F. (2008). *Argumentation Schemes.* Cambridge University Press, Cambridge, UK.

Wigmore, J.H. (1913). *The principles of judicial proof: as given by logic, psychology, and general experience, and illustrated in judicial trials.* Little, Brown & Co., Boston MA, US.

MuST10, Sydney 1-3 March 2017

"Exact replication or varied evidence? Reliability, robustness and the reproducibility problem"

The "Reproducibility Project: Psychology" by the Open Science Collaboration caused some stir among psychologists, methodologists as well as scientists, since less than half of the replicated studies succeeded in reproducing the results of the original ones. The APA has attributed this result to hidden moderators that rendered the replications ineffective. Also publication bias and low power have been identified as possible sources for such mismatch. While some analysts have provided formal confirmation for the plausibility of such explanations (Etz and Vandekerkhove, 2016), others have further insisted on the problem of noisy data and suggested that "to resolve the replication crisis in science we may need to consider each individual study in the context of an implicit meta-analysis" (Gelman, 2015).

I suggest that this meta-analytic approach should keep track of the distinct roles that different epistemic dimensions, such as reliability, reproducibility and robustness of results, play in causal inference in the medical domain. In particular, I propose a 3-layers perspective to modelling epistemic dynamics in probabilistic causal assessment:

1. A basic level of evidential support to the hypothesis at hand (and various evidence aggregation/amalgamation techniques) - the traditional focus of philosophy of science, and of statistical inference;
2. A higher order level of "meta-evidential" dimensions related to the body of evidence itself: coherence of reports, (in)dependence structure; reliability, relevance - this level has been mainly investigated in Bayesian and formal epistemology. However, also standard statistical techniques developed to detect biases focus on these aspects (see Wood et al. 2008, Rising et al. 2008, Krauth et al. 2013);
3. A further level related to the information/evidence concerning these meta-epistemic dimensions, e.g. grounds for judging a given source as reliable, or for assuming specific (in)dependency relations. These grounds may be of epistemic and/or pragmatic nature, such as financial or professional incentives, and derive from the social ontology of a given scientific eco-system. Socially epistemology has been traditionally investigating the interaction of these dimensions with scientific practice.

These three levels have been working relatively independently so far. Furthermore, at the methodological level, the Bayesian and the frequentist paradigms have adopted different techniques to (locally) address these issues and evaluate accuracy and reliability of data.

In Landes et al. (forthcoming), as well as Osimani and Landes (forthcoming), we started to outline a unifying project where all these aspects can be considered in their reciprocal interactions by looking at the entire body of evidence available.

We adopt Bayesian epistemic nets as a platform for such an enterprise and break down the inferential pattern into an epistemic level – where various indicators of causality, such as difference making, correlation, or probabilistic dependence, may contribute to the confirmation of the causal hypothesis, and an empirical level, where concrete evidence, coming from different kinds of studies (RCTs, cohorts, laboratory studies), may be used to support the hypothesis that such diagnostic relationships hold. To the different items of incoming evidence, a "relevance" and a "reliability" node are attached: these constitutes "weights" which mediate the confirmatory force of the evidence.

The major challenge to this project is the manifest incommensurability of such an approach with the standard methodology developed and consolidated in medical research, where frequentist statistics is the canon.
This incommensurability obviously roots in the antipodal foundations of the two paradigms, but also emerges at the practical level (although there have been efforts to reconcile them: see Efron, 2005)
I start investigating possible ways to sidestep this problem, by analyzing the role of reliability issues in the two settings. In particular, I distinguish two notions of reliability: Reliability$_1$ and Reliability$_2$. The former refers to the minimization of random error, the latter to the minimization of systematic error (whether due to bias or confounding or both). I describe its central role in the Evidence Based Medicine approach, as a consequence of the adoption of frequentist statistics. Furthermore, by making this distinction:

a. the "reliability" node in the Bayesian network is let represent reliability$_2$ exclusively. This move allows to let reliability$_1$ be directly incorporated in the likelihood of the evidence on the hypothesis (of the specific relationship holding);
b. By using frequentist measures of reliability$_1$, such as significance level and power, as proxies for such likelihood measures, one can try to include "frequentist evidence" into the epistemic framework and make it interact with other pieces of evidence;
c. Frequentist estimates of reliability$_2$ such as evidence rankings and other measures of bias, can also be incorporated in the epistemic framework by letting them flow into the reliability node.

I also present various "degrees" of replication, from exact replication, to sensitivity analysis, to robustness analysis and their role in ensuring the two sorts of reliability. At a theoretical level, modeling evidence acquisition on this basis goes in the "meta-analytic" direction advocated by Gelman (2015), and casts the reproducibility debate in a new light by detailing its rationales. Furthermore, on the basis of this analysis, some shortcomings in Bovens and Hartmann`s (2003) as well as Claveau's (2013) models of dependence of observations may be identified and corrected.
At a practical level, these analyses provide the basis for the development of guidelines for best research strategies. Furthermore, on a more concrete level, the Bayesian network thus developed, allows for an overall assessment of the causal hypothesis on the basis of all the available evidence, and the related meta-evidential dimensions. This will be illustrated by using simulations and case studies.

Market Crashes as Critical Phenomena?
Explanation, Idealization, and Universality in Econophysics

Jennifer Jhun
jennifersjhun@gmail.com
Department of Philosophy
Lake Forest College

Patricia Palacios
Patricia.Palacios@lrz.uni-muenchen.de
Munich Center for Mathematical Philosophy
Ludwig-Maximilians Universität München

James Owen Weatherall
jim.weatherall@gmail.com
Department of Logic and Philosophy of Science
University of California, Irvine

Mainstream economic models of financial markets have long been criticized on the grounds that they fail to accurately account for the frequency of extreme events, including market crashes. Mandelbrot and Hudson (2004) put the point starkly in their discussion of the August 1998 crash: "By the conventional wisdom, August 1998 simply should never have happened. The standard theories estimate the odds of that final, August 31, collapse, at one in 20 million, an event that, if you traded daily for nearly 100,000 years, you would not expect to see even once. The odds of getting three such declines in the same month were even more minute: about one in 500 billion" (p. 4). Similar critiques have been mounted in connection with the October 1987 "Black Monday" crash, the October 1997 "mini-crash", and the 2007-08 financial crisis, as well as other large drawdowns over the last thirty years. By the lights of ordinary economic reasoning, such events simply should not occur.

Motivated in part by the October 1987 crash, and in part by new interdisciplinary initiatives in nonlinear dynamics during the late 1980s, the last thirty years has seen an upswell in alternative approaches to economic modeling, many of which have been inspired by analogies with statistical physics. Work in this tradition has come to be known as *econophysics*, a term coined by H. Eugene Stanley in 1996. One goal of econophysics has been to develop new financial models that can accurately describe, and perhaps even predict, extreme events such as financial crises.

Despite the apparent empirical successes of some models in econophysics, the field has not been widely embraced by economists. In this contribution, we make the case that econophysics deserves more credit than these critiques give it, as an enterprise that is at least sometimes successful in its main goals of predicting and explaining

economic phenomena of certain kinds. Our strategy will not be to address the general criticisms just described head-on, and we do not mean to argue that all models from econophysics, or even most or many models, are successful. Instead, we will focus on just one model that, we will argue, has two features of interest: it (1) draws on a significant analogy with statistical physics, in a way that goes beyond standard modeling methods in economics; and (2) has real explanatory, and possibly predictive, power. Our principal goal is to elaborate and defend how we take the model to work, including where and how the analogy with statistical physics enters, and to articulate what sorts of novel insights into market behavior we believe it offers. In this sense, we take the model we consider as "proof of concept", while simultaneously providing a case-study for the sorts of explanatory goals that arise in econophysics.

The model we consider is the Johansen-Ledoit-Sornette (JLS) model of "critical" market crashes (Johansen, Ledoit, and Sornette 2000), which uses methods from the theory of critical phase transitions in physics to provide a predictive framework for financial market crashes. This model is of particular interest because it aims both to predict and describe market-level phenomena – crashes – and to provide microscopic foundations that explain how that behavior can result from interactions between individual agents. More specifically, in addition to its predictive role, the JLS model aims to explain two "stylized facts" associated market crashes. The first is the fact that stock market returns seem to exhibit power law behavior in the vicinity of a crash, and the second is so-called *volatility clustering*, which is the fact that market returns seems to exhibit dramatic, oscillating behavior before crashes, with large changes followed by large changes.

The plan of the contribution is as follows. We will first introduce the JLS model itself, focusing on the role the analogy with critical phase transitions plays in the model. Then, in we will argue against one tempting way of understanding how the model works, and instead defend a somewhat different understanding. Central to our argument will be the observation that although the analogy with critical phase transitions is crucial in motivating and developing the model, in the end the analogy is only partial. In particular, although the model fruitfully draws on the renormalization group theory of critical exponents, financial crashes do not seem to constitute a universality class in the strict sense that one encounters in that area of physics. Nonetheless, we argue, there is a weaker sense in which crashes exhibit universal features. This weaker notion of universality allows one to draw novel inferences about the microscopic mechanisms that might underlie crashes. Since the model helps make salient the possible microscopic mechanisms that could explain the occurrence of a crash, we claim that the model provides an explanation of crashes that is both causal (in the sense of Woodward (2003)) and reductive.

Finally we will explore how the argument just sketched relates to recent debates in philosophy of science concerning explanatory uses of idealized models. We will argue that the JLS model is naturally understood as a "minimal model" in the sense of Batterman and Rice (2014); moreover, it is a model that employs infinite idealizations

in apparently essential ways (Batterman 2001; Batterman 2005; Batterman 2009). Nonetheless, we claim, contra Batterman and Rice, that it provides both a causal and reductive explanation of market crashes.

We conclude with some remarks about possible policy suggestions. In particular, we argue that our interpretation of the JLS model as one that yields causal explanations suggests possible methods by which policymakers could intervene on the economy in order to prevent crashes or to halt the spread of one. Indeed, this seems to be the way economic policy is often motivated in the real world; the JLS model can thus be used as a diagnostic tool, allowing economists and regulators to formulate new measures or to assess the performance of ones are already in place.

Emergence of scale-free characteristics in socio-ecological systems with bounded rationality

Dharshana Kasthurirathna and Mahendra Piraveenan*

Complex Systems Research Group, The University of Sydney
*** mahendrarajah.piraveenan@sydney.edu.au**

Network based games are increasingly used to understand critical phenomena in socio-ecological systems. The concept of Nash equilibrium has been an important cornerstone in understanding the dynamics of such systems. While Nash equilibrium assumes that all players in a system are fully rational, most real-world strategic decision making scenarios involve players with non-optimal or bounded rationality, resulting in their strategies and behaviour deviating from those predicted by the Nash equilibrium. The possible limitations, such as the amount of information at hand, cognitive capacity, and the computational time available, may force a self-interested autonomous player or agent to have bounded rationality and therefore to make non-optimal decisions.

Numerous theories have been presented to model the non-optimal rationality of players in strategic games, including the concepts of the near-rationality equilibrium and the quantal response equilibrium. However, these models do not attempt to quantify, in a predictive manner, the levels of rationality prevalent in individual players based on their observable characteristics. Meanwhile, studies in psychology and cognitive science have conjectured that the rationality of individuals is correlated to the level of their social interactions. Here we propose a topological model of bounded rationality in socio-ecological systems, based on this conjecture. Using this model, we investigate how such systems could topologically evolve to have higher system rationality, given a heterogeneous bounded rationality distribution. Since the calculation of Nash equilibrium assumes perfect rationality of all players, we use the average Kullback–Leibler divergence between Nash and Quantal Response equilibria of each game played within the system as an indicator of overall system rationality. The higher this divergence, the lower the system rationality.

We show that when a socio-ecological system with a random topology is optimised towards higher system rationality (the system on average is driven towards Nash equilibrium), scale-free and small world features emerge. This result is true for games with single or multiple equilibria. In the case of games with multiple equilibria, the fraction of links in a network where multiple equilibria are actually prevalent is topologically dependent. When average rationality is lower, the scale-freeness of the socio-ecological network aids in increasing the fraction of links with multiple equilibria. However, when the average rationality is higher, the scale-freeness actually aids in decreasing this fraction. In fact, we demonstrate that the correlation between the ‘scale-freeness’ and the fraction of links with multiple equilibria goes through a phase transition when average network rationality is increased.

Our experiments could be summarized as follows: First of all, we compared a number of network classes, and showed that among these classes, it is the scale-free networks which facilitate the best convergence towards Nash equilibrium (highest system rationality), on average. We argued that this might be one reason why many real-world social systems are scale-free. Seeking further evidence for this conjecture, we simulated the topological

evolution of social systems using the simulated annealing technique, beginning from a random network topology. We showed that when evolutionary pressure is applied on social systems to converge, on average, towards Nash equilibria, scale-free and small world features emerge. Following this, we turned our focus on games with multiple equilibria. Again, we demonstrated that when evolutionary pressure is applied on systems to converge, on average, towards Nash-equilibria (regardless of which equilibrium state a particular pair of players converge towards), scale-free and small world features emerge. We also considered the likelihood of the existence of multiple equilibria among the players of a system with a bounded heterogeneous rationality distribution, and found that a delicate balance exists: when the average rationality (this must be distinguished from what we call the 'system rationality', which is computed from the KL divergence between QRE and Nash equilibria) is low, the scale-free nature of the system encourages the emergence of multiple equilibria, while when the average rationality is high, the scale-free character in fact hinders the existence of multiple equilibria. Therefore, the number of rational choices available to players, from which they cannot deviate without loss, depends on the social network topology as well as the level of rationality prevalent in the system.

It is important to understand that 'rationality' of players and that of a system have been defined in a very specific way in our work. It could be argued that 'rational' players are those who try to maximize their average individual pay-offs. If players attempted to do this within a heterogeneous system, they may well make choices that are contrary to that suggested by Nash equilibrium. Therefore, a system which converges towards Nash equilibrium will not necessarily have increasing average pay-offs. Indeed, in the case of Prisoners Dilemma game, the convergence towards Nash equilibrium results in decreasing average pay-offs. However, in an environment where there is a lot of 'mistrust' and/or competition, the priority of the players will be to make sure that their average pay-offs are better than other players with whom they compete. The self-interest, and the relative well-being in the system, therefore gains prominence over the 'absolute well-being'. In such systems, the convergence towards Nash equilibria, on average, means the players are getting better at preserving their 'relative' self-interest, and thus becoming more 'raional' in a selfish sense. However, in games other than Prisoners Dilemma (for example, in the stag-hunt game), we find that the average pay-off indeed could increase as the system converges towards the (multiple) Nash equilibria, depending on the actual values of pay-offs for each scenario. Thus, the 'public good' of the system matches with the selfish rationality of players. Therefore, it is important to realise that the results we have obtained are applicable in terms of average selfish rationality of players, which often but not always matches with the 'public good' of the system.

Our results provide a possible explanation for the prevalence of scale-freeness in real world socio-ecological systems, and explore how the scale-freeness in turn affects the cognitive decision making behaviour of such systems.

Keywords: *scientific inference, pharmacology, epistemology, Bayesian confirmation, evidence, relevance, similarity, analogy, computer simulation*

Submission to MuST 2017

Learning from Relevant Evidence: Similarity and Analogy in Pharmacology – Abstract –

Roland Poellinger[1]

Pharmacological research is often driven by many forces at once: Cost effectiveness must be balanced against extensive data-collecting, potential risk against probable benefit, and breadth of applicability against well-documented higher confidence for smaller target groups. Many such decisions must be taken during the development stages of a certain drug before the desired effectiveness and safety level is reached and the drug is allowed to be marketed. A language for expressing both benefit and safety is found in the probabilistic language of expected utilities and dis-utilities. Nevertheless, the formalization of a given decision problem in such vocabulary can only be as informative about future drug users as the evidence it is rooted in. Yet, whether it is all the evidence about the drug's effects that is to be taken into consideration, or only the best evidence available, is the subject of an ongoing discussion in the philosophy of medicine. A recent paper by Landes, Osimani, and Poellinger (Landes *et al.* [forthcoming]) explores the possibility of amalgamating all available evidence in a Bayesian reconstruction of scientific inference for the integrated probabilistic assessment of a drug's causal (side-)effects.

One important justification of the confirmatory support a piece of evidence lends to a given hypothesis (by virtue of it being evidence for an indicator of the very hypothesis) is the postulate (or implicit assumption) of analogy between the circumstances generating the evidence and the hypothesis' intended (future) scope of application. Sir Austin Bradford Hill lists analogy as one of his famous guidelines towards an informed assessment of potential causes in epidemiology:

> *In some circumstances it would be fair to judge by analogy. With the effects of thalidomide and rubella before us we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy. ([Hill 1965, p. 11])*

Key to employing the concept of analogy in the context of scientific reasoning is the distinction of the conceptual levels involved (along the lines of [Bovens & Hartmann 2003]): (i) the causal hypothesis, *Hyp*, (ii) testable indicators of the causal claim, i.e., theoretical consequences of the causal hypothesis, *Ind*, and (iii) concrete data, *Rep*, interpreted as evidence speaking for or against the respective indicators (thereby indirectly supporting the hypothesis, or not) at a certain *level of significance* relative to the question asked (iv). Moreover, the causal modeling framework provides the inner structure for the hypothesis and allows for expansion of *Hyp* to "drug D causes side-effect E in causal model M within context U". Within this multi-layered reconstruction

[1]Munich Center for Mathematical Philosophy (MCMP/LMU) · r.poellinger@lmu.de

of scientific inference, the concept of Bayesian confirmation can be utilized to formulate precisely how a causal hypothesis about a drug's potentially harmful side-effects is confirmed or disconfirmed. The aim of this paper is to explore different analogy-based, confirmatory inference patterns with respect to their justification in pharmacological risk assessment. By relating formal explications of similarity, analogy, and analog simulation, three sources of confirmatory support for a causal hypothesis are distinguished in reconstruction:

1. **Inference from relevant reports:** When the conditions of a given study correspond to the intended application of the investigated hypothesis, reports about the study are marked as relevant for the hypothesis, thus facilitating knowledge transfer from evidence to hypothesis. Towards that aim, study and target conditions are broken into three components: the drug itself (D), the causal model implicit in the hypothesis (M), and the respective population (U). Pair-wise comparison is based on a similarity measure to be chosen w.r.t. the nature of the investigation; e.g., the components might be compared using a geometric measure of similarity as given by the distance between property vectors like in the example case.

2. **Inference from established causal knowledge:** In the case of analogical inference from a second well-tested hypothesis Hyp^*, the connection between source and target is not established via similarity but across a syntactic isomorphism between the hypotheses' consequence sets, i.e., a theoretical mapping on the indicator level (along the concept of *analog simulation*, as discussed in [Dardashti *et al.* 2016]). Once this bridge is defined (motivated and justified by model-external empirically grounded arguments), evidence for Hyp^* (the established hypothesis) will also boost confidence in Hyp (the hypothesis under investigation).

3. **Inference from computational models:** If the analog system is a virtual, computational model of the investigated hypothesis, the bridge between source and target is not motivated by model-external considerations but much rather by model-internal constraints. The concept of analogy (analog simulation, respectively) can also be utilized to illuminate how an *artificial* system can possibly provide support for a causal hypothesis about an *actual* drug with *real* risk.

While the paper focuses on hypothesis testing for the purpose of risk assessment in pharmacology, the second and third pattern in the list above make the role of analogical reasoning in the *formulation of a hypothesis* obvious. In tracing the confirmatory support of heterogeneous evidence across distinct epistemological categories, the collection of modules presented in this paper may precisely serve as a toolbox for scientific justification in the dialog between hypothesis testing and theory revision.

References

Bovens, Luc and Stephan Hartmann. 2003. *Bayesian Epistemology*. Oxford University Press.

Dardashti, Radin, Karim Thébaut, and Eric Winsberg. 2016. Confirmation via Analogue Simulation: What Dumb Holes Can Tell us About Gravity. *British Journal for the Philosophy of Science*. forthcoming.

Hill, Austin Bradford. 1965. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, **58**(5), 295–300.

Landes, J., B. Osimani, and R. Poellinger. forthcoming. Epistemology of Causal Inference in Pharmacology. Towards an Epistemological Framework for the Assessment of Harms.

Information Thermodynamics of Complex Computation

Mikhail Prokopenko
Centre for Complex Systems
Faculty of Engineering and Information Technologies
University of Sydney

Any computational process can be defined via primitive information-processing functions, constructing the generic capacity to support complex computation. In studying the fundamental physical limits and properties of computational processes, one is faced with the challenges of interpreting these primitives through well-defined information-theoretic as well as thermodynamic quantities [Prokopenko & Lizier, 2014]. In general, these computational elements affect state transitions of the system, creating information dynamics which may approach critical regimes and undergo phase changes. For example, Fisher information, characterising computational sensitivity, is known to peak at critical regimes [Prokopenko et al., 2011; Prokopenko et al., 2015]. Importantly, the thermodynamic interpretations of the information-processing functions help to explain the underlying critical behaviour, by associating information flows with particular physical fluxes.

A search for fundamental connections between computation and maximisation of suitably defined measures of information dynamics is a vigorous research topic, pursued within the field of information thermodynamics [Spinney et al., 2016]. In this talk, we discuss several novel relationships between well-known information-theoretic and thermodynamic quantities. In particular, in isothermal systems near thermodynamic equilibrium, the curvature of internal entropy is shown to be dynamically related to Fisher information [Prokopenko & Einav, 2015]. This relationship explicitly connects entropy production with sensitivity and uncertainty of computational processes intrinsic to complex systems, and allows us to consider thermodynamic interpretations of several important extreme cases and trade-offs.

References

1. Prokopenko, M., Barnett, L., Harré, M., Lizier, J. T., Obst, O. and Wang, X. R. Fisher Transfer Entropy: Quantifying the gain in transient sensitivity, *Proceedings of The Royal Society A*, vol. 471(2184), 20150610, 2015.
2. Prokopenko, M., Einav, I. Information thermodynamics of near-equilibrium computation, *Physical Review E*, 91, 062143, 2015.
3. M. Prokopenko, J. T. Lizier, Transfer Entropy and Transient Limits of Computation, *Scientific Reports*, 4:5394, 2014.
4. M. Prokopenko, J. T. Lizier, O. Obst, and X. R. Wang, Relating Fisher information to order parameters, *Physical Review E*, 84, 041116, 2011.
5. Spinney, R. E., Lizier, J. T., and Prokopenko, M. Transfer entropy in physical systems and the arrow of time, *Physical Review E*, 94, 022135, 2016.

Quantum causal modelling

Sally Shrapnel[1, 2, *]

[1] *School of Historical and Philosophical Inquiry, The University of Queensland, St Lucia, QLD 4072, Australia*
[2] *Centre for Engineered Quantum Systems, School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia*

(Dated: 17th January 2017)

In this talk I will provide an introduction to the fledgling field of quantum causal modelling. I present a brief history of the field; outline key challenges and detail recent progress. The quantum causal modelling framework of Costa and Shrapnel (2016) is introduced and I consider whether these recently characterised formal tools can provide causal explanations for Bell-type correlations.

References

1. Fabio Costa and Sally Shrapnel (2016), "Quantum Causal Modelling", *New Journal of Physics*, 18(6) 063032.

* s.shrapnel@uq.edu.au

Transfer entropy in physical systems and the arrow of time

Richard E. Spinney*[1], Joseph T. Lizier[1], and Mikhail Prokopenko[1]

[1]Centre for Complex Systems, The University of Sydney, Sydney, New South Wales, Australia, 2006.

December 2, 2016

Abstract

Recent developments have cemented the realization that many concepts and quantities in thermodynamics and information theory are shared. We consider a highly relevant quantity in information theory and complex systems, the transfer entropy, and explore its thermodynamic role by considering the implications of time reversal upon it. By doing so we highlight the role of information dynamics on the nuanced question of observer perspective within thermodynamics by relating the temporal irreversibility in the information dynamics to the configurational (or spatial) resolution of the thermodynamics. We then highlight its role in perhaps the most enduring paradox in modern physics, the manifestation of a (thermodynamic) arrow of time. We find that for systems that process information such as those undergoing feedback, a robust arrow of time can be formulated by considering both the apparent physical behaviour which leads to conventional entropy production and the information dynamics which leads to a newly defined quantity we call the information theoretic arrow of time.

In recent years great progress has been made in describing the thermodynamics of small systems, now increasingly experimentally realizable, through frameworks such as stochastic thermodynamics [1, 2, 3, 4]. These frameworks and, more broadly, the constituent work and fluctuation theorems [5, 6, 7, 8, 9, 10, 11] have deeply connected entropy production, dissipation, statistical irreversibility and the arrow of time [12, 13, 14, 15]. Meanwhile, the field of information theory has enjoyed great success in identifying meaningful components of computation in complex systems strongly implicating transfer entropy amongst other measures [16, 17, 18, 19, 20, 21, 22]. More recently still it has become apparent that these two fields, and the concepts they employ, are deeply connected. Through modern consideration of the thermodynamics of systems which were once only explored in the imagination as thought experiments, such as the enduringly compelling Maxwell's demon, information theoretic measures, such as transfer entropy, have taken a central role in *physical* systems [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 25, 33, 34, 26, 35].

Such developments raise intriguing questions: Is information physical? Do the concepts developed studying computation have a role to play in thermodynamics and do thermodynamic concepts have a role in understanding computation? And what of the arrow of time? We observe clouds of smoke billowing from chimneys, not collecting in the atmosphere and flowing into them, plates fall and smash when they hit the floor and the pieces do not spontaneously recombine. This flow of time is considered synonymous with the thermodynamic arrow of time, entropy increases, it does not decrease. Recent advances have formalized the known generalization that such a law, especially for small systems, is only valid statistically [36, 5, 37]. However, even these statistical laws need not hold under feedback [24] or more generally under different

*richard.spinney@sydney.edu.au

observer perspectives of coupled systems. Information theoretic quantities that characterize the information gained by an observer through measurement, or more generally the information that flows out of the the system must be taken into account. These quantities have been used as corrections to the second law, but so far a treatment or generalization of the arrow of time, identified for a system coupled to some feedback mechanism as an involutive quantity that increases, on average, only in the forwards time direction, but decreases by the same amount if viewed in reverse, has not been offered.

We contribute to these ideas by making further connections between these fields. To do so we seek to understand transfer entropy as completely as possible within physical systems, its role within the question of observer perspective, and to implicate it within a generalized form of the arrow of time. We first contrast the temporal phenomena of irreversibility in the information dynamics with the difference in the implied thermodynamics that results from different configurational resolutions, implicating transfer entropy in the question of observer perspective, uncertainty and irreversibility. This result uniquely considers the difference between transfer entropies in the same direction between partitioned constituents of the system, but with the inclusion of time reversal as opposed to the transfer entropy in different directions as in, for example, [28]. We go on to offer a novel information theoretic interpretation of transfer entropy, relating it to the ability of an observer to encode measured behaviour, identifying it with both a suitable difference in predictive capacity on the target and on the source. Finally, we then bring these concepts together to form a generalized thermodynamic arrow of time and examine the place of such a formalism within the field of information thermodynamics.

References

[1] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).

[2] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[3] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).

[4] C. Van den Broeck and M. Esposito, Physica A: Statistical Mechanics and its Applications **418**, 6 (2015).

[5] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).

[6] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).

[7] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[8] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[9] J. L. Lebowitz and H. Spohn, Journal of Statistical Physics **95**, 333 (1999).

[10] R. J. Harris and G. M. Schütz, J. Stat. Mech. **2007**, P07020 (2007).

[11] R. Spinney and I. Ford, Fluctuation Relations: A Pedagogical Overview, in *Nonequilibrium Statistical Physics of Small Systems*, edited by R. Klages, W. Just, and C. Jarzynski, pp. 3–56, Wiley-VCH Verlag GmbH & Co. KGaA, 2013.

[12] J. M. R. Parrondo, C. V. den Broeck, and R. Kawai, New J. Phys. **11**, 073008 (2009).

[13] C. Maes and K. Netočný, Journal of Statistical Physics **110**, 269 (2003).

[14] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007).

[15] P. Gaspard, Journal of Statistical Physics **117**, 599 (2004).

[16] L. Barnett and T. Bossomaier, Phys. Rev. Lett. **109**, 138105 (2012).

[17] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Phys. Rev. E **77**, 026110 (2008).

[18] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Chaos **20**, 037109 (2010).

[19] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Information Sciences **208**, 39 (2012).

[20] J. T. Lizier, Measuring the Dynamics of Information Processing on a Local Scale in Time and Space, in *Directed Information Measures in Neuroscience*, edited by M. Wibral, R. Vicente, and J. T. Lizier, Understanding Complex Systems, pp. 161–193, Springer Berlin Heidelberg, 2014.

[21] M. Lungarella and O. Sporns, PLoS Comput Biol **2** (2006).

[22] X. R. Wang, J. T. Lizier, and M. Prokopenko, Artificial Life **17**, 315 (Fall 2011).

[23] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Nat Phys **11**, 131 (2015).

[24] T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010).

[25] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat Phys **6**, 988 (2010).

[26] D. Abreu and U. Seifert, Phys. Rev. Lett. **108**, 030601 (2012).

[27] T. Sagawa and M. Ueda, Phys. Rev. E **85**, 021104 (2012).

[28] G. Diana and M. Esposito, Journal of Statistical Mechanics: Theory and Experiment **2014**, P04010 (2014).

[29] D. Hartich, A. C. Barato, and U. Seifert, J. Stat. Mech. **2014**, P02016 (2014).

[30] J. M. Horowitz and M. Esposito, Phys. Rev. X **4**, 031015 (2014).

[31] J. M. Horowitz and H. Sandberg, New J. Phys. **16**, 125007 (2014).

[32] D. Mandal and C. Jarzynski, PNAS **109**, 11641 (2012).

[33] J. M. Horowitz, T. Sagawa, and J. M. R. Parrondo, Phys. Rev. Lett. **111**, 010602 (2013).

[34] D. Abreu and U. Seifert, EPL **94**, 10001 (2011).

[35] M. Bauer, D. Abreu, and U. Seifert, J. Phys. A: Math. Theor. **45**, 162001 (2012).

[36] E. Gerstner, Nature News (2002).

[37] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).

Jan Sprenger (Tilburg)

Foundations of a Probabilistic Theory of Causal Strength

This contribution develops axiomatic foundations for a probabilistic theory of causal strength. I proceed in three steps: First, I motivate the choice of causal Bayes nets as a framework for defining and comparing measures of causal strength. Second, I prove several representation theorems for probabilistic measures of causal strength---that is, I demonstrate how these measures can be derived from a set of plausible adequacy conditions. Third, I compare these measures on the basis of their characteristic properties, including an application to quantifying causal effect in medicine. Finally, I use the above results to argue for a specific measure of causal strength and I outline future research avenues.

Taking Control: How Our Sense Of Agency Can Be A Reliable Sense For Causation

Henning Strandin

Abstract

Manipulations are the paradigmatic examples of *interventions*, as these are elaborated and defined in interventionist theories of causation (Pearl 2009, Spirtes et al. 2000, Woodward 2004). Given this, it is natural to assume that "many voluntary actions do, as a matter of empirical fact, satisfy the conditions for an intervention" (Woodward 2007, p. 30). This gives manipulations a special role in the discovery of causal relations, and this feature is essential to the interventionist view. The present paper tries to elaborate on this suggestion, by first relating it to the *sense of agency*—the "experience of controlling one's own actions, and, through them, events in the outside world" (Haggard and Chambon 2012), and the possibility that the sense of agency is an empirical, non-inferential source of causal knowledge about manipulated systems. The result is then applied to an explanation of the epistemic difference between true experiments on the one hand, and natural experiments and observational studies on the other.

A semi-formal argument is provided to the effect that our sense of agency *can* be a reliable, implicit sense *for* causation, in the context of manipulations and given certain independently plausible assumptions about the causal relation and the causal properties of agents and the environments in which they operate. The argument depends on a quantitative notion of causation, that provides a measure of causal influence between a causal factor and an effect of it, and the theoretical framework of *causal equations* is employed for this purpose (Pearl 2009, Spirtes et al. 2000). The form of the argument is abductive, and intends to broadly mirror the way in which we defend the reliability and reality of other complex senses, with our *depth perception* as a particular example. A naturalistic epistemology is also assumed, in which an agent has direct, experiential knowledge of the presence of X if they believe that X is present, X really

is present, and their belief is due to a sense such that, sufficiently often when the belief is induced in the agent by the sense, X is present. Finally, it is a desideratum of the theory that agents are treated as a class of physical and causal systems. Thus, there is no appeal to a special kind of "agent causation"—causation by the actions of agents is here a subclass of event causation—and this differentiates this theory from some manipulationist theories of causation, which have either assumed that free actions are metaphysically special (e.g., von Wright 1971) or remained silent on the issue.

The paper relates the results of this investigation to two broad classes of explanations of the epistemic role of experiment in science: the manipulationist view and the regularist view. It is argued that, if our defense of the reliability of the sense of agency is correct, then the regularist view misses something vital in its explanation of why we give true experiments higher evidential value than natural experiments and observations studies, and that this is due to unreasonable demands on an argument as to whether we can have direct, experiential knowledge of causation, that are not satisfied by for example our usual explanation of depth perception, and to a failure to consider agents as causal systems in their own right, with certain causal properties that make them, essentially, interventions waiting to happen. However, the arguments given here in no way commits us to giving possible manipulations or interventions a special *ontological* role in our understanding of causation, and as such it is not committed to an interventionist theory of what causation *is*.

Henning Strandin (henning.strandin@philosophy.su.se)
Department of Philosophy, Stockholm University

References

Haggard, P. and Chambon, V. (2012). Sense of agency. *Current Biology*, 22(10):R390–R392.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press.

von Wright, G. H. (1971). *Explanation and Understanding*. Ithaca.

Woodward, J. (2004). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

Woodward, J. (2007). Interventionist Theories of Causation in Psychological Perspective. In Gopnik, A. and Schulz, L., editors, *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press.

MuST10: Causation and Complexity

Give me some space: Causal explanation, population thinking, and agent-based models

Lachlan Douglas Walmsley
Australian National University
Lachlan.walmsley@anu.edu.au

Abstract

If you needed to build a railway network, who would you ask to do it? You have two options. The first option is a team of engineers and the second option is a brainless ameboid, the slime mould *Physarum polycephalum*. On their side, the engineers have their years of combined experience, their university educations, and their mastery of mathematical tools and computer simulations. On its side, this unicellular blob has the pulsating tissue of which it is composed, expanding and contracting so the slime mould can explore its surrounds in search of food. Who do you choose? Surprisingly, the humble slime mould does about as well as the engineers when it comes to building networks like those connecting railway stations (Tero et al. 2010), striking a balance between efficiency and redundancy. So impressive are the slime mould's abilities that computer scientists have tried to replicate their abilities so they can exploit them in the future (Tsompanas et al. 2016). In the end, you may not have had much of a choice. If you picked the team of engineers to build your railway network, you may have effectively picked the slime mould (or at least a model of the slime mould).

The slime mould is a striking illustration of the power of more-or-less independent units behaving according to simple rules. In this case, the units are the bits of expanding and contracting tissue from which the slime mould is made. But in other cases, the units—or *agents*—might be the ants in a colony, the fish in a school, the birds in a flock, or even the people in an economy. In each of these cases, interesting and seemingly directed behaviour emerges from a system composed of many simple parts, each interacting with each other and

their environment on the basis of simple local rules. When we build mathematical models of such phenomena, however, we typically abstract away this behaviour in its formal description.

Mathematical modelling in science is so prevalent and powerful that some philosophers have argued that some explanations in science are mathematical rather than causal. In the philosophy of biology, philosophers like Mohan Matthen (2009) have argued that natural selection may have a purely statistical explanation. In the philosophy of cognitive science, philosophers like Anthony Chemero (2009) argue intelligent behaviour may have a purely dynamical. The responses from those in the causal camp, such as Jun Otsuka (2015, 2016) and David Kaplan (2015), have converged on the same argument. Although we can manipulate a mathematical description to derive interesting conjectures, those formal descriptions don't explain their targets. Rather, the causal relationships represented by those formal descriptions explain the phenomena of interest.

In this paper, I argue that the relevant causal relationships in the statistical-causal debate in the philosophy of biology are those of *population thinking*. According to population thinkers, population-level phenomena are explained in virtue of individual-level phenomena. In these explanations, individuals can be understood as either more-or-less identical or falling into a few distinct types, making individuals (at least within the same type) are spatio-temporally exchangeable. The slime mould's network, for example, is explained in terms of the individual bits of uniform tissue pulsating this way and that, and, in virtue of this uniformity, any given bit of tissue can be exchanged with any other bit of tissue without altering the population-level behaviour that emerges (see Wimsatt 2007). Hence, the parts of a population are unlike the parts a mechanism, which are highly organised and functionally specific.

Although causal and mathematical models struggle to represent spatial distribution, agent-based models present us with a formal system well suited to representing systems in which spatial distribution matters. I argue that, like causal models, agent-based models explicitly represent causal relationships so lend direct support to causal explanations. In addition to being exceptional at representing the bottom-up causal structure and emergent phenomena of interesting to the population thinker, agent-based models excel at making assumptions about spatial distribution explicit. This becomes increasingly important as the as the assumptions of population thinking are relaxed. This is the case in models of the evolution of cooperation, for example. My analysis also gives us a direction for understanding and explaining systems that are both non-trivially population-like and non-trivially mechanism-like, such as neural systems, using agent-based models.

The paper proceeds as follows. In section 2, I present the debate between causal and mathematical explanation in the philosophies of science, biology, and cognitive science. In section 3, I outline the assumptions of population thinking and argue that these assumptions should be maintained even if we reject statistical explanation. I also show that relaxing these assumptions sufficiently, converts them into the assumptions of mechanism thinking. In section 4, I describe agent-based models and show how they can be used with causal models to produce better causal explanations of population behaviour. In section 5, I conclude that agent-based models do not compete with causal models. Instead, the two should be used together to produce robust theorems.

References:

Anthony Chemero (2009). Radical embodied cognitive science. MIT press.

David Michael Kaplan (2015). Moving parts: the natural alliance between dynamical and mechanistic modeling approaches. Biology & Philosophy, 30(6):757–786.

Mohan Matthen (2009). Drift and statistically abstractive explanation. Philosophy of Science, 76(4):464–487.

Jun Otsuka (2015). Using causal models to integrate proximate and ultimate causation. Biology & Philosophy, 30(1):19–37.

Jun Otsuka (2016). A critical review of the statisticalist debate. Biology & Philosophy, pages 1–24.

Atsushi Tero, Seiji Takagi, Tetsu Saigusa, Kentaro Ito, Dan P Bebber, Mark D Fricker, Kenji Yumiki, Ryo Kobayashi, and Toshiyuki Nakagaki (2010). Rules for biologically inspired adaptive network design. Science, 327(5964):439–442.

Michail-Antisthenis I Tsompanas, Georgios Ch Sirakoulis, and Andrew Adamatzky (2016). Cellular automata models simulating slime mould computing. In Advances in Physarum Machines, pages 563–594. Springer.

William C Wimsatt (2007). Re-engineering philosophy for limited beings: Piecewise approximations to reality. Harvard University Press.

Modelling emotional phenomena as dynamic systems

Elena Walsh
University of Sydney / University of New South Wales
elenawalsh@gmail.com

Fifteen years ago, a small number of emotion theorists began to advocate in favour of what has been called the 'dynamic systems approach' to emotion (Lewis & Granic, 2000). A handful of researchers have since outlined conceptual proposals for modelling emotions as emergent products of dynamic processes (Barrett, 2009, 2012; Coan, 2010; Colombetti, 2013; Scherer, 2009a, 2009b). In this literature, a dynamic systems approach denotes the view that emotions are self-organising patterns that emerge over time via nonlinear causal interactions among domain-general system components. The emergence and consolidation of this new approach in affective science marks the continuation of a paradigm shift in the natural and life sciences over the past thirty years toward what Lewin (1999) described as the new 'science of complexity'.

As it stands, the proposal that emotions are emergent properties of dynamic systems is theoretically under-developed, and so does not lend itself to empirical testing, operationalisation, or simulation. This is especially true of the notion of emergence, which refers to a process through which a whole arises out of its component parts. The concept of emergence has played a critical role in debate between natural kind theorists and psychological constructionists in the emotions literature (Barrett, 2009, 2012; Coan, 2010). But the small group of emotions researchers working within a dynamic systems framework deploy the concept of emergence in different ways. There is ambiguity about whether emergent processes are any different to epiphenomenal ones, and whether genuine emergence requires top-down causation.

This paper clarifies how the concepts of emergence and top-down causation ought to feature in such accounts. It shows how these concepts can be deployed in a dynamic systems model for emotion that provides a new way of conceptualising the 'input' emotions provide to decision-making processes. On this model, it becomes possible to say that cognition (including reasoning) is constrained by the type of emotion (i.e., sadness, anger, etc.) that the individual is experiencing at the time of reasoning. The conclusions drawn have implications for moral psychologists and decision theorists interested in the influence of emotions on decision making and reasoning. In addition, the model developed provides an example of how complexity theory may be utilised in the development of multi-level explanatory frameworks for complex behavioural phenotypes traditionally explicated with exclusive reference to single-level explanatory frameworks.

References

Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives on Psychological Science*, 4(4), 326-339.

Barrett, L. F. (2012). Emotions are real. *Emotion*, 12(3), 413.

Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review*.

Colombetti, G. (2013). *The feeling body: Affective science meets the enactive mind*. MIT Press.

Lewin, R. (1999). *Complexity: Life at the edge of chaos*. University of Chicago Press.

Lewis, M. D., & Granic, I. (2000). Emotion, Development and Self-Organisation. *Dynamic Systems Approaches to Emotional Development. Cambridge*.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and emotion*, 23(7), 1307-1351.

Scherer, K. R. (2009a). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3459-3474.

Causation, Information and Specificity

Contact

Brad Weslake
NYU Shanghai
brad.weslake@nyu.edu

Abstract

Recent work in philosophy of biology on what I will call the question of *causal importance* has involved an oscillation between *conceptual* claims and *empirical* claims. Someone advances a conceptual claim about what it means for one cause to be more important than another, and claims that genetic causes are more important than non-genetic causes in this sense. Someone else replies that according to this conception, there is no difference in importance between genetic and non-genetic causes. Yet another person agrees that the conceptual claim does not capture what is important about genetic causes, and proposes a replacement. And around we go[1].

This paper is a contribution to the conceptual side of the discussion. My aim is to diagnose a persistent confusion in theories of causal importance, and thereby help philosophers of biology articulate more plausible concepts with which to approach the empirical facts. The confusion I allege is an instance of an old and familiar one: the conflation of epistemology with metaphysics. In this case, it involves the conflation of the actual frequencies that provide *evidence* for causal importance with the facts that *ground* causal importance. Or so I will argue.

The paper is structured as follows. Section 1 is an introduction. In Section 2 I introduce the concept of causal importance. In Section 3 I criticise Sober (1988) on causal responsibility, and identify the confusion. In Section 4 I criticise Waters (2007) on actual difference making. In Section 5 I criticise Griffiths et al. (2015) on specificity. In Section 6 I propose, for illustration, and inspired by Griffiths et al. (2015), information-theoretic definitions of responsibility and specificity that are immune to the confusion. I conclude in Section 7.

Note to the Organising Committee: An earlier version of this paper was presented at a symposium at the PSA in Atlanta in November 2016. This version of the paper will focus less on the criticisms of alternative theories, and more on the presentation of new information-theoretic measures of causal importance.

References

Griffiths, Paul E., and Karola Stotz. 2013. *Genetics and Philosophy: An Introduction*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511744082.

Griffiths, Paul E., Arnaud Pocheville, Brett Calcott, Karola Stotz, Hyunju Kim, and Rob Knight. 2015. "Measuring Causal Specificity." *Philosophy of Science* 82 (4): 529–55. http://dx.doi.org/10.1086/682914.

Sober, Elliott. 1988. "Apportioning Causal Responsibility." *The Journal of Philosophy* 85 (6): 303–18. http://dx.doi.org/10.2307/2026721.

Waters, C. Kenneth. 2007. "Causes That Make a Difference." *Journal of Philosophy* 104 (11): 551–79. http://dx.doi.org/10.5840/jphil2007104111.

Weber, Marcel. 2016. "Which Kind of Causal Specificity Matters Biologically?" *Philosophy of Science*.

---

1. Many examples could be given, but an especially clear iteration is: Waters (2007), who provides a definition and claims that genes uniquely satisfy it; followed by Griffiths and Stotz (2013 Chapter 4), who reply that it does not discriminate genetic from non-genetic causes; followed by Weber (2016), who claims that genes uniquely satisfy a revised definition.↩

Firm internationalisation as an Emergent Property of a Complex Adaptive System

Yanto Chandra

Department of Public Policy

City University of Hong Kong

Email: ychandra@cityu.edu.hkk

Ian F. Wilkinson

Discipline of Marketing

University of Sydney Business School

University of Sydney

Email: i.wilkinson@econ.sydney.edu.au

25th November 2016

Abstract submitted for 10th Munich-Sydney-Tilburg Conference in the Philosophy of Science

Abstract

Most theories of internationalization are firm-centric, based on methodological individualism, in which they seek to explain internationalization behavior and performance in terms of the characteristics of the individual firm. They are based on the following logic: "We want to believe that X succeeded because it had just the right attributes, but the only attributes we know are the attributes that X possesses; thus we conclude that these attributes must have been responsible for X's success" (Watts 2011 p 27). Theories of this kind are common in business.

Theories of internationalization fall into the same trap. Retrospectively, we can trace the course of events affecting how a particular firm's internationalization behavior (e.g., timing or speed of international market entry) unfolded, and interpret it in terms of a firm's and managers characteristics and the environment and key events taking place. These can include having employees with the right skills, resources and orientations; having the right social networks, the right links to influential firms, a network position that enables it to see and respond to particular opportunities, and a degree of luck.

Here we argue that internationalization behavior of firms is an emergent property of a complex adaptive system comprising a network of interacting people and firms and other types organisations. Aggregate patterns of internationalization emerge in a self-organising bottom-up manner from complex sets of micro actions and interactions taking place over time among various types of actors. In addition, the aggregate patterns emerging have top down feedback effects on the micro actions and interactions taking place. The internationalization behaviour and success of an individual firm cannot be predicted or controlled by the firm because it depends directly and indirectly on a complex web of actions and interactions taking place over time

among many people, firms and other organisations operating in a dynamic environment. It is a highly non-linear system in which seemingly insignificant events and changes can have disproportional effects.

In this paper we explain how complex systems theory applies to firm internationalization and formalize our theory in terms of the social physics Hidden Markov Model of influence networks developed by Alan Pentland and his colleagues at MIT (Pan et al 2012; Pentland 2014).

By viewing and modelling firm internationalization in terms of complex systems science we can gain a deeper understanding of the mechanisms and processes involved, better understand some forms of internationalization behaviour that cannot be explained by traditional theories, and gain a clearer understanding of the challenges facing managers and policy makers. It also opens up whole new lines of inquiry, such as building agent based models of the complex systems involved.

References

Pan, W., Dong, W., Cebrian, M., Kim, T., Fowler, J. H., & Pentland, A. S. (2012). Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems. *IEEE Signal Processing Magazine*, *29*(2): 77-86.

Pentland, A. (2014). *Social Physics: How Good Ideas Spread-The Lessons from a New Science*, Penguin: New York.

Watts, D. J. (2011) *Everything is obvious: How common sense fails us*. New York: Crown Pub.

Models in Systems Medicine

Jon Williamson

Draft of November 14, 2016

Abstract
Systems medicine is a promising new paradigm for discovering associations, causal relationships and mechanisms in medicine. But it faces some tough challenges that arise from the use of big data: in particular, the problem of how to integrate evidence and the problem of how to structure the development of models. I argue that objective Bayesian models offer one way of tackling the evidence integration problem. I also offer a general methodology for structuring the development of models, within which the objective Bayesian approach fits rather naturally.

Partitioning and Explaining Complex Systems: The Brain and Oceanic Gaia

Rasmus Grønfeldt Winther

Idealised causal analysis has received significant attention in philosophical reflection about complex systems: In which ways are randomisation, control, and intervention crucial to causal inference (Cartwright 2007, Woodward 2003)? A second thematic receiving less recent attention is the organisation or the "natural history" of parts or components of complex systems (but see Kauffman 1971, Wimsatt 2007): Of which structures and relations are systems such as the brain or Earth systems composed? Questions of classification of parts (e.g., neuronal networks; biogeographical regions)—and the kinds, frequencies, and topologies of part interaction—come to the fore in dissecting the *partitioning frames* (Winther 2006, 2011) of distinct research programs.

Do not think of a natural history of system partitionings as a philosophical investigation opposed to the philosophical exploration of the rulebook of causal inference. Rather, see the operation of partitioning frames (and philosophical analysis thereof) as establishing a descriptive pre-explanatory ontology of the complex system. Statistical methods, network analysis, experimental set-ups, etc. are only subsequently applied to this classificatory ontology, in acts of causal explanation (Carrillo, Martínez, and Winther in prep, Craver 2007). Admittedly, there is feedback between partitioning and causal analysis stages. Two research domains illustrating the pragmatics of partitioning frames are neuroscience (Anderson 2014, Poldrack 2010) and Gaia or Earth/Ocean Sciences (Lovelock 1979/2000, Dutreuil 2016).

Anderson, M. 2014. *After Phrenology. Neural Reuse and the Interactive Brain*. MIT Press.

Cartwright, N. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press.

Craver C. 2007. *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.

Dutreuil, S. 2016. *Gaïa. Hypothèse, programme de recherche pour le système terre, ou philosophie de la nature?* PhD Philosophy Dissertation, Université Paris 1 Panthéon-Sorbonne.

Kauffman, SA. 1971. Articulation of Parts Explanation in Biology and the Rational Search for Them. *Boston Studies in the Philosophy of Science* 8: 257-72.

Lovelock, J. 1979/2000. *Gaia. A New Look at Life on Earth*. Oxford University Press.

Poldrack, RA. 2010. Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed? *Perspectives on Psychological Science* 5: 753-761, 755.

Wimsatt, WC. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.

Winther, RG. 2006. Parts and Theories in Compositional Biology. *Biology and Philosophy* 21: 471-499.

____. 2011. Part-Whole Science. *Synthese* 178: 397-427.

Woodward, J. 2003. *Making Things Happen. A Theory of Causal Explanation*. Oxford University Press.