INFORMATION, CAUSATION AND COMPUTATION

John Collier

Durban, South Africa
Email: collierj@ukzn.ac.za

Abstract:
Causation can be understood as a computational process once we understand causation in informational terms. I argue that if we see processes as information channels, then causal processes are most readily interpreted as the transfer of information from one state to another. This directly implies that the later state is a computation from the earlier state, given causal laws, which can also be interpreted computationally. This approach unifies the ideas of causation and computation. A complication is the irreducible nature of many complexly organized systems. I offer a solution to this problem for the information transfer interpretation of causation.

1. Introduction

I make two basic assumptions. First, all information has a physical form, and second, that everything that is real is dynamical or can be explained in dynamical terms. Something is dynamical only if it involves nothing but forces and flows. Mathematically, dynamical systems are ones that involve a delta function or a time derivative. I assume a somewhat stronger assumption that dynamical systems must involve some physical properties that can be interpreted as forces or flows. From the dynamical assumption of the physical basis of information, it follows that information must be explicable in terms of forces and flows. At first this is counterintuitive, since information is typically thought of as a cognitive, computational or logical notion. However it is possible to bring logic and causation together through a specific analysis if the logic
of information flow and a reasonable definition of what it is to make a physical difference, given that information is well characterized as "a distinction that makes a difference" (MacKay 1969), or "a difference that makes a difference" (Bateson 1973: 428). Unlike energy flow, information has to do with the constraints on the boundary conditions of energy flow. It has to do with the form, grounded in the symmetries and asymmetries of a given structure, and their relations to each other that can be given a physical interpretation in terms of the flows of information from one form to another. Information is thus neither a measure of energetic relations, or of transfers of matter (Weiner 1948) Information theory, then, is fundamentally the rigorous study of distinctions and their relations, inasmuch as they make a difference. The physicalist assumption implies that these distinctions are physical, and the dynamical assumption implies that they make a difference to forces and/or flows. Fundamentally, information is a constraint on energy flows. In some systems, such as biological and cognitive systems, these constraints (boundary conditions) are much more significant than the energetic and material relations. Such systems are information systems. However, more generally, all systems can be understood from this perspective, and all causal systems can be understood or interpreted as information systems.

By bringing together causation (which makes a dynamical difference) with the logic of information flow, it becomes possible to see causation as a sort of computation. Inasmuch as there are regular relations between an initial state and a later state of a system, the relations are interpretable in computational terms. There may be random relations between the two states as well, but there is not much we can say about these relations except for there expected size.

2. Information

Causation can be understood as the transfer of information, if information is understood in the proper way as a physical mode (Collier 1999). In order to do this, we need a clear notion of the information in a
thing1, a notion of transfer of information, and what it is for information to be physical. I start with the information in a thing.

In the static case, the information in an object or property can be derived by asking a set of canonical questions that classify the object uniquely (possibly up to some error factor for continuous or vague objects) with yes or no answers, giving a 1-1 mapping from the questions and object to the answers. This gives a string of 1s and 0s (see figure).

![Figure 1 Information in a structure](image)

There are standard methods to compress these strings (though whether we have the maximally compressed string is in general non-computable). The compressed form is a line in a truth table, and is a generator of everything true of the thing required to classify it. There need not be a unique shortest string, but the set will be a linear space of logically equivalent propositions. The dimensionality of this space is the amount of information in the original object. The equivalence class of propositions in this space is the factual content of the object. This is

---

1 By ‘thing’ I mean anything objective, including properties, systems, states of systems, and objects. See (Ladyman et al 2007).
reminiscent of Wittgenstein in the *Tractatus*, except that he thought there was a basic set of propositions that constitute the facts of the world, propositions being distinguished by their form. I would say that there is just basic factual content about things, which would, and usually does, generate multiple equivalent propositions, if one prefers to consider propositions as by their encoding rather than their content.

It has often been claimed, and I think that this is the most widespread view, that the information in something is relative to interests, or at least to interactions with other things. For example, the great information theorist Jaynes took the position that information is a property of an epistemic state (Jaynes 1975). Others, such as Carnap and Bar-Hillel (Bar-Hillel 1964), Barwise and Perry (1983) Israel and Perry 1990 would make it a property of sentences or statements, a kind of semantic property. I have argued, however, that information originates in symmetry breaking (Collier 1996), a physical process. Prior to this I had argued that representation (and hence measurement, or detection) requires that there be intrinsic information out there in the world to be detected, and gave an account of how this might be possible (Collier 1990). Although information transfer is relative to a channel and a receiver (especially the state of the receiver), information itself need not be defined in these terms, and it is convenient to ground information in a non-relative version that allows us to talk of intrinsic information. My former student, Scott Muller, has worked these ideas out in careful and original detail to give a reply to Jaynes, while taking over most of Jaynes’ formalism and basic ideas, using the resources of group theory to provide a mathematical basis for the study of symmetry and symmetry breaking (Muller 2007).

The dynamic case is more general, the static case being a special version, but we seem to find it easier to start with the static case. In principle Figure 1 also describes how we could treat the dynamical case, but obviously there is something more going on: information isn’t merely there; it flows. We need an account not just of information in a state, but of information flow in an interconnected system.

So how does it work? First, we need an account of information flow. I start with Barwise and Seligman (1997) *Information Flow: The Logic of Distributed Systems*. I use it because it is the best worked out account
available, and avoids the mistakes many have made by taking the Shannon formulation of information theory as canonical and fundamental, which ignores semantic aspects of information. Barwise and Seligman developed a theory of information channels, something missing from most other accounts of information. Shannon invokes channels, but does not explain what they are, or what their properties are. Barwise and Seligman give a formal definition of information channels. Ironically, the basic ideas, grounded in the idea of distributed systems, were developed in the 1930s, well before Shannon’s work. Perhaps he was taking these ideas for granted. Measures of the quantity (and quality) of information follow naturally from their approach.

Their model of information flow is based on four principles. It is a mathematical model, and thus should apply to anything that can be represented according to the principles. Their examples are primarily from the cognitive and biological world, and they make no mention of causation as a form of information flow, but they do mention information in “every fluctuation of the natural world” (Barwise and Seligman 1997: 4). However, they use causation in examples given in the preliminary discussion of their model. This presents a potential for circularity in any application of the model to causal connection, so I will have to state their approach so as to be sure to avoid any vicious circularity. As they say, their model “is not an analysis of all and sundry accounts of information and information flow”, but it is well adapted to my purposes.

The Four Principles of Information Flow

**Principle 1**: Information flow results from regularities in a distributed system.

**Principle 2**: Information flow crucially involves both types and their particulars.

**Principle 3**: It is by virtue of regularities among connections that information about components of distributed system carries information about the other components.

**Principle 4**: The regularities of a given distributed system are relative to its analysis in terms of information channels.

---

2 Attempts to combine Shannon information measures with the Barwise-Seligman approach can be found in (Awien 2004, Moskowitz et al 2004, and Seligman 1991).
The first principle of information flow

Note that information flow requires a distributed system. Information flow occurs in a context in which the parts are connected. Furthermore, the connections must be regular. Note that this does not require invoking causality, since the regularity can be determined from redundancies in the information system itself. On the other hand, the distribution of the system seems to require causality. In a sense it does, but I will argue below that causality does not have to be invoked explicitly. The connections, I will argue, can be understood as the identical information. Given other conditions, this identity can be understood as a causal relation. The regularities can be nomic, conventional, abstract, or of some other form, depending on the sort of system involved. The principle is completely general. Furthermore, the parts of the distributed system can be of any kind. The kind will determine the type of connection required.

Accidental relations cannot carry information. Generally, the more random a system, the less information it can carry (this is not to be confused with the fact that an equiprobable code carries maximal information in a Shannon coding – the coding is part of the system, and there must be some regularity between the coding and uncoding for information to be carried by a channel). Even if there are statistical regularities between A and B, if these are chance, no information is carried by these regularities between A and B (though, by chance, the result may be useful).

The second principle of information flow

Mathematics and theories deal with types (i.e. classes, or categories or other abstract kinds). For example, if we consider the information that two dice were rolled, then we might consider the probability of getting a seven. Mathematically this is the outcome which has six possible forms, and the probability is 1/6. The information that a seven was rolled therefore is 1/6log₆, or about .43 bits. On the other hand, information is carried by particulars, or tokens. A token is an instantiation of something that is classified, and a type is a particular classification. Specifically, a classification A is an ordered triple \( \langle A, \Sigma_A, \hat{A} \rangle \) of a set A of
objects to be classified (the tokens of A), a set of objects used for classification and binary relation between the two \( \Sigma_A \) that tells which tokens are classified as which type, \( \mathcal{F}_A \). For the dice example, the classification is made up of the total of each possible roll from two to twelve, the particular rolls, and the assignment of each roll to a number. The classification \( \mathcal{F}_A \) constrains the assignment of tokens through constraints if and only if the classification assigns some token \( a \) in the set of interest \( A \) to a type \( \alpha \) within a set of types \( \Sigma_A \) in the classification. The complete set of constraints is called the theory of the classification, \( \Theta_A \). It represents all the regularities of the system modeled by classification A.

An infomorphism is a pair \( \mathbf{f} \) of functions \( \langle f^\wedge, f^- \rangle \) between two classifications \( A \) and \( B \), one from the set of objects used to classify \( A \) to the set of objects used to classify \( B \), and the other from \( B \) to \( A \), such that the biconditional relating the second function to the inverse of the first function holds for all tokens \( b \) of \( B \) and all types of \( A \), \( f^- (c) \mathcal{F}_A \alpha \) if and only if \( c \mathcal{F}_C f^- (\alpha) \). The biconditional is called the fundamental property of infomorphisms.
An information channel for a distributed system is an indexed family of infomorphisms with a common core codomain $C$. The infomorphisms allow information to be carried from one part of the system to another. For example, in a flashlight, the components might be a bulb, battery, switch and case. The channel is basically a connected series of infomorphisms from switch to bulb through the mediation of battery and case. A channel, perhaps obviously, does not need to be sequential as a whole, but various parts do, e.g., the switch sends information to allow the flow of electricity through the case to the bulb, so that the bulb has information about the state of the switch. In this case, if the channel is sound (functioning), the switch also has information about the bulb, but this sort of converse relationship need not hold. For example the relation between a nuclear reactor and its control room should be a channel, so that the control room gives information about the state of the reactor. The reactor, however, does not have information about the state of the control room, except through (we hope) certain switch that should control activities in the reactor. If the reactor, however, gets outside of certain parameters, all hope of controlling it is lost. The readings on the dials and indicators of the control room might still record faithfully the state of the reactor, but the converse channel has become unsound. Many channels are of this one way sort (for example, in typical measurements of physical systems, with quantum mechanical systems perhaps a bizarre exception).

If any of the infomorphisms fails in some distributed system, the transfer of information fails (redundancy of channels could help to avoid failure of information transfer). Note that a common cause $c$ of $a$ and $b$ can allow $a$ and $b$ to contain information about each other, even though they are not sequential. Thus an infomorphism can hold by virtue of a channel even if $a$ and $b$ are not sequential.

**The third principle of information flow**

The classification $\mathcal{F}_C$, and its associated theory $\Theta_C$ give us a way to model the regularities and capture the basic principles of information flow.
flow within the system. \( \Theta_c \) is just the classification of the tokens under the infomorphisms of the channel, such that each token relation is classified under the relations required by the channel. Again, note that the regularities can be determined by the redundancies in the system, and the connections, to be empirically determined, depend on the specific system, or at least its type. For example, a Newtonian and a general relativistic system will share many classifications, but some will be different. For example, spatial movement in a Newtonian space merely by configuration changes is impossible (no information can be conveyed spatially by this means alone), but in general relativistic spacetime, it is possible, no doubt surprisingly (Guéron 2009). The infomorphisms for the two theories are different.

The first principle focuses attention on distributed systems, while the second focuses on their tokens. The tokens allow us to track which things are connected to which.

Here is Barwise and Seligman’s *First proposal* for the third principle. I’m not going to get into the refinements, but this simplified version deals with the simplest nontrivial case in which there are two components \( a \) and \( b \). Supposing \( a \) is of type \( \alpha \) and it carries the information that \( b \) is of type \( \beta \) relative to a certain channel if \( a \) and \( b \) are connected in the channel, and the translation \( \alpha' \) of \( \alpha \) entails the translation \( \beta' \) of \( \beta \) in the theory of \( C \) where \( C \) is the core of the channel.

This proposal ensures veridicality (the information is entailed), and also the *Xerox principle*:

**Xerox Principle**: If \( r \) being \( F \) carries the information that \( s \) is \( G \), and \( s \) is \( G \) carries the information that \( t \) is \( H \), then \( r \) being \( F \) carries the information that \( t \) is \( H \) (small letters indicate tokens, large, their types).

These are both virtues because they guarantee that information that is transferred is reliable, and that information about something can be chained, permitting chains of veridical information. This is quite useful for epistemology, but that is not my concern here. Two shortcomings of the proposal are, first, that it does not directly identify the regularities of the components of the system (this can be resolved by using the inverse function mapping the questions to the answers). A more substantial problem is that the first proposal requires complete information about the system, \( i.e. \), a complete theory of the classification, which we seldom
have. This will not be a concern in the application to causation, since it deals with ontology, with only the requirement that the ontology be epistemologically accessible.

**The fourth principle of information flow**

This principle is designed to deal with exceptions by introducing a relativity to channel of information. For example, what you consider noise on your TV might be a signal to a TV repairman. Notice that this does not imply a relativity of information to interests, but that interests can lead to paying attention to different channels. The information in the respective channels is objective in each case, the noise relative to the (non-functioning or poorly functioning television channel, and the noise as a product of a noise producing channel – the problem for the TV repairman is to diagnose the source of the noise via its channel properties). This sort of relativity might appear to be of little importance in what follows, though it is useful in separating the different laws a causal system might obey, and I also use it to restrict causal channels to dynamical channels to ensure dynamical realism.

3. Causation

The proposal I made in (Collier 1999) was that P is a causal connection in a system from time $t_0$ to $t_1$ if and only if some particular part of its form is preserved between states $s_0$ and $s_1$ from $t_0$ to $t_1$. For physical systems, I restrict the form to structural (roughly, invariant) information as determined by isomorphic mapping and compression described at the beginning of this section, which I call *enformation*. The restriction to enformation is motivated by the requirement that chance regularities cannot carry information. Furthermore I constrain physical information with the Negentropy Principle of Information in order to rule out connected series of preserved information like the trace of a light beam across the face of the moon.

**Negentropy Principle of Information**
**NPI:** \[ I_P = H_{\text{MAX}} - H_{\text{ACT}} \]

\( H_{\text{ACT}} \) is the actual entropy; whereas \( H_{\text{MAX}} \) is the entropy the system would have if its microstates are all equally probable (this is not generally the same as the entropy a system will reach as it approaches equilibrium). This is a definition.\(^3\)

The enformation of a state will be some past of its \( I_P \). NPI rules out cases like the light beam case, since they will not satisfy the Second Law of Thermodynamics, which is presupposed by NPI. In the light beam case, for example, the temperature (and thus the entropy) will vary as it moves across the moon. This means that the information in the light beam is not conserved, even though other aspects of its form are conserved. There may be possible counterexamples to my use of the NPI Principle, but if there are, I am not clear enough about them to be able to evaluate them. In most cases (such as light beam across a suitable curved perfect mirror) the conditions are unrealistic. In other cases it seems that there must have been a connivance that is more probable than that some noncausal relation holds.

The result notion of causal process is: \( P \) is a physical causal process in system over a series of states \( s_i \) from time \( t_0 \) to \( t_1 \) if and only if some part of the enformation is transferred from \( t_0 \) to \( t_1 \), consistent with NPI, and over every intermediate state. The light beam is a special case of common cause rather than direct causation. NPI rules out all cases of such indirect causation (except for a very small number of cases in which the entropies fit NPI, but by chance only. Basically, NPI is a principle that describes the force conditions relevant to information flow in a causal process.

Physical states are tokens of the types of a state space such that each state \( s \) is a token of exactly one type. Macrostates are types of states (microstates) that are equivalent for certain state properties, such as energy, pressure and other properties that are insensitive to the specific

\(^3\) Details on the motivation for the use of NPI as a sort of operational definition in the sense of Mach’s (1960: 264-271) definition of mass are given in (Collier 1999).
state (microstate) of the system. Macrostates are relevant for the computation and/or measurement of entropy. The $\Omega_s$ in Figure 3 can be taken to be macrostates without loss of generality (the macrostates are more general, at the very least) to give a macroscopic theory of the system. I will have bit to say about their reality in my conclusion below.
With these considerations in mind, the original statement of causal connection can by rephrased in Barwise and Seligman’s terms as P is a causal connection in a system from time \( t_0 \) to \( t_1 \) if and only if there is an channel between \( s_0 \) and \( s_1 \) from \( t_0 \) to \( t_1 \) that preserves some part of the information in the first state. Furthermore, P is a physical causal process in system over a series of states \( s_i \) from time \( t_0 \) to \( t_1 \) if and only there is a channel through the states from \( t_0 \) to \( t_1 \), consistent with NPI, and over every intermediate state.

The conditions on a causal process are much more severe than on a mere causal connection, and are therefore easier to verify. In fact it is unclear how we would check that there is a causal connection when there is no underlying causal process or processes. The connection between several discrete spots of light in succession across the surface of the moon, if there were no causal processes like a pulsating laser beam reflecting off the moon connecting them, would be very mysterious, even if they satisfied NPI. If we did discover such a thing, perhaps looking at the constraints (force) of how NPI applies would help us to understand

Figure 3 Infomorphism between states

With these considerations in mind, the original statement of causal connection can by rephrased in Barwise and Seligman’s terms as P is a causal connection in a system from time \( t_0 \) to \( t_1 \) if and only if there is an channel between \( s_0 \) and \( s_1 \) from \( t_0 \) to \( t_1 \) that preserves some part of the information in the first state. Furthermore, P is a physical causal process in system over a series of states \( s_i \) from time \( t_0 \) to \( t_1 \) if and only there is a channel through the states from \( t_0 \) to \( t_1 \), consistent with NPI, and over every intermediate state.

The conditions on a causal process are much more severe than on a mere causal connection, and are therefore easier to verify. In fact it is unclear how we would check that there is a causal connection when there is no underlying causal process or processes. The connection between several discrete spots of light in succession across the surface of the moon, if there were no causal processes like a pulsating laser beam reflecting off the moon connecting them, would be very mysterious, even if they satisfied NPI. If we did discover such a thing, perhaps looking at the constraints (force) of how NPI applies would help us to understand
what is going on. In the case of nonphysical causes, if there are such things, we don’t even know where to begin. This does not, however, mean that the causal connection principle does not apply. It might still be that in some mysterious way that there is an infomorphism between \( s_0 \) and \( s_1 \) from \( t_0 \) to \( t_1 \), in which case there would be a causal connection (the connection would not be coincidental because of the no accidents condition on infomorphisms). Such possibilities are remote to our experience, however. Causal processes, however, under the channel characterization, are eminently discoverable and investigatable because they involve unbroken channels through continuous space and time that must conform to the Second Law of Thermodynamics.

The problem now is to show how the above accounts of causal connection and causal process relate to the four principles of Barwise and Seligman, and then to show why we should accept that all information flow is either dynamical or else explicable in dynamical terms. **Principle 1**: a causal system involves regularities (the enformation) in a distributed system, satisfying the first principle. **Principle 2**: causation involves both types and particulars. The particulars in this case are the tokens of information. The types are specific to the cases, but the most fundamental type is dynamical. **Principle 3**: the connection in this case is identity, which is perhaps the strongest connection one can have, and requires information transmission across time: it is the identical token of information. Epistemologically, we can infer this, since it is much more parsimonious to infer one identical instance of information rather than a series of similar tokens. Again, the regularities are determined by the redundancy of the information token. This is enough to establish that the basic theory of causation fits the Barwise/Seligman model. Note that their first proposal ensures veridicality as well as transitivity of the information in causal events.

The **fourth principle** is of considerable interest. I have already invoked it by using the specific type dynamical, but further classification can allow one to take a particular causal process or interaction and classify it more specifically to allow for specific causal laws. This relativisation is useful for scientific investigation, and also at least mitigates Nancy Cartwright’s (1983) arguments about the laws of
physics lying. It is legitimate to look at specific channels, or types of interactions. For example, inasmuch as the theories of gravitation and electrodynamics are separate, we can separate the gravitational and electrodynamic information in a system. More complex material that I have not introduced allows the additivity of channels to get a net effect.

4. Computation

There is a simple sense in which causation on the above account is like a computation: the information in a later state contains information in the earlier state of the system, entailment depending precisely on the causal influence of the earlier state on the later. Inasmuch as entailment by way of a physical process is a computation, then the account guarantees that causation is computational.

Things are not quite so simple, however. There are dynamical systems whose solutions are not analytically computable, including ones as simple as three-body gravitational systems. Bertalanffy (1968: 20), following Franks, breaks down the classification of dynamical systems as shown in Figure 4 below. Large numbers of real systems have nonlinear equations and many parameters, making their solution analytically impossible, though some such systems can be modelled with numerical computer models. Things are more difficult, however, when there is self-organization and multiple attractors, in which case divergent solutions to the equations of motion cannot be isolated to spatiotemporally local conditions.

Robert Rosen (1991) distinguished between analytic and synthetic systems. Synthetic systems are ones in which the observational basis can be combined in some linear way to obtain and analytic model (thus making the analytic models reducible to their evidence), whereas analytic models are ones having a mathematical dynamical form. They need not be reducible. In other words, there are analytic models whose solutions are not computable. An important case of nonreducible models are those of living systems. They are typified by loops of what Rosen called

---

4 Rosen’s use of analytic here is somewhat confusing and idiosyncratic.
efficient causation. Such systems, I maintain, are exactly those that cannot be fully dealt with via analytical computations or numerical approximations, in particular where they are not in steady state, but show the sort of bifurcations typical of growth and development. These are the complexly organized systems (Collier and Hooker 1999).

<table>
<thead>
<tr>
<th>Equation:</th>
<th>Algebraic</th>
<th>Ordinary Differential</th>
<th>Partial Differential</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Parameter</td>
<td>Trivial</td>
<td>Easy</td>
<td>Difficult</td>
</tr>
<tr>
<td>Several Parameters</td>
<td>Easy</td>
<td>Difficult</td>
<td>Intractable</td>
</tr>
<tr>
<td>Many Parameters</td>
<td>Intractable</td>
<td>Intractable</td>
<td>Impossible</td>
</tr>
<tr>
<td>Nonlinear Equations</td>
<td>One Parameter</td>
<td>Very Difficult</td>
<td>Very Difficult</td>
</tr>
<tr>
<td>Several Parameters</td>
<td>Very Difficult</td>
<td>Impossible</td>
<td>Impossible</td>
</tr>
<tr>
<td>Many Parameters</td>
<td>Impossible</td>
<td>Impossible</td>
<td>Impossible</td>
</tr>
</tbody>
</table>

**Figure 4** Classification of mathematical difficulty After Bertalanffy 1968: 20

How can this be reconciled with the view that causation is like a computation? Firstly, the causal networks typical of causally organized systems can be represented by a device introduced by Rosen (1991), though the basic idea seems to go back to Leibniz (1969, see Collier 1999 for discussion). Rosen noted that theoretical models, if accurate, reflect the causal relations in the system modelled, so that the logical relations of the theory are reflected in the causal relations of the system modelled. Understanding causation as information flow implies that the causal relations actually are logical relations, and that the distributed network of information flow can mirror precisely the logical relations of some analytic model, even if that model is not reducible. However, the distributed network flow of information in such systems still fits the Barwise and Seligman approach.
This raises a bit of a paradox. If the system states are sequential, and are information states with entailments between them, is this not a computational relationship? If so, how can the system equations themselves be noncomputable? I think the answer is that we need to distinguish two notions of computability, sequential computability and analytic computability. There are Turing functions that do not terminate. These functions are not analytically computable. Yet each state is determined by the previous state. The sequential states of systems with no synthetic models are similar to these nonanalytic Turing functions (though unlike Turing functions new information can arise in later states due to stochastic processes and information creation via symmetry breaking) in that the information that is transferred is entailed by the earlier state, so causation is accounted without requiring analytic solutions to the system equations.

Conclusions

Information flow in distributed systems can be given an interpretation in dynamical terms. This gives us an account of causation in information theoretic terms. The relations between information theory and logic can then help us to understand the relations between causation and computation. The relations are complicated by the existence of nonreducible models of real systems for which there are no complete analytical solutions. These systems correspond to those having organized complexity (both complex and well-organized) that are the usual subject of complexity theory. The information theoretic account of causation can be seen to fit these systems through an analogue to the distinction between sequentially computable (stepwise computable) Turing functions and analytic (computable) Turing functions.
References


