CAUSATION IS THE TRANSFER OF INFORMATION

John Collier Department of Philosophy University of Newcastle Callaghan, NSW 2308, Australia email: pljdc@alinga.newcastle.edu.au 14 November, 1997

For Howard Sankey (ed) Causation, Natural Laws and Explanation (Dordrecht: Kluwer, 1999)

CAUSATION IS THE TRANSFER OF INFORMATION

1. Introduction

Four general approaches to the metaphysics of causation are current in Australasian philosophy. One is а development of the regularity theory (attributed to Hume) that uses counterfactuals (Lewis, 1973; 1994). A second is based in the relations of universals, which determine laws, which in turn determine causal interactions of particulars (with the possible exception of singular causation, Armstrong, 1983). This broad approach goes back to Plato, and was also held in this century by Russell, who like Plato, but unlike the more recent version of Armstrong (1983), held there were no particulars as such, only universals. A third view, originating with Reichenbach and revived by Salmon (1984), holds that a causal process is one that can be marked. This view relies heavily on ideas about the transfer of information and the relation of information to probability, but it also needs uneliminable counterfactuals. The fourth view was developed recently by Dowe (1992) and Salmon (1994). It holds that a causal process involves the transfer of a non-zero valued conserved quantity. A considerable advantage of this approach over the others is that it requires neither counterfactuals nor abstracta like universals to explain causation.

The theory of causation offered here is a development of the mark approach that entails Dowe's conserved quantity approach. The basic idea is that causation is the transfer of a particular token of a quantity of information from one state of a system to another. Physical causation is a special case in which physical information instances are transferred from one state of a physical system to another. The approach can be interpreted as a Universals approach (depending on ones approach to mathematical objects and qualities), and it sheds some light on the nature of the regularity approach.¹ After motivating and describing this approach, I will sketch how it can be used to ground natural laws and how it relates to the four leading approaches, in particular how each can be conceived as a special case of my approach. Finally, I will show how my approach satisfies the requirements of Humean supervenience. The approach relies on concrete particulars and computational logic alone, and is the second stage of constructing a minimal metaphysics, started in (Collier, 1996a).

The approach is extraordinarily simple and intuitive, once the required technical apparatus is understood. The main problems are to give a precise and adequate account of information, and to avoid explicit reference to causation in the definition of information transfer. To satisfy the first requirement, the approach is based in computational information theory. It applies to all forms of causation, but requires a specific interpretation of information for each category of substance (assuming there is more than one). For the scientifically important case of physical causation I use Schrödinger's

¹ Jack Smart suggested to me that my approach might be a regularity approach in the wider sense that includes his own account of causation. On my account all detectable causation involves compressible relations between cause and effect (§4 below). Inasmuch as, given both compressibility and all other evidence being equal, it is almost always more parsimonious to assume identity of information token rather than coincidence (and never the opposite), compressibility almost always justifies the inference to causation. If meaning is determined by verification conditions (which I doubt) then my theory is indistinguishable from a regularity theory in Smart's wide sense, since the exceptions are not decidable on the basis of any evidence (see §8.1 below for further discussion.

Negentropy Principle of Information (NPI). Causation can be represented as a computational process dynamically embodied in matter or whatever other "stuff" is involved, in which at least some initial information is retained in each stage of the process.² The second requirement, avoiding circularity, is achieved by defining causation in terms of the identity of information tokens.

2. The Role of Form in Causal Explanations

Suppose we want to ensure that someone is the person we believe them to be. We typically rely on distinguishing features such as their face, voice, fingerprints or DNA. These features are complex enough that they can distinguish a person from other people at any given time, and are stable enough that they reliably belong to the same person at different times (science fiction examples excepted). However, if the person should have a dopplegänger (a qualitatively identical counterpart), these indicators would not be enough for identification; we would need to know at least something of the

Causation is the Transfer of Information

spatiotemporal history of the particular instantiations of the qualities of the person we wish to identify. Sameness of person (or at least of their body) requires a causal connection between earlier stages and later stages. We can recognise this connection through identifying features and spatiotemporal continuity. The body transmits its own form from one spatiotemporal location to another. I will argue that not only is this sort of transmission an evidential basis for causal connection, but it can be used to define causal connection itself.

Central to my account is the propagation of form, as measured by information theoretic methods. This is not so foreign to traditional and contemporary views of causation as it might seem. By form, I mean the integrated determinate particular concrete qualities of any thing, of *any* kind.³ Understanding the propagation of form is necessary for understanding contemporary science. If the reader finds this uncontroversial, I suggest they skip directly to §3.

Form includes the geometrised dynamics of grand cosmological theories like geometrodynamics (harking back to Platonic and Cartesian attempts to geometrise dynamics, Graves, 1971) and geometry and symmetry used to explain much of quantum particle physics (Feynman, 1965). It also includes the more common motions and forces of classical mechanics, as expressed in the Hamiltonian formulation with generalised coordinates.⁴ Treatments of complex physical

⁴ On generalised coordinates, see (Goldstein, 1980). On the embedding of classical mechanics as well as more recent non-

² Causal connection is necessary in the same way that a computation or deduction is necessary, but it is not necessary in the sense that it is impossible for things to be The necessity depends otherwise. on contingent conditions analogous to the premises of a valid argument (see §4 below). I proposed this kind of necessity for laws in (Collier, 1996a). Something that is necessary in this way cannot be false or other than it is, but is contingently true; i.e. it is contingent that it is. This theory of causation fills out the uninterpreted use of 'causation' in concrete particular instances in (Collier, 1996a), and is part of a project to produce a minimal metaphysics depending on logic, mathematics and contingent concrete particulars alone.

³ I will give a more precise, mathematical characterisation of form in §3 below. My definition of form may seem very broad. It is. Naturally, if there are any exceptions, my account fails for that sort of case, but I believe there are none, nor can there be.

phenomena such as Bénard cell convection and other phenomena of fluid dynamics rely on knowledge of the form of the resulting convection cells to solve the equations of motion (Chandreshankar, 1961; Collier, Banerjee and Dyck, in press). In more highly nonlinear phenomena, standard mechanical techniques are much harder to apply, and even more knowledge of the idiosyncrasies of the form of particular phenomena are required to apply mechanical methods. Even in mathematics, qualitative formal changes have been invoked to explain "catastrophes" (Thom, 1975). Sudden changes are common in phase transitions in everyday complex phenomena like the weather, as well as in highly nonlinear physical, chemical. developmental, evolutionary, ecological, social and economic processes.

In biology, causal explanations in terms the dynamics of form are especially of common. Because of the complexity of macromolecules and their interactions, it seems likely that biologically oriented chemists will need to rely on form of molecules indefinitely. This has led to information theoretic treatments of biochemical processes (Holzmüller, 1984; Küppers, 1990; Schneider, 1995). Even the reduction of population genetics to molecular genetics has failed to fulfill its promise because of many-many relations between the phenotypic traits of population biology and molecular genes. Although some molecular biology is now done with the aid of mechanics and quantum mechanics, these methods are limited when large interacting molecules are involved. In large scale biology ecology, (systematics, ontogeny and evolution), causal arguments usually concern some aspect of the transitions of form as I have defined it above (D'Arcy Thompson,

Causation is the Transfer of Information

1942; Wiley, 1981; Brooks and Wiley, 1988; Ulanowicz, 1986).

The most dominant current view of cognition is the syntactic computational view, which bases cognitive processes on formal relations between thoughts. Whether or not the theory is true, it shows that psychologists are willing to take it for granted that form (viz., the syntax of representations) can be causal. Fodor (1968) argues that the physical embodiment of mental processes can vary widely, if the syntactic relations among ideas are functionally the same. To understand the embodiment of mind, if we accept that cognitive processes derive their formal relations from underlying dynamics, we need an account of the role of form and information in dynamics.⁵

Traditional linear and reductionist mechanical views of causation have had limited success in these emerging areas of study. Since the traditional views are well established, any adequate account of causation may initially seem counterintuitive. Causal studies using ideas of form, broadly construed as I have described it, have been more successful than mechanical approaches in the sciences of complex systems, but we need a precise account of the causal role of form to unify and normalise these studies. We need this because there is no hope that mechanical accounts will ever fully replace their currently less regarded competitors. The mechanical view is demonstrably too restrictive to deal with many kinds of possible systems that we are likely to encounter (see Collier and Hooker, submitted, for details).

The view I propose is not entirely without precedent. Except for its idealism, Leibniz' account of causation is in spirit the most developed precursor of the account I

mechanical physics in the dynamics of form, see (Collier and Hooker, submitted).

⁵ This argument is developed in (Collier, 1990a) and (Christensen et al, in preparation).

will give.⁶ The case is complicated because of Leibniz' three level ontology (Gale, 1994). At the observable level Leibniz' physics was mechanical, however this dynamics was explained by the properties of monads, whose substantial form implied a primitive active force (encoded by the logico-mathematical structure of the form in a way similar to the compressed form of causal properties I will discuss later). This primitive active force produces the observable varieties of derivative active force through strictly logical and mathematical relations. At the metaphysical level, the substantial form is based in "clear perceptions" which are articulated in the structure of the substantial form of corporeal substance. A similar hierarchy exits for passive forces, which are similar to Hobbes' material cause.⁷ The

⁷ Although Hobbes attributed causation to the mechanical collisions of contiguous bodies, at

derivative passive force is a consequence at the corporeal level of Prime Matter, which is metaphysically based in the monad's confused perceptions (which, because unarticulated, cannot act as agent; see Christensen et al, in preparation).

Leibniz expressed many of these ideas in "On the elements of natural science" (Ca. 1682-1684, Leibniz, 1969: 277-279). The following quotation illustrates the importance of form in Leibniz' philosophy:

> And the operation of a body cannot be understood adequately unless we know what its parts contribute; hence we cannot hope for the explanation of any corporeal phenomenon without taking up the arrangement of its parts. (Leibniz, 1969: 289).

An earlier version of this view can be found in the 1677 paper "On the method of arriving at a true analysis of bodies and the causes of natural things" The paper emphasises the importance of empirical observation, but the final paragraph makes clear the role of form in causal explanation:

> Analysis is of two kinds – one of bodies into various qualities, through phenomena or experiments, the other of

least one of which is in motion, he believed that causation was necessary. If the total cause is present, then the effect must occur; if the effect is present, then the total cause must have existed (Hobbes, 1839: 9.3). The total cause is made of both the efficient cause (being accidents in the agent) and the material cause (being accidents in the patient). Form played no role for Hobbes' view of causation, except in the geometry of the accidental motions of the agent and patient. On the other hand, despite this extreme mechanism, there is no cause unless the geometries are precisely correct to necessitate the effect.

⁶ Some other possible precursors are Plato, Aristotle, the geometric forms of the atomists' atoms, Descartes' geometric view of dynamics, and Spinoza's theory of perception. I mention them mostly to avoid being accused of thinking myself especially original. These are failed attempts that imported unnecessary metaphysical elements to fill gaps in the accounts that disappear with a proper understanding of computational processes and their relation to physical processes. My position differs from Wittgenstein's position in the Tractatus (1961, see 2.0 to 2.063 especially) in using computational logic broadly construed. I also differ with Wittgenstein on 2.062, in which he says that a state of affairs cannot be inferred from another state of affairs (see §4 below). States of affairs which are unanalysable distinctions or differences may be the only exception, and might satisfy the requirements for Wittgenstein's elementary propositions (for reasons to think not, see Bell and Demopoulos, 1996).

sensible qualities into their causes or reasons, by ratiocination. So when undertaking accurate reasoning, we must seek the formal and universal qualities that are common to all hypotheses ... If we combine these analyses with experiments, we shall discover in any substance whatever the cause of its properties. (Leibniz, 1969: 175-76)

Again we see that for Leibniz, grouping observable phenomena by of their qualities and changes in qualities is but a prelude to explanation in terms of substantial form. The full explanation from metaphysics to physics to phenomena should be *entirely* mathematical. I shall take advantage of the fact that mathematics is neutral to collapse Leibniz' three levels into one involving only concrete particulars. The first step is the quantification of form using recent developments in the logic of complexity.

3. Quantification of Form Via Complexity Theory

A precise mathematical characterisation of form (more precisely, the common core of all possible conceptions of form) can be formulated in computational information theory (algorithmic complexity theory). This will provide the resources for a general account of causation as information transfer (whether physical or not) in §4. In §5 I will connect information to physical dynamics in an intuitive way through Schrödinger's Negentropy Principle of Information (NPI), defines materially embodied which information. Physical causation is defined in §6, using the resources of the previous three sections. A method of quantifying the hierarchical structure of a thing is given in §7, to distinguish between mere complexity and

Causation is the Transfer of Information

organised complexity. This is done through Charles Bennett's notion of logical depth. The notion of logical depth can be used make sense of the account of laws as abstractions from particular cases of causation given in (Collier, 1996a) by showing how laws organise the superficial disorder of particular events and their relations. This completes the technical part of the chapter. The final sections look at some potential objections to the formal approach to causation, and the implications for the four current approaches to causation.

The quantification of form is a quantification of the complexity of a thing. Complexity has proven difficult to define. Different investigators, even in the same fields, use different notions. The Latin word means "to mutually entwine or pleat or weave together". In the clothing industry one fold (e.g. in a pleat) is a *simplex*, while multiple folds comprise a complex. The most fundamental type of complexity is informational complexity. It is fundamental in the sense that anything that is complex in any other way must also be informationally complex. A complex object requires more information to specify than a simple one. Even the sartorial origins of the word illustrate this relation: a complex pleat requires more information to specify than a simplex: one must specify at least that the folds are in a certain multiple, so a repeat specification is required in addition to the "produce fold" specifications. Further information might be required to specify any differences among the folds, and their relations to each other.

Two things of the same size or made from the same components might have very different informational complexities if one of them is more regular than the other. For example, a frame cube and a spatial structure composed of eight irregularly placed nodes with straight line connections between each node may encompass the same volume with the same number of components, but the

regularity of the cube reduces the amount of information required to specify it. This information reduction results from the mutual constraints on values in the system implied by the regularities in the cube - all the sides, angles and nodes must be the same. This redundancy reduces the amount of information required in a program that draws the cube over that required by a program that draws the arbitrary eight node shape. Similarly, a sequence of 32 '7's requires a shorter program to produce than does an arbitrary sequence of decimal digits. The program merely needs to repeat the output of '7' 32 times, and 32 itself can be reduced to 2^{5} , indicating 5 doublings of an initial output of '7'. To take a less obvious case, any specific sequence of digits in the expansion of the transcendental number π =3.14159... can be produced with a short program, despite the apparent randomness of expansions of π . The information required unambiguously to describe ordered and organised structures can be compressed due to the redundant information they contain; other structures cannot be so compressed. This is a property of the redundancy of the structures, not directly of any particular description of the structures, or language used for description.

The specification of the information content of a form or structure is analogous to an extended game of "twenty questions", in which each question is answered yes or no to identify some target. Each accurate answer makes a distinction⁸ corresponding to some difference between the thing in question and at least one other object. The answers to the questions encode the distinct structure of the target of the questions. Every determinate

Causation is the Transfer of Information

aspect of the form of any thing is included in its encoding. Thus, the encoding from questions and target perfectly represents the form of the target. Nothing else is left to encode, and the form can be recovered without loss from the encoding by examining the questions and decoding the answers (assuming the questions to be well formed, and the answers to be accurate). Such an encoding is an isomorphic map of the form of an entity like an object, property or system onto a string in which each entry is a "yes" or a "no", or a "1" or a "0". This string is an object to which computational complexity theory (a branch of mathematics) can be applied. The method is analogous to the use of complex numbers (the Cauchy-Riemann technique) to solve certain difficult problems in mathematical physics. The form is first converted to a tractable encoding, certain results can be derived, and then these can be applied to the original form in the knowledge that the form can be recovered with the inverse function. There is no implication that forms are strings of 1s and 0s any more than that the physical systems to which complex analysis of energy or other relations is applied really involve imaginary numbers.

Let s be mapped isomorphically onto some binary string σ_s (i.e. so that s and only s can be recovered from the inverse mapping), then the informational complexity of s is the length in bits of the shortest selfdelimiting computer program on a reference universal Turing machine that produces σ_s , minus any computational overhead required to run the program, i.e. $C_I = \text{length}(\sigma_s) - O(1)$.⁹ The first (positive) part of this measure

⁸ The logic of distinctions has been worked out by George Spencer Brown (1969) and is provably equivalent to the propositional calculus (Banaschewski, 1977). This is the basis of the binary (Boolean) logic of conventional computers.

⁹ On the original definition, length(σ_s) = min{|p|: $p \in \{0,1\}^* \& M(p) = \sigma_s\}$ = min{|p|: $p \in \{0,1\}^* \& f(p) = s\}$, |p| being the length of p, which is a binary string (i.e. $p \in \{0,1\}^*$, the set of all strings formed from the elements 1 and 0), and M being a specific Turing machine, and f being the *decoding function* to

is often called *algorithmic complexity*, or Kolmogorov complexity. The second part of the measure, O(1), is a constant (order of magnitude 1) representing the computational overhead required to produce the string σ_s . This is the complexity of the program that computes σ_s . It is machine dependent, but can be reduced to an arbitrarily small value, mitigating the machine dependence.¹⁰ I

recover σ_s from p and then s from σ_s . This definition requires an **O**(logn) correction for a number of standard information theoretic functions. The newer definition, now standard, sets length(σ_s) to be the input of the shortest program to produce σ_s for a selfdelimited reference universal Turing machine. This approach avoids **O**(logn) corrections in most cases, and also makes the relation between complexity and randomness more direct (Li and Vitányi, 1990).

¹⁰ For a technical review of the logic of algorithmic complexity and related concepts, see (Li and Vitànyi, 1990 and 1993). The complexity of a program is itself a matter for algorithmic complexity theory. Since a universal Turing machine can duplicate each program on any other Turing machine M, there is a partial recursive function f_0 for which the algorithmic complexity is less than or equal to the algorithmic complexity, plus a constant involving the computational overhead of duplicating the particular M, calculated using any other f. This is called the Invariance Theorem, a fundamental result of algorithmic complexity theory (for discussion of this, to some, counterintuitive theorem, see Li and Vitànyi, 1993: 90-95). Since there is a clear sense in which f_0 is optimal, the Invariance Theorem justifies ignoring the language dependence of length(σ_{s}), an this is now common practice for theoretical work. String maps of highly complex structures can be computed, in general, with the same computational overhead as those of simple structures (the computational overhead is

Causation is the Transfer of Information

deduct it to define the informational complexity to get a machine independent measure that is directly numerically comparable to Shannon information. permitting identification of algorithmic complexity and combinatorial and probabilistic measures of information.¹¹ The resulting value of the informational complexity is the information in the original thing, a measure of its form. Nothing additional needed to specify the form of anything. Consequently, I propose that the information, as measured by complexity theory, is the form measured, despite disparate approaches to form in differing sciences and philosophies. Nothing determinate remains to specify. Any proposed further distinctions that go beyond this are distinctions without a difference, to use a Scholastic saw. The language of information theory is as precise a language as we can have. Once all distinctions are made, nothing else we could say about something that gives any more information about it.

All noncomputable strings are algorithmically random (Li and Vitànyi,

nearly constant), so for complex structures (large C_I) the negative component of informational complexity is negligible. Furthermore, in comparisons of algorithmic complexity, the overhead drops out except for a very small part required to make comparisons of complexity (even this drops out in comparisons of comparisons of complexity), so the relative algorithmic complexity is almost a direct measure of the relative informational complexity, especially for large C_I .

¹¹ The more operational approach that retains the constant achieves only correspondence in the infinite limit, which is the only case in which the computational overhead, being a constant, is infinitesimal in proportion and is therefore strictly negligible (Kolmogorov, 1968; Li and Vitànyi, 1990).

1990). They cannot be compressed, by definition; so they contain no detectable overall order, and cannot be distinguished from random strings by any effective statistical test. This notion of randomness can be generalised to finite strings with the notion of *effective randomness*: a string is effectively random if it cannot be compressed.¹² Random strings do not contain information in earlier parts of the sequence that determines later members of the sequence in any way (or else they could be compressed).¹³ Thus any system

¹² Since it is possible to change an effectively random string into a compressible string with the change of one digit and yet, intuitively, the change of one digit should not affect whether a string is random, randomness of finite strings of length n is loosely defined as incompressibility within O(logn) (Li and Vitányi, 1990: 201). By far the greatest proportion of strings are random and in the infinite case the set of non-random strings has measure 1. It is also worth noting that there are infinite binary strings whose frequency of 1s in the long run is .5, even though the strings are compressible, e.g. an alternation of 1s and 0s. These strings cannot be distinguished by any effective statistical procedure (see above). If probability requires randomness, probability is not identical to frequency in the long run. It seems unreasonable, e.g. to assign .5 to probability of a 1 at a given point in the sequence because the frequency of 1s in the long run is .5, if the chance of getting a 1 at any point in the sequence can be determined exactly to be 1 or 0.

¹³ The converse is *not* true. Arbitrarily long substrings of non-computable strings (and, for that matter, incompressible finite strings) can be highly ordered, and therefore computable, but the location and length of these highly ordered sub-strings cannot be predicted from earlier or later elements in the string. In general, the incompressibility of a string does

Causation is the Transfer of Information

or process whose trajectory cannot be specified in a way that can be compressed is dynamically disorganised and effectively random. Such a system or process can have specific consequences, but cannot control anything, since these effects are indistinguishable from random by any effective procedure: no pattern (form) can be generated except by chance.

Algorithmic information theory can be used to quantitatively examine relations of information, and thus of form. The assignment of an information value to a system, state, object or property is similar to the assignment of an energy to a state of a system, and allows us to talk unambiguously of both the form and its value.¹⁴ We can then compare the form of two states, and of the transfer of form between states. In addition, and unlike for energy (whose relations also require dynamical laws), there are necessary relations between information instances that depend on whether a second instance is a theorem of the theory comprising the first instance and computation theory. Except for noncomputable cases, this relation is equivalent to there being a Turing type computation from the first information to the second. There are several relations of note: the information I_{A} contained in A contains the information in B iff I_B is logically entailed by I_A, and vice versa. This implies that the information in A is equivalent to the information in B if and only if each contains the other. The information in B given the information in A, and their mutual information can be expressed in a similar way. These relations are all standard in algorithmic complexity theory (Li and Vitànyi, 1993: 87ff). They allow us to talk

not imply the incompressibility of its substrings.

¹⁴ As with assigning energy values to real systems, assigning information values for practical purposes is not always easy.

about changes of form from one state of a system to another (processes), and between states of different systems (interactions). A relation between the information in A and that in B is not causal unless A and B have mutual information (they must be correlated). Sufficiency is not guaranteed, since the formal content of A and B can overlap by chance.

4. Causation is the Transfer of Information

The previous section placed the existence of mutual information as a necessary condition on the dynamics of form, but left open the possibility of chance coincidence of forms. An account of information as transfer of form must rule this possibility out non-circularly. I do this by stipulating that the information transferred must be *the same particular information*. Thus, we have the first definition of a causal process:

P is a causal process in system S from time t_0 to t_1 iff some particular part of the form of S involved in stages of P is preserved from t_0 to t_1 .

The preservation of form, here, implies that the form preserved is the identical form, not only in value and logical structure, but in fact. This definition is temporally symmetrical, in keeping with the temporal symmetry of fundamental laws of mechanics and quantum mechanics. Temporal asymmetry is an additional condition that will be considered in §6.1. The definition is similar to Salmon's PCI (1984: 155) according to which a process that transmits its own structure can propagate a causal influence from one spacetime locale to another. The main difference is that Salmon defines the transmission of structure as the capacity to carry a mark. This notion requires an irreducible counterfactual formulation (Kitcher, 1989; Dowe, 1992), as

Salmon has admitted (1994). This violates my methodological assumption that all fundamental concepts should refer only to logic, mathematics and concrete particulars (see footnote 2 above).

Another serious problem for the mark approach is that some causal processes cannot be marked. An electron, for example, has too few fundamental properties to be easily marked (though polarisation is a possibility). Other fundamental particles (photons and neutrinos, for example) cannot be marked without changing them into other particles. Equilibrium statistical mechanical systems cannot be marked without changing them from equilibrium. The mark will disappear relatively quickly through dissipation unless the mark puts the system quite far from equilibrium, which changes the dynamical properties of the system substantially. Since dynamical processes such basic as fundamental particle propagation and equilibrium processes are clearly causal, but cannot be marked, the mark method fails for reasons independent of any problem of the counterfactual dependence of the approach. In addition, it is unclear how nonphysical processes (if such exist, like the thought processes of God or other immaterial minds) could be marked. I avoid both the problem of unmarkable causal processes and the problem of counterfactual dependence by dropping marking entirely in favour of the transfer of form in general, rather than just of marks. The ability to mark a causal process remains important as an intuition leading to the definition above, however.

Critics might complain that my minimised definition of causation is wildly circular, since the identity of preserved form entails a causal connection. I am glad they share this intuition with me, since an adequate analysis of causation should correspond to intuitive ideas about causation. Any suspicion that causation has been surreptitiously imported in the above definition must involve the preservation of information, since form in

a state of a system has been defined with purely logical notions. Preservation, though, I have stipulated to be identity, which is also a logical notion.¹⁵ There is no direct reference to causation in the definition. This is perhaps more clear in the following variant:

> P is a causal process in system S from time t_0 to t_1 iff some particular part of the information of S involved in stages of P is identical at t_0 and t_1 .

This may seem like a trick, and indeed it would come to very little unless there is a way to determine the identity of information over time without using causal notions explicitly. This is an epistemological problem, which I defer until later. It turns out that there are simple methods for many interesting cases. From a strictly ontological view, the above definition is all that is needed, though the metaphysics of identity will depend on the substantial categories involved. Information tokens are temporal particulars. In physics with spatio-temporal locality, they are space-time "worms" (see §6.1).

The notion of transfer of information is useful:

Information I is transferred from t_0 to t_1 iff the same (particular) information exists at t_0 and t_1 .

The definition of causal process can then be revised to:

P is a causal process in system S from time t_0 to t_1 iff some part of the information of S

Causation is the Transfer of Information

involved in stages of P is

transferred from t_0 to t_1 .¹⁶

Interactive causation can now be defined easily:

F is a causal interaction between S_1 and S_2 iff F involves the transfer of information from S_1 to S_2 , and/or vice versa.

This allows a straightforward definition of causal forks, which are central to discussions of common cause and temporal asymmetry (Salmon, 1984):

F is an interactive fork iff F is a causal interaction, and F has distinct past branches and distinct future branches.

and,

F is a conjunctive fork iff F is a causal interaction, and F has one distinct past branch and multiple distinct future branches, or vice versa.

Interactive forks are X-shaped, being open to the future and past, while conjunctive forks are Y-shaped, being open in only one temporal direction. The probability relations Riechenbach used to define conjunctive forks follow from these definitions, the mathematics of conditional information,

¹⁵ The nature of identity is not important here, as long as identicals are indiscernable, i.e. if a=b, then there is no way in which a is distinct from b, i.e. they contain the same information.

¹⁶ It is tempting to define a cause as the origin of the information in a causal process. Quite aside from problems of which end of a causal process to look for the origin, the usual continuity of causal processes makes this notion poorly defined. Our usual conception(s) of cause has(ve) a pragmatic character that defies simple analysis because of the explanatory and utilitarian goals it (they) presuppose(s). Nonetheless, I am confident that my minimalist notion of causal process is presupposed by both vulgar and scientific uses of the term 'cause'. Transfer of information is necessary for causation, and is sufficient except for pragmatic concerns.

temporal asymmetry, and the probabilities derived from the information from the mathematical relation between informational complexity and probability-based definitions of information (justified by the definability of randomness within complexity theory), as do the probabilities for interactive forks. There is no room to prove this here, since apart from the mathematics we need a satisfactory account of the individuation and identity of dynamical processes that is beyond the scope of this chapter. It should be obvious, though, given that a causal process preserves information, that a past common cause (shared information) makes later correlation more certain than a present correlation of two events makes later interaction probable, though the reasons for this are not presently transparent by any means (Horwich, 1988). Likewise, an interactive fork gives information about both past and future probabilities, because the identity of the information in the interaction restricts the possibilities at both of the open ends of the forks.

Pure interactions between independent processes are rare, if not nonexistent. Interaction through potential fields (like gravity) occurs among all bodies continuously. If gravity and other fields are involved in the dynamics of interacting systems, enlarging the system to include all interactions is better than to talk of interacting systems. This is standard practice in much of modern physics, for example, when using the Hamiltonian formulation of Newtonian mechanics.

According to the information theoretic definition of causality, the necessity of causal relations follows easily, since the informational relations are computational. The information transferred must be in the effect and it must be in the cause, therefore the relevant information is entailed by both the cause and the effect. Furthermore, the existence of the identical information (token) in both the cause and effect is both a

Causation is the Transfer of Information

necessary and a sufficient condition for causation. We can think of a causal process as a computation (though perhaps not a Turing computation or equivalent) in which the information in the initial state determines information in the final state. The effect, inasmuch as it is determined, is necessitated by the cause, and the cause must contain the determined information in the effect. Although the causal relation is necessary, its conditions are contingent, so it is necessary only in the sense that given the relata it cannot be false that it holds, not that it must hold (see Collier, 1996a for more on this form of necessitation, and its role in explaining the necessity of natural kinds and laws). Note that the only necessity needed to explain causal necessity is logical entailment. This is one great advantage of the information theoretic approach to causation, since it avoids direct appeals to modalities. Counterfactual causal reasoning fixes some counterfactual conditions in distinction to the actual conditions either implicitly or explicitly through either context or conventions of language. Counterfactual causal reasoning is thus grounded in hypothetical variations of actual conditions.

Locality, both spatial and temporal, is a common constraint on causation. Hume's "constant conjunction" is usually interpreted this way. While it is unclear how causation could be propagated nonlocally, some recent approaches to the interpretation of quantum mechanics (e.g. Bohm, 1980) permit something like nonlocal causation by allowing the same information (in Bohm's case "the implicate order") to appear in spatially disparate places with no spatially continuous connection. Temporally nonlocal causation is even more difficult to understand. but following its suggestion to me (by C.B. Martin) I have been able to see no way to rule it out. Like spatially nonlocal causation, temporally nonlocal causation is possible only if the same information is transferred from one time to another without the information

existing at all times in between. Any problems in applying this idea are purely epistemological: we need to know it is the same information, and not an independent chance or otherwise determined convergence. Resolving these problems, however, requires an appropriate notion of information and identity for the appropriate metaphysical category.

The epistemological problems are diminished immensely if temporal locality is required. If there is a sequence of temporal stages between the start and end of a candidate causal process for which there is no stage at which the apparently transferred information does not exist, the candidate process is temporally local. All other things being equal, it is far more parsimonious to assume that the identical information exists at each stage of a candidate local process than that the information at each stage arises independently. The odds against a candidate local causal process being noncausal (i.e. apparently but not actually transferring the identical information) are astronomical. The main exception is an independent common cause, as in epiphenomena like Leibniz' universal harmony. There are difficulties distinguishing epiphenomena from direct causal phenomena, but in many cases intervention or further knowledge can provide the information needed to make the distinction. For example, we can tell that the apparent flow of lights on a theatre marquee is not causal by examining the circuitry. The null hypothesis, though, despite these possibilities, would be that candidate causal processes are causal processes. Lacking other information, that hypothesis is always the most parsimonious and the most probable. Unfortunately, it can't be shown conclusively that any apparent causal process is really causal, but this sort of problem is to be expected of contingent hypotheses. The important thing to note is that (ignoring pragmatic considerations) any talk of of the transfer of the same information throughout the apparent process. It is interesting to note that my approach to causation permits an effectively random system to be a cause. A large random system will have ordered parts, and an infinite

random system will have ordered parts of arbitrarily large size (see footnote 13 above). If the universe originated as an infinite random system, as suggested by David Layzer (1990), then ordered random fluctuations would be expected, and our observable world could be caused by a particularly large fluctuation that later differentiates through phase transitions into the variety that we observe today. This cosmological theory requires the preexistence of a random "stuff" with the capability of self interaction. No intelligence or pre-existing order is required to explain the causal origin of the order and organisation in the observable world. This is contrary to the views of my rationalist predecessors like Aristotle, Descartes and Leibniz.

So far, this account of causation has very little flesh; it is just a formal framework. This will be remedied in the next two sections in which I apply the framework to physical causation.

5. The Negentropy Principle of Information

To connect information theory to physical causation, it is useful to define the notions of order and disorder in a system in terms of informational complexity. The idea of disorder is connected to the idea of entropy, which has its origins in thermodynamics, but is now largely explained via statistical mechanics. The statistical notion of entropy has allowed the extension of the idea in a number of directions, directions that do not always sit happily with each other. In particular, the entropy in

Causation is the Transfer of Information

causation can be eliminated in favour of talk

mathematical communications theory (Shannon and Weaver, 1949), identified with information, should not be confused with physical entropy (though they are not completely unrelated). Incompatibilities between formal mathematical conceptions of entropy and the thermodynamic entropy of physics have the potential to cause much confusion over what applications of the ideas of entropy and information are proper (e.g. Wicken, 1987; Brooks et al, 1986).

To prevent such problems I adopt the interpretive heuristic known as NPI, according to which the information in a specific state of a physical system is a measure of the capacity of the system in that state to do work (Schrödinger, 1944; Brillouin, 1962: 153), where work is defined as the application of a force in a specific direction, through a specific distance.¹⁷ Work capacity is the ability to control a physical process, and is thus closely related to causality. Nevertheless, it is a state variable of a system, and involves no external relations, especially to effects. So the concept of work capacity is not explicitly causal (though the concept of work is).¹⁸ Through

Causation is the Transfer of Information

the connection with work, NPI ties information, and so complexity and order, to dynamics. NPI implies that *physical information* (Brillouin, 1962)¹⁹ has the opposite sign to physical entropy, and represents the difference between the maximal possible entropy of the system (its entropy after all constraints internal to the system have been removed and the system has fully relaxed, i.e. has gone to equilibrium) and the actual entropy, i.e.,

NPI: $I_P = H_{MAX} - H_{ACT}$

where the environment of the system and the set of external constraints on the system are presumed to be constant. The actual entropy, H_{ACT} , is a specific physical value that can in principle be measured directly (Atkins, 1994), while the maximal entropy, H_{MAX} , of the system is also unique, since it is a fundamental theorem of thermodynamics that the order of removal of constraints does not affect the value of the state variables at equilibrium (Kestin, 1968). This implies that the equilibrium state contains no trace of the history of the system, but is determined entirely by synchronic boundary conditions. Physical information, then, is a unique and

¹⁹ Brillouin (1962: 152) refers to physical information as bound information but in the light of my distinction between intropy and enformation (see below), I avoid this term (since in one obvious sense intropy, being unconstrained by the system, is not bound). Brillouin defines bound information as a special case of free information, which is abstract, and takes no regard of the physical significance of possible cases. Bound information occurs when the possible cases can be regarded as the complexions of a single physical system.

¹⁷ Work has dimensions of energy in standard mechanics, and thus has no direction. However, since it is the result of a force applied through a distance, it must be directed. Surely, undirected force is useless. However, this changes the units of work, since energy is not a vector. Interestingly, Schrödinger (1944) considered exergy as a measure of physical information, but rejected it because people were easily confused about energy concepts. This is remarkable, since exergy and entropy do not have the same dimensions.

¹⁸ Though work capacity is a dispositional concept, it is defined through NPI in terms of the state variables of a system, which can be understood categorically. The causal power of a system is determined by its work capacity.

The details are relatively simple, but are beyond the scope of this paper, since explaining them requires clearing up some common misconceptions about statistical mechanics.

dynamically fundamental measure of the amount of form, order or regularity in a state of physical system. Its value is non-zero only if the system is not at equilibrium with its environment. It is a measure of the deviation of the system from that equilibrium. It is important to remember that NPI is not a formal or operational definition and, given the current proliferation of formalisms for entropy and information, it needs to be interpreted as appropriate for a given formalism and for a given physical system and its environment.²⁰

On the other hand, NPI is an implicit definition, since it determines how terms like entropy and information are to be used in a physical context. As in mathematics, central definitions in empirical theory should be supported with an existence proof. This is done by showing that violating the definition would violate any known or theoretically projected observations (Mach, 1960: 264ff).

²⁰ With respect to the need to interpret the principle in relation to the system and environment under consideration, the situation is exactly paralleled by that for energy and momentum. By referring information to the system environment the need to define some absolute reference point where all constraints of any kind are relaxed, which is not obviously a well defined condition is avoided. Just as there are very different formulae for all the forms of potential energy in different systems, so too are there for forms of entropy and information. The 0th Law of Thermodynamics suggests an absolute measure of entropy, but in practice the "freezing out" of complex order in the hierarchy of energy levels precludes strict application of this "law", except to ideal gases. For the 0th Law to apply, all degrees of freedom of a system must be equally flexible. This is very unlikely to be true in any real physical system (see also Yagil, 1993b).

Causation is the Transfer of Information

If we assume NPI, then reliable production or reproduction of one bit of information requires a degradation of at least kTln2 exergy (available energy), where k is Boltzmann's constant in a purely numerical form (Brillouin, 1962: 3), and T is temperature measured in energy units. This relation must hold, or Maxwell's demon will come to haunt us, and the Second Law of Thermodynamics will come tumbling down. NPI, then, reflects the strongly confirmed intuitions of physicists and engineers that the physical form of things cannot be used in some tricky way to control effectively random physical processes. There are strong reasons to believe that this is logically impossible in a world restricted to physical causation (Collier, 1990b). I will return to this later in §6.1. NPI is empirically justified; we know, for example, that violation of NPI, which would amount to using information to reduce the entropy of an isolated system, violates our most common experiences of physical systems. NPI implies that a bit of information can be identified with the minute but non-negligible physical value $k \ln 2$ and that its transfer from one system or part of a system to another will require the *transfer* of at least kTln2 exergy (see Brillouin, 1962 for details). This gives us a quantitative physical measure of form that is directly related to exergy and entropy, the central concepts in nonequilibrium processes. These relations allow us to study complexity changes in physical processes, and permit principled extensions of the concepts of entropy and information.²¹

²¹ It is worth noting at this point that logical processes, such as computations, obey the Second Law as well, in the sense that a computation can produce only as much information as it starts with, and generally will produce less. There are theoretically possible reversible computers, but they produce vast amounts of waste stored bits if they compute practical results. Consequently,

NPI can be motivated more directly from information theory. This might be useful to those who find themselves on the wrong side of C.P. Snow's two cultures, the divide being the understanding of entropy. Entropy cannot be explained simply without loss of content²², but the following explanation will give the main details, though it will give no idea of how to apply the ideas (unlike the way I introduced NPI above, which rigorously connects information to known physical principles and their common applications). H_{MAX} represents a possible state of the system in which there is no internal structure except for random fluctuations. All possible microstates of the system are equally likely. There is no physical information within the system, and it cannot do any work internally, since it is statistically uniform except for random fluctuations, which, because of their random nature, cannot be harnessed for any purpose from within the system. The actual entropy, however, except for systems in equilibrium, permits internal work, since there is energy available in the nonuniformities that can be used to guide other energy. The information equivalent to this ordered energy is just that we would obtain with a perfect "game of twenty questions" that determines the information gap between the information of the macrostate and the information of the microstate, and hence the probability

Causation is the Transfer of Information

distribution of microstates. It therefore represents the form (nonrandom component) of the system, according to the definitions of §3. This justifies the connection between form and capacity for work. Any other consideration of dynamics and physical information will have to be consistent with this connection (however subtle) between dynamics and form, i.e. any physical system must satisfy NPI.

There are two ways that entropy is significant in physical systems, sorting and energy availability, though they are really extremes of one set of principles. To take a simple example of sorting, imagine that we start with a container of m "red" and n "white" molecules in an ideal gas at equilibrium, S_0 , and it ends in a state, S_1 , in which all the red molecules are on the right side of the container, and the white molecules are on the left side, so that we could move a frictionless screen into the container to separate completely the red and white molecules without doing any additional work. The entropy of S_0 is $-\Sigma P_0 k \ln P_0$, and the entropy of S_1 is $-\Sigma P_1 k \ln P_1$, where P_0 is the inverse of the number of complexions in the initial state, and P_1 is the inverse of the number of complexions in the final state. Simplifying again, assume the $m = n = 1.^{23}$ Then the entropy of the final state is obviously 0, since there is only one possibility, in which the red molecule is on the right, and the white molecule is on the left, so $P_1 = 1$. The entropy of the initial state is higher: both molecules can be either on the right or the left, or there can be a red on the left or a red on the right, giving four distinct possibilities, and $P_0 = .25$. If we know that the system is in S_1 , we have 2 bits more information than if we knew merely that it was in S_0 . For example, we might have the

arguments concerning the dynamics of physical complexity also apply to any sort of creature governed by logic. This places some limits on the role of gods as counterexamples to causal analyses, unless the gods act inconsistently with logic. We might as well just assume uncaused events in these supposed counterexamples (see §6.2).

²² Many bright students have taken more than one full course on the subject at university without coming to understand entropy properly.

²³ This is not quite as simple as Szillard's case (see Brillouin, 1962: 176ff), which uses only one molecule!

information that no two molecules are on the same side, and that a red molecule is on the right, requiring two binary discriminations. To slide the screen in at an appropriate time, we need the information that the system is in S_1 , i.e. we need the information difference between S_0 and S_1 . This is exactly equivalent to the minimum entropy produced in a physical process that moves the system from S_0 to S_1 , as can be seen by setting k to 1, and using base 2 logarithms to get the entropy in bits. To move the system from S_0 to S_1 , then, requires at least 2T work. This is a very small amount; the actual work input would be larger to cover any energy stored and/or dissipated. Alternatively, a system in S_1 can do at most 2T work before it has dissipated all its available energy from this source. Putting this in other words, the system can make at most two binary distinctions, as can be seen by reversing the process.²⁴ These two

Causation is the Transfer of Information

bits measure the maximal controlling potential of the system: implemented as a controller, controlling either itself or another system, the system could function as at most two binary switches. Calculating the physical information for each case from the definition above, $I_P(S_0) = 0$, while $I_P(S_1) = 2$. As it should, the difference gives us the amount of information lost or gained in going from one state to the other. A number of years ago it was confirmed that the entropy production of the kidneys above what could be attributed to the basal metabolism of its cells, could be attributed to the entropy produced in sorting molecules for elimination. Presumably, more subtle measurements would also confirm a physical realisation of the molecule example.

The relations between information and energetic work capacity are somewhat subtle, since they involve the correct application of NPI, which is not yet a canonical part of physics.²⁵ The physical information in a given system state, its capacity to do work, breaks into two components, one that is not constrained by the cohesion in the system, and one that is. The former, called *intropy*, ι , is defined by $\Delta \iota = \Delta(\text{exergy})/T$, so that $\int T \Delta \iota$ measures the available energy to do work, while the latter, called *enformation*, ϵ , measures the structural constraints internal to the system that can guide energy to do work (Collier, 1990a). Enformation determines the additional energy that would be obtained in a system S if all cohesive constraints on S were released. Intropy measures the ordered energy that is not controlled by cohesive system processes, i.e. by system laws, it is unconstrained and so free to do work. For this reason, though ordered, both intropy and exergy are system statistical properties in this

²⁴ NPI is assumed throughout, as is the impossibility of a Maxwellian demon (Brillouin, 1962; Bennett, 1982; Collier, 1990b). Szillard's original argument makes the connection to work more obvious by using a molecule pushing on a cylinder in a piston, but the more general arguments by Bennett and Collier examine (in different ways) the computational problem the demon is supposed to solve. The connection to work is implied by thermodynamics and NPI. Szillard used thermodynamics explicitly, but NPI only implicitly, which meant that his exorcism of the demon could not be general. Denbigh and Denbigh (1985) argue that information is not required for the exorcism, since thermodynamics can be used in each instance. It seems to have escaped them that proving this requires something at least as strong as NPI. The problem of Maxwell's demon is especially important because it forces us to be explicit about the relations between control and physical activity. A demon that could store information in some

non-physical form could perform its sorting job, though at the expense of producing waste (unusable) information in this storage.

²⁵ That would require the equivalent of the acceptance of the ideas in this section.

sense: their condition cannot be computed from the cohesive or constrained system state, the cohesive state information determines the micro state underlying the intropy only up to ensemble of intropy-equivalent an microstates. There is another system statistical quantity, entropy, S, but it is completely disordered or random, it cannot be finitely computed from any finite system information.²⁶ Entropy is expressed by equiprobable states, and so appears as heat which has no capacity to do work; $\Delta S =$ $\Delta Q/T$, where Q is heat, and $\int T\Delta S$ measures heat. Enformation is required for work to be accomplished, since unguided energy cannot do work.²⁷ Intropy is required for work in dissipative systems, to balance dissipation (S production).

Consider, for example, a system S with heat Q as its only unconstrained energy. If S is at equilibrium then the only enformation is the existence of a system temperature (not that it is of some specific value T), for only that follows from the system constraints, and

²⁷ Archimedes lever with which he could move the world, like any other machine, must have a specific form: it must be rigid, it must be long enough, there must be a fixed fulcrum, and there must be a force applied in the right direction. If any of these are lacking, the lever would not work. No amount of energy applied without regard to the form in which it is applied can do work, except by accident.

Causation is the Transfer of Information

 $\iota = 0$ and Q is entropic since Q cannot do work on S. If S nomicly maintains an internal temperature gradient G then G is enformation for S since it cannot be released to do work without first altering the cohesive structures of S. If G is unconstrained by S then G expresses intropy in S since G is an ordering of the heat energy and work can be done in S because of G. (In fact S will dissipate G, creating entropy, until equilibrium is reached.) Further, note that if S, even if at internal equilibrium with G = 0, is made part of a larger system S_s where it is in contact with another sub-system P of S_s at a lower temperature, then there is now a new temperature gradient G_s unconstrained by S_s so S will do work on P with heat flowing between them until equilibrium is reached (G_s = 0) at some intermediate temperature; hence G_s is intropic in S_s even though S has no intropy and S's temperature, which serves in part to determine G_s , is enformation in S.²⁸ These analyses carry over to all other physical forms of energy.

The main difference between intropy and enformation is the spatial and temporal scale of the dynamical processes that underlie them.²⁹ The dynamics underlying intropy

²⁸ There is nothing arbitrary about these system-relative distinctions; each is grounded in the system dynamics. Relational properties, like intropy, entropy and enformation, necessarily produce relativised applications across relationally distinct contexts, e.g. S and S_s here, and it is an error (albeit a common one) to equate this to relativism, which is the absence of any principled basis for distinguishing conflicting claims across contexts.

²⁹ All enformation except perhaps the enformation in some fundamental particles, like protons, will eventually decay, which means that at some temporal scale all, or at least most, enformation behaves as intropy. The scale is set by natural properties of the

²⁶ One obvious information basis to consider is a complete microscopic description of a system. However, behind this statement lies the vexed issue of a principled resolution of the relations between mechanics and thermodynamics that respects the irreversibility of the latter despite the reversibility of the former. While the analysis offered here represents a small step toward greater clarity about this complex issue, I do not pursue it here.

have a scale smaller than that of the whole system, and involve no long term or spatially extended constraints, except those that govern the system as a whole, which in turn constitute the system enformation. The intropy of a system S is by definition equal to the difference between S's actual entropy and its maximal entropy when exergy has been fully dissipated (given enformation invariant, i.e. S's constraints remaining unchanged, and environment invariant); so, $\iota = I_P = H_{MAX}(S)$ - $H_{ACT}(S)$, all at constant environment and constraints. The enformation is just the additional information equal to the difference between $H_{ACT}(S)$ and the entropy of the set of system components that result when the constraints on S are fully dissipated and S comes to equilibrium with its environment (assumed to remain otherwise invariant); $\epsilon =$ $I_E = H_{MAX}(S_E) - H_{MAX}(S)$. Note that $I_P(S) = \iota +$ $\epsilon = H_{MAX}(S_E) - H_{ACT}(S)$ as required by NPI. This is perhaps more clear with an example. A steam engine has an intropy determined by the thermodynamic potential generated in its steam generator, due to the temperature and pressure differences between the generator and the condenser. Unless exergy is applied to the generator, the intropy drops as the engine does work, and the generator and condenser temperatures and pressures gradually equilibrate with each other. The enformation of the engine is its structural design, which guides the steam and the piston the steam pushes to do work. The design confines the steam in a regular way over time and place. If the engine rusts into unrecoverable waste, its enformation is completely gone (as is its intropy, which can no longer be contained), and it has become one with its supersystem, i.e. its surroundings. Such is life.

Causation is the Transfer of Information

As noted, NPI allows us to divide a physical system into a regular, ordered part, represented by the physical information of the system, i.e. $\iota + \epsilon$, and a random, disordered part, represented by the system entropy. The orderedness of the system is its information content divided by the equilibrium (i.e. maximal) entropy, i.e.; $O = I_P/H_{MAX}$, while the disorderedness is the actual entropy divided by the equilibrium entropy, i.e. D =H_{ACT}/H_{MAX} (Layzer, 1975; Landsberg, 1984); it follows from NPI that O+D = 1. The informational complexity of the information in the system, C_I (I_P), is equal to the information required to distinguish the macrostate of the system from other macrostates of the system, and from those of all other systems made from the same components.³⁰ The mathematical relations between statistical entropy and algorithmic information (Kolmogorov, 1965, 1968; Li and Vitányi, 1993) ensure that $C_{I}(I_{\mathbf{P}}) = H_{MAX}$ - H_{ACT} , so $C_{I}(I_{P}) = I_{P}$. This is so since the

system in question. Specifically, the extent of the cohesion of the system implies a natural scale (Collier, 1988, Collier and Hooker, submitted; Christensen et al, in preparation).

³⁰ A complete physical specification would amount to a maximally efficient physical procedure for preparing the system, S, in the macrostate in question from raw resources, R (Collier, 1990a). Furthermore, the procedure should be self-delimiting (it finishes when S is assembled, and only when S is assembled). The information content of this specification is just I_P plus any intropy that must be dissipated in the process. The latter is called the thermodynamic depth of the state of the system, and is equal to $H_{ACT}(R) - H_{ACT}(S)$ if there are no practical restrictions on possible physical processes. The algorithmic complexity analogue of thermodynamic depth is the complexity decrease between the initial and final states of a computation (through memory erasure). This quantity is often ignored in algorithmic complexity theory, but see (Bennett, 1985; Collier, 1990b; also Fredkin and Toffoli, 1982), who would hold that the analogy is a physical identity.

physical information of a system determines its regularity and this regularity can be neither more nor less informationally complex than is required to specify the regularity. (The informational complexity of the disordered part is equal to the entropy of the system, i.e. $C_{I}(H_{MAX} - I_{P}) = C_{I}(H_{ACT}) = H_{ACT}$ and since O $= I_P/H_{MAX}$, the ordered content of $S = H_{MAX}O$ = I_p as required.) These identities allow us to use the resources of algorithmic complexity theory to discuss physical information, in particular to apply computation theory to the regularities of physical systems. This move has always been implicit in the use of deductive reasoning to make physical predictions, and should be non-controversial. The main achievement here is to tie together explicitly computational and causal reasoning within a common mathematical language (see also Landauer, 1961, 1987; Bennett, 1988).³¹

Causation is the Transfer of Information

It is important to note, however, that NPI can be stated entirely in terms of state descriptions and relations between state descriptions, and involves no explicit importation of causal notions.

6. Physical Causation

The analysis of causation in §4 is very abstract and perhaps hard to comprehend. In this section I use NPI to give an account of physical causation, the target of most contemporary metaphysical accounts of causation. My account divides into ontological and epistemological issues.

6.1 Ontological Issues

The mark approach fails because of its dependence on counterfactuals, and the inability of some obviously causal processes to be marked (see §4). This problem can be overcome if we take the form of the states of a physical process to itself be a mark, where the information in the mark is given by the methods of §5. The mark approach is attractive, since we can make a recognisable mark, and then check it later. A paradigmatic example is signing or sealing across the flap of an envelope so we, or someone else, can check that it is the original envelope, and that it has not been tampered with. Modern computer security methods using open and

³¹ There is one further terminological issue concerning physical information that should be noted. By NPI, the disordered part of the system does not contain information (because it cannot contribute to work), but the information required to specify the complete microstate of the system is equal to the information in the macrostate plus the information required to specify the disordered part. Layzer (1975, 1990) speaks of the information required to specify the disordered part as the "microinformation" of microstates, as if the information were actually in the microstate. This information can do work only if it is somehow expressed macroscopically. For this reason, I prefer to regard unexpressed microinformation as a form of potential information (Gatlin, 1972; Collier, 1986; Brooks and Wiley, 1988). Expressed information is sometimes called stored information (Gatlin, 1972; Brooks and Wiley, 1988). Potential information can also be *directly expressed* as intropy, e.g. in the Brownian motion of a particle, as opposed to at the expense of enformation, e.g. when

micro fluctuations disrupt structure. Although expression as intropy is physically possible, it cannot be physically controlled (Collier, 1990b). Control of this process would imply the existence of a Maxwellian demon. In dissipative structures, especially those formed in systems with multiple attractors, in which the branch system followed in a phase change is determined by random fluctuations, potential information can be expressed macroscopically at the expense of dissipation outside the macroscopic structure.

private keys are directly analogous, and much more secure. Unfortunately, many causal processes are too simple to mark at all, let alone permit the complex mathematical methods of open key security. Security and recognition, however, are more epistemological problems than an ontological ones, and I will postpone this issue until section §6.2. For now I will concentrate on the ontology of the transfer of physical form.

Information preserved in physical will have constrained causation (enformational) and may have unconstrained (intropic) components. For example, a steam locomotive retains its structure as it moves along the tracks, but it also burns fuel for its intropy, part of which is converted into motion. It is only the enformation that is essential to a dynamical process, since the intropy is statistical and its microscopic basis is possibly chaotically variable, whereas the enformation guides the dynamical process, and constitutes the structure of the system at a given time. Therefore we might try:

> P is a physical causal process in system S from time t_0 to t_1 iff some part of the enformation in S is transferred from t_0 to t_1 .

We may or may not want to add locality requirements. Familiar cases of physical causality are both temporally and spatially local.

Unfortunately, pseudoprocesses like the passing of a beam of laser light across the face of the moon satisfy this definition, but the causal process involved is actually a good deal more complicated. NPI can help us here. First, though, it helps to invoke Russell's "atat" theory of causal propagation (Salmon, 1984: 147ff) to ensure locality:

P is a causal process in system S from time t_0 to t_1 iff some part of the enformation in S is identical from t_0 to t_1 , and at all times between t_0 and t_1 .

Causation is the Transfer of Information

As mentioned in §4, the at-at approach to locality makes the information token a spacetime "worm". Locality disallows causation over a temporal gap, but it is very much in tune with the intuitions of physicists and other scientists that all physical causation is local. The main exception might arise in quantum mechanics on Bohm's implicate order approach, which is controversial, and nonetheless requires locality of a different sort through the enfolding of the universe. The above definition can be revised, if necessary, to take into account this different sort of locality. Of course the intuitions of physicists may be false, but at this time they are our best experts on how to interpret causality and cognate terms.

The pseudoprocess problem is then the underlying problem for the information transfer theory, as it is for the mark approach. I attack this problem by invoking NPI explicitly:

P is a physical causal process in system S from time t_0 to t_1 iff some part of the enformation in S is transferred from t_0 to t_1 , and at all times between t_0 and t_1 , all consistent with NPI.

Consistency with NPI is a fairly strong constraint. It requires that causal processes be consistent with entropy changes in the processes. This is enough to rule out the flashlight beam across the moon pseudoprocess, since the information in the spot comes from nowhere, and goes to nowhere, if the movement is all there is to the process. This violates not only the Second Law of Thermodynamics, but also strong physical intuitions that embodied order cannot just appear and disappear. Quantum mechanics and the emergence of dissipative structures seem to violate this intuition, but on closer study symmetry requirements in quantum mechanics and the influence of the order in microscopic fluctuations ensure that no new information is generated.

The Second Law itself has an interesting status. Although reversible systems are possible in nature (apparently reversible systems can be designed in the laboratory as thermodynamic branch systems, but in fact they obey the Second Law when it is properly interpreted), it is impossible for any physical device or physical intervention to control the overall direction of the entropy gradient because to do so is computationally impossible (Bennett, 1987; Collier, 1990b). Reversal of the normal increase in entropy requires very special conditions that can be detected by examining the history, boundary conditions and dynamics of the system. Consistency with the Second Law is not merely an empirical requirement; it is closer to a logical constraint, and holds for abstract computational systems as much as for physical systems (Landauer, 1961, 1987; Bennett, 1988; Li and Vitànyi, 1993). Processes can be distinguished from pseudoprocesses, then, by their consistency with the Second Law, if we take care to ensure that special conditions allowing spontaneous reversal of entropy increase do not hold. It is possible (though highly unlikely) that a pseudoprocess could by chance mimic a real process with respect to the constraints of NPI, but experimental intervention could detect this mimicry to a high degree of probability.

NPI ensures that if information is not lost, the causal process is temporally symmetrical, and there is no internally defined temporal direction. If dissipation occurs, however, the information in the final state is less than in the initial state and the initial state cannot be recovered from the final state. Consequently, dissipative causal processes are temporally directed. The complete nature of dissipation is not yet completely understood (Sklar, 1986, 1993), but we know that it occurs regularly.

I cannot give a complete account of chance causation here, but I will give a brief sketch. If the information in the effect cannot

Causation is the Transfer of Information

be computed from the information in the cause, the system is not predictable, even though it may be deterministic in the sense that the same cause would produce the same effect. In either deterministic or indeterministic cases with this sort of informational gap between cause and effect, the probability of the effect can be computed by the informational size of the gap by using the standard relations between information and probability. Perhaps the most interesting case is deterministic chance, which at first appears to be an oxymoron. Consider a coin toss. Suppose that the coin's side is narrow compared with the roughness of the surface on which it lands, so it comes up either heads or tails. Suppose further that it's trajectory takes it through a chaotic region in the phase space of the coin toss in which the head and tail attractor basins are arbitrarily close to each other (the definition of a chaotic region). The path of the coin to its eventual end in one of the attractors in this case cannot be computed with any finite resources (by any effective procedure). This means that the attractor the coin ends up in is irreducibly statistical, in the sense that there is no effective statistical procedure that could distinguish the attractor selected (however much it is determined) from a chance occurrence (see end of section 3.1). The actual odds can be computed by the size of the information gaps in prediction of each of the outcomes, since some of the form can be tracked (e.g. the coin keeps its shape). If the coin has the right form (i.e. it is balanced and symmetrical), the odds will be roughly 50-50 for heads or tails. Note that no counterfactuals are required for this account of chance, nor is the chance in any way subjective.

Some readers might resist the idea of a deterministic system being a chance system, but since no effective statistical procedure can distinguish a chaotic system from a chance system, the difference is certainly beyond our reach. The decision to call systems like the coin toss chance systems is somewhat

arbitrary, but it is consistent with anything we could know about statistics or probability. The distinction between unpredictable systems and indeterministic systems forces us to choose which to call intrinsically chancy. Since chance has long been associated with limits on predictability, even by those like Hume who considered it to be a purely subjective matter, I believe that the association of chance with intrinsic unpredictability rather than with indeterminism is justified. The difference between chance deterministic systems and chance indeterministic systems, then, is that the information gap in the former is only in the computability of the information transferred, while in the latter the gap is in information transferred. Deterministic are systems entirely causal, but indeterministic systems are not. A completely random system might still be completely determined. Our universe might be such a system (see §4 above), showing only local, but not global order beyond the constraints of logic.

6.2 Epistemological issues

One problem with the information theoretic approach is that it requires precise assessments of the quantity of form. This is difficult even for simple molecules, though techniques are being developed using informational complexity (Holzmüller, 1984; Küppers, 1990; Schneider, 1995; Yagil, 1993a, 1993b, 1995). A further problem is that the maximally compressed form of an arbitrary string is not computable in general, though again, methods have been developed for special cases, and approximation methods have been developed that work for a wide range of cases (Rissanen, 1989; Wallace and Freeman, 1987). This problem does not affect the metaphysical explanation of causation in terms of information transfer, however.

Perhaps a more serious problem is determining the identity of information in a system trajectory. For example, apparent

Causation is the Transfer of Information

causation might really be Leibnizian preestablished harmony. We might not be able to tell the difference, but the information flow would be different if God or some demon is the cause of the apparent direct causation. This situation does not violate the informational metaphysics, however, since the information flow in the preestablished harmony case would be from God to individual monads, with the form originating in God.³² The problem of the intervention of gods in a specific causal process is just a special case of the Leibniz case, and can be handled the same way, as long as they are subject to the constraints of logic. If they are not, and the effects of their actions bear no determinate relation to the cause, the effects are chance, and can be handled as such.

What appears to be a causal process from the information theoretic point of view might be a chance sequence, or contain a chance element that breaks the causal chain. For example, at one instant the causal chain might end indeterministically, and a new chain might start at the next instant, also indeterministically, where the form types involved are identical to the types if the chain were unbroken. Phil Dowe's chapter in this volume deals with the identity across time issue fairly effectively by showing that other approaches to causation also suffer from the problem. I see no conclusive way around it. I think we just have to live with this possibility. On the other hand, if locality holds, and NPI

³² It seems to me that Leibniz had something like this in mind, but it is unclear to me how the generation of new information could be possible without God suffering from the waste problem of the computational version of Maxwell's demon. God could solve the problem by storing huge amounts of waste storage someplace otherwhere, but it would certainly complicate the metaphysics. I believe my one levelled approach is more parsimonious.

is applied, we can reduce the probability that what appears to be a transfer of the same information is actually a chance configuration to a minuscule consideration, as argued in §3. The lack of certainty should not be philosophically bothersome, since we cannot be certain of contingencies in any case.

7. Organisation, Logical Depth and Causal Laws

One last technical idea will be useful for connecting causality to causal laws. The redundancy in a system (physically, its I_{P}) can be decomposed into orders n based on the number of components, n, required to detect the redundancy of order n (Shannon and Weaver 1949).³³ Complex chaotic (and nearly chaotic) conservative systems, e.g. a steel ball pendulum swung over a pair of magnets under frictionless conditions, typically show relatively little low order redundancy, but a significant amount of high order redundancy, while living systems typically show significant redundancies at both low and high orders (Christensen et al, in preparation). It is ultimately an empirical matter just how local and global redundancies interrelate to lower and higher order redundancies in particular classes of systems, though usually higher order redundancies will also have large temporal or physical scale, or both.

Causation is the Transfer of Information

In many of these cases the higher order redundancy is hidden or buried, in the sense that it is not evident from inspecting small parts of the system or local segments of the dynamic trajectory of the system. Nevertheless, it can be seen in the overall structure of the system, and/or in the statistics of its trajectory. For example, the trajectory of a chaotic system is locally chaotic, but it is (probably) confined to spatially restricted attractor basins. Because the information in such systems involves large numbers of components considered together without any possibility of simplification to logically additive combinations of subsystems (the systems are nonlinear), computation of the surface form from the maximally compressed form (typically an equation) requires many individual steps, i.e. it has considerable logical depth (Bennett, 1985; Li and Vitányi 1990, 238). Bennett has proposed that logical depth, a measure of buried redundancy, is a suitable measure of the organisation in a system.

Formally, logical depth is a measure of the least computation time (in number of computational steps) required to compute an uncompressed string from its maximally compressed form.³⁴ Physically, the logical

³³ This is a strictly mathematical decomposition. Physical decomposability is not required. A level of organisation is a dynamically grounded real structural feature of a complex system which occurs when (and only when) cohesive structures emerge and operate to create organisation (Collier, 1988). The same level may manifest or support many different orders of organisation and the same order of organisation may be manifested or supported at many different organisational levels.

³⁴ Some adjustments are required to the definition to get a reasonable value of depth for finite strings. We want to rule out cases in which the most compressed program to produce a string is slow, but a slightly longer program can produce the string much more quickly. To accommodate this problem, the depth is defined relative to a significance level s, so that the depth of a string at significance level s is the time required to compute the string by a program no more than s bits longer than the minimal program. A second refinement, depth of a sequence relative to the depth of the length of the sequence, is required to eliminate another artefact of the definition of depth. All

depth of a system places a lower limit on how quickly the system can form from disassembled resources.³⁵ Organisation requires complex large scale correlations in the diverse local dynamics of a system. This, in turn, requires considerable high order redundancy, and a relatively lower low order redundancy. This implies a high degree of physically manifested logical depth. Whether or not organisation requires anything else is somewhat unclear right now. For present purposes, high level redundancy implied by logical depth will be a more important consideration than organisation or dynamical time, since it will be shown to explicate natural laws as described in (Collier, 1996a). A deep system is not maximally complex, because of the buried redundancy (more internally ordered than a gas), but it is not maximally ordered either, because of its surface complexity (less ordered than a crystal).

Logical depth requires correlation (redundancy), but is silent about dynamics.

³⁵ Since computation is a formal concept, while time is a dynamical concept, it isn't completely clear how we can get a dynamical measure of computation time. Generally, the minimal assembly time of a system will be less than the expected assembly time for assembly through random collisions, which we can compute from physical and chemical principles. Maximally complex systems are an exception, since they can be produced only by comparing randomly produced structures with a non-compressible template.

Causation is the Transfer of Information

No dynamical interconnections among the parts of the system are implied, because of the formal nature of the concept (which, like all purely formal concepts, ignores dynamics). Logical depth needs to be supplemented with a dynamical account of depth, within the context of NPI. How to do this is not presently entirely clear (because the dynamical grounding of logical depth requires a way to physically quantify the notion of computational time, or, equivalently, of a computational step, and how to do this properly is not clear). But when we do observe organisation we can reasonably infer that it is the result of a dynamical process that can produce depth. The most likely source of the complex connections in an organised system is an historically long dynamical process. Bennett recognised this in the following conjecture:

> A structure is deep, if it is superficially random but subtly redundant, in other words, if almost all its algorithmic probability is contributed by slow-running programs. ... A priori the most probable explanation of 'organized information' such as the sequence of bases in a naturally occurring DNA molecule is that it is the product of an extremely long biological process. (Bennett, 1985; quoted in Li and Vitányi, 1990: 238)

However we should also note that higher order redundancy could arise accidentally as an epiphenomenon (a mere correlation), but then it would not be based on a cohesive structure (cf. Collier, 1988) and so its emergence can't be controlled and it will not persist.

Entrenchment is physically embodied depth *per se*, with no direct implications concerning the historical origins of the depth. *Canalisation*, on the other hand is

sequences of n 0s are intuitively equally trivial, however the depth of each string depends on the depth of n itself. The additional depth due to sequence of 0s is small. The depth of a sequence of n 0s relative to the depth of the length of the sequence itself is always small. This relative depth correctly indicates the triviality of sequences of the same symbol.

entrenchment resulting from a deep historical process (and also describes the process). Bennett's conjecture is, then, that cases of entrenchment are, most likely, cases of canalisation. This is an empirical claim. Natural laws are usually taken to be entrenched, but not canalised. Future studies in cosmology may prove this wrong. On the information theoretic account, the same historical origin for the same forms in different systems, including law-like behaviour, is an attractive hypothesis. In any case, logical depth implies high order redundancy, whether it took a long time to form or not. This high order redundancy is a measure of organisation. Natural laws are at the maximal level (or levels, if specificity of information is not linearly ordered) that there is redundancy within a system (cf. Collier, 1996a), and are specified by this redundancy (information) and the inverse mapping function. As such, they serve as constraints on the behaviour of any system. They are thus abstractions from concrete particulars. System laws are not always true scientific laws, which must be general. This can be assured by taking as the system the physical world. This is a standard scientific practice, according to which a purported scientific law that fails under some conditions is thereby shown not to be a law after all.

The information theoretic approach to causation can be used, then, to interpret natural laws in the minimalist metaphysics described in (Collier, 1996a), according to which laws are relations between natural kinds, which are in turn the least determinate classes related by the mathematical relation required to ensure particular instances of the laws hold. These classes are defined in terms of their information, and the mathematical relation is computational consequence, ensuring necessity, given the existence of the particular informational structures (i.e.

Causation is the Transfer of Information

forms).³⁶ The laws turn out to be computationally deep, in the sense that the phenomena obeying the laws show high order redundancy, and the computation of the surface phenomena is relatively long (Collier and Hooker, submitted, also Collier, 1996b). The explication of causation, laws and counterfactuals, then, requires only logic with identity (computation theory) and particular concrete circumstances. This is the sort of metaphysics the logical empiricists were looking for, but they made the mistake of relying too heavily on natural language and phenomenal classifications (i.e., they put epistemology before ontology). Of course computation theory was poorly developed before their program was undermined by their mistakes, so they had no way to recover from those mistakes see.37

8. Information Theoretic Causation and Other Approaches

Some aspects of the information theoretic approach to causation can be clarified by comparing it with other accounts of causation. I will deal with each of the major current approaches in turn. Not surprisingly, as a minimalist approach, my approach can be interpreted as a version of each of the others with suitable additional assumptions.

8.1 The regularity approach

The regularity approach to causation is widely supported by philosophers of science, since it seems to represent well how scientists actually establish correlations and causal

³⁶ For an explanation of how this supports counterfactuals, see (Collier, 1996a) and §4 above.

³⁷ See (Collier, 1990a) for a discussion of the inadequacy of Carnap's attempt to determine the information in a proposition.

influence through controlled experiments using statistical methods (Giere, 1984). Information content (compressibility and depth) is a measure of regularity. It is more reliable than the constant conjunction approach: 1) constant conjunction fails for accidental generalisations, whereas the information transfer model does not because it requires computational necessitation, and 2) samples of chaotic systems appear irregular, thus unlawlike, but have a simple generating function that can often be recovered by computational methods.³⁸ For systems in chaotic regions, random sampling of data will give results indistinguishable from chance events, even though the generating function for the data points can be quite simple. Minor deviations in initial or boundary conditions can lead to wildly different behaviour, so experiments are not repeatable. Constant conjunction as a means to decide regularity is unworkable. Time series analysis can improve the chances of finding the generating function, especially if the basic dynamics of the system can be guessed from analogy to more tractable systems. There is still a problem of determining the dimensionality of the phase space of the system, and wrong guesses can lead investigators far astray. Testing guesses with computer models is helpful, but the mathematics of chaos ensures

that there is no reliable method for finding the generating function of a given time series: the problem is computationally intractable.

The alternative Humean approach uses counterfactuals (Lewis, 1973). This presents problems of its own. Any attempt to distinguish laws from accidental generalisations using counterfactuals without going beyond particulars by using a possible worlds ontology is plagued by the lack of a computable similarity metric across deterministically chaotic worlds. The phenomena in such worlds might as well be related by chance, since by any effective statistical procedure, their relations are chance. This objection is not telling, however, except for verificationists. There is a deeper problem for anyone who is not a verificationist, or anyone who is а metaphysical realist. It seems we can imagine two worlds, one with deterministic chaotic generators, and one produced solely by chance, which are nonetheless identical in the form of all their particulars. Either these worlds are distinguished by the separate existence of laws, which undermines the reason for inferring possible worlds, or else the two worlds must be the same. This latter assumption seems to me to be arbitrarily verificationist. especially given that unrestricted verificationism directly undermines the possible worlds approach, and is also contrary to metaphysical realism. If one is willing to swallow these consequences, then I can see no objection. The same argument can be applied to chance and nonchance worlds that are not chaotic, which is perhaps more telling.

These problems are also telling against any attempt to reduce causation to probability, since it is question begging to infer common cause from probability considerations if causation is defined in terms of probality relations. Probability considerations alone cannot distinguish between a world with chance regularities and one in which the regularlites are caused. On

³⁸ It is worth noting that our solar system, the epitome of regularity in classical physics, is stable for only relatively short periods. Over longer periods it is difficult to predict its evolution (physicist Philip Morrison calls this the "Poincaré Shuffle". In a world with infinite time, it is mathematically impossible to predict the evolution of the solar system. On the other hand, dissipative processes like tidal dissipation probably explain the regularity that we observe in the solar system. A world in which all processes are in a chaotic regime would need to lack such dissipative processes that produce regularity.

the informational approach, there are no such problems. The metaphysical distinction between chance correlations and causal correlations depends on the identity of information. The sensible hypothesis, on any reliable evidence that there is a possibility that the world is not a chance world, would be that the world has natural causal laws that preserve information and determine the probabilities. But that hypothesis could be wrong.

8.2 Universals and Natural Kinds

A distinction might be considered to be the only required universal. I see no great advantage in making distinction a universal, but it is certainly a natural kind in the sense of (Collier, 1996a). The issue of whether or not it is a universal ultimately depends on the ontological status of mathematical objects. Other natural kinds are forms that can be analysed in terms of their informational structure, perhaps in terms of their distinctions from other natural kinds. In any case, the information theoretic approach to causation does not *need* to invoke either universals or natural kinds except as abstractions from particular systems and their particular properties. If mathematical objects are universals, then so are natural kinds, but so are a lot of other forms as well. Invoking universals seems to have no additional explanatory value in the case of causation, since all possible distinctions are already presupposed by the information theoretic account.39

8.3 The conserved quantity approach

Causation is the Transfer of Information

Salmon has dropped the mark approach, and has adopted Dowe's conserved quantity approach (Dowe, 1992; Salmon, 1994). The idea is that causal processes, unlike pseudocausal processes (like the spot of a flashlight crossing the face of the moon), involve a non-zero conserved quantities. The problem with this approach is that it does not allow for causation in dissipative systems, which predominate in this world. It is possible that dissipative processes can be reduced to underlying conservative approaches, but how to do this is not presently known (Sklar, 1986, 1993: 297ff). Energy and momentum are conserved in dissipative processes in our world (so far as we know). Nevertheless, it seems to be possible to have physical systems in which not even energy and momentum are conserved (e.g. through the spontaneous disappearance of energy/momentum from the physical world as it dissipates, or through the appearance of matter through Hoyle's empirically refuted but seemingly possible continuous creation). Dissipative systems of this sort still preserve information, even if they do not conserve the quantity of information or any other non-zero physical quantity.

The information approach and the conserved quantity approach are equivalent in conservative causal processes and in causal interactions because conservation laws are equivalent to informational symmetries (cf. Collier, 1996b and references therein). In any conservative process, no information is lost, and no information is gained. However, the quantity of information is not necessarily conserved in causal processes (by definition it is not conserved in dissipative processes), though some information is preserved in any causal process. I suppose that focusing on the preserved information in the process as the conserved quantity might make the two approaches identical, but this seems a bit stretched to me. In our world, there are conserved quantities in all causal processes

³⁹ There is an ingenious but not entirely convincing argument by Russell that nominalists are committed to at least one universal, similarity. I take it that all distinctions are particular, and depend only on the existence of distinct particulars.

(energy/momentum and charge/parity/time, so far as we know), but the information theoretic approach will work in worlds in which there are no conservation laws as well, if such worlds are indeed possible. At the very least, the information theoretic approach and the conserved quantity approach mutually support each other in worlds in which all causal processes involve some conserved quantity. The main practical advantages that I see for the informational approach is that it explains the longstanding importance of form in accounts of causation, and it does not rule out dissipative worlds. Theoretically, the approach gives a deeper insight into the nature of conservation as a form of symmetry.

8.4 Humean supervenience

Humean supervenience requires that causation, natural kinds and natural laws are supervenient on particulars. It is satisfied trivially by the information theoretic approach that I have proposed. All that is required for the my general account is the particular information in particular things and their computational relations, and a natural way to specify the information in things that is epistemologically accessible. For physical things, NPI provides the last condition. Although NPI cannot be completely specified right now, if ever, the practices of scientists and engineers allow us to use information as unambiguously as any concept in science. I cannot say exactly what NPI means, but I can show how it is used. Unfortunately, it is my experience that the learning process takes at least many weeks at least. It takes much longer if the student keeps asking for explanations in English. Information theory is the most precise language we can have. Asking for a clearer natural language explanation is pointless. To paraphrase Wittgenstein, we must learn by practice what cannot be said.

Causation is the Transfer of Information 9. Conclusion

The identification of causation with information transfer permits a minimalist metaphysics using only computational logic and the identity through time of contingent particulars. It also helps to account for the persistence of causal explanations involving form from the beginnings of science to the present. It needs no possible worlds or universals, so it is ontologically parsimonious. Necessitation arises naturally from the computational basis of the approach: causal processes can be thought of as analogue computations that can, when recursively definable, be mapped onto digital computations for tractability. There are some epistemological difficulties with the approach, but it shares these difficulties with more elaborate approaches. The more elaborate approaches can be seen as special cases of the information theoretic approach involving at least one further methodological or empirical assumption, so it can be seen as the core of the other current approaches. Furthermore, it is immediately compatible with modern computational techniques, and be applied directly using thus can conventional methods that have been developed by physicists. NPI and the quantitative methods of Statistical Mechanics permit the quantification of the strength of causal interactions. I doubt these advantages can be gained by any other philosophical approach to causation.

Acknowledgements

This work was undertaken with the generous support of Cliff Hooker and the Complex Dynamical Systems Group at the University of Newcastle, financially, intellectually and emotionally. I would also like to thank Howard Sankey for his encouragement in producing this chapter. Penetrating questions by David Armstrong,

Neil Thompson, John Clendinnen and Tim O'Meara, who were present at the first public presentation of the ideas in this chapter have greatly improved the final version. Comments by Gad Yagil on some ideas contained in this chapter helped me to clarify some central issues concerning NPI. Ric Arthur and Jonathan D.H. Smith made helpful suggestions on historical background and mathematical details, respectively. A very audience interactive version of this paper given at the Department of Philosophy at Newcastle University helped me to clear up three central sticking points of an early version of this chapter. John Wright's acute questions were especially helpful. Malcolm Forster and Steve Savitt made some useful suggestions on presentation. Finally, I would like to thank two anonymous referees, some of whose suggestions I adopted. Other of their suggestions helped me to see that I had not expressed my line of argument concerning the significance of form as clearly as I might have. I hope that the present version dissolves the basis of these suggestions.

Bibliography

- Armstrong, David (1983). What is a Law of Nature? Cambridge: Cambridge University Press.
- Atkins, P. W. 1994. *The Second Law: Energy, Chaos, and Form.* New York: Scientific American Library.
- Banaschewski, B. (1977). On G. Spencer Brown's Laws of Form, *Notre Dame Journal of Formal Logic* **18**: 507-509.
- Bell, John L. and William Demopoulos (1996). Elementary Propositions and Independence, *Notre Dame Journal of Formal Logic* **37**: 112-124.
- Bennett, C.H. (1982). The Thermodynamics of Computation: A Review, International Review of Theoretical Physics 21: 905-940. Reprinted in

Causation is the Transfer of Information

Lef and Rex (eds) *Maxwell's Demon*.

Bennett, C. H. (1985). Dissipation, Information, Computational Complexity and the Definition of Organization, in D. Pines (ed.), *Emerging Syntheses In Science. Proceedings of the Founding Workshops of the Santa Fe Institute*: 297-313.

Bennett, C.H. (1987). Demons, Engines and the Second Law, *Scientific American* **257**, no. 5: 108-116.

Bennett, C.H. (1988). Notes on the History of Reversible Computation, *IBM Journal of Research and Development* **32**: 16-23. Reprinted in Lef and Rex (eds) *Maxwell's Demon*.

- Bohm, David (1980). *Wholeness and the Implicate Order*. London: Routledge & Kegan Paul.
- Brillouin, L. (1962). Science and Information Theory, second edition. New York: Academic Press.
- Brooks, Daniel R. and Edward O. Wiley (1988). Evolution as Entropy: Toward a Unified Theory of Biology, 2nd edition. Chicago: University of Chicago Press.
- Brooks, D.R., E.O. Wiley and John Collier (1986). Definitions of Terms and the Essence of Theories: A Rejoinder to Wicken, *Systematic Zoology* **35**: 640-647.
- Christensen, W.D., Collier, John and Hooker, C.A. (in preparation). Autonomy, Adaptiveness and Anticipation: Towards Foundations for Life and Intelligence in Complex, Adaptive, Self-organising Systems.
- Collier, John (1988). Supervenience and Reduction in Biological Hierarchies, in M. Matthen and B. Linsky (eds) *Philosophy and Biology: Canadian*

Journal of Philosophy Supplementary Volume 14: 209-234. Collier, John. (1990a). Intrinsic Information, in Philip Hanson (ed) Information, Language and

Cognition: Vancouver Studies in Cognitive Science, Vol. 1. Oxford: University of Oxford Press: 390-409.

Collier, John D. (1990b). Two Faces of Maxwell's Demon Reveal the Nature of Irreversibility. *Studies in the History and Philosophy of Science* **21**: 257-268.

- Collier, John (1993). Out of Equilibrium: New Approaches to Biological and Social Change. *Biology and Philosophy* **8**: 445-456.
- Collier, John (1996a). On the Necessity of Natural Kinds, in Peter Riggs (ed) *Natural Kinds, Laws of Nature and Scientific Reasoning*. Dordrecht: Kluwer: 1-10.

Collier, John. (1996b). Information Originates in Symmetry Breaking" *Symmetry: Culture and Science* **7**: 247-56.

Collier, John, S. Banerjee and Len Dyck (in press). A Non-equilibrium Perspective Linking Development and Evolution, in John Collier and Douglas Siege Causey (eds) *Between Order and Chaos: Studies in Non-Equilibrium Biology*.

Collier, John, E. O. Wiley and D.R. Brooks (in press). Bridging the Gap Between Pattern and Process, in John Collier and Douglas Siege Causey (eds) *Between Order and Chaos: Studies in Non-Equilibrium Biology*.

Collier, John and CA Hooker (submitted). Complexly Organised Dynamical Systems.

Collier, John and Scott Muller (in preparation). Emergence in Natural Hierarchies.

Causation is the Transfer of Information

Collier, John and Douglas Siegel-Causey (in press). *Between Order and Chaos:* Studies in Non-Equilibrium Biology.

Denbigh, K.G. and J.S. Denbigh (1985). *Entropy in Relation to Incomplete Knowledge*. Cambridge: Cambridge University Press.

Dowe, P. (1992). Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science* **59**: 195-216.

Fodor, Jerry A. (1968). *Psychological Explanation; An Introduction to the Philosophy of Psychology*. New York: Random House.

Feynman, Richard P. (1965). *the Character of Physical Law*. Cambridge, MA: MIT Press.

Fredkin E. and T. Toffoli (1982). International Journal of Theoretical Physics **21**: 219.

Gale, George (1994). The Physical Theory of Leibniz, in Roger Woolhouse (ed) , G. W. Leibniz: Critical Assessments. London: Routledge & Kegan Paul: 227-239.

Gatlin, Lyla L. (1972). Information Theory and the Living System. Columbia University Press, New York.

Giere, Ronald N. (1984). Understanding Scientific Reasoning, 2nd ed. New York: Holt, Rinehart, and Winston.

Goldstein, Herbert (1980). *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley.

Graves, John Cowperthwaite (1971). The Conceptual Foundations of Contemporary Relativity Theory. Cambridge, MA: MIT Press.

Hobbes, Thomas (1839). *Collected Works*, *Volume 1*. William Molesworth (ed). London: John Bohn.

Holzmüller, Werner (1984). Information in Biological Systems: The Role of Macromolecules, translated by Manfred Hecker. Cambridge: Cambridge University Press.

- Horwich, Paul (1988). Asymmetries in Time. Cambridge, MA: MIT Press.
- Kestin, Joseph (1968). A Course in Thermodynamics. Waltham, MA: Blaisdell.
- Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World, in P. Kitcher and W.C. Salmon (eds) *Minnesota Studies in the Philosophy of Science, Vol. 13 Scientific Explanation*. Minneapolis: University of Minnesota Press: 410-505.
- Kolmogorov, A.N. (1965). Three Approaches to the Quantitative Definition of Information. *Problems* of Inform. Transmission 1: 1-7.
- Kolmogorov, A.N. (1968). Logical Basis for Information Theory and Probability Theory. *IEEE Transactions on Information Theory* **14**: 662-664.
- Küppers, Bernd-Olaf (1990). *Information and the Origin of Life*. Cambridge: MIT Press.
- Landauer, Rolf (1961). Irreversibility and Heat Generation in the Computing Process. *IBM J. Res. Dev.* **5**: 183-191. Reprinted in Lef and Rex (eds) *Maxwell's Demon.*
- Landauer, Rolf (1987). Computation: A Fundamental Physical View. *Phys. Scr.* **35**: 88-95. Reprinted in Lef and Rex (eds) *Maxwell's Demon*.
- Landsberg, P.T. (1984). Can Entropy and 'Order' Increase Together? *Physics Letters* 102A: 171-173.
- Layzer, D. (1975). the Arrow of Time. Scientific American 233: 56-69.
- Layzer, David (1990). Cosmogenesis: the Growth of Order in the Universe. New York: Oxford University Press.
- Lef, Harvey S. and Andrew F. Rex (1990). *Maxwell's Demon: Entropy, Information, Computing.* Princeton: Princeton University Press.

Causation is the Transfer of Information

Leibniz, W.G. (1969). *The Yale Leibniz*, translated by G.H.R. Parkinson. New Haven: Yale University Press.

- Li, Ming and Paul Vitànyi (1990). Kolmogorov Complexity and its Applications, in *Handbook of Theoretical Computer Science*, edited by J. van Leeuwen. Dordrecht: Elsevier.
- Li, Ming and Paul Vitànyi (1993). An Introduction to Kolmogorov Complexity and its Applications, 2nd edition. New York: Springer-Verlag.
- Lewis, David (1973). Causation. *Journal of Philosophy* **70**: 556-67.
- Lewis, David (1994). Chance and Credence: Humean Supervenience Debugged. *Mind* **103**: 473-90.
- Mach, Ernst (1960). *The Science of Mechanics*. Lasalle: Open Court.
- Rissanen, Jorma (1989). *Stochastic Complexity in Statistical Inquiry*. Teaneck, NJ: World Scientific.
- Russell, Bertrand (1913). On the Notion of Cause. *Proceedings of the Aristotelian Society, New Series*, **13**: 1-26.
- Salmon, Wesley C. (1984). Scientific Explanation and the Causal Structure of the World. Princeton: Princeton University Press.
- Salmon, Wesley C. (1994). Causality Without Counterfactuals. *Philosophy* of Science **61**: 297-312.
- Schneider, T.S. (1995). An Equation for the Second Law of Thermodynamics. Word Wide Web URL: <http://www-lmmb.ncifcrf.gov/~tom s/paper/secondlaw/index.html>
- Schrödinger, Irwin, (1944). What is Life?, reprinted in What is Life? And Mind and Matter. Cambridge: Cambridge University Press.
- Shannon, C.E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

- Sklar, Larry (1986). The Elusive Object of Desire, in Arthur Fine and Peter Machamer (eds) PSA 1986: Proceedings of the 1986 Bienneial Meeting of the Philosophy of Science Association, volume 2. East Lansing: Philosophy of Science Association: 209-225, reprinted in Steven F. Savitt (ed) Time's Arrows Today Cambridge: Cambridge University Press: 209-225.
- Sklar, Larry (1993). *Physics and Chance*. Cambridge: Cambridge University Press.
- Spencer-Brown, G. (1969). *Laws of Form*. London: Allen & Unwin.
- Thom, René (1975). *Structural Stability and Morphogenesis*. Reading, MA: W.A. Benjamin.
- Thompson, D'Arcy Wentworth (1942). On Growth and Form, 2nd ed. Cambridge: Cambridge University Press.
- Ulanowicz, R.E. (1986). Growth and Development: Ecosystems Phenomenology. New York: Springer Verlag.
- Wallace, C.S. and P.R. Freeman (1987).
 Estimation and Inference by Compact Coding. *Journal of the Royal Statistical Society, Series B, Methodology* 49: 240-265.
- Wicken, Jeffrey S. (1987). Evolution, Thermodynamics and Information: Extending the Darwinian Paradigm. New York: Oxford University Press.
- Wiley, E.O. (1981). *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York: Wiley-Interscience.
- Wittgenstein, Ludwig (1961). *Tractatus Logico-Philosophicus*, translated by D.F. Pears and B.F McGuiness. London: Routledge & Kegan Paul.
- Yagil, Gad (1993a). Complexity Analysis of a Protein Molecule, in J. Demongeot, and V. Capesso (eds)

Causation is the Transfer of Information

Mathematics Applied to Biology and Medicine. Winnipeg: Wuerz Publishing: 305-313.

- Yagil, Gad (1993b). On the Structural Complexity of Templated Systems, in L. Nadel and D. Stein (eds) 1992 Lectures in Complex Systems. Reading, MA: Addison-Wesley.
- Yagil, Gad (1995). Complexity Analysis of a Self-Organizing vs. a Template-Directed System, in F. Moran, A. Moreno, J.J. Morleo, and P. Chacón (eds.) Advances in Artificial Life. New York: Springer: 179-187.