Intrinsic Information

John D. Collier
Department of Philosophy
University of Calgary
Calgary, Alberta T2N 1N4
February 25, 1988
(revised September 14, 1988
and February 22, 1989)

In P. P. Hanson (ed) *Information, Language and Cognition: Vancouver Studies in Cognitive Science, Vol. 1* (originally University of British Columbia Press, now Oxford University Press, 1990): 390-409

1. Introduction

In everyday usage, information is knowledge or facts acquired or derived from study, instruction or observation. Information is presumed to be both meaningful and veridical, and to have some appropriate connection to its object. Information might be misleading, but it can never be false. Standard information theory, on the other hand, as developed for communications (Shannon and Weaver, 1949), measurement (Brillouin, 1962) and computation (Solomonoff, 1964; Kolmogorov, 1968; Chaitin, 1975), entirely ignores the semantic aspects of information. Thus it might seem to have little relevance to our common notion of information. This is especially true considering the range of applications of information theory found in the literature of a variety of fields. Assuming, however, that the mind works computationally and can get information about things via physical channels, then technical accounts of information strongly restrict any plausible account of the vulgar notion. Some recent information-oriented approaches to epistemology and semantics go further.

The usual sense of information is intentional: information is a property of representations that are meaningful to some subject. Systematic studies in epistemology (Dretske, 1981) and semantics (Barwise and Perry, 1983), though, suggest that information can be conveyed from the world to our minds. If so, concrete situations in the world can bear information. Despite this, we don't normally think of concrete situations as either representational or intrinsically meaningful. Dretske tries to resolve this problem by basing information on causal laws, which he claims *are* intentional. His claim is dubious unless the concept of intentionality is stretched beyond the limits of cognitive penetrability. Barwise and Perry also connect information to causality, holding that meaning exists in the world of inanimate objects, as well as in our thoughts and ideas. Whether or not the move to place meaning out in the world is accepted, if we assume even a modest realism there must be some property of objects that allows us to have information about them. This property must be causally based and causally communicable to us, as well as being commensurate with information in the vulgar sense. This paper explores the nature of this property.

I argue below that even if Dretske, Barwise and Perry are right in thinking that meaning is out in the world, the property of interest must be characterizable in non-intentional terms. In the following sections I largely ignore questions of meaning, focusing only on non-intentional aspects of information. I will use the resources of contemporary information theory to sketch an account of information intrinsic to external objects. My project is modest. I wish to specify the characteristics of things in the world that allow them to be objects of representations.

There are reasons to suppose the concept of information can be usefully extended to the non-intentional world. First, cognitive systems are physical (and also biological). This is not analytic, but it is a "deep" fact about our world. Whatever allows meaningful information at the cognitive level is constructed from physical resources. Physical reality constrains the way cognitive systems can work, including how they can process and interpret information. It is possible that there are no interesting correlates of cognitive information outside of representations, but I believe this is empirically false.

The second consideration is less direct. Any explanation of the veridicality of our representations must appeal to a connection between them and the world. This connection, if knowledge is not purely coincidental, must be fairly regular and predictable. The only plausible candidate is some sort of causal relation. The connection need not be direct, since indirect or common causes would serve as well for information transmission (Dretske, 1981: 38). The world must contain either information itself, or else something that when properly connected to our cognitive processes is converted into information. This something is transmitted by physical and biological means; received information at the cognitive level must interface in a law-like manner with the transmissions.

I am going to stipulate that what is transmitted *is* information, irrespective of whether there is a cognitive receiver. This notion of information is more general than the common one, containing it as a species. It might be less confusing to use a new term for this broader notion, but technical usage has already extended information to the non-intentional realm. This move has produced insights into the interpretation of randomness and probability theory (Kolmogorov, 1965, 1968; Chaitin, 1975), as well as to computational complexity and the thermodynamics of computation (previous references, and Bennett and Landauer, 1985; Bennett, 1982, 1987). The extension of the concept is both relatively harmless and highly productive. Whether this indicates some deep truth about the world, I leave to the reader.

2. Information Viewed from the Top and from the Bottom

There are two approaches to understanding information which might be loosely labelled "top-down" and "bottom-up". The top-down approach starts with statements or other representations for which intentionality is taken for granted, and tries to specify what determines their information content. This approach has both formal and "informal" versions. The bottom-up approach starts with an account of the information content of concrete objects and works up to intentionality and beliefs. I will argue that the purely formal top-down approach fails because there is no unequivocal way to assign empirically relevant *a priori* probabilities to statements. The informal top-down approach is more promising, but needs to be supplemented by the bottom-up approach. A pure bottom-up approach is likely possible (evolution seems to have accomplished it), but impractical at present.

An early formal version of the top-down approach is the Carnap/Bar-Hillel account of semantic information (Bar-Hillel, 1964). They use the resources of inductive logic to define the information content of a statement in a given language in terms of the possible states it rules out. For "technical reasons" they calculate the states ruled out as a number of *state descriptions*. A state description is a conjunction of atomic statements assigning each primitive monadic predicate or its negation (but never both) to each individual constant of the language. The information content of a statement is thus relative to a language. Evidence, in the form of *observation* *statements*, contains information in virtue of the class of state descriptions the evidence rules out. (They assumed that observation statements can be connected to experience unambiguously.) Information content, then, is inversely related to probability, as intuition would suggest.

It turns out, though, that our pre-systematic intuitions confuse two different measures of information content, both of which have plausible but incompatible properties. The first measure of the information content of statement i is called the *content measure*, cont(i). It is defined as the complement of the *a priori* probability that i is true:

$$\mathrm{cont}(i) = 1 - p(i) \qquad [1]$$

This measure fails the *additivity condition*, according to which the combined information content of two inductively independent statements[1] should be the sum of their individual information contents (Bar-Hillel, 1964: 302). It also fails some natural assumptions about conditional information. These problems motivated the introduction of another measure, called the *information measure*, inf(i):

$$\mathrm{inf}(i) = \log_2 (1/(1 - \mathrm{cont}(i))) = -\log_2 p(i) \qquad [2]$$

The value of this measure is in bits. Although inf satisfies additivity and conditionalization requirements, it has a property that some people find counter-intuitive. If some evidence e is negatively relevant to a statement i, then the information measure of i conditional on e will be greater than the absolute information measure of i. This violates a common intuition that the information of i given e must be less than or equal to the absolute information of i. The content measure, cont(i), does satisfy this intuition (Bar-Hillel, 1964: 306-7). Personally, I do not share this widespread intuition since it requires effort to correct the inference based on e that i is less likely. The issue requires further study, but is not relevant in what follows.

Elegant though the Carnap/Bar-Hillel account may be, it has problems. Because they use methods restricted to applied first order languages with identity, Carnap and Bar-Hillel cannot deal with complex scientific predicates like mass, temperature and energy. This could perhaps be corrected with a more sophisticated inductive logic using a more holistic view of theories. Any move towards a more holistic view, however, undermines the use of basic sentences to determine *a priori* probabilities. The probabilities of individual sentences of a theory become dependent on the overall *a priori* probability of the theory. At the least, this will greatly complicate calculations of information content. The demise of the analytic/synthetic distinction and the implausibility of pure observation statements presents a similar but even more fundamental problem.

This problem arises from what has misleadingly been called the theory-ladenness of observation. Observation statements, if they are not purely demonstrative, assign predicates to evidence. This amounts to a classification of the evidence. Individual bits of evidence are particulars that don't carry with them the rules for their own classification. If the information content of a statement is information about the world, not an artifact of the way evidence is classified, the assignment of predicates to experiences must reflect some sort of regularity in the evidence itself. Any non-arbitrary classification of evidence under some predicate presumes its similarity (at least) to other evidence or possible evidence (see Collier, 1987). Barring astonishingly good luck, this presumes some information about appropriate similarities. Thus

---

[1] Inductive independence means that the conditional probability of each statement given the other is the same as its initial probability.

knowledge is presupposed in determining the information content of a piece of evidence; the information is not determined purely by syntactic relations. Something about the evidence itself, the workings of the mind, or both, must determine this knowledge. Since none of these are purely formal, the formal top-down approach cannot work.

The informal top-down approach resembles the structuralist approach to scientific theories, which uses informal model theory and empirical constraints to determine acceptable interpretations of scientific theories (see Stegmüller, 1979 for a review). The approach starts with meaningful representations and tries to specify their interpretation, making use of available empirical constraints. Barwise and Perry (1983) are engaged in this project. On their view, the interpretation of a representation is given in terms of the information it conveys. Unlike the formal top-down approach, in which information content is determined entirely by the structure of language (or other representational system), information in this approach is the content (or factual content) of a representation (or information-report) (Israel and Perry, this volume).

The goal of this approach is to connect meaningful representations to the concrete situations represented. Perry and Barwise base this connection on nomic regularities, which they call "constraints" (Israel and Perry, this volume). Information is conveyed to us by causal chains connecting situations in a lawful way. The information indicated by a situation is relative to the causal chains connecting the indicating situation both to our beliefs and to the situation the information is about. Thus, "The information a factual state of affairs carries is relative to a constraint". Complete determination of the reference of a representation (at least in cases involving indexical terms) also requires specific circumstances. The information content of a representation available to us is delimited by our knowledge of these constraints and circumstances. Further evidence could modify this available information, and it may be that complete knowledge would determine the absolute information content of a representation.

There is something "out there in the world" that can be transmitted to intelligent beings who can understand the information it contains, and pass it around among themselves. What the transmitted something is called at various stages along the route is not all that important. Either this something is pre-existing meaningful information or else it can be converted by cognition into meaningful information. If conversion is necessary, since the content of representations is characterized in intentional language, whereas the referents of veridical representations don't require intentional characterization, an account of reference should characterize referents in non-intentional terms commensurate with the characterization of representational content, otherwise there will be a gap. Barwise and Perry (1983: 94) held that meaning pre-exists in the world, hoping thereby to avoid the gap. Their choice, though, is question-begging unless there is a characterization of things out in the world that is commensurate with intentional language. Normally, things in the external world are not described in intentional terms. The Barwise/Perry approach needs an information-theoretic account of nomic regularities and causal interactions, and of the transmission of the information these nomic regularities and causal interactions contain. This account will be largely empirical. If nothing else, factual constraints are the only way to rule out clever but fanciful exceptions. The selection of an appropriate characterization, though, is a philosophical matter.

Superficially, Dretske's (1981) approach to information resembles the Carnap/Bar-Hillel approach. He also defines the information content of a piece of evidence in terms of the cases ruled out. A major difference is that Dretske does not try to specify representations purely syntactically. Rather than calculating the information content of statements, he uses states of affairs directly. His measure of information is similar to the inf definition (Dretske, 1981: 52):

$$I(s) = -\log_2 p(s), \quad \text{in bits,} \qquad [3]$$

where p(s) is the probability of the state of affairs s.

From the examples and discussion of the first part of his book, Dretske appears to adopt a variety of the informal top-down approach. He describes perceived situations that rule out possibilities until only one relevant possibility is left (e.g., 1981: 4-6, 48, 53, 78, 95). Dretske's treatment of intentionality in this part of the book supports this interpretation: "The ultimate source of intentionality inherent in the transmission and receipt of information is, of course, the *nomic regularities* on which the transmission of information depends" (1981: 76). This is similar to Barwise and Perry's placement of meaning in the world. Although it is doubtful that the presumed intentionality of laws[2] satisfies Dretske's later characterization of intentionality (Dretske, 1981: 172-173), there is no need to pursue the matter here. Even if we accept that natural laws are intentional, the concrete situations that manifest them don't need intentional characterization, and they are not normally characterized intentionally. The gap in the Barwise/Perry treatment exists for Dretske as well.

Dretske's definition of information in terms of the cases ruled out might seem to fall afoul of the problem of background knowledge that plagues the Carnap/Bar-Hillel approach. Dretske admits that information received, as commonly understood, is relative to background knowledge (1981: 80-81), but argues (1981: 87) that the background knowledge can be treated in the same way as the received knowledge. This process is iterated until all background knowledge is accounted for. The problem with this solution is that prior knowledge of what possibilities need to be eliminated still seems to be necessary. Dretske meets this objection by turning in the second part of his book to the bottom-up approach.[3]

The idea behind the bottom-up approach is that concrete situations are sources of information that can be transmitted through channels that don't themselves generate new information (Dretske, 1981: 115; see also below, section 4). Information is transmitted (perhaps indirectly) from structure to structure according to causal laws. If it is transmitted to a structure with the right order of intentionality[4] there is knowledge. The causal processes producing beliefs eliminate other possibilities from consideration; what is important is not that we have reasons for rejecting other possibilities, but that the correct causal connections are made. This approach is bottom-up because we start with information out in the world, and ask what makes it into the content of a belief.

---

[2] Dretske argues from the modal quality of laws.

[3] Remnants of the top-down approach of the first part continue to appear throughout his book.

[4] See (Dretske, 1981: 171-189) for a discussion of how higher orders of intentionality are formed from structures of lower order.

There is a lot of information in the world. Much of it isn't very interesting to creatures like us. Some information is interesting and useful, though, and we are attuned to picking it up (see Israel and Perry, this volume). As long as there is a reliable causal connection (possibly indirect) between the source of the information and the higher order intentional structures Dretske describes, we have knowledge. We may have sceptical doubts about the efficacy of the channels involved, leading to doubts about the nature of the source, but these doubts are irrelevant to whether or not we really have knowledge. The bottom-up approach allows us to have knowledge without requiring justification. We do not need to eliminate all possibilities other than the correct one, or to know what the alternative possibilities are. The correct causal connections through reliable channels do all this work for us.

The bottom-up approach disarms sceptical doubts as an objection against the possibility of knowledge. This gain has its cost, though. Our pre-systematic intuitions about information are not suited to the bottom-up approach. We can no longer think of information in terms of the reduction of pre-determined possibilities. The reduction is the *effect* of the receipt of information on us. The information source must have some intrinsic property that produces this effect. The whole approach is somewhat fanciful if the source does not really contain information (or some equivalent) capable of having this effect. The bottom-up approach needs a definition of intrinsic information content.

3. Intrinsic Information

Physical things have properties that give them a definite structure and causal capabilities.[5] If information is an intrinsic property of physical objects, then it seems likely that it is contained in their physical structure. Brillouin (1962: 152) distinguished two types of information: free and bound. The two forms of information differ in how they are regarded; Brillouin did not rule out the possibility that they might refer to the same thing. Free information depends on possible cases which are regarded as abstract and have no specific physical significance. Bound information depends on possible cases which are the complexions[6] of a physical system; it is a special case of free information. The information of a concrete situation, then, if it is either of these, is bound information.

On Brillouin's account, the bound information of a macroscopic system is inversely related to its entropy and directly related to its negentropy, the difference between the maximal entropy of the system and its actual entropy. Negentropy and information are reversibly interconvertible, and Brillouin may have considered them to be identical. A measurement, which produces information, must also produce negentropy, which must, according to the second law of thermodynamics, be produced at the expense of producing entropy someplace else. The duplication of information requires the conversion of available energy into entropy someplace in

---

[5] I will leave the question of whether it is possible for the same thing to have different causal capabilities under different natural laws open, since this question is irrelevant to intrinsic information content.

[6] The complexions of a system are the possible microscopic states it can be in, given its macrostate.

the system or its surroundings. Likewise, the transfer of information must be compensated for by the production of entropy in the system from which it is transferred. If a measured system is open, its entropy can remain constant (effectively duplicating the information transmitted), but the entropy of its surroundings must increase.

Landsberg (1984), following Layzer (1975), defined the order of a system in terms of the difference between its actual entropy and its maximal entropy:

$$\text{Order} = S_{max} - S_{act} \qquad [4]$$

where

$$S_{max} = k\log P, \qquad [5]$$

and

$$S_{act} = -k\sum p(m_i)\log(p(m_i)), \qquad [6]$$

where $p(m_i)$ is the probability of the i^{th} microstate, and P is the number of elementary complexions of the system. The constant k can be eliminated by choosing base 2 for the logarithms. This gives the entropy in dimensionless entropy units (Brillouin, 1962: 118). Order, in this case, has the same form as the inf definition of information content. S_{max} is reached when all complexions are equally likely. This is the equilibrium condition. The order is non-zero only if the system is not in thermodynamic equilibrium, i.e., if some elementary complexions are more probable than others.

The order is sometimes called the "intropy" of the system (in contrast to its entropy). It is well defined only if the actual entropy S_{act} and the equilibrium entropy S_{max} are well defined. Both entropies are well defined only if the elementary complexions (the possible microstates) of the system are well defined. This requires that the macrostate be well-defined, or else it is not clear which microstates should be included as possible. As in the top-down approach, information in the statistical version of the bottom-up approach requires reference to possibilities. In this case, however, we do not need to know what the possibilities are; they are determined by the physical circumstances that prevail.

Although many writers have assumed that macrostates are subjective, I doubt they are right. Irreversibility is a consequence of the second law of thermodynamics under non-equilibrium conditions. The explanation of the second law is in terms of statistical mechanics, which assumes the objectivity of macrostates. Thus irreversible chemical reactions depend (factually) on the objectivity of macroscopic phenomena. The existence of our bodies depends on such irreversible reactions; its maintenance in the living state depends on its not being at thermodynamic equilibrium. The modest assumption that the mind depends on the body, and that subjectivity depends on the mind, leads immediately to the conclusion that if macrostates are subjective, then a precondition for the mind's existence is its own existence. This is patently absurd, as Prigogine (Prigogine and Stengers, 1984) has pointed out. One of these assumptions must go.

I maintain that macrostates have a mind-independent existence. I have argued elsewhere (Collier, 1988b) that the basis of the objectivity of macrostates is *cohesion*.[7] Cohesion is produced by the causal interactions among the parts of a system that make it insensitive to fluctuations in

---

[7] It is possible to imagine non-cohesive macrostates. These are artificial, and have no macroscopic causal properties. An example would be the state of the gas molecules in an arbitrary unbounded one metre sphere.

its microstates. This insensitivity is not absolute; large enough perturbations can destroy a cohesive macroscopic system. Nonetheless, if cohesion is strong enough, there is a causal basis for macroscopic entities and their macroscopic properties. Cohesion keeps the macrostate stable while allowing minor external influences to change the microstate unpredictably, given the macrostate. These external influences are empirically likely, and justify a statistical treatment of microstates (Prigogine, 1962: 265-269), in turn justifying the use of probabilities in equation [6] above.

Cohesion acts as a constraint on the range of possible states of the system, determining not only the objectivity of the macrostate, but also the value of S_{max} for the system. Thus the basis of the objectivity of entropy is also the basis of the objectivity of intropy. Cohesive entities at one level can serve as the basis for the complexions at higher levels, allowing application of the statistical methods of modern thermodynamics to very large scale entities (e.g., Ulanowicz, 1986; Brooks and Wiley, 1986, 1988; Collier, 1986, 1988b). This allows the notion of intropy and its related information to be applied to a wide range of systems. Since cohesion guarantees the irrelevance of lower level fluctuations to the macrostate (Collier, 1988b), the intropy of a cohesive system depends only on the states of the next lower cohesive level. Fluctuations at still lower levels are screened off by the cohesion of the elements at this level. This dramatically simplifies the calculation of intropy. For example, to determine the information in a gas we don't need to consider the nuclear microstates of its atoms.

Many situations are not cohesive, since they lack the necessary internal causal connections. These non-cohesive situations are epiphenomenal.[8] Epiphenomenal situations are either part of a larger cohesive situation that provides their causal basis, or are non-cohesive combinations of cohesive parts. In the first case, the intropy of the situation is determined by how it alters the probability of the microstates of the inclusive situation from the equilibrium expectation, i.e., according to equation [6]. Equation [4] then gives the intropy of the situation. The intropy of the partial situation will generally be lower than the intropy of the inclusive cohesive situation. In the second case, the intropy of the situation is the sum of the intropies of its cohesive parts. (Any apparent order of the parts has no macroscopic effect, since by assumption there is no cohesion amongst the parts, therefore this order has no macroscopic effects.) In both cases, the intropy of the situation is due entirely to the factual conditions that make it up. This is as required for a physical basis for the information of concrete situations.

Unfortunately, things are not quite this simple. The intropy is not the only part of the system that bears information. The causal relations producing cohesion also contain information. (This information is called *constraints* in information theory. It is not internal to the system in a special technical sense. See Collier, 1986 for a discussion of internal information and constraints). For many purposes, like analysis of the self-organization and evolution of systems, the internal information is all that need be considered in order to understand system behaviour. For perceptual and semantic purposes, however, we must presume an external perspective. The constraints on the system are part of its complete description.

---

[8] Often our interactions with situations will create a larger system with closed loops which can generate cohesion, as when we sort a deck of cards into suits, but this is certainly not necessary for something to be a situation.

Some examples will be helpful here. The statistical measure of information (intropy) applies best to the idealized case of an ideal gas made up of non-interacting point sized molecules. The intropy ranges from zero to some maximal value depending solely on the position and momentum distribution of the molecules. The intropy is equivalent to the ability of the system to do work internally. It is zero at equilibrium, and non-zero under non-equilibrium conditions. Another extreme type of system is a crystal. A perfect crystal has only one possible microstate for its macrostate. It is completely ordered. Its information content is very low, as is its entropy. There can be no difference between S_{act} and S_{max}. The order of the crystal comes from causal interactions between its molecules, which confine it to an entirely rigid structure, so that only its size and location remain undetermined. Of course there are no perfect crystals, but actual systems can approximate this state. A more interesting case is a configuration which has a great deal of variety in its structure, but is causally constrained to take on only one possible state. An example is a state of the physical embodiment of a message with low redundancy, such as the signal in a telephone cable. In this case, as in a crystal, S_{act} and S_{max} are numerically close, but the information content of the state is intuitively very high.

The low intropy of both a low information crystal and a high information message shows that the information due to constraints cannot be intropy. Because it is based in the form of the system, I call it "enformation". Production and loss of enformation also involves loss and production of entropy, just as for intropy. The entropy increase, however, must be external to the system, whereas for intropy the changes can be internal as well.[9] The intrinsic information of a system, relative to its causal constraints, is the sum of its intropy and enformation:

$$I_{int} = \text{Intropy} + \text{Enformation} \qquad [7]$$

Changes in intrinsic information depend on changes in entropy, but I_{int} depends on whatever determines the quantity of enformation as well as on S_{act} and S_{max}.

The concept of enformation could be avoided by enlarging the system to include the observer and the processes by which information is conveyed from the object to the observer. All information would be internal to the system, so that it could be treated as intropy. This approach has merit, but it runs into both practical and methodological problems. The approach is complicated and requires knowledge of physical and physiological processes we don't yet understand very well. Furthermore, information about an object would be contained entirely in the perceptual processes; the object would create only external constraints on this information. This goes against the idea that information contained in the object is transmitted to the observer.

Another way to avoid enformation is to use an absolute entropy not relative to any constraints. This approach also runs into trouble. In order to define an absolute entropy for a system, we need to know its state when all physical constraints have fully relaxed (i.e., the system has reached equilibrium). This state cannot be determined empirically for two reasons. First, it isn't known if all physical structures fully relax; for example, it is not known whether protons decay. It might be possible to use the equilibrium state of the energy equivalent to the matter in a system to define its S_{max}, but there is no guarantee that this is the highest entropy state. Furthermore, it is unclear in what form the energy should be assumed to be. Second, the system

---

[9] Although he does not distinguish explicitly between intropy and enformation, Holzmüller (1984) discusses the relations of enformation and entropy in macromolecules.

must be assumed to be closed. This assumption is unrealistic for anything less than the whole universe. The universe is expanding, so even the whole universe does not have an absolute S_{max}; S_{max} is always relative to the size of the universe at some time. We don't know enough to specify absolute entropies, and even if we did, there is reason to think they do not exist. These considerations are somewhat esoteric. I include them to indicate the difficulties in giving a purely statistical account of intrinsic information.

Even if we could define absolute entropies, the intropy obtained would not be very useful since we are usually concerned with small parts of the universe that are heavily constrained (in the short run, at least). It would be much more practical to find a measure of enformation. Standard approaches to information theory (e.g., Shannon and Weaver, 1949; Brillouin, 1962) use combinatorial or probabalistic definitions of the amount of information (Kolmogorov, 1965). These definitions can be applied to intropy, but not to enformation. The standard approaches apply only to ensembles of systems; we need a conception of information which is applicable to individual cases. Algorithmic information theory (Kolmogorov, 1965, 1968; Chaitin, 1975) has this property.

The fundamental hypothesis of algorithmic information theory is that the information content of something is the length of the shortest program in binary form that can produce it. The idea behind this hypothesis is that a thing can be specified by making a series of binary distinctions (Spencer Brown, 1972). The minimum number of distinctions required is its information content in bits.[10] The algorithmic definition of information content is equivalent to the combinatorial and probabalistic information contents except for an additive constant representing computational overhead (Kolmogorov, 1968) that can be made arbitrarily small (Chaitin, 1975). The basic concepts of information theory can be defined without recourse to probability theory, and are applicable to individual cases (Kolmogorov, 1968). Furthermore, the relations between information and probability allow probability theory to be based on algorithmic information theory.

Algorithmic information theory is very abstract in its formulation. By analogy to Brillouin's (1962) definitions, we can distinguish between free algorithmic information and bound algorithmic information. The latter depends on computations which are physically possible. Computations, in this interpretation, are causal processes obeying natural laws. The information content of a state is determined by the most parsimonious causal process that can produce it from disorganized constituents. This is equal to its causal power, which is the measure of the number of distinctions the state can causally produce. The enformation is the bound algorithmic information of the constraints.

If we consider the total initial and total final states of any causal process, the information content of the final state must be less than or equal to the information content of the initial state. This is just a restatement of the second law of thermodynamics. If the process is 100% efficient (i.e., there is no dissipation of available energy into lower grade forms), the total information content remains constant. If a state is causally effective, but is itself unchanged in a causal process,

---

[10] The logic of distinctions can be shown to be equivalent to Boolean algebra (Banaschewski, 1977).

either the resulting final state contains only components of unchanged information content, or else information must be dissipated in one part of the total system in order to produce it elsewhere. This is the computational analogue to Brillouin's discussion of the negentropy budget. A system can causally maintain and change its form without dissipation of information, but this requires total efficiency of the processes involved. This will occur only if there is no friction involved, e.g., for static conditions, or if the system does not deviate from equilibrium.

The enformation and intropy of a system are independent if there is no dissipation. Enformation is converted into intropy if the rate of equilibration (relaxation time) of the system is greater than the rate of loss of enformation, i.e., under non-equilibrium conditions. Equilibrium and/or frictionless conditions are rarely if ever encountered in nature. Alternatively, intropy can do work on the system, sometimes producing enformation. I mention the close theoretical relations between enformation and intropy in order to strengthen the intuition that there is a common property of which they are different forms.

To summarize, the causal power of a state of a system has two components, its enformation and its intropy. Together, these constitute the information content of the state. Intropy is the ability to do work within the system; it represents the energetic aspect of causal power. Enformation, on the other hand, is the ability to alter the form of things; it represents the organizational aspect of causation, the ability to guide energy. Both are required for useful work. This simple distinction is complicated by the inter-convertibility of intropy and enformation. Work produces new enformation. Enformation can also produce intropy at expense of its own destruction if it is dissipated under non-equilibrium conditions. If friction is present in any causal process some information is lost; it becomes equivocation, or physical entropy. The general idea should be clear by now: *Intrinsic information is a measure of causal power*. The less information, the less causal power.

It is now possible to define the intrinsic information of a situation. Situations contain both enformation and intropy. Many situations are not cohesive; they do not represent independent objects or systems, what might be termed individuals. Instead, they are epiphenomenal. The intropy and enformation of cohesive situations is not problematic. The intropy of epiphenomenal situations can be defined with respect to the cohesive situations (see above). Similarly, the enformation of an epiphenomenal situation is a part of the enformation of any inclusive cohesive system. Its enformation is the sum of the enformation of the cohesive parts that compose it. This enformation is equal to the sum of the algorithmic information of the most efficient causal processes that can make the parts. In other words, the enformation of an epiphenomenal situation is derivative from the enformation of the parts that make it up.

The direct relationship between information and causal power may seem paradoxical from the top-down view of information as a reduction of possibilities. It would seem that a situation with less information would allow more possibilities than one with more information, and consequently would have more causal capabilities. This, however, conflicts with our intuition that more complex things are capable of doing more than simpler things. The bottom-up approach is more in accord with this intuition. A situation with less intropy is capable of doing less work than one with more. Likewise, a situation with less enformation is capable of producing fewer distinctions than one with more enformation. It isn't the range of different abstract possibilities which determine intrinsic information content, but the productive power of a situation in optimal cases.

4. Information Transmission

The basics of information transmission are fairly easy to understand, given the above account of intrinsic information. Since intrinsic information is a measure of causal capacity, it can be detected through its effects. An information channel is affected by a source, according to its ability to do so, but is not significantly affected by other sources whose effects on the channel might interfere with those of the information source (see Dretske, 1981: 115). If there is extraneous interference, it produces equivocation. Equivocation results in lost information, because the effect on the channel is the net result of the various causal influences acting on it. Since there is only one effect for various causes, the individual causes cannot be discriminated on the basis of the intrinsic information in the channel alone. The desired information from a source can sometimes be detected by *tuning* the receiver correctly. This involves finding a channel with minimal interference from other sources. Such a channel can be termed *reliable*.

Our senses act as filters that select generally reliable channels. Presumably evolution has selected our particular sensory mechanisms because, among other things, they are reliable. At higher levels of cognition we learn to discriminate reliable channels from unreliable ones, ignoring the unreliable ones by using classifications that don't suit them. This sort of tuning allows us to detect abstract and general properties. This selection process allows us to receive information about particular situations (within the limits of the available reliable channels). Tuning for abstract and general properties creates classification systems that reflect natural classifications.

A consequence of this view of sensation and classification is that the selection of reliable channels will maximize transmitted information. The meaning conveyed from the world will thus be greatest. If unreliable (i.e., noisy or equivocal) channels are selected, the meaning conveyed will be diffuse and equivocal, even though it might give the illusion of clarity.

A channel cannot transmit more information than the intrinsic information of its states. However, a situation distant from, but causally connected to an information source can indicate more information about the source than it contains itself (Dretske, 1981; Barwise and Perry, 1983). For example, the decay of tritium might be the only thing which produces a gamma ray of a certain energy, so a photon with the correct frequency would carry the information that a tritium atom had decayed, even though the enformation of the photon is less than the enformation of either a tritium atom or its other decay products. The extra information is contained in the natural laws of radioactive decay and gamma ray production. How is this information transmitted?

Dretske (1981), as I mentioned in section 2, tried first to deal with background information by assuming it is obtained in the same way as the information that is directly present. This leaves open the objection that background information about the field of possibilities is still required. Dretske turned to the bottom-up approach to allow the causal conditions involved in the information transmitted to have the reduction of alternative possibilities as an effect. We don't, therefore, need information about the eliminated possibilities. The case of tritium decay, though, is not obviously amenable to this sort of treatment. Dretske's treatment requires that the information transmitted from the indicating situation (the gamma rays) together with previous transmitted information that we have retained determines the presence of a tritium decay. But if natural laws are the only thing that ensures that there was a tritium decay, then part of the information required is information transmitted to us from natural laws. It isn't clear how the information in a natural law can be transmitted. The universality of natural laws implies that information transmitted by any particular cases is always deficient. It seems that an inference beyond the available information is always required. This seems to open a gap into which the wedge of scepticism can be driven.

If we consider theories to be systems for classifying and organizing empirical evidence, then a true theory is one that selects reliable channels for all cases within its scope. If so, every instance of a natural law within this scope will be transmitted without corruption. This eliminates the need for inference. Usually, though, selected channels are less than completely reliable. This doesn't mean that we can't be right; some information will probably get through uncorrupted, except for very bad choices. It is true that choice of theory, and thus channel selection, involves inferences beyond immediate evidence, but if reliable channels are selected, knowledge is the result. On this approach inference plays a role in the production and identification of knowledge, but is not a constituent of knowledge.

5. Further Development of the Bottom-Up Approach

As I mentioned above (section 2), the bottom-up approach requires an account of intrinsic information, an account of information transmission, and an account of the conversion of information into a cognitive form. In this paper I have focused on intrinsic information in particular concrete situations. I have also sketched an account of information transmission, but have had little to say about its conversion into cognitive form.

Further development of the account of transmission would involve a better characterization of noise and its avoidance, and of the selection process called tuning. Israel and Perry (this volume) discuss tuning from a biological perspective. Natural selection no doubt plays a central role in determining our basic repertoire of channels. The extension of biological concepts to cognition is most naturally carried out in the context of developmental psychology, along the general lines proposed by Piaget (1952). Piaget sees cognitive development as continuous with both ontogenetic and phylogenetic development. As Matthen (1988) has pointed out, ways of representing the world that are innate for the individual organism must be acquired by the lineage of which it is a member. Biological conditions provide not only an analogy for cognitive development, but are its foundation. Top-down philosophical and empirical studies of cognition, intentionality, and inductive logic place further constraints on acceptable accounts.

The transformation of raw information into cognitive content remains fairly mysterious. The continuity between biological selection and cognitive selection suggests that the roots of the transformation lie in evolution. Matthen (1988), for example, discusses biological functions from an evolutionary perspective, and sketches how it might shed light on how cognitive content appears. Adaptations are a form of tuning selected by the environment. On the present account, learning is also a process of tuning, suggesting that it is analogous to adaptation. This is in line with the general trends in Piaget's thought. Problems arise, though, from the fact that what is useful is not necessarily true. For example, Grandy (1987) used an ecological context to discuss perceptual content as mutual information of a representation and an organism's environment. He observed that optimal behaviour in conditions with less than complete information may favour having mistaken beliefs. This suggests that the selection of reliable channels is neither direct nor inevitable.

A full information-theoretic account from the biological perspective is required to make connections between raw physical information and cognitive content. Initial studies on this project have been done by Brooks and Wiley (1986; 1988), Collier (1986), and Ulanowicz (1986). These studies are still in their infancy, and there is much to be clarified about the interactions of biological organization, ontogeny, ecology, and evolution. The approach seems promising, since the basis of contemporary systematic biology, the genetic code, has both physical and representational properties that are amenable to information-theoretic analysis (Collier, 1988a). Ulanowicz (1986) has used the connections between information and entropy to do information-theoretic analyses of ecology and fitness. This work needs to be connected to information-theoretic studies of ontogeny and evolution before connections can be made between intrinsic information, adaptation, and eventually learning.

The current state of information-theoretic approaches to knowledge and semantics is very rudimentary. The applicability of related concepts at a variety of levels and to a number of apparently disparate applications suggests a large consilience. The major problem is to find general concepts that are broad enough to apply to a variety of applications, but concrete enough to avoid being vacuous. I believe that keeping close to the physical and causal basis of information will avoid vacuity. Generality will come from resolving problems that arise at the interface of different applications.

Acknowledgements
I am grateful to the History and Philosophy of Science Department at Indiana University for providing me with office space and research facilities during the preparation of this paper. Final revisions were done during the tenure of a SSHRCC Canada Research Fellowship at the University of Calgary.

References

Banaschewski, B. (1977) `On G. Spencer Brown's Laws of Form', *Notre Dame Journal of Formal Logic 18*: 507-509.

Bar-Hillel, Y. (1964) *Language and Information*, especially Chapters 15-17, Reading, MA: Addison-Wesley.

Barwise, J. and J. Perry (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press.

Bennett, C.H. (1982) `The Thermodynamics of Computation -- A Review', *Int. Jour. Theor. Phys. 21*: 905-940.

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

Bennett, C.H. and R. Landauer (1985) `The Fundamental Physical Limits of Computation', *Scientific American 256*, No. 7: 48-56.

Brillouin, L. (1962) *Science and Information Theory*. New York: Academic Press.

Brooks, D.R. and E.O. Wiley (1986) *Evolution as Entropy,* Chicago: University of Chicago Press.

Brooks, D.R. and E.O. Wiley (1988) *Evolution as Entropy, 2nd edition*, Chicago: University of Chicago Press.

Chaitin, G.J. (1975) "A Theory of Program Size Formally Identical to Information Theory", *J. ACM 22*: 329-340.

Collier, J.D. (1986) `Entropy in Evolution', *Biology and Philosophy 1*: 5-24.

Collier, J.D. (1987) `Theory Ladenness', in P. Weingartner and G. Schurz (eds) *Logic, Epistemology and Philosophy of Science, Proceedings of the 11th Annual Wittgenstein Congress*: 116-118.

Collier, J.D. (1988a) `The Dynamics of Biological Order', in Depew, Weber and Smith (1988).

Collier, J.D. (1988b) `Supervenience and Reduction in Biological Hierarchies', in M. Matthen and B. Linsky (eds) *Biology and Philosophy, Canadian Journal of Philosophy Supplementary Volume 14*.

Depew, D.J., B.H. Weber and J.D. Smith (1988) *Entropy, Information and Evolution*, Cambridge, MA: MIT Press.

Dretske, F. (1981) *Knowledge and the Flow of Information,* Cambridge, MA: MIT Press.

Grandy, R. (1987) `Information-Based Epistemology, Ecological Epistemology and Epistemology Naturalized', *Synthese 70*: 171-204.

Holzmüller, W. (1984) *Information in Biological Systems: the Role of Macromolecules*. Cambridge: Cambridge University Press.

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.

Landsberg, P.T. (1984) `Can Entropy and "Order" Increase Together?', *Physics Letters 102A*: 171-173.

Layzer, D. (1975) *Scientific American 233*: 56-69.

Matthen, M. (1988) `Biological Functions and Perceptual Content', *J. Phil 85*: 5-27.

Piaget, J. (1963) *The Origins of Intelligence in Children*, New York: Norton.

Prigogine, I. (1962) *Non-Equilibrium Statistical Mechanics*, New York: John Wiley and Sons.

Prigogine, I. and Stengers, I. (1984) *Order out of Chaos*, New York: Bantam.

Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Solomonoff, R.J. (1964) "A Formal Theory of Inductive Inference", *Inform. and Contr. 7*: 1-22.

Spencer Brown, G. (1972) *Laws of Form*, New York: Bantam.

Stegmüller, W. (1979) *The Structuralist View of Theories*, New York: Springer-Verlag.

Ulanowicz, R.E. (1986) *Growth and Development*, New York: Springer-Verlag.