THE BEHAVIORAL AND BRAIN SCIENCES (1986) 9, 121–166
Printed in the United States of America

Intentionality and information processing: An alternative model for cognitive science

Kenneth M. Sayre
Department of Philosophy, University of Notre Dame, Notre Dame, Ind. 46556

Abstract: This article responds to two unresolved and crucial problems of cognitive science: (1) What is actually accomplished by functions of the nervous system that we ordinarily describe in the intentional idiom? and (2) What makes the information processing involved in these functions semantic? It is argued that, contrary to the assumptions of many cognitive theorists, the computational approach does not provide coherent answers to these problems, and that a more promising start would be to fall back on mathematical communication theory and, with the help of evolutionary biology and neurophysiology, to attempt a characterization of the adaptive processes involved in visual perception. Visual representations are explained as patterns of cortical activity that are enabled to focus on objects in the changing visual environment by constantly adjusting to maintain levels of mutual information between pattern and object that are adequate for continuing perceptual control. In these terms, the answer proposed to (1) is that the intentional functions of vision are those involved in the establishment and maintenance of such representations, and to (2) that semantic features are added to the information processes of vision with the focus on objects that these representations accomplish. The article concludes with proposals for extending this account of intentionality to the higher domains of conceptualization and reason, and with speculation about how semantic information-processing might be achieved in mechanical systems.

Keywords: artificial intelligence; cognitive science; communication theory; computational models; information processing; intentionality; mental representation; philosophy of mind; vision

1. The need for a theory of intentionality

In his comment on the *BBS* target article by Dennett (1983), Bennett (1983) takes its author to task for failing to explain what intentionality is, and thereby failing to provide a theoretical basis for adopting the intentional stance in ethology and psychology. What is needed, Bennett urges, is an account of intentionality as a feature of "one kind of function from sensory inputs to behavioral outputs" – that is, a "description of what these functions are, of how they actually work" (p. 357).

Another key perspective on the type of theory needed is provided by the target article itself. As Dennett points out quite candidly (p. 344), the "decision to conduct one's science in terms of beliefs, desires, and other 'mentalistic' notions" is not usual in science. It involves a risk. "It banks on the soundness of some as yet imperfectly described concept of information . . . the concept of what is often called *semantic information*" (author's emphasis). The concept of semantic information is not that of Shannon and Weaver, (1949), but rather is what is "recounted in ordinary talk in terms of beliefs and desires and the other states and acts philosophers call *intentional*" (author's emphasis). The contents (Dennett's term again) of the mental states cited in the intentional idiom can be spoken of in terms of the semantic information they represent; and what imparts informational content to a representation is its being directed upon a particular object or state of affairs – that is, its intentional character. So another requirement of the account of intentionality needed by cognitive science is that the account provide a precise description of what it is that makes semantic information semantic – what it is that lends content to the internal representations involved in cognitive activity.

First, what do the functions of the nervous system that we ordinarily describe in the intentional idiom actually do? What do they accomplish? How do they work? And second, what exactly is it about the information involved in these functions that makes it semantic? These two are among the most pressing questions a scientific account of intentionality should be equipped to answer. Yet, as Bennett, Dennett, and various others have seen, they are questions to which cognitive science (CS) at present has no clear answers. Although the notion of semantic information, and the notion of intentionality standing behind it, are wholly indispensable to CS, its theorists have yet to produce an account of these features that explains their presence in cognitive organisms.

The primary purpose of this essay is to lay out the rudiments of an account of intentionality that shows promise of providing intelligible answers to these two questions, and to illustrate how these answers would go within a limited range of cognitive activity. Caveats are in order to prevent an already ambitious goal from further inflation. There is no completely worked out general theory of intentionality waiting in the wings to be pre-

 0140-525X/86/010121-46/$06.00

sented in capsule form, and no detailed answers to the questions above that are ready to be put forward in more than tentative fashion. What is said in this target article is ready for discussion, but surely not ready for the scientific archives.

Another important caveat is that this essay deals with intentionality primarily within the domain of visual perception and will not have much to say by way of extending the account to more complex forms of cognitive activity. The justification for this narrow focus, beyond obvious considerations of space, is (1) that visual perception also has been the main focus of other authors (notably Gibson, 1979, and Marr, 1982) with whom comparisons naturally will arise, and (2) that from my perspective the intentionality of perception is both prior to and constituent of the intentionality of the "higher" activities of language and reason, and that first things should be taken first. Nonetheless, a complete theory of intentionality would presumably have a great deal to say about the relations between perception and more complex forms of cognition. A very brief sketch of how these relations look from my perspective is provided in the final sections of this essay.

2. The "official" stance of cognitive science

What CS needs, it has been urged, is an account of intentionality that explains (1) what it is that the input–output functions producing intentionally described activity actually do, and (2) what it is about the information involved in these functions that lends semantic content to our mental representations. But it might be objected that CS has the answers to these questions already at hand, in the form of what Fodor (1980) calls "*the* basic idea of modern cognitive theory" (p. 68, author's emphasis). In barest outline, the "basic idea" is to conceive the brain as a computing device that accomplishes major cognitive tasks as the accumulated effect of many subtasks, all intentional features of which are paired with physical constraints of the computing mechanisms.

The "official" view behind this idea, equally condensed, is that propositional attitudes (perceptions, beliefs, decisions, etc.) are relationships between the overall physical organism and certain of its internal states, that these internal states are representations that provide the content of the attitudes in question, and further that one propositional attitude can function as the cause of another (as perceiving that the book is on the table can result in believing the same) in the manner in which one computer state can cause another. This means, first, that the complex representations through which the system interacts with its environment must be broken down into simple representations with which the system is equipped to deal computationally and, second, that the causal processes through which these canonical representations interact are just the physical procedures of computation. Accordingly, the way to understand how the human organism functions in a given cognitive task is to analyze that task under its intentional description, and to proceed to a level of analysis at which the intentionally specified subtasks can be performed by computations on canonical representations. Under their intentional description these representations have cognitive content; and under their computational (physical) description they are capable of causal interaction.

The question to which this view obviously leads is, what reason do we have for thinking that there is a level in the analysis of a cognitive task at which intentional and computational descriptions come together in this fashion? Like the 17th-century metaphysician Malebranche, cognitive scientists view the intentional (the "mental") and the computational (the "physical") as held together by a principle of overarching correspondence. Without relying on Malebranche's notion of Divine Contrivance, what reason do we have, in the interests of science, to think that such correspondence is a real possibility? The "official" answer, simply, is that we have actual computer systems that illustrate the principle. The "basic idea" is that the human brain is just another example of a computing system that "pairs physical states of the device with formulae in a computing language in such a fashion as to preserve desired semantic relations among the formulae" (Fodor 1975, p. 73).

But what computer systems are supposed to illustrate this principle? Not those guiding aircraft, or running heating systems, or assembling parts on a production line. Although systems of this sort ("number crunchers") may be as complex as you like, they have no truck with semantic content (i.e., technical explanations of how they function assign no essential role to computations on semantically laden representations). The all-time favorite examples of computing systems that do invite intentional description are chess players and theorem provers. Of these two, the logical theorem prover is most compelling, for it seems actually to require an intentional description. Since logical inference is by definition truth-preserving, and since truth is a paradigmatic property of intentional expressions, systems that prove logical theorems by computational procedures are proof also of the possibility of the needed correspondence. (Some readers may not be content with theorem provers as an illustrative case, feeling that such systems are old hat in artificial intelligence [AI] research. According to other viewpoints within AI, however, topics of automatic deduction have remained central up to the present and are currently enjoying a renaissance [Barr and Feigenbaum 1981, p. 228; Cohen and Feigenbaum 1982, pp. 77–79]. In view of the fact that theorem-proving mechanisms seem almost to beg for intentional description, choice of this illustration should not call for apology.)

So, to return to the query at the beginning of this section, do we not already have a theory of intentionality that explains what has to be explained about the mind's dealing with representations and semantic information? The functions from sensory input to behavioral output of which Bennett speaks, although varying in particulars from case to case, fall generally under the category of computational operations and can be studied by studying analogous operations in computer simulations. They work the way other computational operations work. As to the second question of where the semantic information involved in these operations comes from, the answer is that it derives from the functional relationship between internal representations and the things that they symbolize (Fodor 1981a, p. 123), a relationship secured by the principle of correspondence between semantical and physical properties.

3. Criticism of the "official" stance

There has been a good deal of criticism of the cognitive program since it began to receive coherent articulation in the early 1970s (with Fodor, 1975, as a sort of watershed between pre- and postarticulate periods), matched by elaborate defenses by its major theoreticians (including articles by Fodor, 1980; Pylyshyn, 1980a, 1980b; and Dennett, 1983, in recent issues of *BBS*). Although the evidence is far from complete, it seems fair to say on balance that there are both deep insights that will be preserved as CS matures and deep problems to be worked out in the process of maturing. One of the deepest problems, I believe, lies with the notion of semantic information-processing upon which the computational model of cognition is based. To see this, in effect, is to see why Bennett is right in claiming that CS still needs a coherent account of how intentionally described functions "actually work," and why Dennett is right in observing that the concept of information involved in these functions has yet to be precisely described.

The heart of the problem is that computers do not operate on symbols with semantic content. Not even computers programmed to prove logical theorems do so. Hence pointing to symbolic operations performed by digital computers is no help in understanding how minds can operate on meaning-laden symbols, or can perform any sort of semantic information-processing whatever. But what is the problem with theorem provers?

What a theorem prover does, put very roughly, is to start with axioms which it breaks up into canonical notation and, following various heuristic procedures, to generate theorems from the axioms by prescribed rules of inference. Put somewhat less roughly, representations of instructions in the programmer's language are translated via compiler and interpreter into machine-language instructions, which govern the internal workings of the machine in such a fashion that machine states corresponding to representations of logical axioms at the input generate other machine states in sequence, leading eventually to states producing formulae at the putput that are interpretable as representing theorems according to the conventions of the programmer's language. If the programmer, the compiler, the machine-language instructions, and so forth, all do their jobs properly, the formulae at the output are properly interpreted as true in just the same circumstances and under the same interpretive conventions as those under which the input axioms are so interpreted. When this happens, we say the program has proven the output formulae, in the sense of showing that they follow logically from the axioms at the input. The catch to all this is that none of the representations internal to the machine has meaning, or truth, or external reference, just in and by itself. Whatever meaning, truth, or reference they have is derivative (the term is borrowed from Haugeland, 1981, p. 32), tracing back to interpretations imposed by the programmers and users of the system.

There is a sense, to be sure, in which rules of inference built into the program are "truth preserving," but it is no different from the sense in which equivalent rules would "preserve truth" when applied by human logicians to symbols on paper or blackboard. As long as the person applying the rules holds to the same interpretive conventions (e.g., those of Polish notation), the evaluation "true" will be warranted for formulae at the end of the properly applied inference procedure if that evaluation is warranted for formulae at the beginning. As far as the uninterpreted formulae themselves are concerned, however, no truth evaluation is warranted whatever, because uninterpreted formulae bear no semantical relation to the world at large. In like fashion, the sense in which a computer can be programmed to prove logical theorems, and in the process to "preserve the truth of its representations," is a sense in which various symbolic representations go through various stages of transformation carefully controlled to permit the same semantic interpretations at input and output. Of intrinsic intentional properties, however, these representations are entirely innocent.

This predicament of the theorem prover admits generalization. There is no purely formal system – automated or otherwise – that is endowed with semantic features independent of interpretation. Such is the case even with symbols in a natural language. There is nothing about the symbolic structure of the term "cat" (type or token, oral or written), for instance, that establishes reference to feline animals. Inasmuch as the English word "cat" refers to cats, the word consists of more than can be uttered or written on paper. It consists of the symbolic form CAT (which can be instantiated in many ways in speech and writing) plus interpretive conventions by which instances of that form are to be taken as referring to cats. Similarly, the symbolic form GO means the opposite of STOP (or COME, etc.) by appropriate interpretive conventions of English, while by those of Japanese it means a board game played with black and white stones. But without interpretive conventions it means nothing at all. Such is the case with formal symbols generally, including the formulae involved in machine computation. And since interpretive conventions are rules applied by the users of the symbol system, it follows that formulae involved in computation – whatever their character as formal symbols – have reference only derivatively from user interpretation. In terms of the useful distinction of Searle (1980a), the intentionality attaching to machine representations is not "intrinsic" but "observer-relative," being entirely dependent upon the "intrinsic" intentions of the programmer or user.

This seems manifestly at odds with certain lines of current AI theory, particularly in those areas overlapping with CS. Accordingly, it is relevant to indicate briefly why recent work in AI provides no exceptions to the general remarks above. AI is often depicted as having originated with the realization that digital computers (unlike desk calculators) are not limited to the processing of numerical data, but are also capable of symbol processing (Barr and Feigenbaum 1981, p. 4). This means, roughly, that computers can be programmed so that operations on their internal states have the effect of working out implications and ramifications of symbol structures presented at their inputs. Broadly perceived, the major emphasis in AI over the past dozen or so years has been an attempt to devise formats for the representation of nonnumerical data, and procedures for the processing of such representations, that will enable machines to operate with symbol systems in a manner approximating that of human cognition. Perhaps the two most influential approaches to that goal have been those of procedural representations and of semantic networks.

The key idea behind procedural representation is that semantic features of natural language can be represented in the machine by procedures in the form of microprograms. The function of these procedures is to compute relationships among input symbols corresponding to semantic relationships in natural language. A paradigm example of this approach is Winograd's SHRDLU (Winograd 1972), which accepts commands and answers questions about a simulated "microworld" consisting of blocks and a robot manipulator. Among other components, the SHRDLU program incorporates a syntactic parser for decoding input sentences in natural language, and a semantic analyzer for converting input information into commands for the robot and queries to the data base. Semantic analysis is based on "definations" of linguistic structures, such as conjunctions and noun phrases, in the form of LISP programs incorporating information about implications and other meaning relationships. With the help of such internal procedures, a human interlocutor is able to direct the operations of the simulated robot, and to monitor the results on a cathode-ray tube.

The function of the microprograms in a system of this sort is to represent semantic relationships among words and concepts in a natural language. But to represent a semantic relationship is not thereby to possess semantic features of the sort represented. Given that the term "cat" refers to cats, and that microprogram C represents "cat," it does not follow that C itself refers to cats. (A red light on a dashboard may warn of [designate, refer to] inadequate oil pressure, while a photocopy of the light warns of nothing whatever.) Similarly, from the fact that a microprogram represents a semantic relationship among terms in a natural language (e.g., "bird" implying "winged"), it does not follow that the microprogram itself posseses that same semantic feature. This is not to say that the microprograms of this sort are totally innocent of meaning relationships. By contrivance of the programmer, they are representations – and representation itself is a semantic feature. But they are representations only by programmer contrivance, which means that they are "observer-relative" in Searle's sense above. The microprogram represents what it represents by the programmer's intentions, and this representation is not intrinsic to the operation of the program. Presumably SHRDLU would have been written differently if Winograd had worked according to different intentions; but once written, the operation of its microprograms is indifferent to what they happen to represent.

A more prominent recent approach is that of semantic networks, using a notation in which meanings are expressed by a configuration of nodes and links. Nodes typically represent objects and situations, whereas links stand for relationships between the things thus represented (see Winston, 1984, for an up-to-date discussion). Connecting two nodes representing cats and animals with a directional link representing class membership, for example, can represent the fact that all cats are animals. In so-called knowledge representation research in AI, networks of this sort are useful as formats both for storing information relevant to a given subject domain and for tracing inferences to facts not directly entered in the data base. A simple illustration of one available form of inference is provided by a pair of networks representing the facts that all cats are animals and that all animals are mortal which, when associated by the common term "animal," yield the conclusion that all cats are mortal. With respect to AI applications, it is important to note that meanings are assigned to nodes and links not solely by interpretive conventions, but also by the nature of the particular computational procedures that manipulate the network structures (Barr and Feigenbaum 1981, pp. 157, 186–87). Due perhaps to this continued emphasis on procedures, coupled with the fact that networks of this sort stand behind other current approaches using frames and scripts (pioneered by work reported in Minsky, 1975, and Schank and Abelson, 1977), representation formats provided by semantic networks find wide use in current AI research.

But do semantic networks in themselves possess intrinsic content? No more than the procedural representations examined above. To be sure, the network format allows greater flexibility and greater detail in the articulation of meaning relations. A node–link representation of the fact that all cats are animals can be embellished by the addition of a wide range of other relations – reflecting, for example, the physical makeup of cats and their average lifetimes – something not so easily accomplished by procedural representations of a simpler variety. But no amount of detail will convert a computerized version of a node-and-link network into an "intrinsic" representation of a particular meaning or meaning relationship. The reason is that exactly the same network could represent any number of alternative meanings. As Pylyshyn observes, insofar as the computer itself is viewed as a formal symbol processor, "we have a great deal of latitude in assigning semantic interpretations to states" (1980a, p. 443). As is the case with formalisms of the propositional calculus, for purposes of manipulating symbols within the system it makes no difference how the symbols are interpreted. It is even irrelevant whether they are interpreted at all.

From the fact that a computer program is intended (by the programmer) to represent meanings, it does not follow that the program has intentional features, any more than it follows that the keys and levers of a typewriter have intentional properties because they represent the typist's meanings as they are symbolized on paper. To the extent that CS relies upon the model of a symbol-processing system that "behaves in a certain way *because* certain expressions represent certain things" (Pylyshyn 1980a, p. 443, my emphasis), the representation schemes of AI are poorly suited to its purpose.

Observations of this sort are not new in the literature. Searle (1980a), for example, draws upon similar considerations as part of an ingenious (and much debated) argument to show that agents (brains or computers) do not acquire intentionality merely by instantiating the right kind of program, thus disenfranchising the approach of what he calls "strong AI." My point is similar, but with a different consequence. My point is that computers, just in and by themselves, no matter how programmed, do not exhibit intentionality at all, and hence that the attempt by CS to devise an explanation of the intentional character of human mental activities by thinking of such activities basically as computations on internal representations does not have so much as a "ghost of a chance" of succeeding. Maybe computations of some sort are involved in all or most mental activities; but this is of no help in explaining the nature of the intentionality those activities exhibit.

Can CS be rescued from this impasse? I believe a way has been indicated by Pylyshyn himself. This way has to do with environmental constraints upon representational procedures, and hence upon the meanings these procedures can represent. As far as computer intentionality is concerned, Pylyshyn suggests that "if we equip the machine with transducers and allow it to interact freely with both natural and linguistic environments" (1980a, p. 443), then representations may be set up in its internal circuitry that do not depend upon user interpretation. But CS is (or should be) primarily concerned with human intentionality. And since human organisms already are equipped with transducers of the sort required, there is reason to expect that environmental interaction has a good deal to do with human intentionality as well. If it can be shown that basic forms of human intentionality can be understood in terms of environmental interactions, without direct reference to computational procedures, then CS may be relieved of its dependency upon AI modeling.

Since it is not the concern of this paper to advance the cause of AI as such, little more will be said about computer intentionality in the pages that follow. Suffice it to say that if we can discover how intentionality operates in the human organism, we will have a much better idea than anyone has at present of how to build intentional features into a computer system.

My concern is with the intentionality of human perception, and with how this intentionality depends upon environmental constraints. Since perception seems naturally to go together with phenomena of "problem solving, language processing . . . and so on" (Pylyshyn 1980a, p. 443), there is reason to hope that what can be found out in the domain of perception might eventually prove fruitful in these other domains as well. But apart from some brief remarks at the end of the paper, these other concerns will be left for other occasions.

4. Resources for a theory of information processing

In his earlier target article Pylyshyn observed that there are "clear indications in the history of science . . . that periods of progress are coincident with major new technical and conceptual developments or. . . with taking an existing formalism seriously as a way of understanding the world" (1978, p. 93). Stressing the "technical developments" angle, Pylyshyn argues that progress in computer science within the past twenty years provides "reasons for viewing the potential contribution of the computational approach" to cognitive theory with optimism (p. 93), although he adds at the end of the article that it remains to be seen whether that optimism is warranted. Having examined reasons why optimism is not warranted, with respect to a computational theory of intentionality at least, I propose to pick up on Pylyshyn's second alternative – that of "taking an existing formalism seriously." Specifically, I wish to advocate the usefulness of the mathematical theory of communication (MTC) for clarifying the intentionality of certain cognitive activities.

There are antecedent reasons for expecting MTC to have significant application in the study of cognition. For one, MTC is a general theory covering all forms of communicational transactions, and some cognitive activities are obviously communicational transactions (perception and language at the very least). Another reason is that cognitive activities involve informational interchanges of various sorts, and MTC is a formal theory of informational interchanges. An immediate problem, of course, is that information in the *technical* sense of MTC (which I will label "info(t)") has no direct connection with semantic information ("info(s)"). The info(s) behind the intentional stance, as Dennett (1983) makes clear, serves as the content of representations that are directed upon objects. In a word, info(s) has intentional features. Info(t), on the other hand, is a statistical concept, defined precisely as the inverse of the logarithm (to the base 2, when measured in bits) of the probability of the event bearing the information. Intuitively conceived, the info(t) of a given event is the number of times the prior probability of that event must be doubled to reach the 100% probability it enjoys after occurrence. Info(t) is thus merely the opposite of statistical uncertainty and is entirely devoid of semantic features. A more or less standard way of summarizing the difference between these two concepts (e.g., Dennett 1983) is to say that info(t) measures the capacity of communication channels, while info(s) concerns the content of what is communicated.

Despite the uncertain connection between these two concepts of information, there is a tradition of confidence that MTC will ultimately prove applicable to the study of semantic structures. With the first publication of Shannon's theory under a single cover (Shannon and Weaver 1949), in fact, there was appended a hopeful article by Weaver suggesting that Shannon's "analysis has so penetratingly cleared the air that one is now, perhaps for the first time, ready for a real theory of meaning" (p. 116). Soon after, MacKay (1950) appeared with a version of "information theory" attempting to graft semantical features onto MTC, which version was adopted without further ado into the influential works by Broadbent (1958) and Garner (1962). This may help account for the assumption not uncommon among cognitive theorists that MTC by itself constitutes a theory of info(s) processing (Gibson 1979, p. 238; see preface of Sayre, 1976, for other instances), which certainly it does not.

Another way of telling the story is that of Cohen and Feigenbaum, who identify the concept of information as the key to the shift from behaviorism to cognitive psychology, and then cheerfully describe a transition from the precise sense of MTC to "a more relaxed, and more appropriate, conception of information" (1982, p. 5) that began to emerge by the 1960s. The same "relaxed" spirit is reflected in the insistence of Garner that psychologists can use informational concepts in any way that helps them, and that if "going beyond or even distorting established usage" (that of MTC) helps solve behavioral problems, then they "should feel free to do so" (1962, p. 15). The most significant recent instance of this approach is Dretske (1981), which attempts "to rescue information from the clutches of the statisticians and put it to work in mathematically less purified surroundings" (Dretske 1983b, p. 83; see my commentary on Dretske, 1983a for details and criticism of Dretske's departure from MTC). Salutary as this later work may prove to be, the "relaxed" approach has not provided much help in clarifying the nature of the information that is processed in cognition. As matters stand, Dennett is exactly right in pointing out that the intellectual solvency of cognitive theory depends upon arriving at a sound account of an as yet imprecisely understood concept of info(s).

In speaking reprovingly of a "relaxed" approach, I do not mean to suggest that an account of info(s) processing can be derived by strict deduction from the formalism of MTC, or that the concepts of MTC by themselves provide adequate resources for the type of account called for. Additional resources are clearly needed. And since our concern in developing such an account is to understand better the intentionality of certain processes in the nervous system of a cognitively endowed organism, it is reasonable to expect neurophysiology to make a substantial contribution. A paradigm to heed in this respect is Marr (1982), who brings mathematical rigor to the empirical study of the mechanisms of vision in a novel and highly creative way. The present approach attempts to parallel Marr's in several respects. An important difference, by way of anticipation, is that Marr's approach involves virtually no application of MTC as such, and in fact relies upon the computational model to back up his extensive use of information-processing terminology.

Inasmuch as cognitive activities are functions between sensory input and behavioral output (as Bennett puts it succinctly), they also involve responses to energy transformations ("nerve hits") at the sensory periphery. An adequate account of these activities may thus be expected to use certain resources from thermodynamics, particularly regarding the relationship between information and entropy. I have in mind particularly the intimate formal relationship between info(t) in MTC and entropy (the opposite of energy) in thermodynamics, which is accessible in outline to scientists without special training in either discipline.

As a final resource there are the basic concepts of evolution and natural selection, again on a level accessible without special training. An assumption behind the present approach is that cognitively endowed organisms have evolved from simpler life forms and that the neuronal mechanisms of the cognitive functions have evolved from relatively less complex neuromuscular control systems. One angle exploited by this approach is to ask, in connection with a given cognitive function, what empirical constraints probably guided its phylogenetic development, and to take those constraints as clues to how that function actually operates. In application to visual perception specifically, this emphasis upon evolutionary constraints is in the spirit of Gibson (1966). My hope is that the application of MTC, which Gibson all but ignored, will help us say more about how the visual system processes information from the objective environment than Gibson apparently was able to articulate.

5. Technical concepts of communication theory (once-over with summary)

Communication theory is a mathematical discipline, conceived originally for application in the design of communication circuitry. For present purposes, fortunately, the technical concepts needed are relatively few, and they can be defined with minimal use of symbols. (For readers who want additional detail, the classic source is Shannon, 1948; Abramson, 1963, is one of the clearest technical treatments available; a reliable source in the psychological literature is Luce, 1960.)

Information (here called info(t)) is a characteristic of a single event (a) within a set (A) of alternative events, one and only one of which occurs at a given point in a sequence (temporal, spatial, or otherwise). Given probability of occurrence $P(a)$, the information content in bits of event a is the logarithm (base 2) of the inverse of its probability (log $1/P(a)$; equivalently, $-\log P(a)$). The average information ($H(A)$) of the set A is the sum of the products, for each member taken severally, of information content times probability of occurrence ($\sum_A P(a) \log 1/P(a)$; equivalently, $-\sum_A P(a) \log P(a)$). This quantity is often called the *entropy* of set A. (I will label it "entropy(c)" to distinguish it from the entropy of thermodynamics.) In a set of three equiprobable events, for example, the information content of each is 1.58 bits, and the entropy(c) of the set is the same ($0.33 \times 1.58 \times 3$). In a set of three events with probabilities 0.1, 0.3, and 0.6, however, the entropy(c) is only 1.30 bits ((0.1×3.32) + (0.3×1.74) + (0.6×0.74)). In general, the closer that the events within a set approach equiprobability, the closer its entropy(c) approaches maximum value.

All communication takes place across an information channel (of which communication equipment provides only one class of example). An information channel, defined with complete generality, is a pair of event sets the members of which are statistically nonindependent. In the context of a given communication transaction, one set A is source (or input) and the other B is terminus (or output). Since the sets are nonindependent, events at B are to some extent indicative of events at A. In a sense (not a semantic sense), events at B contain information about events at A – in the sense, exactly, that B events add information with respect to A events. To see what this means, consider the limiting case in which event b at the output provides a completely reliable indication of what event a occurred at the input. Before b, all that can be known at B with respect to the occurrence of a is its prior probability (less than 100%). After b, given its complete reliability, it can be known that a occurred with 100% probability. The occurrence of b has changed the probability of a relative to the output from its prior value to a posterior value of 100%. The occurrence of b thus contains info(t) with respect to a (in the amount of 100% minus the prior probability of a), which in this sense (again not a semantic sense) may be said to be information about a.

The value $H(A)$ fails to reflect the statistical dependence between A and B and is thus sometimes called the *a priori entropy(c)* of the input (the entropy(c) of A prior to B). Another entropy(c) value of A can be calculated when statistical dependency is taken into account. The *a posteriori entropy(c)* of A relative to B is defined as the sum over all members of B of the *a posteriori* values of each b with respect to each a. This quantity is symbolized $H(A/B)$, and represents the amount of uncertainty at the output about events at the input remaining after occurrence of all associated output events. When each event of A is indicated by events in B with complete reliability, then no uncertainty remains and $H(A/B)$ is zero. In real-life information channels, however, some uncertainty typically remains at B with respect to A, which means that the situation of the input is to some extent ambiguous at the output. ($H(A/B)$ is accordingly called the equivocation of A with respect to B.)

One further concept is needed, that of mutual information. It should be obvious with a little thought that channels capable of receiving more info(t) at the input are generally capable of passing more info(t) through to the output. The capacity of channel A-B as a reliable communicator of info(t) is thus directly proportional to $H(A)$. Since equivocation is the opposite of reliable communication, moreover, this capacity is inversely proportional to $H(A/B)$. The quantity measuring a channel's capacity for the reliable communication of info(t) is called its *mutual information* and symbolized $I(A;B)$. The equation $I(A;B) = H(A) - H(A/B)$ is an important formula of MTC and will figure prominently in the following discussion of visual perception.

By way of summary, events that are less than 100% probable (i.e., are uncertain) in advance convey info(t) by occurrence, in an amount proportional to their advance uncertainty. The info(t) of a given event is thus the change in probability attending its occurrence. The entropy(c), or average uncertainty, of a set of events is the average info(t) of its members weighted by advance probability of occurrence. An information channel is any pair of statistically nonindependent sets of events, conventionally distinguished as input and output. The a priori entropy(c) $H(A)$ of the input A of channel A-B is the average uncertainty of A prior to occurrences at B, in contrast with its a posteriori entropy(c) $H(A/B)$ relative to B which is its average uncertainty after occurrences at B. $H(A/B)$ measures the average uncertainty at the output with respect to the input and is hence referred to as channel equivocation. The overall capacity $I(A;B)$ of the channel as a reliable communicator of info(t), finally, is equal to the quantity $H(A) - H(A/B)$ and is referred to as the channel's mutual information – that is, the info(t) shared at input and output.

These concepts from MTC are completely precise and general, and they are available for service in a nonmetaphorical account of information processing in organic systems. In their mathematical form, however, these concepts have no bearing on meaning or intention; they have no bearing on info(s) processing. What is needed to assure the solvency of the information-processing approach to cognition, as Dennett saw so clearly, is an account that bridges the gap between the concepts of info(t) from MTC and of info(s) with intentional features. The endeavor to develop such an account is of first-order importance; and there is room in this endeavor for as many different approaches as can be coherently formulated. The approach I wish to recommend in the present essay depends upon the connection between info(t) and the concept of entropy in thermodynamics.

6. The connection between energy and info(t)

Ever since Von Neumann in 1952 proclaimed the "identity" of MTC and thermodynamics (Bar-Hillel 1964, p. 12), there has been speculation among highly competent scientists about the relationship between these two fundamental disciplines (see Sayre, 1976, Ch. 3, for survey). Although the relationship is still not completely understood, a signal contribution was made by Tribus (1961). Drawing upon earlier work by Maxwell and Boltzmann, and more recent work by Gibbs, Planck, and Jaynes, Tribus was able to show how classical thermodynamics can be derived from MTC with the help of a few unproblematic axioms. Substantial help also came from Brillouin (1962). Leaving aside aspects of this relationship that still provoke disagreement, we can outline what is needed here in a relatively straightforward manner.

Entropy in thermodynamics is a mathematical measure of the disorganization of a closed physical system. Any physical system consists on the microlevel of elements that might exist in other configurations, or complexions (Planck's terminology; see Brillouin, 1962, p. 120), and that are not empirically distinguishable. For any one distinguishable macrostate of the system there are many complexions that might produce it; and the a priori probability of a macrostate is equal to the proportion of complexions that might produce it to all possible complexions of the system overall. If P is the priori probability of a given macrostate, and k is Boltzmann's constant, then the thermodynamic entropy (S) of the system in that macrostate is defined as k $\log_n P$ ("$\log_n$" means natural logarithm). As a series of interrelated general facts, the higher the proportion of complexions correlated with a given macrostate, (1) the more disorder among its elements, (2) the greater its measure of thermodynamic entropy, and (3) the less energy it makes available for useful work. In a state of maximum disorder, a system contains only thermal energy evenly distributed throughout and thus contains no energy for work at all – hence the figurative reference to its highest entropy state as the "heat death" of the universe.

Like the entropy(c) H of MTC, thermodynamic entropy S is a function of the quantity $\log P$. The major difference is that, while S is directly proportional to $\log_n P$, H is directly proportional to the negative of $\log_2 P$. The way the relationship works out, fine points aside (for which see Whitrow, 1967, or Sayre, 1976, Ch. 3), is that the entropy of MTC is the opposite of thermodynamic entropy. Intuitively, disorder on the microlevel is randomness in the arrangement of microstates; and the more random a system's microstructure, the more uncertainty remains regarding that structure after empirical examination, and the less info(t) the system can convey to a receptive observer.

The upshot is that all three quantities, energy (from thermodynamics), structure (from physics and chemistry), and info(t) (from MTC), vary inversely with thermodynamic entropy. All three are forms of negative entropy or negentropy for short (following Brillouin, 1962). As such, each can be converted to one of the other. As gravitational motion is the conversion of structure to energy, for example, and measurement the conversion of structure to info(t), so info(t) can be converted to either structure (biological growth governed by genes) or energy ("Maxwell's demon"). In any such conversion, however, the entropy of the system overall tends to increase; in the economy established by the second law of thermodynamics, nothing comes free – not even a measurement (engineer's waggery: "There ain't no free lunch").

7. The evolutionary background

Background from one other area of scientific inquiry must be marshalled before turning specifically to the processes of perception. This area is evolutionary biology, tapering off into ethology and learning theory. One point to be stressed in this section is that human perception is first and foremost a means of adapting to a changing environment. As species evolution is the (relatively slow) adaptation to changing environmental conditions on the level of the reproductive group, and as conditioning is the (relatively quick) adaptation to changing structures of reward and punishment on the level of the individual organism, so perception is the (very rapid) adaptation of the neuronal patterns that guide the organism's response to environmental contingencies.

Another point to be stressed is that each of these processes – evolution, behavioral conditioning, and human perception – can be viewed as a procedure for maximizing the efficiency of the negentropic coupling between organism and environment, thus justifying the deep insight of Marr that evolution and perception, among other of life's mysteries, are "primarily phenomena of information processing" (1982, p. 4).

According to the second law of thermodynamics, any closed system tends to increase in entropy with the forward progression of time. Biological systems, however, are open with respect to their environment, and they tend to decrease (with growth) or to remain constant (at maturity) in entropy level. As Schrödinger puts it picturesquely, a living organism is a device for "sucking orderliness from its environment" (1967, p. 79). This means that an organism is constantly receiving negative entropy from the world around it, in the form of structure (for growth), energy (for metabolism), and info(t) (for guidance). In Schrödinger's words once again, "What an organism feeds upon is negative entropy" (p. 76).

As life is a process of appropriating energy, structure, and info(t) from the environment, and of sloughing off the resulting entropy, so death is cessation of the flow of negentropy from the environment and the beginning of a progressive entropy increase instead. To sustain the life process, an organism must maintain a relationship with its immediate surroundings that enables it to assimilate negentropy in the forms and amounts needed, and to rid itself of the resulting byproducts. I refer to this relationship as a negentropic coupling. Since the environmental side of this relationship is always subject to change, the organism must be equipped to adjust its behavior accordingly if it is to accommodate these changes with a chance of survival. Evolution, conditioning, and perception are all ways of adjusting, but on different time scales and on different levels of activity.

The general drift of evolution is to accumulate gene pools that produce organisms capable of forming stable negentropic couplings with their environments within the normal range of prevailing living conditions. The general drift of natural selection, in turn, is to single out gene pools producing negentropic couplings that are most efficient under prevailing conditions or that are most capable of adjusting under conditions of change. I shall refer to these capacities in combination as negentropic flexibility. The upshot of competitive adaptation is to establish species with relatively high degrees of negentropic flexibility – not through some "entelechy" or "vital force," but as a result of the dynamics of its natural operation.

One tactic for enhancing flexibility of negentropic coupling that was "discovered" during the course of species evolution is the several-faceted process of behavioral conditioning. (Despite the recent devaluation of behaviorism as psychological theory, the careful empirical studies of respondent and operant conditioning during earlier decades remain among psychology's most substantial achievements.) For purposes of (very rough) comparison, we may think of the adaptive modification of behavior as occurring primarily in the association between receptor and effector mechanisms. Invariable receptor–effector associations include those that support reflexes, autonomic activity, and various species-specific forms of behavior like web-building by the spider and the dance of the bee. Although associations of this sort are adjustable on the level of species evolution, they do not vary with the experience of the individual organism. In adaptive behavior, by contrast, the working association between receptor and effector mechanisms is subject to adjustment, enabling the organism to alter its responses to repeated stimulus conditions on the basis of its individual past experience. An organism capable of adjusting its response in this fashion – capable, that is, of behavioral conditioning – is clearly superior in negentropic flexibility to other organisms that adapt only through species evolution.

Whereas adaptive alteration of dominant gene structures (species evolution) requires several generations, adaptive alteration of neuronal associations governing behavior (behavioral conditioning) can occur many times within an individual lifetime. There is little novelty today in the suggestion of substantial parallels between these two adaptive processes (tracing back at least to Skinner, 1969). [see special issue on canonical papers of B. F. Skinner: *BBS* 7(4) 1984.] The suggestion of a similar parallel with perceptual awareness is more adventurous. Nonetheless, another giant step in adaptive flexibility came with the development of patterned perception, enabling an organism to adjust its behavior in a changing environment on a scale marked in milliseconds rather than portions of lifetimes. What adapts in perceptual awareness, of course, is not gene pools or afferent-efferent mechanism pairings, but rather the structure of the afferent neuronal patterns that generally guide our discretionary behavior. A leading theme in the present account of visual perception is that the processes by which these patterns are adaptively altered can be characterized as forms of info(t) processing.

8. Info(t) processes in the evolution of visual mechanisms

If a one-sentence description of human vision is in the cards, I believe it would go something like this: Vision is a process occurring over a hierarchy of neuronal functions between retina and cortex by which representations are adaptively altered in response to changing stimulus configurations, in a fashion maintaining a channel of communication between representation and environmental object with sufficiently high mutual information to enable the organism to respond selectively to a wide range of environmental circumstances that affect its interests.

This emphasis on the adaptive alteration of representations constitutes a major departure from the theory of Marr (1982), which is the most carefully worked out account of vision yet produced by the information-processing approach. By Marr's account, the same type of representation is used in the visual systems of spiders, house flies, and frogs as in those of human percipients, with differences only in complexity born of different purposes (pp. 32–34). The main drift in the evolution of visual processes is toward representation of "progressively more objective aspects of the visual world" (p. 340), and toward deriving representations with increasing rapidity (p. 105). According to the present account, by contrast, evolution tends toward producing increasingly adaptive types of representation, so that the representations involved in human vision differ radically from those of "lower" organisms with respect to flexibility. A likely story of how that came about may serve as a prologue to this account of how human vision functions.

Any sensory system is (among other things) a control mechanism for regulating behavioral interactions between the sensing organism and its sensible environment. The course of evolution leading to the human visual system is a progression toward control mechanisms that are increasingly flexible, and increasingly efficient in their info(t)-processing capacities. The reason can be indicated in terms of an important theorem of MTC – the 10th theorem of Shannon (1948).

Shannon's 10th theorem states, in paraphrase, that the ability of a control system to correct deviations from an optimal mode of operation cannot exceed the amount of info(t) at the system's input (for a more exact paraphrase, see Sayre, 1976, p. 159). Suppose that it is optimal for an organism to emit behavior b_0 under environmental circumstances e_0, and b_1 under e_1, but that in the absence of control it emits b_0 and b_1 randomly with 40% chance of the right (optimal) response. Because the deviation or error rate of the behaving system then in effect is its equivocation of E with respect to B, the deviation from optimal operation is measured by the quantity $(0.4 \times \log 1/0.4) + (0.6 \times \log 1/0.6)$, or approximately 0.97. This means that an error-correcting system (such as a system of perception) must be capable of representing an average of at least 0.97 bits of info(t) at its input to bring the organism to optimal performance. A biological corollary to this important theorem is that the development of organisms capable of adapting their responses selectively to a progressively wider range of environmental circumstances (hence with progressively greater potential for error) must be supported by the development of sensory mechanisms with progressively greater info(t)-handling capacities. Put as directly as possible, the implication is that organisms under selective pressure to increase the range of their adaptive behavior will be under pressure as well to increase the mutual information of their sensory channels.

Since mutual information is a joint function of equivocation and input entropy(c), there are two distinct strategies for increasing this quantity. One is to increase the capacity of the organism's receptor mechanisms (increasing their entropy(c)), the other to increase the reliability of the sensory system overall (decreasing its equivocation). Both appear to have been exploited in the course of evolution.

As an example of the first, consider what might be done to increase the capacity of a noninteractive group of photosensitive receptor cells, such as those interspersed with nonreceptors in the skin of an earthworm (Walls 1942). One obvious expedient would be to increase the sheer number of receptors involved – an alternative that would be attractive only if bulk were not a factor. A less obvious but more efficient alternative would be to make the receptors interactive. From the communication-theoretic point of view, this amounts to introducing conditional probabilities among neuronal events, in effect making the uncertainty associated with a given input event (and hence its info(t) content) depend upon the occurrence of other input events. (Various mechanisms that would accomplish this physiologically are discussed in Hebb, 1949, and Kandel, 1970.) What is interesting for our purposes is that MTC can show why making receptors interactive increases their info(t) capacity as a group. The capacity in bits (C^s) of a combined channel consisting of two otherwise independent channels with capacities C^1 and C^2 is $\log_2$ of the sum (2 raised to the power C^1 and 2 raised to the power C^2) – in symbols, $2^{C^s} = 2^{C^1} + 2^{C^2}$. Inspection of this equation shows that when C^1 and C^2 are both less than unity, their sum is less than C^s (recall that the sum of the logarithms of two numbers is the logarithm of their product). This means that when two channels with independent capacities less than unity (single-event channels with equivocation) are combined by making their input sets interactive, the capacity of the resulting integrated channel is greater than the sum of the capacities of the two channels acting separately. In general, it follows that the receptive capacities of a group of sensory channels can be used more efficiently if the channels operate jointly rather than individually. Given the selective advantages attendant upon perceptual efficiency, it is predictable on the basis of MTC that the visual systems of the "higher" organisms will be highly integrated.

The second strategy for increasing mutual information of a sensory system is to decrease its equivocation, which is equivalent to making it less susceptible to noise. Because a noiseless channel is by definition one in which each output event correlates invariably with a single event at the input, this can be accomplished physically by "tightening up" the connection between input and output, in the sense of making a given input indispensable for a given output event. An instructive illustration of how this might work is afforded by the much-cited work on the visual system of frogs reported in Lettvin, Maturana, McCulloch, and Pitts (1959). One thing these experiments indicated is that the frog's nervous system discriminates only a few specific configurations of visual stimulation – sharp discontinuity in brightness level (e.g., a silhouette against a bright sky), overall reduction in illumination (e.g., a shadow cast by a large predator nearby), a small object about the size of a fly moving a few inches away, and so on. Each of these configurations can be described in terms of its info(t) characteristics, and each can be detected on the basis of the characteristics described. For example, the "sharp discontinuity" detector responds to sharp edges of differential retinal activity, corresponding to lines of light–dark contrast cast upon the retina. One side of such an edge is characterized by relatively infrequent firings, hence by events with relatively high levels of average info(t), and can be picked out on the basis of random samplings of receptor events showing consistently high averages. A series of random samplings on the more highly stimulated side of the edge, similaly, would show consistent measures of average info(t) at a lower level due to more frequent firings. Detection of the light–dark contrast could be accomplished by a mechanism for registering different levels of average info(t) within a local area of retinal activity, the line being indicated by a sequence of markedly different readings among neighboring areas. (I am not proposing that lines of light–dark contrast are detected in this fashion by the frog's visual system, but only that they could be. How they are actually detected is a matter for empirical determination – once we know exactly what to look for.)

Another discovery of these experiments is that the frog's visual system provides pathways of excitation that appear to be assigned exclusively to one visual format or another. In effect, each configuration is signaled to the brain by a preestablished channel, which is reserved for messages of that particular format. Dedicated channels of this sort may be presumed to be highly reliable (noise-free). Inasmuch as a frog can conduct its life successfully at the edge of a pond on the basis of only a few discriminable visual formats, this particular organization of its afferent nervous system represents an efficient use of its limited info(t)-processing resources. Evolution of the frog's visual system, in effect, appears to have exploited the second general strategy for increasing mutual information, that of reducing the equivocation of the information channels conveying messages from the retina.

Yet another strategy must have been used in the "design" of the human visual system, for reasons that appear weighty from the evolutionary standpoint. Unlike the frog, which can get by with predetermined responses to a relatively few distinct visual configurations, the human organism has become dependent upon the ability to discriminate a practically indefinite number of distinct visual situations. This means in turn that the human visual system must be equipped to process indefinitely many different message configurations. It is obviously impractical from the evolutionary standpoint to provide separate channels for each configuration. What apparently happened instead was the development of a novel and highly flexible set of procedures for increasing the mutual information of spatially confined afferent channels. What appears to distinguish the human visual system from that of frogs or cats or other organisms with similar retinas is that it incorporates procedures for adjusting the formats of messages passing through the system, in response to changes in the objective environment. These procedures provide a form of short-term adaptation that I believe to be the essence of human vision.

9. Functional stages of vision as info(t) processes

Any message can be symbolized in a variety of codes, differing in length, configuration, number of code elements, and so forth (see Abramson, 1963, for definitions and discussion). Among the practical applications of MTC is finding the code best suited for a given technological purpose. For the biological purpose of genetic transmission, evolution is to be credited with the "finding" that 4 fixed code elements (nucleotides) are an optimal set for encoding the 20 or so amino acids that must be manufactured for biological growth (see Beadle and Beadle, 1966, for discussion in explicitly communication-theoretic terms). For purposes of visual info(t) processing, the "finding" was of a different sort – namely, that a variable code format provides the best means of symbolizing the many different visual circumstances to which the human organism must respond selectively. The symbols of this remarkably flexible code are the variable patterns that structure our visual perceptions. It is with reference to these patterns, I shall argue shortly, that the intentionality of vision becomes intelligible.

The processes of human vision take place over a cascade (an end-to-end series) of information channels. The cascade extends from sets of physical events at the object end of the series, through sets of electromagnetic events in the adjacent media, hence through sets of electrochemical events in the percipient's retina, to sets of neuronal events in the cortex by which the sequence is completed. Between perceptual object and sense organ are situated as many different channels as there are distinct media of light-wave transmission (distinct air masses, optical devices, etc.). Within the sense organ proper there are further channels, from cornea to aqueous humor to lens and retina; which in turn lead into an upper series of channels, with junctures at optic chiasma, lateral geniculate body, and various levels of the visual cortex. Theoretically the output of the cascade is its juncture with the efferent system, to which it must convey info(t) properly processed for the effective guidance of the organism's behavior. It is somewhere between the retina and this efferent juncture that visual patterns are formed, and that info(t) is transformed into info(s). Although a full-fledged account of what goes on at these various stages is nowhere near reach at present, it should be possible to conjecture reasonably about the types of info(t) processing that might occur at the major junctions. Nothing more is attempted in the account that follows.

The key to this account is the concept of pattern. What I mean by "pattern," informally defined, is a relationship among elements in a set such that when an arrangement of a subset of these elements is specified, the arrangement of the remainder is indicated at a level of probability increasing generally with the size of the subset. As an intuitively accessible example, consider the pattern of stones in an archway. If the location of only the cornerstone is known, there are many different ways imaginable in which the other stones could be arranged. As the arrangement of more and more stones becomes known, however, the number of ways in which the structure might be completed diminishes rapidly; and when all but a few have been set, the location of those remaining is all but determined. Defined in communication-theoretic terms, a pattern is a set of elements the probabilities of occurrence of which (at specific times or places, or in specific sequences) are so interrelated that, as the arrangement of progressively larger subsets of element-occurrences is given, the info(t) conveyed by the remaining occurrences progressively diminishes.

In the account that follows, the functions served by the higher levels of the visual system are all aimed at the formulation of patterns of neuronal activity that serve in the guidance of the organism's behavior, and the elements of these patterns are neuronal events (firings, inhibitions, summations) made interactive by conditioning or inheritance. From the neurophysiological perspective, these patterns are configurations of neuronal events so interrelated that the occurrence of a subset increases the probability of the remainder's occurring, in proportion to the size of the occurring subset (see Sayre, 1976, pp. 130–33, for physiological evidence about how this might occur). From the perspective of visual perception, on the other hand, these patterns serve as representations of salient features of the objective environment. Our task is to account for the origin of these representations – to tell a coherent story about the info(t) processes by which the patterns in question emerge.

9.1. The Retina. In rough estimate, it can be calculated that between 10^5 and 10^8 bits of info(t) are presented on the retina during a typical second of activity. Measured with comparable tolerance, only about 10^2 or 10^3 bits survive on the level of visual awareness (see Sayre, 1976, for calculations and data sources). This amounts to a difference of somewhere between 10^2 and 10^6 bits of info(t) between retina and cortex. In return for this considerable amount of info(t) lost (or expended), there is a gain (from none to some) of info(s). For the thing of topical interest about the patterns of vision is that they are representations of – are directed upon – specific features of the objective environment. A natural beginning of an account of how this happens is with the weeding out of info(t) at the retinal level.

An expedient available to engineers for reducing the load on an artificial info(t) channel is the reduction of noise. In the normally functioning visual system, noise primarily takes the form of retinal stimulation that is irrelevant to the guidance of current behavior (what is irrelevant at one moment may not be so at the next). One technique for reducing noise is smoothing, or averaging, illustrated schematically by a line drawn to fit untidy data points on a graph. A comparable engineering application is the smoothing of hits on a radar display, by way of establishing a vector of aircraft movement. An example pertaining directly to retinal processes is the procedure discussed above in the case of the frog for detecting a line of contrast among many disparate "nerve hits." As already stressed in that discussion, this procedure relies essentially upon the info(t) features of those retinal events.

Another technique of load reduction is the elimination of redundancy. In an engineering context, redundancy is often useful to achieve reliable transmission in the presence of noise, and one of the fine points of communication-system design is to find the fine line between reliability and waste. In the visual system, similarly, redundancy may be assumed to be essential to reliable operation, but in excess it is a drain on limited info(t)-handling capacities. In one's normal view of a clear sky, a large building, or a familiar road, by way of example, the greater proportion of incoming info(t) deriving from these objects amounts to duplication and is suppressed at the periphery without further processing. One available means for accomplishing this is an edge-tracing procedure that "abstracts" the boundaries of areas of relatively invariable retinal activity. (A computer program for tracing edges on the basis of info(t) features was developed as part of a working pattern recognition program reported in Sayre, 1973, and various other methods are also available.)

Presumably these are not the only info(t)-reduction procedures operating at the retina; and probably the ways such procedures are accomplished at the retina bear little discernible similarity to the ways they are accomplished in human technology. Although what is accomplished may be identical (e.g., elimination of redundant info(t), the means of accomplishment of man and nature may be very different.

In this regard it is interesting to reflect upon the remarkable assumption underlying the account of Marr (1982), to the effect that the nervous system actually calculates (p. 149), or infers (p. 44, passim), or computes (p. 297, passim) its representations; and that it does so according to mathematical formulae specified by the author, in the manner of the author's own mathematical inferences. At first glance this line seems distinctly eccentric – like speaking of subatomic particles computing the equations of quantum theory. A more credible line would seem to be that this or that neuronal mechanism in the retina or optic tract operates on visual data passing through it in such a manner as to approximate the results of this or that mathematical transformation, much as the operation of a radio filter on an incoming signal is approximately described by some set of equations. In Marr's behalf, however, it must be pointed out that he is working within the context of computational psychology. Indeed, he makes explicit that in his use of technical terms, he wants to restrict "attention to the meanings associated with machines that are carrying out information-processing tasks" (p. 22). If computers can calculate mathematical functions (which in a limited sense they obviously can), and if the brain is a computer (the germinal thesis of computational psychology), then there should be no problem about the brain – or some part of the visual system – doing mathematics as Marr's manner of speaking requires.

The problem for Marr, however, is that in his view the brain is not merely carrying out operations on formal symbols – formal in the sense of Fodor's formality condition (1980, p. 65) – and hence doing so in a manner independent of what the symbols might be interpreted as being about. According to Marr's view, rather, the calculations the brain is supposed to be carrying out are about various features of external objects (their surfaces, pp. 42, 102; their motion, pp. 162, 166; etc.); and given the position of visual realism that Marr repeatedly stresses, it seems essential that these calculations have semantic reference. Where this approach gets hung up, however, is on the simple point argued earlier in this essay, that the symbols involved in machine calculation in and by themselves are not about anything at all. So neither Marr nor anyone else can insure intelligibility in the talk about intentional information processing by nodding in the direction of machine computation. In a nutshell, buying into the computational vocabulary provides no clarification of how the brain deals with info(s). Since Marr never defines the sense of "information" his mathematical analysis is dealing with, we have reason to remain dissatisfied with his free use of the computational metaphor.

Leaving aside questions of who or what calculates according to exactly which formulae, however, the present approach (much encouraged by the "Pythagorean" urgings of Churchland, 1979 agrees with Marr's in treating mathematics as a major resource for understanding the processes of vision. According to this account, vision begins with the detection via their info(t) content of relatively simple configurations of retinal stimulation, such as lines and shapes and angular orientation (cf. the work with cats reported in Hubel and Wiesel, 1962). The detection of these simple formats is the work of the Retina, with capital "R" to distinguish the function from the particular sense organ.

9.2. The Accumulator. The next step is to group these elementary formats into complex patterns, serving as potential delineations or sketches of features of the objective environment. What I have in mind here is something similar to Marr's "2½-D sketches," but not yet focused on or matched up with particular objects. The role of these sketches is to assemble the elementary formats picked out by the Retina into combinations that reflect persistent configurations of retinal stimulation and hence are candidates for representing particular features of objects.

On one side of the retina exists a determinate set of physical circumstances, revealing its presence in an array of electromagnetic signals. On the other side, I am conjecturing, there are many sets of temporary sketches, being synthesized anew as retinal stimulation changes. One or more at a given moment might match events at the other side of the retina ("match" will be defined shortly in terms of mutual information); but further processing is required before that can be determined.

It should not be assumed, as Marr's account encourages, that these preliminary sketches take the form of a two-dimensional picture. Almost surely they have no pictorial form at all, and they bear no physical resemblance to the neat photographs and diagrams with which Marr's book is so profusely illustrated. But this is as it should be, for what are being reproduced and synthesized in these sketches are not physical features of the perceptual object, but info(t) structures impressed upon the retina. The media in which these structures are momentarily reproduced may be scattered throughout the midreaches of the visual system. What is required basically is a field of loosely interactive neuronal circuits, the activity of which can duplicate and summate the dominant patterns of activity detected by the Retina. Because it would be idle to try to identify a physiological juncture at which something like this might take place, given our current state of knowledge, let us refer to it figuratively as the Accumulator.

In the retina of an organism sweeping its gaze across an unfamiliar scene, many different elementary formats will be detected (by the Retina) from moment to moment, but few will be retained within the Accumulator. When the organism's gaze is directed upon a fixed set of objects, however, the elementary formats issuing from retinal stimulation will remain similar through successive moments and will emerge in the Accumulator as temporary sketches. An info(t) process by which this preliminary sorting could be accomplished is based upon a familiar concept of MTC, that of a Markov source. A Markov source is equivalent to an information channel with input and output drawn from the same set of events (the source alphabet) and related by specific conditional probabilities of occurrence. In a first-order Markov source, the probability of occurrence of a given member of the source alphabet (an output event) is a probabilistic function of the one event immediately preceding it (an input event). In general, with an nth-order Markov source, any given output event is conditional upon the occurrence of the n events preceding. Application of channel measures (like mutual information) to a Markov source is assured by the equivalence of an nth-order Markov source to an n-membered cascade of information channels with inputs and outputs drawn from the same alphabet (see Sayre, 1976, pp. 29–30, for reasons).

By treating the series of elementary formats issuing from the Retina as a Markov source, the Accumulator can test for repetition by applying mutual information measures over successive members of the series. A reading of high mutual information over n successive members would indicate n repetitions of similar configurations on the retina. By appropriate adjustment of n, the Accumulator can make available for further processing a set of more or less persistent retinal formats.

9.3. The Discriminator. This discussion of operations on info(t) structures at various stages of the visual cascade is obviously very schematic. Its purpose is not to provide even an approximate physiological description, but rather to indicate a sequence of info(t) processes that could yield representations with intentional features. To characterize the next stage in the sequence, it will be necessary to bring the effector capacities of the organism into the picture. It has been suggested by several commentators (e.g., Haugeland 1981; Searle 1980a) that consideration of mechanisms through which an organism interacts with its environment might be necessary for an adequate account of intentionality. For the present account, we want to consider not only the effector mechanisms, but also the overall perceptual–behavioral control loop guiding the organism's responses to a changing environment.

The perceptual side of this control loop includes the visual system, channeling info(t) from retina to cortex, as well as channels feeding into the retina that link the visual system to the perceptual object. The behavioral side of the control loop begins with the various stages of the efferent nervous system, initiating control signals that are translated into bodily activity undertaken with respect to objects in the proximate environment. In standard behavioral and perceptual circumstances, when the organism is active and its control system is functioning normally, its bodily activity is directed toward, or with respect to, the same object, or set of objects, that produces the signals at the input of the perceptual side of the loop. The primary function of the control loop is to enable the organism to adjust its behavior with respect to those objects in a manner responsive to its needs and interests.

For the behavioral side of the control loop to serve its function adequately with respect to a given object, the perceptual side must provide an adequate representation of that object at their cortical interface. Considerations of adequacy in representation bring us back to the third major stage of visual info(t) processing.

The overall function of the third stage of info(t) processing is to formulate stable patterns of neuronal activity, based on the sketches of the Accumulator, that match in their info(t) characteristics the features of the objects with which the organism is interacting. As was the case also at the Accumulator, the representations at this higher level have no orderly spatial characteristics, and hence do not match in the sense of image or picture. Their representation is with respect to info(t) features that have been preserved through the many channels beginning with the perceptual object. Because the role of this stage, in a broad sense, is to highlight the particular features of the object that are implicated in current behavior, we may refer to this stage as the visual Discriminator.

The Discriminator is the main locus of visual adaptation. Its input is a set of composite patterns of varying degrees of completion and detail – that is, the temporary sketches composed by the Accumulator. Its output is a stable pattern highlighting those features with which the organism is dealing in its objective environment. The transformation it accomplishes between input and output is a trial-and-error procedure, in which various temporary sketches in various combinations are tested for success in the guidance of behavior.

There are two types of case that must be considered separately in illustrating the function of this info(t)-processing system. The first assumes the existence of a pattern-storage faculty, like long-term memory except in not being propositional. This is the faculty of Percept Storage, which contains patterns proven successful in past perceptual ventures. In the first type of case, the Discriminator cooperates with Percept Storage in finding stable patterns that will both (a) serve well in the guidance of current behavior and (b) maintain a high degree of mutual information with events at the Retina. Application of criterion (b) involves treating the series of information channels between Retina and Discriminator as an nth-order Markov source, with n's value adjusted according to prevailing rate of change in retinal stimulation (lower n for higher rate of change ultimately in the visual field). Patterns meeting this criterion exhibit a high degree of correspondence in info(t) features with repeating patterns of retinal activity. Application of criterion (a), on the other hand, involves temporarily "advancing" a pattern to a position of control over the activity of the efferent system, and testing out its effects on the guidance of behavior. If the pattern functions well (as measured by impending degree of satisfaction of the organism's immediate needs), it continues to serve in this capacity until change in circumstances dictates a change in perceptual focus. If it serves moderately well, but needs fine tuning, the pattern will be augmented by greater detail, or changes in detail, from resources available in the current stock of the Accumulator. If the pattern serves poorly or not at all, it will be replaced quickly by another at the position of control.

A case of the second type is one in which Percept Storage has no patterns available that exhibit high mutual information with events at the Retina. This would typically be a case in which the organism is presented with objective circumstances it had not encountered previously and hence cannot draw on experience for visual orientation. In such a case, the Discriminator vacillates rapidly between criteria (a) and (b), seeking out a combination of Accumulator sketches that will maintain a moderately high level of correspondence (mutual information) with events at the Retina while at the same time showing promise in the guidance of behavior. When a pattern emerges that promises to satisfy both criteria at once, it is refined by testing changes in detail available from current Accumulator output, and by engaging the object perceptually from different angles and perspective. If and when a pattern is found that enables the organism to "lock" its perception firmly on the circumstances of the object – when, that is to say, the pattern is reinforced by success – it will be set aside for future use in Percept Storage. These are the major outlines of the info(t) process of percept learning.

A comparison with Piaget at this point may be instructive. Piaget (1954) distinguishes two operations with sensory-motor schemata that he considers essential to our perception of objective reality. One is "assimilation," which is a matter of organizing sensory inputs under previously established schemata; the other is "accomodation," which is a matter of developing new schemata when the old prove inadequate. As Piaget describes the contrast, "assimilation is conservative and tends to subordinate the environment to the organism as it is, whereas accommodation is the source of changes and bends the organism to the successive constraints of the environment" (p. 352). Piaget's sensory-motor schemata are similar in obvious respects to what I have characterized as patterns controlling the efferent mechanisms, his assimilation corresponding to the process of bringing incoming retinal stimulation into relationships of high mutual information with patterns already in Percept Storage, and his accommodation to what I have called percept learning. Insofar as the communication-theoretic framework is more exact in its basic concepts than the one in which Piaget was working, it may provide a context in which the latter can be more precisely formulated. By reverse token, happy to say, Piaget's findings may provide material by which the former can be refined and expanded.

9.4. The Abstractor. One further stage of info(t) processing is required before visual patterns take on features familiar to introspection. The role of the Discriminator, summarized, is to adjust the details of the afferent patterns brought to the effector mechanisms for the guidance of behavior until a set of patterns is found that enables the organism's perceptual–behavioral control system to "lock on" to appropriate features of the objective environment. Once the control system is locked on and the organism's behavior is under effective perceptual control, it will typically turn out that effective control can be maintained with patterns of appreciably less detail. Although considerable perceptual detail may be required to discern a walking stick on a bare branch, or a quail in the underbrush, or a stop sign among billboards and neon lights, once the object has been picked out and identified it can be held in view with less attention to details. The function of this final level of info(t) processing is to cut back on the details of perceptual patterns without compromising their effectiveness in the control of behavior. From the negentropic viewpoint specifically, its role is to enable the organism to retain an effective environmental coupling with minimum drain upon its info(t)-processing resources. This it accomplishes by trimming perceptual patterns of unnecessary detail; in effect, by abstracting those features that prove essential for guidance. We may refer to this function as the visual Abstractor.

The info(t) processes behind the operation of the Abstractor are described with reference to the equation defining mutual information $I(A;B)$ as input entropy(c) $H(A)$ minus equivocation $H(A/B)$. To apply this equation, let $H(A)$ be the average info(t) of the set of elementary formats produced by the Retina at a given moment, and $H(A/B)$ the equivocation of the Discriminator with respect to the Retina. The more detail on the average among patterns at the Discriminator, the less equivocation between these patterns and the elementary formats of the Retina. The less the Discriminator reflects the diversity of detail at the Retina, on the other hand, the greater the level of equivocation ($H(A/B)$. As a matter of working definition, when the control system is locked on to the appropriate environmental object, the mutual information $I(A;B)$ of the overall channel A-B (Retina to Discriminator) is holding at an adequate level. Quite simply put, the role of the Abstractor is to reduce the level of detail at the Abstractor (increase the level of $H(A/B)$) to limits compatible with maintaining adequate mutual information $I(A;B)$ of the visual tract, given the current entropy(c) $H(A)$ of events at the Retina.

Sketchy as this operational description of the Abstractor inevitably has been, it is now possible to sample the type of explanation it offers of familiar Gestalt phenomena. The "vase-face" illusion can be understood as a consequence of the role played by patterns from Percept Storage in the shaping of the perceptual field. Illusions of this sort occur invariably in connection with shapes and contours that are familiar to the observer from past experience – that is, with shapes and contours well established in Percept Storage. In typical cases of patterning by stored percepts, the info(t) configurations arriving from Retina and Accumulator have been derived from objects with clearcut identity, so that any momentary ambivalence about the proper percepts to be used in further shaping of these configurations is quickly resolved during further interaction between organism and object. With a display engineered to trigger "vase-face" reversal, however, the stimulus configuration is designed to produce Retinal activity standing in equally high mutual information with both face and vase percepts, and to yield no cues that would give precedence to either display under further scrutiny by the observer. In such a case the Discriminator will oscillate between equally dominant patterns as it attempts to organize the visual field.

The phenomena of figural constancy illustrate the role of the Abstractor function. Phenomenologically considered, figure constancy is the tendency of representations within the visual field to retain a fixed configuration, despite variations in visual perspective and momentary interruptions in retinal stimulation. Considered with reference to info(t)-processing operations, figural constancy results from the tendency of the visual system to produce patterns incorporating as little detail in info(t) structure as is compatible with effective operation of the perceptual–behavioral control loop. In the moment-by-moment functioning of the Abstractor, this amounts to suppressing detailed changes in perceptual patterns as long as these patterns remain adequate in their control of behavior. Inasmuch as our day-to-day dealings with familiar objects do not rely upon fine discriminations of differences due to perspective and distance and are not thwarted when objects drop momentarily from view, details of this sort are typically not registered in the patterning of our visual fields. They are abstracted from the patterns produced by the Discriminator unless they prove essential for the guidance of the organism's current behavior. The effect is a constancy in patterning from moment to moment, until changing circumstances dictate more attention to detail.

Observing that the stimulation of the receptor and resulting sensations are "variable and changing in the extreme," Gibson posed as *the* "unanswered question of sense perception . . . how an observer . . . can obtain constant perceptions in everyday life on the basis of these continually changing sensations" (1966, p. 3). Gibson's answer, much simplified, is that the senses constitute perceptual systems that detect invariant structures in the energy flux surrounding the organism which provide information about the permanent environment. What remains deeply problematic about Gibson's view is that even in its unsimplified version it provides few details about how exactly the perceptual systems accomplish this, or about the nature of the information they are supposed to provide. Marr (1982) is on target with the complaint that although Gibson "asked the critically important question" of how constant perceptions can be obtained on the basis of ever-changing retinal stimulation, "he did not understand properly what information processing was, which led him to seriously underestimate the complexity of the information-processing problems involved in vision" (p. 29). Although Marr certainly did not underestimate the complexity of these problems, there remains room for serious doubt whether Marr himself has understood the informational character of these processes. For not only does Marr provide no discussion of the sense of "information" in which vision is an information-processing phenomenon, but moreover there is the deep unclarity already noted about how neuronal mechanisms can perform calculations about external objects. Without ruling Marr's approach out of court in its entirety, it seems fair to claim as an advantage of the present account not only that it offers a fairly straightforward explanation of how "constant perceptions can be obtained from continually changing sensations," but moreover that it does so in terms of a precisely defined sense of "information."

10. Intentionality

The account has been silent thus far about info(s); and intentionality has been mentioned only in anticipation. The stage is almost set for these two main actors, however. I have only to recapitulate a few essentials, by way of introducing a few new labels.

It is natural to think of the perceptual side of the perceptual–behavioral control loop as divided into an anterior branch, extending from object to retina, and a posterior branch, from retina to visual cortex. The input to the anterior branch, and hence to the visual cascade overall, is a set of physical events at the surface of the object – the events responsible for the particular structure of light radiation reflected from the object in the direction of the observer. Let us refer to this set of object-centered events as the input set O. The output of the anterior branch is the set of events R at the observer's retina, which at the same time constitutes the input of the posterior branch. In normal circumstances the retinal set R possesses high mutual information with respect to the object set O, which is to say both that O is capable of issuing a wide variety of distinct signals – that is, that it is characterized by high entropy(c) $H(O)$ – and that R by and large provides reliable indications of events at O – that is, that channel O-R is characterized by low equivocation $H(O/R)$. Given the normally high mutual information $I(O;R)$ of the channel O-R, there is a sense in which the info(t) present at R is literally the same as info(t) present at O. In their probabilistic features, the structures established at R faithfully reproduce (at least some of) the structures at O. Hence (at least some of) the info(t) present at O is also present at R.

What is important for the perceptual process ensuing is not the two-dimensional image set up in the retina, inasmuch as spatial features are lost in the higher reaches of the visual system. What is important are those info(t) structures of retinal events that are identical to corresponding structures in the objective environment. This identity is retained at subsequent stages of visual processing, but only for structures actively involved in perceptual guidance.

The posterior branch of the visual cascade extends from R to a set of cortical events C. Unlike the anterior branch, however, whose function is to get as much reliable info(t) from O to R as the retina can handle, the posterior branch has as its main function to cut back radically on the amount of info(t) passed on to the cortex, while at the same time retaining just those info(t) structures that are important for the guidance of the organism's current behavior. The processes by which this is accomplished are those described in the preceeding section (9.2–9.4). When these processes are working properly the organism is enabled to adapt its behavior to a variable environment, under the guidance of perceptual patterns that impose a relatively light drain on its info(t)-processing resources while maintaining a high level of mutual information $I(R;C)$.

Since the set of cortical events C stands in a relationship of high mutual information $I(R;C)$ with respect to the retinal set R, and since R shares high mutual information $I(O;R)$ with O, it follows that the mutual information $I(O;C)$ between object and cortex is comparably high. In brief, if both channels O-R and R-C are characterized by high mutual information, then so is the combined channel O-C. (Except under special conditions, $I(O;C)$ will nonetheless fall somewhat short of $I(O;R)$; for these conditions see Abramson, 1963, p. 115.) This means that by and large the info(t) structures present at C are the same as (some of) those at O – that the neuronal events in the observer's cortex faithfully reproduce, in probabilistic structure, certain events at the object end of the visual cascade. These events of C are identical with events at O.

The identity in question is not pictorial. There is no picture at C of events at O – no duplication of color, shape (pace Locke), or "logical form" (pace Wittgenstein). The identity in question is of info(t) structure, of a sort akin to the identity of mathematics (as $\frac{1}{3} = \frac{2}{6}$). (Historical note: perceptual realism is the position that events in the mind of the perceiving organism are somehow identical with events in the object perceived. The present theory is a version of perceptual realism, maintaining an identity of info(t) structure at O and C.)

Looking backward to the topics of Section 7, we see that because of the high level of $I(O;C)$ sustained by the perceptual process, the perceiving organism enjoys the benefits of a highly efficient negentropic coupling with its objective environment. And because of the indefinite variety of patterns this process makes available for guidance, the organism rates high in negentropic flexibility. This throws light upon the emergence of patterned vision in the selective processes by which the human organism was shaped. For, as pointed out above, the general drift of natural selection is to single out organic processes with a high degree of negentropic flexibility.

But what about intentionality and info(s)? The answer in general is that the relationship of identity in info(t) structure between O and C *is* the intentionality of perception, and that a structure at C enjoying this relationship possesses the corresponding structure at O as its semantic content. This is the answer in general; but qualifications are needed.

First, it should be clear that not just any relationship of info(t) identity between input and output structures across a cascade of information channels constitutes a relationship of intentionality. It is not enough even that the channel in question be involved in the perceptual guidance of an organism's behavior. The channel controlling pupillary contraction might possess high mutual information with respect to illumination level, for instance, but does not thereby exhibit intentionality. What is required is a perceptual process that locks on to particular structures in the objective environment and tracks those structures through changing perceptual circumstances. The point of stressing identity of structure between cortex and object is not merely to adopt a familiar (philosophic) manner of speaking about the perceptual process, but rather to emphasize the fact that cortical structures thus engaged are literally focused upon particular features of an objective environment. The relationship of "being focused on" is a necessary component of perceptual intentionality.

A second requirement is that the focus of cortical upon objective structure must be maintained by adaptive procedures that enable the former to adjust to relevant changes in the latter. In terms used above, info(t) reaching the upper levels of the posterior branch of the visual cascade is processed in a variable coding format, the configurations of which are continuously being adjusted in interests of efficiency. The relatively static formats in which the frog's brain receives reports from its visual sensors do not qualify for intentionality. Perceptual intentionality is a dynamic relationship. While the perceptual process maintains a steady match between cortical and objective structures, both poles of the relationship are constantly changing.

A helpful analogy might be drawn here with the television tuner (just the tuner; there are only misleading analogies with the video-tube display). The upper levels of the human visual system deal with info(t) impressed across the retina in something like the fashion in which a television tuner deals with info(t) impressed across its antenna. The antenna picks up signals on many different wavelengths; but only those get through for amplification that correspond to the resonant frequency of the variable tuning circuits. Similarly, signals from many sources are presented to the retina; but only those signals become encoded for further processing that respond to the organism's current needs for perceptual guidance. What this amounts to in either case, quite literally, is the intervention of circuits at the upper levels of the respective cascades whose activities possess a high degree of mutual information with respect to selected sets of input signals. An important disanalogy, however, comes with the manner in which these upper level circuits are adjusted; for the tuner is adjusted by external dialing. In the human visual system, by contrast, the structures with which the info(t) processes of the system culminate are selected for their ability (1) to focus the percipient's behavior on changing features of the objective environment, (2) with minimum outlay of information-processing capacity. The tuning of the visual system is a process of continued adjustment, by which the organism preserves balance with an ever-changing environment.

A third requirement standing behind these two is that the system must be under load to exhibit perceptual intentionality. This means that the system must be operating in relation to an objective environment so that its info(t) coupling makes a practical difference in the quality of its continued operation. The case may be otherwise with the intentionality, say, of speculative reason. But perceptual intentionality is a feature of info(t) processing in certain behaviorally involved organisms, as they operate under constraints from a demanding environment.

With this, the way is paved for info(s). In visual perception, to recapitulate, the relationship of intentionality is quite literally a relationship of high mutual information between a set of objective circumstances and a representation in the cortex of the perceiving organism. The representation picks out that particular set of circumstances, by virtue of its being the only object in the perceptual scene with which the representation shares that relationship. Through such a representation the organism's perception is directed upon a specific object, which is thereby the object to which the representation refers. By sharing identically in its particular structure, the representation is true of the corresponding object.

But these are precisely the characteristics – intentionality, reference, truth, direction upon an object – that serve as paradigms of semantic features in CS literature (see Section 1). Beginning with the concept of info(t), we have traced the origins of info(s).

Although this account is far from complete and may require correction in numerous respects, it is fashioned to capture a set of necessary requirements that together are sufficient for intentionality. The effect intended is that any system constructed so as to be able to process info(t) from its environment, in the manner and with the results laid out above, would for that reason exhibit perceptual intentionality. To put it another way, if perceptual intentionality were understood as proposed in this essay, then (allowing for inevitable refinements and corrections that will be required as the theory develops) we would never find ourselves in the position of confronting a system of this description and having to admit that it was devoid of intentional features.

It is not essential that the system be a human organism. Although the account has been developed with the help of illustrations drawn from human experience, the human character of these experiences has been incidental. What is essential is an info(t)-processing system capable of formulating representations that (1) are identical in info(t) structure with specific features of the system's perceptual environment, (2) vary adaptively in their formulation with variations in their corresponding objective structures, and (3) make the difference between success and failure in the system's dealings with a hazardous environment. There are undoubtedly many respects in which the human organism differs in its info(t)-processing capacities from other organisms, but intentionality of perception is not necessarily among them. Although the issue is one for empirical determination, there is no reason by the present account why animals other than human should not possess this feature. Although fishes and reptiles are probably excluded, higher primates and felines presumably qualify.

It is not essential even that the system be biological in origin. Although the account alludes to the likely evolutionary origin of the adaptive processes involved in perceptual intentionality, there is no reason in principle why info(t) processes of this general sort might not be designed into an entirely mechanical system. I have no idea how close to the present state of the robotic arts such a system might be. But advances of two sorts at least appear to be in order before an intentional robot is feasible. One is the perfection of data-processing systems that can compute statistics on their own operations, in real time and in a manner reflecting fluctuations in the environment from which their data are derived. Statistical computations of this sort would be necessary to determine the continually changing value levels of the communication–theoretic parameters ($H(A)$, $H(A/B)$, $I(A;B)$, etc.) upon which the info(t) processes described above are based. The second is a much fuller understanding than we can muster at present of the physiological mechanisms by which these info(t) processes are accomplished in biological organisms. As Pylyshyn has remarked, discussion of machine intentionality is often handicapped by the fact that "we cannot state with any degree of precision what it is that enables us to claim that *people* refer" (1980a, p. 444). Although Pylyshyn might not look favorably upon various features of the account developed above, he would undoubtedly agree that knowing how intentionality works in humans would aid considerably in building intentional machines. If the account above is basically correct, at any rate, advances in our understanding of human perception may be expected to precede our design of mechanical counterparts.

At the same time, experiments in the design of mechanical counterparts will surely help us test out and improve our theories of how the original operates. So AI still has a role to play in the alternative approach laid out in this essay. The time may even come when we learn enough about other forms of intentional info(t) processing in the human organism – for example, the processes of language and reason – to be able to build similar capacities into computer-based systems. And presumably such systems, when they arrive, will manipulate symbols. But the reason for their intentionality will be their special forms of info(t) processing, and not the mere fact that they are symbol manipulators.

Two questions were posed at the beginning of the essay: (1) What do the functions of the nervous system that we ordinarily describe in the intentional idiom actually do? and (2) What is it about the information involved in these functions that makes it semantic? Visual perception is a particular function of the nervous system ordinarily described in the intentional idiom. The answer to (1) proposed in this regard is that visual perception establishes variably coded representations in the cortex of the perceiving organism that share a high degree of mutual information with changing events in the objective environment. The answer to (2), in turn, is that the info(t) involved in the various stages of the visual process becomes info(s) when a representation, by virtue of its high mutual information with a particular object, picks out that object uniquely – becomes directed upon it – and hence refers to that object specifically in the visual environment.

11. Postscript

This account of intentionality is obviously incomplete, not only in requiring additional conceptual and empirical development, but also in being limited to intentionality of a rather low-level sort. It tells us nothing as it stands about language or reasoning, in which CS is no less interested than in the processes of perception. To show that the present account is not bereft of implications for these higher level activities, I wish to sketch out very briefly the view it opens up on language and reasoning. (These final remarks are not intended for critical discussion, being much too scant for that purpose, but only as a gesture indicating directions worth exploring.)

In Section 9.3, stable patterns of neuronal activity involved in the processing of perceptual info(t) were referred to as percepts. Percepts are normally activated by stimulation at the sensory periphery and are thus controlled by the external environment. A related account can be given of meanings, which are in effect percepts freed from stimulus control. The meanings in question, of course, are not abstract entities, but rather are cortical patterns actually functioning in the info(t) processing of linguistically competent organisms. Because meanings, like percepts, are patterns of neuronal activity, their structure can be defined in terms of MTC. And because the same info(t) structures can be present in different sets of activities (as with the identical structures at O and at C cited in the discussion above), the same meanings can be present in different cortices. Linguistically competent organisms can be conditioned to activate a given meaning structure on the occasion of a given auditory signal. And linguistic communities emerge when many individuals learn to activate the same meanings upon the same signal presentations.

Although initially under the control of verbal stimuli, meanings also come to control each other through relationships of redundancy (see Section 9.1). The meaning "ripe," applied to bananas for instance, is compounded of the meanings "yellow" and "handlength" in such a fashion that application of "ripe" in the organism's info(t)-processing activities ensures applicability of the other two meanings. Application of "ripe" thus renders the application of "yellow" redundant. By reverse token, applicability of "yellow" is necessary for the application of "ripe," thus controlling the latter in a restrictive sense.

Concepts, in turn, may be conceived as meanings thus removed from stimulus control (by either vocalization or object perception) and brought under the control of other meanings. Whereas the percept "yellow" is normally activated only by yellow objects, and the meaning "yellow" by either objects or auditory signals, the concept "yellow" can be activated not only by sounds and objects but also by the meaning "ripe" itself. Thus understood, percepts, meanings, and concepts alike are structures of neuronal activity, distinguished with respect to source of control.

The redundancy relationships controlling concepts are basic factors in human reasoning, in two respects at least. One respect has to do with the notion of a conceptual linkage. Concepts belong to a shared linkage if they are mutually relevant to each other's application, perhaps through the mediation of other concepts within the same linkage. The concepts "ripe," "yellow," and "handlength" thus share a linkage with the concept "banana," an association resulting from the coincidence of color and size properties discovered during standard experience with edible bananas. If a variety of bananas were encountered that are not ready to eat when yellow, on the other hand, the relevant conceptual linkages on the part of the individual or individuals attempting to eat this fruit would soon adapt to the novel set of circumstances. Due to the public nature of the language from which concepts are derived, however, it is not necessary for a given individual to undergo the relevant experiences in order for his conceptual linkages to be appropriately adjusted. They can be adjusted through discourse with the affected individuals. Conceptual linkages are thus media of information storage, subject to augmentation and correction through the experience of subgroups within the linguistic community. In effect, conceptual linkages are abstract mappings of percept associations, adjusted by continued encounters with a shared living environment. By using these maps to plot the course of anticipated behavior, a rational agent can explore alternatives before committing himself to action.

A second respect in which conceptual redundancy serves reason has to do with the structure of logical inference. As any student of elementary logic should know, the paradigmatic form of deduction was long considered to be the syllogism: if all M is P, and all S is M, then all S is P. Because the variables in question are usually interpreted as standing for classes or concepts, an equivalent formulation would be: if M renders P redundant, and S renders M redundant, then S renders P redundant. Since the redundancy relationship can be precisely defined in communication–theoretic terms, deduction in this basic form can be understood as an info(t) process in the cortex of a rational organism. Reason is thereby brought within the causal matrix, where it contributes directly to rational behavior.

Unlike the case with percepts, the info(t)-processing role of meanings is not primarily the guidance of ongoing behavior, but the coordination of activity among different individuals. In the case of concepts in turn, the role is maintenance of the organism in a state of readiness to deal with future experiences. With these higher level cognitive functions, intentionality is a relationship of high mutual information with objects that may or may not be perceptually present. The intentionality of percepts is initially established through a connected series of info(t) exchanges extending between the cortex and a perceptually present object. But there is nothing in the relationship of mutual information that requires a direct causal connection between participating structures. And once a relationship of high mutual information has been set up with a particular object or set of objects, that relationship may continue when the object is no longer sensibly present. The intentionality of language and reason is thus derivative from that of perception. Such is the case also with cognitive activities like searching and hoping where the object in question is expressly absent.

These summary observations on the nature of language and reason are further elaborated in Chapters 11 and 12 of Sayre (1976), which also show how the relationship of conceptual redundancy can be precisely defined in terms of MTC. Sayre (1979) contains a mathematical proof of the validity of the basic syllogistic form expressed in terms of that technical theory.

Further development of these notions, as well as further refinement of the account of perception that underlies them, depends essentially upon the criticism and help of the CS community. I hope this help might be forthcoming despite differences over the fruitfulness of the computational paradigm. Among substantial points of agreement with the computational approach as matters stand are (1) an emphasis upon the functional analysis of cognitive phenomena, (2) a favorable attitude toward computer modeling, and (3) a predilection for the terminology of information processing. Among notable divergences, on the other hand, are (4) the central role played in the present account by the biological sciences (contrary to protestations that psychology ought to be autonomous; for example, Fodor, 1981b, introduction), and (5) its stress on the technical sense of information as treated in MTC. However, it should be noted regarding (4) that the present account is not reductionistic, and regarding (5) that if computers are ever endowed with intentional features, communication theory might be expected to help substantially. So maybe even (4) and (5) are not unforgivable.

At the very least, it should no longer be credible to remark, with Pylyshyn, that "only one remotely plausible answer has ever been tendered" (the computational answer), "compatible with a monistic" view of causation, to the question of what properties an entity must have "to function in a manner that depends on what it represents" (1980b, p. 159). The deep-lying purpose of the present account has been to tender an alternative answer to just that question.

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Not an alternative model for intentionality in vision

R. Brown, D. C. Earle, and S. E. G. Lea
Department of Psychology; Washington Singer Laboratories, University of Exeter, Exeter EX4 4QG, England

We will consider four questions about Sayre's argument.

1. Has Sayre resolved the problem of intentionality in visual perception? We think not, for two reasons. First, the (admittedly imprecise) terminology of "picking out" implies an essentially *efferent* process in object recognition: the semantic system (which Sayre does not specify) is given an active role in determining what is perceived. A theory of the intentionality of perception, however, should presumably offer a way in which visual input can specify uniquely what a given percept is "about." Sayre points to the ambiguous figures, but from our point of view, what is remarkable about vision is not that it is occasionally ambiguous, but that it is so nearly always well determined.

Secondly, Sayre's proposed solution is so underspecified as to be impossible to evaluate. His hypothesised stages of processing probably contain more free parameters than the processes they are trying to account for. The formalism of his account, in terms of communication theory, is illusory: the "Retina," "Accumulator," "Discriminator," and "Abstractor" are homuncular entities whose behavior is quite unspecified, and it is impossible to tell what intentional characteristics are implicitly embedded in them.

2. Is Sayre's approach novel? In important respects it is not. The idea of a semantic resultant of perception helping to determine what is perceived is not new. It is, for example, the key feature of the "analysis by synthesis" model put forward by Licklider (1952) and popularised by Neisser (1967). Sayre says little about the nature of his info(s), except to reject, explicitly, procedures and networks as vehicles for meaning. But so far as we can tell (largely from Section 11, which is supposed to be off-limits for criticism), his model for the activity of representations involves the activation of particular nodes in some kind of associational network not very different from the classic proposal of, for example, Collins and Quillian (1969). This strengthens the analogy to the analysis by synthesis model.

3. Is the proposed model truly an alternative? We have to ask, alternative to what? Explicitly, Sayre is opposing what he calls the "official" stance of cognitive science, of which "the 'basic idea' is to conceive the brain as a computing device that accomplishes major cognitive tasks as the accumulated effect of many subtasks." Implicitly, Sayre's model must be taken as an alternative to that of Marr (1982), currently the most influential general theory of vision. Of these two models, Marr's is much the more detailed and precise; but is Sayre saying anything that is really different?

Sayre ignores Marr's proposed three distinct levels of explanation (computational theory, algorithm and representation, and neural implementation). Sayre's criticisms, if they apply at all, apply to the second of these, and Marr himself (p. 28) criticises conventional artificial intelligence theories for their exclusive attention to algorithms. Marr's computational theories are designed to specify what must be computed, and why this and only this must be computed, in order to construct a representation of aspects of the world (e.g., shape, space, spatial arrangement, p. 36), and thus to provide information *about* the world. The theories draw on real-world constraints incorporated into an evolved biological system. Insofar as Marr is successful in producing such theories, they provide the analysis of the precise relationship between environment and representation that Sayre requires; and they do so without Sayre's apparently efferent, "picking-out" process.

We can conclude that Sayre has indeed proposed an alternative to current views, specifically those of Marr. But it does not seem to us to be a better one. Marr's account provides a positive heuristic for empirical investigation of vision. Sayre's information-theoretic account is far too general to make straightforward empirical predictions. As an example, consider Marr's computational theory of stereopsis. It may well be incorrect (Mayhew & Frisby 1981); but it specifies what must be done in order to solve the correspondence problem uniquely, and why that solution will be physically correct (Marr, pp. 111–16). Marr gives a choice of algorithms by which to implement the computational theory; they specify how it is possible to construct a unique representation of the orientation and depth of visible surfaces, which must then be the correct representation of these surfaces. This seems to us to be a better solution to the problem of intentionality of vision than Sayre is able to offer. If Marr's account is wrong, a different account within the same general framework is more likely to bring progress than Sayre's retreat to ideas that failed a quarter of a century ago.

Can Sayre's account of the intentionality of vision be extended? Sayre mentions the need to extend his account to language and reason; in the broad context of adaptive behavior that he uses, it might be more relevant to consider its extension to the intentionality of action.

As regards reason, however, we have only Sayre's concluding remarks to suggest the kind of extension he would favor. He says, "As any student of elementary logic should know, the paradigmatic form of deduction was long considered to be the *syllogism*," and he proceeds to draw a parallel between (presumably) the set-theoretic notation of the syllogism and the set-theoretic translation of shared information or redundancy. He concludes that "reason is thereby brought within the causal matrix, where it contributes directly to rational behavior." This conclusion is obviously circular, but in any case any student who has gone much beyond elementary logic should know that the syllogism is far from being paradigmatic. Indeed, the explanation of syllogistic inference can be seen as a test case for cognitive psychology, involving far more than some simple set-theoretic mental logic (Johnson-Laird 1983).

As for the intentionality of action, it is a traditional crux of philosophical psychology; yet somehow psychology has managed to make progress as a science without the dispute ever being resolved. Indeed, the philosophical battle has been waged most fiercely over questions in conditioning and learning (e.g., Taylor 1964), yet the analysis of these phenomena is, as Sayre notes, among the most solid achievements of psychology.

In conclusion, we find the formalism of Sayre's presentation misleading. The ideas he expresses are loose; they are not totally original; and when they have been used in the past, they have failed. As an account of intentionality of visual perception, his target article seems to us to offer nothing to make us prefer it to existing theories, in particular that of Marr (1982), and as an account of intentionality in general it seems to have very little potential.

Semantic content: In defense of a network approach

Paul M. Churchland

Department of Philosophy, University of California, San Diego, La Jolla, Calif. 92093

Sayre's applications of the general mathematical theory of communications (MTC) to the special case of perceiving, cognizing, acting creatures are as illuminating and compelling as any I have encountered. Considered as natural objects with a hierarchy of internal complexions coupled very sensitively to each other and to the environment, sentient creatures emerge as clear cases of information channels, and hence as special instances of the general phenomena addressed by the MTC. And given the cognitive complexity of many living creatures, it is fair to expect that we will turn out to be, in various ways, *interesting* instances of the mathematical concepts at issue. The picture Sayre paints in answer to his question (1) seems to bear out this expectation. I have no criticisms of this part of his discussion. It inspired me, and I hope it inspires others to continue exploring the same approach.

On the narrower matter of question (2), however, I must decline to accept Sayre's answer. I do not think his well-turned sketch of tight negentropic coupling will provide a very illuminating or faithful account of the common-sense notion of semantic information, nor will it provide a solution to "Searle's Problem" (Searle 1980). For an adequate solution to both of these, I believe, we must for the following reasons return to the general area of the "procedural" or "network" approaches rejected by Sayre.

One of the problems with trying to explicate a state's semantic content in terms of its peculiar relations with the external environment is that we can always contrive counterexamples where the relevant environmental relations are widely divergent, and yet the semantic content of the states at issue remains the same.

The simplest illustration of this is the thought experiment in which a sleeping man has his brain removed from his body and placed in a vat, there to be stimulated at all afferents and monitored at all efferents by a supercomputer, one which stimulates in every detail a normal but "fictive" life for the unknowing victim. The point of the example is that the brain's informational coupling to the environment has thus been radically changed. It is now interacting with the ebb and flow of a computer's insides, instead of with the ebb and flow of the man's home and job and family. But has the semantic content of its perceptions and thoughts and intentions been changed as a result? I take it the answer is no. Were we to listen in on the man's subjective adventures by monitoring his fictive speech, we would ascribe semantic content just as before ("Now he thinks he's at a baseball game; now he's watching a double play; now he's asking for a hot dog"; etc.). In particular, we would not reconstrue his fictive speech as being systematically about the shifting configurations of flip-flops in a computer. Our victim knows nothing about computers. He is not even aware that his situation has changed.

Examples like this illustrate the divergence between the semantic content of a state and its nomological or statistical relations to environmental objects and states of affairs: One can have the same semantic content even where one has different environmental relations. A divergence shows up in the other direction as well, for there are real cases of people enjoying the same coupling to the same environment, but whose semantic contents are very different from each other. We need only think of people and cultures who bring very different concepts and theories to their perception and understanding of the same external phenomena. Where you see lightning and hear thunder, a primitive sees heavenly fire and hears god shouting, and a physicist sees a sudden flux of electrons and hears the sudden thermal expansion of the air within its path. Same external phenomena, same coupling, but different semantic content in all three cases.

The lesson of these examples is that the intentional states of cognitive creatures enjoy nomic and statistical connections not only with the environment, but also with the enormous range of *other* intentional states, actual and possible, to which the creature is subject. Intentional states are part of an intricate and ongoing economy of such states, and the semantic content of any such state is determined primarily, and perhaps exhaustively, by its peculiar causal or computational role *within* that complex cognitive economy. This returns us to a network or conceptual role theory of semantic content (as outlined in Churchland, 1979).

Sayre rejects this approach on basically Searlean grounds: A purely formal set of internal relations, no matter how configured, seems to him inadequate to capture genuine semantic content. Let me close with some comments on this. The proper strategy with which to address Searle's Problem is not to try to augment or inflate such formal economies in hopes of finally achieving the golden fleece of True Intentionality. The proper strategy is to deflate the inflated conception of the goal we are after.

The fact is, when one assigns a sentence, a sentence of one's own idiolect, as the "semantic content" of another's internal state or overt utterance, one is not thereby showing how the target state relates to the external *world.* One is showing how the target state relates to *one's own states.* Talk of reference and meaning masquerades widely as talk about how language and thought relate to the world, but the cash value of such talk is always a matter of relating words and utterances to other words and utterances. In trying to divine another's content, one is always trying to find a semantic analogue within one's own cognitive economy, a concept or sentence that plays a sufficiently similar inferential role.

It therefore seems wrong to me to assume that the states of a genuinely cognitive creature enjoy some elusive "intentional" relation to the external world, whereas the states of a systematic computer mockup of that cognitive system must inevitably fail to enjoy such a relation, and must therefore fail to be a genuine cognitive system. We humans enjoy no relations to the world that such a system would not also enjoy. We, and the systematic functional mockup, are just intricate internal economies in causal interaction with an external world. Given a rough isomorphism between the relevant economies, there is as much, and as little, sense in ascribing semantic content to the artifact's states as there is in ascribing it to the states of your own brother. If intuition syas otherwise, then intuition has been misschooled.

Communication theory and intentionality

John G. Daugman
Engineering and Computer Science and Psychology Departments, Harvard University, Cambridge, Mass. 02138

Sayre's proposal is that the intentionality of mental events originates in certain feedback and control processes which he characterizes in terms of communication theory. In principle, such a formulation is compatible with the fundamental intuition (originally articulated by the 19th-century Austrian Franz Brentano) that the distinguishing feature of mental events, which imparts to them intentionality rather than physical extension, is their character of referring to or representing other objects or states. Because communication theory provides a framework for the scientific study of signification and representation, Sayre's basic premise has a certain fresh appeal.

Taking Sayre's project in its own terms, I think its central weakness is the underlying view that the goal of perception is to maintain high mutual information (essentially fidelity) between object and representation. I will argue on various grounds that this is surely not an appropriate description of perceptual processes beyond the earliest retinal levels, if indeed it is applicable even there.

(1) In a formal sense, there is a basic incoherence in Sayre's use of Shannon's (1948) analysis of the capacity, equivocation entropy, and mutual information of a communication channel. Focusing on visual perception as a goal-directed process that clearly exemplifies intentionality, Sayre treats the sequence Object-Retina-Cortex as forming communication channels $H(O/R)$ and $H(O/C)$ with associated mutual information measures $I(O;R)$ and $I(O;C)$ (in Sections 5, 8, and 10); but this springs from a vernacular sense of information structure that is entirely different from the statistical sense of information structure for which Shannon's formulation applies. Whereas Sayre acknowledges the lack of direct connection between the technical sense of information ("info(t)") and semantic information ("info(s)"), his investment in communication theory commits him to a primary concern with fidelity. As Shannon stresses at the beginning of his classic monograph, "The fundamental problem of communication is that of reproducing at some point either exactly or approximately a message selected at another point" (1948, p. 379).

But fidelity analysis either misses the whole point about perception or else it must be formulated so loosely that it is incoherent, as suggested by the following simple considerations. First, An awesome feat of visual perception is the inference of three-dimensional (3D) world properties from a two-dimensional (2D) image – a kind of "inverse optics." As Poggio and Torre (1984) have stressed, this task is computationally an ill-posed problem whose solution is indeterminate. Remarkably, this basic problem confronting the visual system is completely lost sight of in Sayre's fidelity-based analysis of the "communication channel" between 3D object and 2D representation of it in Retina or Cortex. How can one define or compute the fidelity of a representation in which dimensionality is changed? The units are incommensurate; the event sets are of different rank. Because no relationship of identity is possible between incommensurate structures, it is difficult to give coherent meaning to Sayre's basic claim that "the relationship of identity in info(t) structure between O and C *is* the intentionality of perception, and that a structure at C enjoying this relationship possesses the corresponding structure at O as its semantic content" Section 10).

Second, The "information content" of an object or its image, defined in Shannon's sense of possible states, is infinite until some grain structure is imposed (such as the molecular structure of matter or the quantal nature of light). The bizarre thing is that Sayre would have to specify such quantities of information content in order for the definition of mutual information (or equivocation entropy, etc.) to have any meaning, and then such measures would of course be irrelevant. That is the trouble with trying to put objects into communication channels: It is doubtful that one could specify in any useful way the "information content" of an object.

(2) In communication theory, the assignment of "input" and "output" ends of a channel is irrelevant; the theory is about the relationship between two probability distributions, and there is no intrinsic difference between them except in statistical terms. But this is inadequate for Sayre's intentions, which require an asymmetry in direction of reference between object and representation in order for the intentionality of perception to be captured by its referentiality. Within Sayre's "mutual information" account we would have to say equally that a stone is also a representation of our percept of it, and then the desired distinction between extension and intention would be lost.

(3) It is generally agreed that perceptual processes are not in any sense photographic, but rather involve such things as data compression and summarization of image information by the extraction of salient features. But these are processes of data reduction and hence yield low mutual information, contrary to Sayre's characterization of maximizing mutual information. (The best way to maximize mutual information, which Sayre says is the ground of intentionality in perception, is to take a very high quality photograph of the retinal image!) Making explicit the semantically relevant features of the visual environment has less to do with the verisimilitude of signals (which Sayre's account of "information structure" is about) than with their deconstruction. And of these processes, the explanations proposed by Sayre (Section 9) amount only to Cognitive Psychobabble (positing an Accumulator, a Discriminator, an Abstractor). Such explanatory strategies, which one sees all too often in cognitive psychology, are reminiscent of the well-known medieval explanation of the mechanism by which sleeping powder has its effect: namely, through its Dormative Potency.

(4) Finally, in support of Sayre's basic goal of guaranteeing the

intellectual solvency of cognitive theory by showing how info(t) is transformed into info(s), I think that some encouragement might be found through the example of a well-understood perceptual faculty in which a meaningful object-property is inferred at an early level from a less meaningful image-property. In the perceptual phenomenon of color constancy (the Land Effect), the perceived colors in a Mondrian patchwork are largely independent of the actual wavelength distributions of light reflected from any given patch. Rather, the colors that we see are governed directly by the pigments in the object regardless of illuminant. The ability to "abstract away" the irrelevant properties of illuminating light conditions (wavelength mix and intensity) and to infer directly the object's true pigment properties despite the unfaithful signals reaching our retinae is the remarkable inference whose retinal mechanisms are specified by Retinex Theory. Perhaps this phenomenon illustrates concretely how even very low-level mechanisms make info(s) inferences about invariant object-properties that are not explicit in the info(t) content of the signals received.

Engineering's baby

Daniel C. Dennett
Center for Cognitive Studies, Tufts University, Medford, Mass. 02155

First let me tell you about a wonderful invention of mine, the Life-of-the-Party Machine. Here's how it works: As each guest arrives at your party, the Vigilator "locks on" to that guest for a few minutes, establishing a communication channel of high mutual information, so that the structures of some of its own internal states come to be identical to the objective structures of the guests – that's phase one. In phase two, the Extrapolator selects a subset of those structures "for their ability" to anticipate the guest's reactions to the other guests; and finally, in phase three, this subset of structures guides the Life-of-the-Party Machine in the selection of the beverage for each guest that will optimally "tune" the "impending" behaviors of all the guests, so that everybody ends up having a wonderful time. That's how it works, and of course it's still a bit sketchy and no doubt will need some modifications along the way, but now that I've worked out the basic idea, from here on out it's engineering's baby.

This won't do, alas, for even if the basic idea were fine, all the hard problems have been kicked downstairs to engineering, with scarcely a clue about how they are to be solved. I proffer this parody of Sayre's target article in a tentative and apologetic spirit. I am probably missing the point and conjuring up a strawman. It is now Sayre's turn to explain why my parody is unfair, why he is not confusing the "specs" of a system with its design, why his various proposals take us a step or two forward, instead of backward, as they seem to me to do. What puzzles me is that Sayre apparently levels the very charge I have just expressed against him against Gibson (1966) – correctly, in my opinion: "even in its unsimplified version it provides few details about how exactly the perceptual systems accomplish this, or about the nature of the information they are supposed to provide." So apparently Sayre thinks he has moved significantly beyond Gibson in these regards. Apparently he thinks he has an alternative to Marr (1982), for instance, rather than just a theory sketch at a *less* specific level, though I cannot see why.

At one point Sayre offers a telling analogy with the television's tuning circuits, "whose activities possess a high degree of mutual information with respect to selected sets of input signals." This is a somewhat ominous analogy, since it characterizes the *desired effect* of such a tuning circuit, rather than the means by which the effect is achieved. In the case of the tuning circuit, the means are well enough known that this is a legitimate specification. Can the same be said for all the various "info(t)" processes that Sayre must invoke in his account? There seems to me to be a striking difference in explicitness in Sayre's descriptions of processes, and at just the crucial place: where he must turn the corner from what he calls info(t) to info(s).

Sayre begins with the claim that a (desirable) relation between distal objects of perception and the cortical states that represent them is high mutual information. We can all agree on that; everyone from Gibson to Marr to Fodor (1975) to Winograd (1972) supposes that a necessary condition of (good, useful) representation is high fidelity – accuracy plus informativeness – and the mathematical theory of communication (MTC) concept of high mutual information captures this nicely. If others make less of this condition than Sayre does, perhaps it is because they take it for granted. It is quite clear that there are many ways of meeting this necessary condition, as shown by Sayre's relatively detailed account of noise reduction and tuning, occurring between what he calls the Retina and the Accumulator. So long as we are on MTC's home turf of maintaining high mutual information relations, its resources are indeed impressive, but this necessary condition does not distinguish vision from television. We still have to get the Discriminator and the Abstractor working, if we are to have any account at all of vision and the promised info(s), and when Sayre turns to them, his accounts get sketchier.

The Discriminator is to operate rather like an analysis-by-synthesis process, one gathers. The task is to find "stable patterns that will both (a) serve well in the guidance of current behavior and (b) maintain a high degree of mutual information with events at the Retina." Sayre describes an MTC method for meeting the second criterion – but note that it is crucial that the relata are events at the Retina, not distal events, which the Discriminator cannot observe, on pain of regress. But what MTC process can select patterns meeting criterion (a)? We get a rosy description of a process that tests the effects on the guidance of behavior of "advancing" a pattern to a position of control, and while this sounds right phenomenologically, Sayre can hardly claim to be cashing out his phenomenology in the good hard coin of MTC. And when we turn to the Abstractor, we get more handwaving – acknowledged, to be sure, but then where is the contribution of MTC to cognitive science?

The way to turn the corner and distinguish vision from television, Sayre notes, is to design a system that will have the right sort of negentropic coupling and negentropic flexibility. There is nothing to quarrel with here, and nothing new except the language. We must figure out how we get constancy of perception out of all the variability on the retina, but of course it has to be the right constancy – the right mix of negentropic coupling and negentropic flexibility. As Gibson saw (and Sayre saw that Marr saw that Gibson saw), this is the central "how" question, but Sayre seems to me to be no closer to giving an answer (right or wrong) than Gibson was. Indeed, in spite of the terminological innovation, he does not seem even to have given a more perspicuous setting of the question. If we had a scaling defined for negentropic coupling or flexibility (as we do for mutual information), we might have some new powers of description, and hence problem setting, but so far as I can see, given several different negentropic couplings, or flexibilities, there is no prescribed way of saying which gets the higher rating. Compare two moths, one with second-rate eyes and a second-rate but lightweight brain, and the other with eagle eyes (more structure for more info(t)), but more weight to lug around, hence greater energy demands – which creature has struck the better negentropy bargain? When does it improve one's negentropy coupling to turn off one's information-gatherers – and sleep, for instance? I do not want to disparage these questions; on the contrary, they strike me as well worth pursuing, but precisely because their answers are not obvious, one cannot simply fix everything but info(t) at some standard value and then declare that (always, or even in general) more info(t), or higher mutual information relations, is a Good Thing. Many increases in info(t) result in decreases in safety, efficiency, and so forth. Moreover, of course, claiming that a certain trade-off would be optimal does not explain by subsumption under a law (there is no law of nature to the effect that design will be optimal), but rather generates a hypothesis to test – and leaves the question of mechanism untouched (Dennett 1983).

Sayre realizes all this, for he draws attention to the familiar point that "the posterior branch [retina to visual cortex] has as its main function to cut back radically on the amount of info(t) passed on to the cortex, while at the same time retaining just those info(t) structures that are important for the guidance of the organism's current behavior (Section 10, paragraph 5). But he has not yet shown that the description of processes for performing this latter task has been advanced by his proposed innovations. Moreover, he has not shown that he has any real alternative to approaches that treat this part of the problem as a "computational" problem in the sense he disparages.

His main criticism of that school of thought, by the way, is curiously undercut by his own proposals. What is wrong, he says, with AI (artificial intelligence), is that "the symbols involved in machine calculation in and by themselves are not about anything at all." But the same would be true of the stable patterns selected by the Discriminator in virtue of their prowess in behavioral control: *In and by themselves* – that is, independently of their relational, nonintrinsic properties such as the property of yielding high mutual information relative to something distal – they are not about anything at all. There are many things wrong with various versions of the "machine calculation" paradigm in AI, but failing to do justice to intrinsic intentionality is not one of them. By Sayre's own lights, intentionality is some sort of extrinsic property – presumably definable in the terms of MTC, after all – so there is no such thing as intrinsic intentionality.

Stalking intentionality

Fred I. Dretske
Philosophy Department, University of Wisconsin, Madison, Wis. 53706

Sayre asks many of the right questions. From where I sit, he even looks in the right places for the answers. He has, of course, been looking in these places a long time. He was, as far as I know, one of the first to explore the distinctively philosophical applications of communication theory. My own interest in information, first in epistemology and later in the philosophy of mind, was stimulated by Sayre's early work in this area. But though we both look in the same places, we seem to find different things. I too think that the key to intentionality lies in that network of organism–environment relations characteristic of information (in the *statistical* sense). Why, then, can't we agree about just how this key unlocks the door?

Because Sayre is explicit about restricting himself to the "low-level" intentionality of perception (and not to "more complex forms of cognitive activity"), and because he devotes merely a speculative postscript to cognitive activity involving meaning and concepts, I assume he intends to avoid asking himself (or having me ask) the question he asked me in his *BBS* commentary on my book (Dretske 1983a): *How does one manage, as all of us with false beliefs do manage, to manufacture false* contents out of info(t) (Sayre 1983)? In other words, how, on an information–theoretical account of things, does one get an account of misrepresentation? Sayre suggests in his postscript that it is all really quite simple: Meanings and concepts are merely percepts that have been progressively freed from stimulus control. As I understand things, though, this means they have been freed from the only thing that, in Sayre's account of things, gives meaning (reference, aboutness) to something. The percept "yellow" gets its intentionality, we are told, by being activated only by yellow objects, by "picking out," "focusing," or "locking" on the color yellow. How, then, do concepts or meanings, being freed from this kind of exclusive control, manage to be about yellow when they do not pick out, focus, or lock on this color? To suggest (as Sayre seems do to in his postscript) that these "higher order" structures get their meaning by coming under the control of, or being activated by, other meanings is an obvious evasion. Where do these other meanings come from?

This is not, by the way, an original criticism. It is the same criticism Sayre made of me. I repeat it, not just to even the score, but because it is an important point, one that any account of intentionality must confront. And the issue is particularly acute for causal or informational theories of intentionality because in tracing the source of intentionality to a structure's causal or informational relations to other conditions – conditions (notice!) that must exist for the structure to stand in these relations to them – it is hard to see how one can have meaning without truth, how one can get an account of misrepresentation. Sayre may not have liked my way of dealing with this problem, but I was surprised to find him so casual when it was his turn to deal with it.

I am, of course, being unfair. Sayre deliberately restricts his account to the intentionality of perception. His speculative remarks in the postscript on "higher order" cognitive activities are "not intended for critical discussion" (though I am curious about what kind of attention he expected them to receive – especially in *this* journal). But what is the intentionality of perception? Is this supposed to differ from the intentionality of knowledge, thought, and belief? Sayre apparently thinks of visual perception as some kind of cognitive activity, as (I assume) a coming to know that something is so by visual means: for example, seeing *that* the flower is yellow or *that* there is a yellow flower nearby. But doesn't this involve the application of concepts, the having of thoughts and beliefs, perhaps even (on some accounts) the use of reason and the making of inferences? Isn't this a red-blooded propositional attitude, something with as fancy a content as we can ever expect to find inhabiting the mind? Are we to understand all this as "low-level" intentionality? Why? It looks pretty robust to me. In fact, it looks suspiciously like the higher level cognitive activities that Sayre wants to exempt from critical discussion.

Of course, in speaking of the intentionality of perception Sayre could have meant the fact that perception is perception of something, the fact that our experience (or percept) is *of* or *about* or *directed upon* a specific object (e.g., a flower). He certainly talks this way at times. But if this is what he meant, a great deal of what he says about the conversion of statistical information (info(t)) into semantic information (info(s)), the transformation that is supposed to explain this form of intentionality, doesn't make sense. For *this* aspect of intentionality clearly doesn't depend on the kind of negentropic coupling (high mutual information) between cortical structures and object that Sayre requires. I can see a flower (my percept can be *of*, and in this sense *directed upon*, a flower) when, because of bad lighting and distance, I get very little information about the flower. My representation of the flower need not (and often does not) pick out the flower in the way Sayre says it must (being the *only* object with which the representation shares that relationship) in order to be *of* the flower. A representation (whether percept or photograph) can be *of* X (e.g., my cousin Clyde), it can exhibit this kind of intentionality (call it referential intentionality) with respect to X, without distinguishing X from Y (e.g., a twin brother), without, that is, representing X *as* X. Without, in other words, exhibiting the kind of *classificatory* intentionality characteristic of knowledge and recognition.

It is not only unclear what aspect of intentionality Sayre means to be explaining, what kind of thing info(s) is supposed to be. (Is it a proposition? A truth condition? An object or circumstance – that which the representation is of or about? Or something else?) It is also unclear exactly what process is supposed to effect the transformation of info(t) into info(s). At one point we are told that the essence of the matter is the loss of info(t) ("in return for this considerable amount of info(t) lost (or expended), there is a gain (from none to some) of info(s)"). At another point the notion of a "pattern" seems to figure prominently ("the key to this account [of how info(t) is converted into info(s)] is the concept of pattern"). Later it is suggested that the important point is that the information present at *O* (the perceptual object) is the same as the information in the cortex (or, and here begins a series of equivocations, the information structures are the same, or the events in the cortex are identical with the events in the environment). And then, of course, we have the ad hoc sounding qualifications designed to rule out simple mechanical systems from having intentionality; doorbell systems, for instance, do not have intentionality, even though the information at the bell is the same as the information generated at the button because the system doesn't satisfy the right kind of "focus," "dynamic relationship," or "load" requirements.

Frankly, I find it hard to tell what is supposed to bear the weight of this account. It sometimes seems as though "key" points are tugging in different directions. If huge chunks of information are lost between the retina and the cortex, and this loss is the key to intentionality, then how can the information in the cortex be the same as that at the retina (or in the environment)? It can't be numerically the same, of course. And despite the way Sayre sometimes talks, the cortical events or structures are clearly not the same as the external circumstances about which they carry information. So how can the relationship of identity in info(t) structures between *O* (object) and *C* (cortex) be the intentionality of perception.

I think I know what Sayre wants to say. He wants to say that the information in the cortex – the information, namely, that there is a yellow flower out there – is the same information that is on the retina and the same information that is generated by the yellow flower's being out there. But he can't say this because this is already to invoke a notion of semantic information (that there is a yellow flower out there) which he doesn't have. And until he has it, he can't claim an identity between what is happening out there and what is happening in the cortex. And no identity, no conversion of info(t) into info(s).

Intentionality and information theory

David P. Ellerman
Computer Science Department, Boston College, Chestnut Hill, Mass. 02167

Sayre begins with an excellent rendition of an old argument that formal symbol manipulation, "symbol crunching," can never by itself capture the semantic dimension of mental processes – an argument forcefully restated in a previous target article by Searle (1980a). How could one build a physicalistic theory of intentionality – of the semantic relation between symbol and referent? This seems particularly unlikely if the referents are past or future events, or are abstract possibilities as in mathematics. But when the referents are actual objects or states of affairs (e.g., in perception), a causal relation can be established between the symbolic representation and referent. Thus the hope for a physical reductive treatment of some intentionality, the hope for "strong AI" (artificial intelligence), lies in adding transducers to symbol crunchers. Computers may be just syntax, but computers-plus-transducers may deliver semantics. That is the crux of the robot-reply to Searle used by Fodor (1980) and others, which was ably but briefly answered by Searle in his response (1980b, especially p. 454).

Sayre's theory is an information-theoretic (I-theoretic) variation on the robot reply. It uses the I-theoretic properties of the transducing channel between objective states of affairs and cortical events in an attempt to account for intentionality in human perception. The intuitive idea is given in an aside by Sayre that a dashboard light may "designate" or "refer to" low oil pressure. The same idea occurs in Dretske's gas gauge example: "Our humble gauge even exhibits the rudiments of intensionality [sic] – representing the amount of gas in my tank" (1983b, p. 82).

The key assertion is that "the relationship of identity in info(t) structure" (given essentially by low channel equivocation) between the object set *O* and cortical events *C* "*is* the intentionality of perception." There is a simple counterexample. Let the input alphabet *O* and the output alphabet *C* both be (0, 1). In channel A, a 1 is received if and only if a 1 is sent, and similarly for 0. Since channel A is noiseless (outputs determine inputs), the equivocation (or conditional entropy) of the inputs *O* with respect to the outputs *C* is zero (see Abramson 1963). Since the channel is deterministic (inputs determine outputs), the equivocation of *C* with respect to *O* is also zero. This is the best possible case for Sayre's assertion about the identity of info(t) structures between *O* and *C*.

To describe the channel using semantic notions, we ascribe the "meaning" of an output signal to be its unique causally antecedent input signal (if such exists). An observer who knew the particular causal properties (conditional probabilities) of channel A would then know that a 1 received means a 1 was sent and similarly for 0. In terms of perceptual intentionality, a 1-received would be perceived as a 1-sent, and similarly for 0. Can these semantic relations be characterized using the equivalence of info(t) structures between *O* and *C*?

Consider another channel B defined by permuting the roles of 0 and 1 in the output alphabet. A 1 (response 0) is received if and only if a 0 (response 1) is sent. All the I-theoretic concepts using averages, such as the entropies, the equivocations, and the mutual information, are identical in channels A and B. In Sayre's terms, the info(t) structure defined on *C* by channel B is also the same as the original info(t) structure on *O*. Yet the semantic meaning or informations(s) carried by the output signals in channel B is totally different. A 1-received means a 0-sent, and a 0-received would be perceived as a 1-sent. In general, scrambling (i.e., permuting) the output alphabet leaves all the average properties of the channel exactly unchanged but scrambles the meaning of the output signals. Thus the attempt to characterize semantic relations and perceptual intentionality using average I-theoretic properties of the channel from the object set *O* to the cortical events *C* (e.g., low or zero equivocation) must fail.

In the examples, the "meaning" of an output signal was defined as its unique causally antecedent input signal, a definition which only makes rigorous sense for noiseless channels. To know the meaning of output signals, an observer must know the particular causal properties that determine the channel. It is not sufficient that the channel have certain average I-theoretic properties. Thus, to know the meaning of a cortical event *r*, an observer must know its causally antecedent objective event *s*. But to assume such observer knowledge in a theory of perception or perceptual intentionality is to assume that which needs to be explained.

The last and most fundamental point is the distinction between the semantic and the causal relations that relate the input and output signals, the distinction between the meaning and the cause of an output signal. So far we have used a mind-to-world direction-of-fit between the semantic relation and the causal relation. That is the stance of perception, of learning about the world. If we perceive a 1-received as a 1-sent (i.e., hypothesize a certain causal antecedent) and then find out the signal was sent over the invertor channel B, then we correct the misperception. We change the hypothesized meaning of 1-received; we adjust the mind to the world.

The opposite stance takes the direction-of-fit to be world-to-mind. That is the stance of intentional action such as constructing an artifact to fit a design. We now specify that by design a 1-received (response: 0-received) shall mean a 1-sent (response: 0-sent), and we construct the channel to fit. If we test it and find that a 1 was received when a 0 was sent, then we change the channel, not the meaning of 1-received.

Thus the meaning and the cause of receiving a signal are not only distinct; they can stand in two different relations. A hypothesized meaning of an output signal can be adjusted to fit its cause, or the cause can be adjusted to fit a specified meaning. If the meaning and cause of output signals were identical, then dashboard lights could not malfunction. And the same holds for robots.

Information is in the eye of the beholder

Rhea T. Eskew, Jr.
Center for Human Information Processing, University of California, San Diego, La Jolla, Calif. 92093

I see two problems with Sayre's application of information theory (info(t)) to visual perception (and, by extension, to the rest of cognition). The first is perhaps only a "technical" difficulty, while the second is more fundamental.

The technical problem is this: How can the brain stages described by Sayre in Sections 9.2–9.4 alter their operations in order to "maintain a high degree of mutual information with events at the Retina" (Section 9.3)? If transmitted or mutual information is indeed simply the input entropy minus the equivocation, how can the system or any of its stages apply "mutual information measures" (Section 9.2)? How can the system "know" (respond to) the entropy at input, even if it somehow "knows" its own equivocation rate? At each moment all it has access to is the pattern of transmitted events; it has no way of knowing which of these are related to input and which are noise. This problem is common to all theories that take the philosophical position of perceptual realism, as Sayre's does; Katz (1983) provides a further discussion of the logical difficulties of realism as a theory of visual perception.

Sayre himself refers to this difficulty, and a possible solution, when discussing the feasibility of a nonbiological system possessing intentional characteristics (Section 10). There he notes that a computer system using info(t) in the way he describes would have to calculate statistics on its own operations over time, and in real time, "in a manner reflecting fluctutations in the environment from which [the] data are derived." If the mathematics that would permit information measures to be derived from such computations have been worked out, I am not aware of them. I suspect, however, that it might be possible for a system, operating in this way, to calculate (respond to) the relevant measures (properties) of info(t), given certain constraints on the input information or entropy and equivocation or noise ($H(A)$ and $H(A/B)$, respectively). One such constraint would surely be that $H(A)$ reflects a stationary process, that is, that its statistical properties do not change (or at least change very slowly) over time. Markov processes may have this property, and Sayre does describe his "Accumulator" (Section 9.2) as responding to a "Markov source." But I do not think he fully emphasizes the critical importance of the sequential properties of stimulation. Sayre says that "the Accumulator can test for repetition by applying mutual information measures over successive members of the series [of Retinal patterns]." I doubt that it is possible to calculate mutual information *at all* without comparisons over time (and possibly not even then). This suggests, in other words, that the temporal comparison must be the primary process, and if info(t) measures are generated they can only be generated by comparison of inputs over time. A "snapshot" of the input cannot be used to calculate the amount of information transmitted, because from a snapshot, the system cannot know $H(A)$. Sayre, on the other hand, seems to treat the temporal comparison as a secondary, though important, process.

There is, however, a more fundamental issue to be raised. After rereading Sayre's target article several times I still do not completely understand in what sense he is suggesting that the nervous system uses info(t). On the one hand, the bulk of the article (and especially his criticisms of David Marr, Section 9.1) indicates that Sayre takes a mechanistic view: that is, to use his own words, "neuronal mechanism[s] in the retina or optic tract [operate] on visual data passing through it in such a manner as to approximate the results of this or that mathematical transformation" (Section 9.1). On the other hand, some parts of the paper – particularly phrases like "applying mutual information measures" (Section 9.2) – suggest instead the same sort of computational view for which Marr is criticized, a view in which parts of the brain are thought actually to calculate "information measures." The fundamental point to be made here is that whether one assumes a mechanistic model or a computational one, "information" (info(t)) must be "observer-relative" (Searle 1980a), just as intentionality is in the case of the logical theorem prover (Section 3).

Take, as a simple mechanistic model, a telegraph system. A telegraph system transmitting Morse code may be described by information theory, but the info(t) is a description that we, as observers, apply to it; the telegraph responds to electrical impulses, not to information (in any sense). The information – info(t) or info(s) – is in the observer, not intrinsic to the system. For an example of a computational system, consider a computer running a routine to calculate information theory measures. The machine may print out a number that the programmer can validate, but the computations are purely formal operations (Fodor 1980). The programmer, not the computer, identifies the output number with "information."

If info(t) is observer-relative, how can intentionality depend on it? Is there some cosmic observer, imbuing us with divine purpose? The answer may lie, not in our stars, but in reaffirming the importance of *un*conscious processing in everyday behavior (cf Dixon 1981). [See also Holender: "Semantic Activation without Conscious Awareness" *BBS* 9(1) 1986.] To the extent that consciousness can be described as self-observation (e.g., Skinner, 1974, Chap. 10), intentionality and semantic information are terms properly applied to consciousness. [See special issue on the work of B. F. Skinner, *BBS* 7(4) 1984.] The meaning we experience in our own behavior is indeed observer-relative, but we are our own observers. Few would claim, however, that they consciously compute anything as they behave, or that any "info(t) measures" reside in consciousness. If information processing of any sort occurs in the human brain it is (I would claim) normally sub- or unconscious, and therefore without intentionality or semantic content. In this view, accounts of intentionality belong to the domain of consciousness (whatever that is), a domain in which information theory can find no place.

ACKNOWLEDGMENTS
I wish to thank Kathleen F. Melia for helpful discussions and a critical reading of this commentary.

On some specific models of intentional behavior

Richard M. Golden
Department of Psychology, Brown University, Providence, R.I. 02912

Sayre properly emphasizes the importance of environmental considerations, principled mathematics, and computer simulation for understanding the problem of intentionality. On the other hand, he places too much faith in communication and information theory to the exclusion of other approaches to this problem.

The mathematical theory of communication. Sayre introduces the mathematical theory of communication (MTC) as a means of exploring the set of relationships characterizing organism-environmental interactions. This approach is certainly of great interest and should result in new insights into the problem of intentionality. To be sure, MTC is very useful for comparing and contrasting broad classes of different types of relatively simple communication system models. Still, before any calculations using MTC can be made, a particular probabilistic model of the relationships between (and within) the organism and its environment is required. Sayre characterizes this modelling problem as finding the correspondence between info(t) and info(s). He further suggests that with some knowledge of evolutionary neurobiology and psychology in conjunction with certain relationships existing between info(t) and energy it should be possible to identify the relationship between info(t) and info(s). I do not believe that the solution to the problem of intentionality will be this simple.

Before MTC can be applied, one must first grapple with the practical problem of defining the concept of an "event," because MTC is based on the assignment of probabilities to specific events. Exactly what is a psychological event, and what is a relevant neurophysiological event? Second, one must identify the nature of the communication channels of interest. If these channels interact, what probabilistic laws should be assumed to capture that interaction? Is the channel output at time t a function of only the channel input at time t and not a function of the channel input at the set of times preceding t (i.e., is the channel memoryless)? And, if so, what probabilistic laws are required to characterize this time-variant system behavior? Sayre suggests that "it should be possible to conjecture reasonably about the types of info(t) processing that might occur at the major junctions." Where are these major junctions? How are they assumed to interact? These questions must be answered before any calculations can take place. Sayre's discussion of how MTC might be applied to the visual system does not address these issues. He does not explicitly motivate why the visual system can be broken up into roughly four serial processing stages (i.e., the Retina, Accumulator, Discriminator, and Abstractor stages), what communication channels link these stages, or even what types of information may flow through these channels.

Moreover, even if the required probabilistic laws were specified, the types of statements that can be made by MTC are rather narrow and limited in scope. For example, although MTC is quite useful for making general statements about broad classes of communication system models (e.g., specific bounds on the system's error rate), it is doubtful that such a theory would be useful for making more detailed predictions (e.g., reaction times). The point is that although MTC is certainly a useful tool that has not really been fully exploited, other approaches are required in the formulation of psychological theories of perception and cognition.

Some additional useful formalisms. Some mathematical tools, similar in spirit to MTC, are now offered as possible candidates for exploring some of the issues discussed by Sayre. A mathematical framework that seems to be of exceptional utility is linear algebra (Luenberger 1979). Using this formalism, specific patterns of neural events are represented by lists of numbers called state vectors. A given state vector, therefore, has a unique correspondence with a specific pattern of neural events (for additional details, see Anderson & Mozer, 1981). Such a framework nicely captures Sayre's suggestion that "patterns" of neuronal events are "representations of salient features of the objective environment."

Sayre also suggests that the basis of visual perception is a matching process between a specific neural pattern of activity and a set of objective events in the world. The similarity measure that this matching process is based upon is given explicitly by the mutual information between some specific set of objective events in the world and a specific cortical representation. More formally, the computational problem the organism must solve is to find that cortical representation that maximizes the mutual information between that representation and that set of objective events in the world. Although this approach may certainly be a useful one, considering our current lack of knowledge, a more general formulation of the problem seems reasonable. In particular, suppose the computational problem the organism must solve is to find the cortical representation that is most probable, given some observed set of objective events in the world (i.e., a partially specified cortical representation). This type of problem is referred to as maximizing the *a posteriori* probability density function (MAP estimation) in the engineering literature. The main advantage of reformulating the problem in this manner is that a certain amount of generality is immediately obtained. The basic difficulty with this approach, however, is that the density function must be identified before it can be maximized. The relatively new theory of Markov random fields (Besag 1974) seems very promising as a means of obtaining useful general forms of such density functions. Assuming now that a general form of the density function is obtained, its parameters can be estimated, following Sayre, based upon knowledge of the effects of evolutionary pressures and past individual experiences upon the organism.

For example, consider the set of state vectors whose elements (representing individual neural events) are binary-valued (true, false) random variables; assume that the organism is capable of detecting only "pair-wise" neural feature correlations. Interestingly enough, in this special case, the resulting form of the above probability density function is intimately connected with thermodynamic concepts of both energy and entropy. The above restrictions upon the class of probability density functions is of even greater interest when one realizes that a number of investigators (Ackley, Hinton & Sejnowski 1985; Anderson, Silverstein, Ritz & Jones 1977; Hopfield 1982) have constructed some simple algorithms (amenable to neurophysiological interpretation) that maximize this specific probability density function. That is, these neural network models are provided with a subset of known state vector element values and effectively generate the most probable configuration of element values for the unknown state vector elements. It is important to note that although each of these models is identical at the computational level (to use Marr's 1982 terminology), they are different in terms of their behavior at the algorithmic level. That is, each network model possesses a unique neurophysiological interpretation and a set of unique practical and theoretical difficulties in carrying out the necessary maximization.

Computers and intentionality. Sayre argues that "there is no purely formal system – automated or otherwise – that is endowed with semantic features independent of interpretation." Since computers are formal systems, to continue the argument, computer simulations of cognition cannot exhibit "intentional" behavior. Sayre, however, does not consider the possibility of modelling both the organism and its environment on a computer.

If an organism and its environment are modelled on a computer, then the simulation model of the organism might be said to exhibit intentional behavior with respect to its model environment. One argument against this approach might be that modelling both organism and environment unnecessarily complicates matters. I see no other possibility, however, because I am sympathetic with Sayre's view that intentionality (and meaning) can exist only with respect to an organism functioning within an environment. Moreover, note that this solution may become more tractable when one realizes that the purpose of a model is to serve as a useful abstraction of a fairly complex situation. Thus, only those aspects of the environment and organism that are relevant to the psychological modelling problem of interest need be considered.

In summary, I commend Sayre for seriously considering formalisms other than those based upon symbol manipulation, but I suggest an even more eclectic approach to the problem of intentionality. MTC, like any other approach, is not sufficiently strong by itself to provide a foundation for cognitive science. The foundations of cognitive science will eventually emerge themselves, if we give them some time.

Uncertainty about information

Ian E. Gordon
Department of Psychology, University of Exeter, Exeter EX4 4QG, England

Sayre's interesting attempt to relate two types of information does not convince. Imagine the following real-life situation: One is on a hillside when the mist suddenly clears. Ahead is a snow-capped mountain, not quite where it was expected to be; nevertheless it is clearly visible.

The question arises as to the informational aspects of this experience. In terms of the formal information-theoretic calculus (info(t)), what was the prior uncertainty associated with the mountain as a stimulus object? Several million rays are entering the eye before and during the first glimpse of the stimulus object. But is one's uncertainty comparably great? And should calculations include the admitted uncertainty over the mountain's position, say to the nearest degree? If there are no rules for applying the informational calculus at this stage, how does one quantify the stimulus array? Analysis of a picture of the scene in terms of pixels or similar units won't do because one was seeing something before the mountain appeared: It is not that the information in the scene has suddenly increased, it has simply changed. As Sayre is aware, a neutral measure such as info(t) does not capture the nature of this qualitative change.

The situation becomes more complex when a companion points out that one slope of the mountain looks like a human face. After a few moments one suddenly sees that this is correct. What was one's uncertainty before and after seeing the face? Is it to be calculated on the basis of all faces previously experienced, or only faces formed from rock and snow?

When we consider the causal chain between light from the object reaching the retina and the eventual formation of a percept, the application of information theory has some attraction. Barlow (1985) illustrates the usefulness of this approach in his speculations concerning the relationship between the number of neurons involved at different stages in vision and the proportions of active to inactive cells at each stage. But this is a quantitative exercise; in Barlow's terms it represents an "informational balance sheet." Even here, however, information is calculated in terms of likelihoods of neural firing, infrequent events being assumed to carry more information. There is no discussion of how a neurone waiting to receive impulses from earlier ones in a chain can be said to manifest *uncertainty*. What would this mean?

When Sayre suggests that representations at his third stage of visual perception have no orderly spatial characteristics and that they function with respect to info(t) features "preserved through the many channels beginning with the perceptual object," I find it impossible to conceive how such a scheme would actually deliver the goods.

Intentionality and the explanation of behavior

John Heil
Department of Philosophy, Virginia Commonwealth University, Richmond, Va. 23284

Sayre believes that we lack at present an account of intentionality of the sort required if we are to understand the behavior of "cognitive organisms." What is needed, he suggests, is a straightforward description of the mechanism of intentionality, one that locates intentional content where it belongs: inside the organism. The latter point is important for Sayre. The computational paradigm now dominant in cognitive science precisely fails to do this. It is not that the internal states of a computing machine lack intentional or semantic properties. The difficulty, rather, is that they possess these properties, to the extent that they do, only derivatively: "None of the representations internal to the machine has meaning, or truth, or external reference, just in and by itself . . . Of intrinsic intentional properties . . . these representations are entirely innocent." This, according to Sayre, is because "the intentionality attaching to machine representations is not 'intrinsic' but 'observer-relative,' being entirely dependent upon the 'intrinsic' intentions of the programmer." Hence if our aim is to explain intentionality in a way that will prove useful in explanations of behavior, an appeal to the internal states of computing machines will be of no help at all. Computing machines possess intentionality only by courtesy.

What might it be for something to have intentional properties intrinsically? Inscriptions on blackboards and in journals clearly do not. The semantic content of such things is bestowed on them by inscribers. Human thoughts, however, seem different. The content of my thought that "*p*" does not, in any obvious way, depend on my or anyone else assigning it that content. In this sense, its intentionality is intrinsic to it. Putting the point this way, however, may be misleading. If one supposes that an intrinsic property is, whatever else it is, nonrelational (in the sense that its possession by some particular does not depend on some other particular), then it is an open question whether anything at all intrinsically possesses intentionality (see, e.g., Heil, 1981). Although Sayre's discussion may lead one to suppose that he has something like this in mind, the account he offers makes it clear that he means only to exclude cases in which a thing's intentional properties are parasitic on the intentional properties of something else. The intrinsic intentionality of neural configurations, for instance, does not arise, in his view, just from features of those configurations "in and of themselves," but from a certain relation of neural to external configurations.

Very roughly, the idea is this. *We* decide on the semantic content of the states of an ordinary computing machine by providing interpretations of its symbolic input and output. Were such a machine linked, however, via appropriate "transducers" directly to the world, the content of its internal states might more plausibly be regarded as intrinsic. In Searle's (1980a) Chinese-room example, the content of inscriptions passing through the system is said to be dependent on the intentionality of those outside the room for whom the inscriptions have meaning. Now imagine a similar room that provides inputs from the outside world without the mediation of intentional agents. The room is equipped with a transducer, perhaps, that converts incoming patterns of energy into inscriptions that are then manipulated as in the original case and, once processed, determine symbolic and behavioral outputs. Sticking to something approximating vision, we may suppose that the transfucer in question consists of a television camera that produces not images on a screen but arrays of Chinese characters *describing* incoming scenes. Behavioral outputs might then include, in addition to further Chinese characters, camera adjustments and the like.

Is a story of this sort a story about intentionality? If the original Chinese room contained no intrinsic intentionality then I think this one does not either; not, however, because we have imagined the room as containing Chinese symbols, hence symbols whose intentionality is derivative. One may substitute any system or code one pleases, and the point remains. If I am in the room manipulating items that have certain external origins and certain "behavioral" effects, I do so without any inkling whatever of their semantic content. The latter plays no role at all in my operation or the room's.

One may be inclined to describe the system – the room, together with its operator and associated transducer – as intrinsically possessing intentionality. There is something to this suggestion. However, a crucial point emerges: The system (describable now as one with something approximating nonderivative intentionality) does not make any use whatever of the semantic content of its internal symbols. These play no discernible role in its operation. (If one is bothered by the fact that the envisaged system is "programmed" by cognitive organisms, then one may imagine an identical device resulting from a cosmic accident. Here there can be no question of parasitic intentionality.)

Cases of this sort suggest, I think, that the prospects of finding a mechanism of intentionality inside intelligent creatures are not encouraging. If it is right to regard intentional states as supervening on ordinary physical states then we will be obliged to say that their "supervenience base" (Kim 1982; Heil 1983) includes more than the current internal features of the creatures possessing those states. It includes, in addition, a heterogeneous assortment of goings on – biological, environmental, historical, and social – external to such creatures. And it is difficult to see how such things could figure in the sorts of behavioral explanation that Sayre is after.

Sayre's attempt to explicate intentionality in terms of the identity of "info(t) structures" raises other questions as well. Semantic content is thought to emerge when the info(t) of a certain neural structure, *C*, is identical to the info(t) of some structure in the world, *O*. When *C* serves to adjust behavior in a particular way toward *O* the info(t) at *C* is converted into info(s) and perceptual intentionality is born. Consider a pair of difficulties that any such account might face.

First, the lack of a causal component is especially troubling. MTC (mathematical theory of communication), as Sayre describes it, does not require causal links between a pair of structures for one to bear info(t) about the other. Without some such link, however, it is difficult to see how Sayre's attempt to derive semantic content from info(t) structures can avoid familiar Twin Earth sorts of counterexample. Smith, an Earthling, is looking at a banana. On Twin Earth, Smith's Doppelganger, Smythe, is undergoing identical neural transformations. Does Smythe see Smith's banana? And if not, why not? Sayre might wish to reply that Smythe's behavior is not directed on the appropriate object, namely, the banana confronted by Smith. It is not merely the identity of info(t) structures that is required for the emergence of info(s), semantic content, but behavior directed toward external objects as well. It is this that is missing in the case of Smythe. Such a response, however, misses the difficulty, for in such cases what determines the appropriate object? The object on which behavior is directed must be identical to the object represented in the relevant intentional state. And there seems to be no independent way of ascertaining which object is that object. Indeed, this is *the* problem of intentionality (in one of its many guises).

Second, Sayre's account of info(s) and the emergence of semantic content from contentless info(t) raises the question posed earlier: Why should it matter? The system as described by Sayre seems only to react to and operate on info(t) structures. When these satisfy certain conditions, they may be described as having a particular semantic content. This is not something that plays any obvious role in the operation of the system, however, at least not if Sayre's discussion is on the right track. (Nor, for all that, does it seem to matter that certain of its neural configurations happen to be info(t) structures. What matters is just that they have certain neurophysiological properties.)

The point is not that intentionality – mental content – has no role to play in the explanation of behavior, only that Sayre's account, despite its focus on intentional substrata, provides it with no such role. Not only does that account fall short of its espoused goal – that of providing a characterization of the mechanism of intentionality – it undermines as well one important motivation for such a characterization.

Information, causality, and intentionality

David Kelley

Cognitive Science Program, Vassar College, Poughkeepsie, N.Y. 12601

Sayre uses information theory to construct a very suggestive hypothesis that does, as he argues, provide an alternative to the computational approach and would help put theories of cognition into a wider biological context. If neuroscientists confirm his hypothesis by finding actual cortical events and patterns that have the info(t) structures he describes, the result might well be a comprehensive theory of how the nervous system gives rise to intentional states. But Sayre makes a stronger claim: "The relationship of intentionality is quite literally is a relationship of high mutual information between a set of objective circumstances and a representation in the cortex." And this identity claim runs into a classic problem that arises for causal theories of perception.

In the case of perception, a theory of intentionality must explain the relationship between the percept and its intentional object – for example, the yellow sheet of paper I am seeing at present. A causal theory, in its crudest form, identifies the intentional with the causal relationship between percept and object. The fact that the sheet of paper is the object of my perception and the fact that the paper caused my experience of it are the same fact. The problem is that the paper is only one stage in a causal sequence leading to the perceptual experience. The other stages – events in the medium of light transmission, in the retina, in the lateral geniculate nucleus – might equally well be described as causes of the percept, yet none of them is an intentional object. I do not *see* any of them. So causality per se is not sufficient for intentionality.

To avoid this problem, causal theorists have imposed further conditions on the causal relation, conditions that will pick out the physical object and exclude the other stages in the causal sequence. Sayre specifies those conditions in terms of information theory. There must be an identity in info(t) structure between the set *C* of cortical events and the set *O* of "object-centered" events responsible for the reflection of light. But this identity in structure is established through the identity each has with a third, intermediary set (*R*) of events at the retina. "In brief, if both channels *O-R* and *R-C* are characterized by high mutual information, then so is the combined channel *O-C*" Then why is the structure at *O* the intentional content of the cortical events, while the structure at *R* is not?

The problem here is not avoided by the further conditions that Sayre says must be satisfied before a relationship of high mutual information can constitute a relationship of intentionality. (1) The perceptual process must "lock on to" and "track" the structures in the objective environment. But presumably it does so by locking on to and tracking the retinal events that carry information about those external structures. (2). The relationship of mutual information between *O* and *C* must be maintained by adaptive procedures, making use of a variable coding format. But once again, the same adaptive procedures are presumably used to maintain the mutual information relation between *R* and *C* as well. Finally, (3) the info(t) coupling between *O* and *C* must make a practical difference to the organism, but of course whenever that is true it will also be true that the coupling between R and C will also make a practical difference.

In short, Sayre does not explain why it is the external object, not the retina or the optic array, that we perceive. The web of relationships he posits may well turn out to play an important role in explaining the genesis of intentional states, but so far as I can see they do not constitute intentionality.

Semantic information: Inference rules + memory

Michael Lebowitz

Department of Computer Science, Columbia University, New York, N.Y. 10027

Sayre has proposed an interesting way of looking at how information is transformed through the stages of an understanding system. As Dennett (1983) and others have pointed out, it is often useful to look at problems using a number of different levels of abstraction, and the one selected by Sayre should be valuable in examining certain aspects of cognition. However, I must disagree with his appraisal of symbolic computer models as having no real semantic content and of "computations on internal representations" as unable to exhibit intentionality.

As Sayre quite accurately points out, the existence of a program that transforms input (a scene or natural language, perhaps) into internal symbols – for example, the phrase "a cat" into the symbol, "CAT" – does not automatically imply semantic significance for the representation. However, symbolic representations in artificial intelligence (AI) systems are more than just isolated symbols. They are, first and foremost, bundles of inference rules. The symbol "CAT" might refer to the facts that such entities walk on four feet, like milk, are mortal Sayre's example), and dozens or hundreds of other rules. While small examples make it seem as if this is a trivial aspect of a system, the rich, interrelated rules in a knowledge base are indeed the semantic core of symbolic information. We must not be confused by the fact that a few simple rules in a system (or even the knowledge bases in existing AI programs) clearly do not exhibit the kinds of complex behavior of biological systems. There has been no reason yet put forth to make us believe that complex sets of symbolic rules cannot encompass all aspects of understanding. Results to date would indeed lead one to believe the opposite.

Conceptual dependency (CD), developed by Schank (1972), is a good example of the true essence of a semantic representation. Consider a simple action such as "John gave Mary a book." This action is represented in CD as:

```
                                    to
                  object            ----> *MARY*
*JOHN* <=> ATRANS <------ *BOOK*
                                    ----< *JOHN*
                                    from
```

This representation includes some elements that are significant semantically and some that are not. Specifically, since CD was more concerned with representing actions than objects, the representations of John, Mary, and the book as *JOHN*, *MARY*, and *BOOK*, respectively, have essentially no semantic content – the asterisks simply indicate that a conceptual representation is needed. ATRANS, a language-independent primitive that stands for abstract transfer of control, is another story. This is not the case simply because we use a symbol other than those in the sample sentence, but rather because the definition of ATRANS includes a number of inference rules about transfer of control. For example, a system might have a rule that the filler of the TO slot of an ATRANS now possesses the filler of the OBJECT slot, another that the ACTOR of the ATRANS wanted the filler of the TO slot to have the OBJECT, and so forth. To the extent that these rules completely describe the effect of ATRANS in the world, the system has semantic information about the action.

AI researchers have sometimes confused the issue of what carries semantic content by using seemingly significant names for concepts (such as the "IS-A link" used in semantic networks to represent class membership and allow inheritance of information). As was pointed out by McDermott (1981), the only semantic significance of representations lies with the rules that they entail. However, these rules do constitute semantically meaningful information. We can, as McDermott points out, test the computer implementations of systems by replacing all the "wishful mnemonics" with arbitrary symbols and convincing ourselves (and others) that the system exhibits aspects of understanding.

Recent AI theories and computer systems have added another key component to the semantic meaningfulness of symbolic representations – long-term memory. Programs such as IPP (Lebowitz 1980; Lebowitz 1983a), CYRUS (Kolodner 1984), and RESEARCHER (Lebowitz 1983b) take streams of input (news articles in the cases of IPP and CYRUS and patent abstracts for RESEARCHER), determine the conceptual content of the input, and add it to a long-term memory. The updating of memory involves comparison with earlier examples for the purpose of generalization. Such systems provide a clear connection between representations and the outside world.

If we gave RESEARCHER (or perhaps a much more advanced version of it) rich enough input about cats instead of disk drives (its current domain), it would build up descriptions of various kinds of felines. It would notice how the simplest (primitive) elements of the various representations relate to each other and would develop higher level concepts. The generalized concepts could then serve as the basis for later understanding – in effect, the rules of understanding in a system of this sort can change in response to inputs received. The dynamic nature of systems that build up memories from their input embodies the programs with another important element of semantic representation.

Having briefly looked at the role of inference rules and dynamic memory in symbolic representations, we can consider the relation of such representations to intentionality. Searle, in describing intentional states, requires that they "be *satisfied* or *not satisfied* depending on whether representative content actually matches or represents anything in reality" (1980c, p. 48). This important element of intentionality can easily be captured by symbolic models. Their rules of inference can be applied to their world (further input) and used to judge the satisfiability of a given representation. The connection of representations to a permanent memory forces such intentionality to be more than isolated symbol pushing.

We can, it would seem, safely conclude that symbolic representations do capture a meaningful part of semantic meaning – as long as we include attached inference rules and the ability for these rules to be dynamically changed in response to input. "CAT" in such systems is not just a meaningless symbol, it is a semantically significant bundle of rules and memories. By looking at systems' abilities to interpret input, predict the consequences of input, and relate new input to old memories, we can see all the elements of intentional states. The only way to deny the intentionality of such symbolic representations is to declare that the input given to computer systems is somehow less a part of "reality" than that received by biological organisms. And there would seem to be no valid reason to do that.

ACKNOWLEDGMENT

Preparation of this commentary was supported in part by the Defense Advanced Research Projects Agency under contract N00039-84-C-0165.

The relationship between information theory, statistical mechanics, evolutionary theory, and cognitive Science

Michael Leyton

Department of Psychology and Social Relations, Harvard University, Cambridge, Mass 02138

Sayre's target article puts together a particular relationship between ideas in at least the four following areas: (1) information theory, (2) statistical mechanics, (3) evolutionary theory, and (4) perceptual processing. I will argue that the particular relationship he puts together is incorrect on both logical and empirical grounds. I will do this by developing a different relationship between the above areas and showing that this alternative is logically more consistent and empirically better supported.

(1) Negentropy and energy flow. After describing his perceptual system, Sayre claims that "because of the high level of $I(O;C)$ sustained by the perceptual process, the perceiving organism enjoys the benefits of a highly efficient negentropic coupling." However, the concept of negentropic coupling is a highly technical one in nonequilibrium (i.e., irreversible) thermodynamics, and Sayre has not substantiated his mention of it. Briefly, the relationship of negentropic coupling to living systems is as follows: In order to prevent the decay of a living system to equilibrium, work has to be performed. However, because work cannot be performed by an isolated system, a crucial requirement is that the living system be an intermediate one between an energy source and an energy sink. The required work is then accomplished as a result of the energy flow from source to sink through the living system. That is, energy flow, not energy itself, is crucial to negentropic coupling. Sayre does not mention energy flow in describing his perceptual system.

(2) Ecology and the second law of thermodynamics. There is in fact an important direct relationship between energy flow and negentropic flexibility. Elton (1927) discovered that an organism that is higher on a food chain is rarer and has more food chains converging on it, thus making it more flexible. The classic paper by Lindeman (1942) provided the following analysis of these phenomena: A fixed amount of energy enters the ecosystem from the sun via photosynthesis. As it is passed up successive ecological levels, each transfer across levels, via feeding, must be less than 100% efficient (in accordance with the second law of thermodynamics). That is, some energy must be degraded into a lower potential state such as heat. Thus there is less and less useful energy available at each successive level. This forces animals on higher trophic levels to be fewer and more versatile. The above argument therefore provides the relationship between negentropic flexibility and the thermodynamics of energy *flow* that is missing from Sayre's article.

(3) The principle of requisite variety. I do agree with Sayre, however, that Shannon's 10th theorem is an important factor bridging ecological flexibility and the demands on the cognitive system. Neither Sayre nor I, however, is at all original in believing this. We are predated by at least thirty years in a famous argument by Ashby (1956), who proved a principle, called the Principle of Requisite Variety, related to Shannon's theorem. However, where I think Sayre is wrong can be expressed in the following argument by Ashby: Flexibility necessarily demands vulnerability to external disturbances, in a way that nonflexibility (e.g., in the form of a hard shell) does not. One consequence of this is that the organism cannot simply wait for disturbances. It must actively seek environmental information for the general purpose of providing a continual description of the environment. Such information I will call *nonspecific.* The factor of higher nonspecific information is a major one that seems to violate Sayre's proposal, as follows:

(4) High-level nonspecific information-seeking. Although Sayre is vague on this point, I have a strong sense, from several aspects of his description of the "perceptual–behavioral" loop, that perception in his system acts in the service of immediate ongoing manipulative tasks. For example, (1) he claims that higher perception "locks onto" or "focuses on" specific objects in the behavioral loop; and (2) he claims that cortical representations retain "just those info(t) structures that are important for the guidance of the organism's current behavior." There is considerable evidence, however, that flexible biological organisms seek information purely for the latter's complexity and novelty, rather than for extrinsic utility. Furthermore, nonspecific information-seeking cannot be consigned to a low level of the perceptual system, as in Sayre's view. For example, three types of higher information-seeking phenomena would seem to cause problems for his view:

(A) Sensory deprivation. A man sitting in solitary confinement does not require the accomplishment of any tasks; for example, food and protection is provided without work. Nevertheless he does report seeing the environment. According to Sayre's view, that high-level perception retains only information relevant to specific tasks, he would not see anything.

(B) Information seeking during goal accomplishment: Several studies show that animals prefer an indirect route rather than a direct route to a goal if the indirect route is perceptually more complex (e.g., Krechevsky 1937). In Sayre's system, in which nontask-oriented information is rejected at higher levels, this presumably should not happen.

(C) Behavior for perceptual goals. Several studies show that animals learn to perform tasks in which a neutral percept is the only reward (e.g., Kish 1955). These studies actually reverse what Sayre calls the perceptual–behavioral loop in that the organism behaves in the service of perceiving, rather than perceiving in the service of behaving.

Furthermore, one could not solve these problems by claiming that information seeking is just one of those tasks that Sayre's perceptual system attempts to facilitate, for this would lead to two irreconcilable demands: (1) According to Sayre, the goal of high-level perception is to stabilize info(t) structures with respect to current function; and (2) the function of high-level information seeking is the destabilization of info(t) structures.

In summary, then, the successive stages of my argument have been: (1) It is energy flow that allows negentropic coupling; (2) thermodynamic constraints on energy flow drive higher organisms to flexibility; (3) Ashby's principle forces flexible organisms to seek high-level nonspecific information; (4) high-level nonspecific information is incompatible with Sayre's proposals.

Intrinsic versus contrived intentionality

Donald M. MacKay

Department of Communication and Neuroscience, University of Keele, Staffs. ST5 5BG, England

The impasse from which Dr. Sayre wants to rescue cognitive science results, I think, from the established practice in artificial intelligence of taking a nonbiological approach to the internal representation of the external world, using intrinsically meaningless symbols. I have long argued (MacKay 1951; 1962) that the key to intrinsically intentional representation is to start by thinking of the artificial system as primarily a goal-pursuing, norm-guided, evaluative agent, as we do a biological system. The environment or field of action can then be thought of as that which sets constraints (both positive and negative) on the planning and organization of goal-directed action, which needs to be represented internally by a matching set of constraints. What is required is not just that the system should interact freely with its environment but specifically that it should probe its world in order to develop a matching state of conditional readiness to reckon with the cognizable features of that world (MacKay 1956: 1962; 1978; 1981). The crucial ingredient in cognitive agency is evaluation, computation of the extent to which and the respects in which the current state of affairs matches or mismatches current goal-criteria or the current state of conditional readiness. Feedback from evaluation drives the agent to seek actively to maximize "match" or minimize "mismatch."

In a system functioning according to this principle, feeling its way into developing an internal structure of statistical constraints to match those imposed by specific features of its world, the internal structure naturally functions as an intentional representation of those features (MacKay 1956). It is not a representation in the sense of an array of arbitrary tokens that require a semantic interpreter to discern their meaning. It is constructed of elements with a nonarbitrary meaning, which function directly at the interpretive level. It embodies implicitly something the agent "knows" about its world, in that setting it up means *ipso facto* setting up the relevant implications of the information it represents where they are needed.

A simple (and ancient) illustration may clarify the point (MacKay 1963). In a car designed to use feedback or feedforward from a sense-organ to adapt the setting of its steering wheels automatically to the curvature of a semicircular drive with the goal of avoiding collisions with the walls, the steering angle of the wheels becomes an implicit representation of the curvature. It is not a scale model of the drive nor a linguistic description of it, but it embodies implicitly what the system needs to know about its shape, with respect to the goal of drive-following. No further step of interpretation is needed to discern its meaning for the goal-pursuing system. By the same token, perception of a change in drive curvature would be represented by the updating of the steering angle in matching response to sensory signals, whether of feedback or feedforward. As such, perception would be qualitatively distinguishable from the reception and analysis of sensory signals that led up to it. Perception of an abstract feature of the shape of the drive (say, a regular alternation of left and right turns) would require the development of appropriate subroutines to set up a matching conditional readiness to turn left after turning right and vice versa (MacKay 1956). On this basis perception would be intentional, in a sense in which sensory signal-analysis *per se* is not.

In this context, I am afraid I do not see Sayre's invocation of (Shannon's) information theory as adding anything helpfully new. His assertion that the joint capacity C^S of two channels with capacities C^1 and C^2 can be increased by making their inputs interactive depends on an unexplained assumption. Why should the equivalent number of equiprobable alternatives $2C'$ for the joint output be the sum of those for each channel $2^{C^1} + 2^{C^2}$, rather than their product $2^{C^1} \times 2^{C^2}$? Does Sayre have in mind some special kind of interaction that would justify this step in the argument? If so, he should specify it more explicitly, and qualify his conclusion accordingly. Introducing redundancy by allowing two input signals to interact would normally be expected to reduce rather than increase overall rates of information transfer, though it could of course be used to increase reliability.

What Sayre calls "info(t)" means simply "unexpectedness." It seems incoherent to talk of this as becoming info(s), because the semantic concept of information itself (that which determines form) is in quite a different category (MacKay 1969). (Incidentally, what I did in 1969 and earlier was to show how Shannon's theory of unexpectedness could be grafted on to a theory of meaningful communication, and not vice versa as Sayre suggests in Section 4, paragraph 3!) Sayre's central emphasis on "constraints from a demanding environment" is generally in line with what I proposed in 1956; but it is not obvious (to me at least) that taking Shannon's "mutual information" as a measure of the coherence between a set of objective circumstances and their cortical representation throws additional light on the basic philosophical problem Sayre is addressing.

Cognitive science and the pragmatics of behavior

Lawrence E. Marks
John B. Pierce Foundation Laboratory, Yale University, New Haven, Conn. 06519

How much should we expect a theory of mind and behavior to explain? Consider the following relatively simple and certainly commonplace event. A man is dressing. His immediate task is to select a necktie appropriate in color, material, and pattern to complement his brown gabardine suit. What does an adequate psychological theory need to account for in the way of the mental and behavioral processes that ensue? More to the point, how much of this explication can cognitive science (CS) hope to accomplish?

CS functions at its best in providing a computational account of the processes that underlie the man's search procedure: He scans the rack of ties (serially); one tie after another he compares both indirectly to a mental prototype of the tie he "desires" and directly to the brown jacket already donned. Yet, as Sayre implicitly argues, is there not much more going on? The constraints on the man's behavior – or, to put it more accurately, the constraints on theories to explain the man's behavior – far transcend, it seems to me, what CS can hope to supply; indeed, the constraints transcend, I suspect, even CS plus the semantic content that Sayre offers as a supplement.

Let us ask, what are the factors that shape choices of colors (leaving out from consideration pattern, material, etc.)? In part, cultural conventions and personal experiences may combine to dictate the range from which the man will choose a suitable color. But some of the constraint in color preference undoubtedly arises from visual biology – from the peripheral, receptor-cell kinetics and subsequent central opponent-process mechanisms of color vision. Such biological constraint makes an orange hue more similar to a slightly yellowish red than to a good yellow, despite the greater physical difference between the dominant wavelengths (or wavenumbers) corresponding to the orange and yellowish red.

But the biological underpinnings to the man's behavior go far deeper than the visual physiology of color. For it is clear that the cognitive processes underlying the man's search for sartorial appropriateness are neither ends in themselves nor processes geared to mediate visual preference; rather, these processes subserve a wide range of activities – many of which can be characterized as being in the broad sense adaptative. If the search for a necktie were to precede an important business meeting, the choice of a particular color might play a modest but perhaps not insignificant role in an outcome deeply desired. It takes no great leap of imagination to construe the selection of a necktie as one small piece of behavior nested within a hierarchical organization of activities with clear purpose, value, and biological function – and I mean here biological not just in the narrow sense of satisfying a short-term physical need but in the broader, Darwinian sense. Behaviors as (relatively) minor as selecting a necktie have their niche, play their adaptative role. More to the point, the underlying cognitive mechanisms – those mechanisms that CS aims to describe – themselves presumably evolved as the result of adaptative pressures. If, as Sayre argues, CS is a theory of the syntactics of mind, lacking a semantics, it also lacks a pragmatics, an account of mind's functional properties (encompassing the intentional ones), as these may be viewed from a broad biological perspective.

Actually, I should clarify my example: The man I have in mind is not preparing for a business meeting, but rather for dinner with a woman to whom he is strongly attracted. The tie he finally chooses has a deep red color that might not be predicted either on the basis of cultural conventions or from an analysis of visual mechanisms. Nor is the man aware of any reason underlying his choice, for he has long forgotten – Freud would say, repressed – the memory of a time when he was four or so years old; in those days, his father often wore a necktie of precisely the same color, and in fact had been wearing it early one particular evening a short time before the boy had entered his parents' bedroom. . . .

A total process approach to perception

Maxine Morphis
Department of Philosophy, University of Hartford, West Hartford, Conn. 06117

Sayre has taken seriously the need to explore the technical nature of the transitions from the quantity of information with which a system deals to the semantic content of that information. Although I think that he has fallen short of clarifying how nonbiological systems can qualify as intentional systems, Sayre's work suggests some rather more important and interesting considerations.

Sayre claims that intentionality arises from the identity of info(t) structures of retinal events with corresponding structures in the objective environment. Unfortunately, Sayre tells us more about what he doesn't mean by "identity" than what he does mean by it. He retreats to a use of identity as "of a sort akin to the identity of mathematics," which tells us little. He then states that a system does not necessarily have to be biological to manifest these characteristics.

Although Sayre distinguishes his theory from Dretske's(1981), it is not clear how his work escapes the criticisms leveled at Dretske's work, for example, by Churchland and Churchland (1983). Moreover, Sayre does not explain how these relations of identity with their functional applications provide us with a view of info(s) that is more acceptable than the ones provided by variations of the robot reply discussed by Searle (1980a, p. 420). Why is Sayre's model not merely reflecting a greater level of mechanical complexity – a difference in degree rather than kind?

Can Sayre, or anyone, give us the answer to the question of whether nonbiological entities can be intentional and have info(s)? As the debate stands, we must suspect that there is something misconceived in the enterprise. Regardless of the complexity of robot behaviors proposed, an objection can always be raised: "I know that you describe the system in question as doing X, Y, and Z, but is it *really* intentional?" Herein lies the questionable strength of the list of disabilities proposed by Turing (1950). Thus I conclude that Sayre has not achieved his main goal – but I am not sure that the goal is achievable as approached at present.

Sayre's more important contributions arise from the implications of his application of MTC (mathematical theory of communication) to studies of the human cognitive system. What we may need is his basic approach, but done with a much broader and more adventurous hand. Sayre has still not quite explained the move from info(t) to info(s) given a *biological* system. Let me offer what might be the next step. Following Sayre's lead, and claiming the same caveats in which Sayre wraps himself, I will center the discussion on biological systems for the moment.

Like other levels of explanation, such as the neurological or the behavioral, MTC provides a way of looking at perception systematically. However, we must keep in mind that to say, for example, that the goal of the system is "maximizing the efficiency of the negentropic coupling between organism and environment" is only one way of expressing it. There is an artificiality to the way in which we keep each level of explanation separate (Norman 1981), pursuing each to its end and then wondering where the others connect. A good example of this is demonstrated by Dennett's (1978) tracing of the subpersonal chart for pain, which allows no slot for the feeling of pain.

My argument is not against the use of the MTC approach but against, for the want of a better term, any linear approach to the explanation of perception. To understand intentionality or semantic information is like the corresponding effort to understand pain. It must be looked at in its totality, not in parts.

Although Sayre grants that intentionality is a dynamic relationship that includes interaction with the environment, he cuts his view of this relationship too short too soon. For example, his discussion of the cascades of information channels may be a move in the right direction, but he does not go far enough. In talking about redundancy, he brings up the subject of memory but then limits his discussion of memory to percept levels. Elsewhere, as he discusses the Accumulator, he again fails to explore complex connections to other cognitive areas. He does mention pattern storage facilities, but why is it useful to try to see the Abstractor as functioning without resources in higher levels of cognitive function? Although Sayre mentions that high levels of mutual information let an organism pursue its interests, how shall we interpret this?

Information theory as an independent measure of person-independent signals has been found irrelevant by psychologists, because information received is dependent on the recipient (Haber 1983). Linear approaches to understanding a perceptual system continually fail to account for many factors. The awareness of an organism of its own states is intimately wound up in these multifaceted connections. Indeed, it has often been suggested that the eyes are part of the brain. Visual events do not happen in isolation, any more than pain events happen in isolation. Visual perception occurs according to the context of emotions, experiences, and expectations. A system that cannot know also cannot see, although it can receive and generate information (Dretske 1982, p. 153). In an emergency, we don't notice pain, nor do we "see" what it is not urgent to see. We would not know what it means to see devoid of thought and behavioral responses any more than we know what it would mean to have pain devoid of the complex variety of responses with which it is commonly associated (Morphis 1980).

How can we isolate the visual perception of a cat from all associated experiences and still be interacting with the environment? In an extreme case, I might not even be able to distinguish a totally new item from other items, but I am not often in this position. More commonly, I am in only somewhat unfamiliar circumstances, which allows me to search for applicable parallels, seeking confirmations and completing partial information through other senses whenever possible. We scan according to our goals and desires, and also scan what seems anomalous.

If we are to identify the "aboutness" of semantic information, it will not be found in one location or series of events any more than the feeling of pain can be so located. Cognitive science must develop a clearer model of the extraordinary complexity of the human perceptual system; the totality of this model will reveal what we want to know about intentionality. Such a model may not immediately do anything to help resolve the issue of whether robots can then be intentional, but one goal at a time. As the concepts of parallel processing develop, perhaps these will give further clues about the ways in which multilayered feedback systems can complete a "total process" model of the intentionality, with MTC forming one of the parts of that totality. [See Ballard: "Cortical Connections and Parallel Processing" *BBS* 9(1) 1986.]

Intentionality as internality

Don Perlis[a] and Rosalie Hall[b]

[a]*Computer Science Department and* [b]*Psychology Department, University of Maryland, College Park, Md. 20742*

Sayre strives valiantly to understand meaning by linking internal processes to external ones. However, this approach seems in conflict with his own argument against the so-called official stance. We suggest that his argument against the official stance can be put in the following "strong" form: If an entity *S* (person, mechanism, etc.) is to be said to attribute a meaning (of its own) to one of its internal objects (process, token, etc.), then this attribution of meaning *by S* must itself occur *within S*. Moreover, since the removal of entities external to *S* does not in general affect whether *S* is attributing meaning to its tokens, such an attribution of meaning by *S* should not depend on anything outside *S*, that is, should be determined solely by phenomena within S. This position we shall call the "internal" stance.

Sayre's solution focuses on the "relationship of identity in info(t) structure between *O* and *C*" – which he says "*is* the intentionality of perception." In fact this identity is external to the mechanism *S* embodying the "structure at *C*" said by Sayre to be "enjoying this relationship." His "structure at *C*" does not have (is ignorant of) *O* or any relationship between C and *O*, at least as far as Sayre describes it. This relationship is there in the world for Sayre to observe and report truly to us, but in what way is it there for *S*? Sayre is silent on this.

For *O* to carry the meaning of C for S, this meaning must, if we hold to the internal stance, be something such that it would make a difference to S were the so-called meaning to be altered.

But as Dennett (1978) and Stitch (1984) have amply illustrated, the external contexts can be modified ad nauseam, to the point that we ourselves are no longer sure how to ascribe meanings or truths.

Sayre might reply that these criticisms depend on a richer cognitive setting than that of visual perception. However, if we restrict ourselves to Sayre's criteria for intentional perception, our strong argument again suggests why they are not adequate. For example, it seems very strained to argue that because a visual mechanism can "track" 3-D mock-ups of roads (as is being done at the Computer Vision Laboratory at the University of Maryland) it thereby "intends" to follow a road of which it has no concept. (More on this mechanism below.) Sayre employs evocative but question-begging terms: "focused on," "tracks," "locks on," and discusses mechanisms that carry out functions intuitively related to these terms, but he does not seem to get away from the underlying quandary.

If we accept the internal stance (that whether *A* means *B* for *S* must depend only on elements that reside wholly within S), then we can propose a partial solution: that a form of self-reference occurs within S, so that, at a rough gloss, we might say "S perceives itself as engaged in a process of focusing on *B* via *A*." Without further elaboration, of course, this would be an idle and circular notion. We are at present working on a theory of intentionality along lines suggested by such a gloss and will in the following discussion try to contrast our approach with that of Sayre.

In our approach, *S*'s attribution of meaning *B* to *A* occurs only if *A* and *B* are entities within S. Any external connection to entities *O* outside S are not part of attributions of meaning properly ascribed to S. (We do grant that external entities may well be part of what it means [to us] for the meanings within S to be realistic, much as our seeing in our mind's eye a unicorn is not realistic although it is certainly meaningful for us.)

For example, with an automatic visual road-follower, it makes all the difference in the world whether the mechanism has a token "*A*" that is its road-concept, to which it compares its road-image token "*B*," as well as its self-image "*I*" comprising tokens for its effectors and sensors. Only then, it would seem, could it make sense of such a thing as being "off-track" or "nearly back on track," etc. Thus, elements of reasoning (in the form of internal assertions) appear absolutely central to the very prospect of intentionality, even in the restricted domain of visual perception.

That is, unlike Sayre's *O*, which is external, the meaning of *A* is another token or process within S, call it *B*, which S represents as something "tracked" or "focused on" by a perceptual comparator *A*. Figure 1 hints at the unavoidably complex scenario this requires:

Here *O* is in fact unknown to S, it cannot be put inside S. Nonetheless, S can "know" of its tracking efforts by having both processes (tracked [*B*] and tracker [*A* and *I*]) present in its tokens, and, by comparing *B* to *A*, S can take steps to rectify any imperfections in *B* (such as being out of focus). Of course, for this to work well in a context of external visual inputs such as from *O*, the effectors and sensors must be well-tuned. Therefore *O* is not necessarily irrelevant to the overall behavior of S, but it is not the criterion for the occurrence of intentionality.

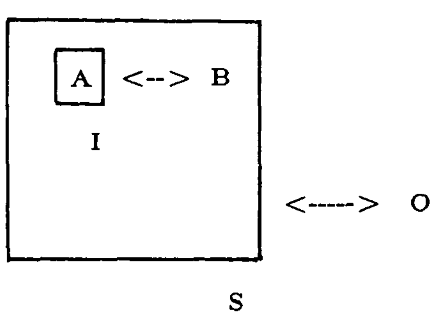

Figure 1 (Perlis and Hall). *A* means *B* to S and loosely corresponds to *O* (see text).

We confess that Figure 1 grossly oversimplifies things. There is more properly no fixed *A* or *B*. Rather, a quotation device should allow fairly unrestricted "reflection" of any element in *S* "into" *I*. Minsky (1968) addresses this issue but seems to think a hierarchy or network is called for. In Perlis (1985) it is suggested that this is unnecessary.

Intentionality: No mystery

William T. Powers
1138 Whitfield, Rd., Northbrook, Ill. 60062

Sayre flirts with the idea that would solve the problem of intentionality, the concept of control, but his treatment of that subject is too superficial and uninformed to succeed. Perhaps this is because, like generations of scientists before him, he assumes from the beginning that intentionality is only a way of speaking, an idiom. With that *a priori* admission of defeat in the background, it is not surprising that his treatment of intentionality wanders into the underbrush of verbal abstraction where neither the trees nor the forest are visible.

Sayre's treatment of control might fool a layman, but any control engineer would find it, to be kind, inadequate. Every time the discussion of control comes to a thicket, Sayre waves his arms and presto! we have popped out on the other side. For example, he speaks of "the ability of a control system to correct deviations from an optimal mode of operation." The word "optimal" aside (optimality has almost nothing to do with control), control systems do indeed have the ability to correct deviations (from a set-point, whether optimal or not), and this ability is a strong hint about intentionality. Whether a given state of the environment constitutes a "deviation" depends on some criterion for the "right" condition, which in control theory is physically embodied as a perfectly real physical reference signal inside the control system. Without a reference signal for comparison, no perception, no input, implies anything about the behavior of a control system.

In the section on the "Discriminator," Sayre says "The behavioral side of the control loop begins with the various stages of the efferent nervous system, initiating control signals that are translated into bodily activity undertaken with respect to objects in the proximate environment." This is, of course, the problem with the normal input–output model. The "control" signals (a misnomer) are translated, somehow, into just the motor actions needed to produce a consistent result in a variable environment, meaning that they must (and do) change according to external disturbances and constraints, even invisible ones. Discovering how this could possibly occur would solve the problem of intentionality. But Sayre simply proceeds: "when the organism is active and its control system [just one?] is functioning normally, its bodily activity is directed toward, or with respect to, the same object, or set of objects, that produces the signals at the input of the perceptual side of the loop."

The phrase "directed toward" again contains the heart of the issue. How is direction established, and by what? This would be an opportunity to discuss orienting as a control phenomenon, but Sayre presses ahead. "The primary function of the control loop is to enable the organism to adjust its behavior with respect to those objects in a manner responsive to its needs and interests." And finally, "For the behavioral side of the control loop to serve its function adequately with respect to a given object, the perceptual side must provide an adequate representation of that object at their cortical interface." Yes, yes, now you've got it!

But he hasn't got it. He has a perceptual representation of the object as it is, a behavioral output that can affect the object (and thus the perceptual representation of it), and the concept that the behavior must affect the object (and the perception) in a way "responsive to needs" rather than in other equally possible ways. But he stops there, within a millimeter of control theory. Suppose the "need" is another signal, representing not the actual state but the "needed" state of the object. And suppose the behavior is driven by the difference between the perception and this reference signal. The result is a standard arrangement for a physically realizable control system that varies its actions to make the "object" approach the "needed" state and remain there despite disturbances. The signal representing the "needed" state is identically one intention regarding that object. The problem is solved. We could build such a system (and have done so hundreds of thousands of times over the past 40 years). This is no metaphor.

Having come that close to the answer he seeks, Sayre abandons the chase and starts pursuing info(t) and info(s) and other creatures of the imagination. Not that I object to them: they may or may not have usefulness in a working model. But they tell us nothing about control, and nothing, therefore, about intention. I do object to Sayre's blithe assumption (like Gibson's) that anyone's brain can compare a perceptual representation with the external object it represents and thus determine its adequacy or verisimilitude. "Ojective circumstances" are, literally, a figment of the imagination. But that is a different subject.

The final proof that Sayre has missed the point is contained in his assumption (which precedes as well as follows the discussion of control) that behavior is somehow "guided" by the information in perceptual representations. While this interpretation preserves the traditional input–output model, it imputes to perception the ability not only to inform the brain about the state of the environment, but to inform it about the proper or intended state of the environment. Without control theory, of course, there is nowhere else one can look for an explanation of the purposiveness of behavior, but with control theory there is no need for this most unlikely postulate. Guidance is a term into which a host of unspoken assumptions are packed. Unpacked, it becomes simply what a control system does, and it is not contained in any one signal. The correct way of describing a control system's behavior is to say that the system varies its output as required in order to maintain its input signal matching its reference signal. The output controls the input, not the other way around. Everyone who really understands control systems understands this description and can cite experimental methods for showing that it is correct and the other model is wrong. Control systems do not control their actions. They control their perceptions, by acting on the source of those perceptions, the outside world.

Intentionality: A problem of multiple reference frames, specificational information, and extraordinary boundary conditions on natural law

M. T. Turvey

Department of Psychology, University of Connecticut, Storrs, Conn. 06268; Haskins Laboratories, New Haven, Conn. 06511

It is refreshing to see a scholar who is largely sympathetic to the so-called information processing or representational/ computational approach to cognitive systems recognizing its fundamental inadequacies. To be blunt, that approach fails to come to terms with either information or intentionality. Sayre's response to these inadequacies, however, keeps close to the received view. He assumes that a biologically and psychologically relevant sense of information can be provided by the mathematical theory of communication; he assumes that intentionality amounts to representation. These assumptions are bolstered by the closely cognate beliefs that intentionality is to be ascribed to some roughly midway state in the classical afferent–efferent link and that there is a metamorphosis from meaningless states to meaningful states. To his credit, Sayre aspires to make the representations genuine. He wants them to stand for real things. He wants the transition from meaningless sensory states to meaningful perceptual states to be (mathematically) principled.

From my perspective as a proponent of the ecological approach to perceiving-acting (see Gibson 1979; Turvey, Shaw, Reed & Mace 1981), Sayre's sentiments are right but his premises are wrong. Not surprisingly, I find his treatment of intentionality disappointing. I concur with Sayre's implicit wish for a concerted effort to *naturalize* (my word) intentionality, but my preference is to keep the deliberations very close to natural science and the search for lawful regularities. Sayre is quite right in his assessment that an attempt to devise an explanation of intentionality in the Turing reductionism/token physicalism perspective of cognitive science (which denigrates intentionality to the states of a computational device) does not have a "ghost of a chance" (Carello, Turvey, Kugler & Shaw 1984; Turvey et al. 1981). But he is quite wrong, I believe, in suggesting that pursuing the purer equation of intentionality with representation (relieved of computational procedures) can fare any better.

Intentionality is directedness toward objects. Locomoting terrestrial animals, including humans, direct themselves through openings and around barriers. They direct their limbs in certain ways with respect to a brink in a surface – directing them one way if the brink is where they can step down and another way if it must be negotiated by jumping. Gibson (1966; 1979; Reed & Jones 1972) advocated mutually constraining theories of animals and environments (see Alley in press; Mace 1977; Michaels & Carello 1981) as the basis for an understanding of perceiving-acting that addressed such mundane intentional behavior. (This central thesis of the ecological approach, the duality of animal and environment [Shaw & Turvey 1982], implies that efforts to ground intentionality only in "environmental constraints" will miss the mark. Duality, by the way, is not dualism.) Gibson pursued a perceptual theory that was fundamentally intentional rather than one that is made intentional as an afterthought. With considerable care he identified how an understanding of intentionality of perceiving poses challenges for science on several fronts, and how these challenges might be met. I will describe two of them.

The first challenge is to describe the layout of surfaces with reference to the animal. This move is continuous with the larger lesson of relativity theory: All state descriptions are frame dependent. Reference frames are substantial and are not to be confused with the coordinate systems that abstractly represent them. The properties of an animal to which surface layout must be referred are basically the animal's magnitudes, its morphology, its metabolism. With regard to a brink, the separation of surfaces is in reference to limb magnitudes. Obviously a given brink can be referred to multiple, equally real frames. One frame is the terrestrial frame with distances and durations measured in arbitrary units. This frame is useful to the physicist but it is, by definition, animal-neutral. (In the received view it is mistakenly adopted as the sole objective frame.) Other frames are individual animals. Consequently, the same brink in the terrestrial frame is a place negotiable by leg extension in the frame provided by one (larger) animal not negotiable in this fashion in the frame provided by another (smaller) animal.

A second challenge is to describe how animals can be informed about these frame-dependent environmental properties (affordances) to which their activities are directed. There are two senses in which the term information is used (cf. Turvey & Kugler 1984). In the indicational/injunctional sense information consists of symbol strings identifying states of affairs ("the situation is so-and-so") or things to be done ("do so-and-so now"). Information in this sense is underconstraining, like a stop sign. The other sense is the specificational sense of Gibson (1979). In the case of vision, information is optical structure *lawfully* generated by facts – properties of surface layout, properties of an animal's movements. This structure does not resemble the facts; rather it is *specific* to them. The ecological argument is that information in the specificational sense meets the above challenge. I will give some examples shortly but I wish to preface them by noting what's at issue in the contrast between the two senses of information.

The indicational/injunctional sense, I believe, fits neatly into a tradition that takes the primary perceptual activity to be discriminating among members of a set and the equilibrium thermodynamics of closed systems as the branch of physics to which discussions of information can be meaningfully referred. In such a system the states are enumerable from the outset. To put it very roughly, the information notion only has to address their individual probabilities, thereby providing a basis for discriminating among them. Living things, however, are open systems. The animal–environment system, in which an animal participates as one of the two mutually tailored components, is open. Significantly, the states of an open system need not be fixed at the outset. Given fluctuations in the microstructure and nonlinearities, a scaling up in one or more variables discontinuously decreases an open system's symmetry. More constraints arise. The system becomes more ordered. New states come into existence. Consequently, the order principle and complexions of Boltzman, and the notion of information that they sustain, are of limited applicability to open physical systems (e.g., Prigogine 1980), including animal–environment systems.

Open (evolving, developing) systems motivate a different notion of information from closed systems (Kugler, Kelso & Turvey 1982; Kugler & Turvey, in press). Sayre makes an offhand remark about the information in the genes and in the phenotype. Efforts to apply classical information theoretic notions to the genotype–phenotype link, conceived as a communication channel, have largely been dismissed. In intuitive terms, the dismissal is based upon a feeling that an information metric should recognize the greater complexity of the full-fledged animal (Waddington 1968). Even where the open–closed distinction is sidestepped, as in Pattee's (1973; 1977) thoroughgoing and celebrated efforts to detail the problem of a physical interpretation of "genetic information," the conceptions of the mathematical theory of communication have proven to be of little value.

The specificational sense of information is consistent with the perspective that takes perceiving the persisting and changing properties of a thing as primary. For Gibson (1966; 1979) the fundamental question is how to characterize the information that supports the perceiving of *P*; the question of how to characterize the information that supports distinguishing *P* from *Q*, *R*, and so on is secondary and derivative. Suppose that *P* is the animal itself. In locomoting, a terrestrial animal generates forces that displace it relative to the surroundings. There are obvious mechanical regularities to be noted. They are ordinarily expressed through Newton's laws. But this situation also exhibits nonmechanical regularities expressed by non-Newtonian laws of wide (though not universal) scope. For instance, all the densely nested optical solid angles, whose bases are the faces and facets of surfaces and whose apex is the point of observation, change concurrently. An optical flow field – crudely, a smooth velocity vector field – is generated. The global form of the flow, or optical morphology, is specific to the configuration of locomotory forces and to the displacements of the animal. Rectilinear forward locomotion, for example, lawfully generates a *dilating parabolic flow; a dilating parabolic flow specifies rectilinear forward locomotion.*

This simple but significant example of information in the specificational sense permits me to make briefly some important points that can be more carefully developed (e.g., Solomon, Carello & Turvey 1984; Turvey & Carello 1985; in press; Turvey et al. 1981). First, optical information in the specificational sense is optical structure whose macroscopic, qualitative properties are nomically dependent upon and specific to (under natural boundary conditions) properties of the animal–environment system. Second, optical information in the specificational sense does not reduce to neural signals in the visual system (see below). Thinking about optical information as alternative (macroscopic, qualitative) descriptions of the photon light field, structured by the layout of material surfaces and defined relative to locations and paths in the transparent medium (air for terrestrial animals), is useful. It aids an understanding of optical information independent of vision and of the kinds of ocular systems that evolved. Optical information in the specificational sense is tied to laws at the ecological scale, laws that relate optical properties to kinetic properties (of the animal–environment system). The ecological approach argues that these laws were the basis for the evolution of, and are the basis for the everyday realization of, locomotor activity and its directedness and intentionality.

Let's extend the example a little. Dilation of an optical solid angle relative to a point of observation specifies the approach of a substantial surface. The inverse of the relative rate of dilation, τ, specifies when the collision will occur if the current kinetic conditions persist (Lee 1980). And the rate at which τ changes has a critical point property below which it specifies that the upcoming collision will be soft and above which it specifies that the upcoming collision will be hard (Kugler, Turvey, Carello & Shaw 1985; Lee 1980). The foregoing are not so much quantities as they are local flowfield morphologies and their changes. They specify pending states. They make possible the synchronizing of acts with events – the prospective control of basic behavior. They are meaningful in a very pragmatic sense of the word. Speaking in Dennett's (1983) terms, information in the specificational sense has "intentional features." And to echo Gibson's (1966, 1979) longstanding gripe, the "meaningless to meaningful" problem with which Sayre struggles is not a problem. (Coming to terms with the laws at the ecological scale on which the intentionality of perceiving-acting is founded, and figuring out how to formulate and systematize them, now that's a problem!).

Said succinctly, there is a description of optical structure under which its detection guarantees the intentionality of perceiving. There are other descriptions of optical structure under which it must be translated or processed or interpreted or embellished to *make* perceiving intentional. Sayre is playing with one such description. In this respect it is important to note that Gibson (1966, 1979) avidly denied that optical information in the specificational sense was the sort of thing that could be "processed." It is bizarre, therefore, for Sayre to claim that Marr (1982) is on target with his criticism that Gibson underestimated the complexity of visual information processing. There is a clash of metaphors here. Marr and Sayre are operating in the orthodox metaphor of the nervous system as an efficient cause; for example, it produces percepts. Gibson (1966) sees the nervous system as functioning vicariously in perceiving. It is a part (albeit extremely rich) of the supportive basis for the expression of natural cum ecological laws (cf. Ben-Zeev 1984). An understanding of the nervous system's role in vision in the support metaphor will be radically different from the processing/producing understanding subscribed to by Marr and Sayre (Kugler & Turvey, in press). At all events, in the ecological view, optical descriptions that invoke processing to render intentionless inputs into intentional percepts are of the wrong kind. They beg too many questions and they cast intentionality as a derivative rather than a primary phenomenon.

The last sentences, of course, are just another way of saying that intentionality should not be reduced to representation. As I remarked above, Sayre's goal of disengaging intentionality from computational procedures is admirable; his insistence on the intentional–representational equation is not. That equation, as I have been trying to stress, diverts us from addressing intentionality in a way that reveals its position in the natural order of things. Consider the following: What are customarily referred to as an animal's or person's intentional contents (cf. Dennett 1969; Searle 1983) constitute extraordinary boundary conditions on natural law (especially those laws that are particularly pertinent to the ecological scale). A flying animal aiming to collide gently with a surface will synchronize its deceleration with one value of τ; an acceleration to produce a timely, violent collision will be generated with respect to another value of τ (e.g., Lee & Reddish 1981; Lee, Young, Reddish, Lough & Clayton 1984; Wagner 1982). In these simple examples the final conditions – the animal's intentional content – specify the initial conditions that a law (relating optical properties to kinetic conditions) must assume. Examples like this abound, and one of them has been investigated quite thoroughly (Kugler & Turvey, in press). They suggest a profound challenge for naturalizing intentionality: understanding the principles by which intentional contents harness natural laws.

Author's Response

Intentionality and communication theory

K. M. Sayre
Department of Philosophy, University of Notre Dame, Notre Dame, Ind. 46556

Whether human cognitive capacities are fashioned by species adaptation (i.e., they are innate) or by adaptation of individual organisms (learned), or by some of both, they evolved from noncognitive information-processing functions. This is the basic assumption behind my approach. It might, of course, be factually wrong ("creationists" presumably would take that line). But if it is even roughly correct, certain methodological consequences follow for the study of cognition – that is, for CS.

First and foremost, it follows that explanations of cognition in any of its various forms – perception, language, and reason, to name a few – should be pursued in an evolutionary setting. This means, for one thing, that an acceptable explanation of a given cognitive capacity should make intelligible how that capacity might have evolved from lower level cognitive or noncognitive functions (for example, visual perception from mere photoreceptivity). It also means that an adequate account of how a given capacity actually operates might be expected to allude to mechanisms functioning elsewhere in organic nature (for example, mechanisms for adjusting negentropic coupling). And it means that there are important priorities to be observed in our explanatory approaches to cognition overall, inasmuch as a reasonably secure understanding of a given low-level capacity (like perception) might be a precondition for even a plausible conjecture regarding the operations of a higher level capacity (like referential language) that probably evolved from it.

Another assumption behind this approach is that the distinction between cognitive and noncognitive neuronal processes is basically a matter of whether info(s) (information with semantic content) or info(t) (information in the technical sense of communication theory) is being processed, and that the mark of an info(s)-processing function is its intentionality – its characteristic of being about some particular object or state of affairs. A consequence of this assumption, together with the first, is that MTC (the mathematical theory of communication) might be expected to make indispensable contributions to our understanding of intentionality – not only because MTC is the theory of info(t) transactions in general (surely including those of visual perception), but also because of the unique help it promises in imparting to CS the mathematical substance characteristic of successful science.

In the target article, these considerations dictated an investigation, incorporating relevant principles of MTC, of the nature of the intentional functions operating in visual perception, of how they might have evolved from info(t)-processing functions typical of noncognitive forms of photoreceptivity, and of the adaptive mechanisms they *might possibly employ. Success in this investigation* would pave the way for a subsequent study of higher info(s)-processing capacities such as language and reason.

Explaining intentionality? A head-on challenge to this project overall was posed by **Turvey** who rejects my approach (and Marr's [1982] as well) in favor of Gibson's (1966; 1979) on the grounds that the former casts intentionality in "a derivative rather than a primary" role while the latter treats perception as "fundamentally intentional." The objection, I take it, is that intentionality is something that does not admit explanation, being found in nature as a primitive phenomenon. My disagreement on this score is too basic to permit any short-order response other than to say that treating intentionality as primitive in a theory of perception is on a par with treating "vital forces" as primitive in biology. From a properly scientific viewpoint, I am convinced, the "meaningless to meaningful" problem which Turvey and Gibson reject is not only a legitimate problem but the single most important problem in CS today. It was Gibson's failure to see this that led to my seconding Marr's claim that he had underestimated the complexity of visual information processing. What I find bizarre about Turvey's commentary, incidentally, is its grouping my approach with Marr's in this respect. As far as can be told from his major work, *Vision*, Marr didn't have an account of intentionality at all, and this is one of my objections to his approach. Whichever may be the side of the angels in this regard, Turvey is the one who appears to be on the same side as Marr. Both find intentional content present in systems (biological or mechanical) where it does not require explanation but is available as it stands for explaining other phenomena. Needless to say, I take the other side on that question.

What computers can contribute. This brings us back to the much belabored issue whether computer states can possess intentionality (info(s)) that is not "observer relative." Only one commentator took exception to my argument that they cannot, which is interesting as a possible indication that the contrary opinion is finally on the wane. In his argument for the old orthodoxy, **Lebowitz** urges that symbolic representations in AI (artificial intelligence) systems are more than isolated symbols, being instead "bundles of inference rules" – that the symbol "CAT," for instance, might refer to such facts as that cats walk on four feet, like milk, are mortal, "and dozens or hundreds of other rules." But surely the degree of complexity in the "rules" behind a symbol's use does nothing to change the stubborn fact that if a symbol refers to anything at all it does so because of the programmer's (or user's) intentions. It may be worth noting that Lebowitz's claim about "inference rules" appears to make sense in the first place only because of an equivocation on the term "rule." There are rules for formal symbol manipulation (which computers can follow) that, being formal, have nothing to do with reference at all, and rules for the interpretation of symbols (which so far only people can follow) that, although clearly having to do with reference, owe their standing as rules to human intentions. In neither case do the rules in question provide reference for the symbols by themselves, regardless of how complex the set of rules involved may be. This stubborn fact is not changed by the addition of capacities such as long-term memory to the system. What might change is the degree of temptation to describe the capacities of the system thus augmented in psychological terms (Lebowitz blithely speaks of "develop[ing] higher level concepts," determining "conceptual content of the input," "understanding," and so forth). "Semantic representation" is just another in the long list of psychological terms computer scientists have appropriated to describe their art. Yet the ability of a computational system to deal with what Lebowitz calls "semantic representations" no more amounts to representation of the kind my article addresses than its ability to "understand" a given programming language amounts to an approximation of human understanding in a cognitively interesting sense.

Golden's thoughtful comments suggest the need to clarify my argument that computational systems of the sort with which CS is currently occupied do not exhibit "intrinsic" intentionality. I have not argued that computer-based systems generally cannot possess intentionality of the sort my account ascribes to human percipients. In Section 10 I briefly discuss what properties systems of that sort might need in order to establish intentional relationships with the environments supplying their data. What I had in mind primarily were environments like those with which human percipients form intentional relationships, but Golden reminds us that another system–environment arrangement is possible – that the environment might be simulated as well as the percipient. I agree that a promising initial step toward the actual construction of an intentional mechanism would be to model both organism and evnironment on a computer, provided the data-processing functions involved were roughly those of the target article. I also agree with **Eskew**, however, that the mathematics needed for the mechanical implementation of these functions may not currently be available in the forms required.

Another (barely) possible arrangement is pointed out by **Churchland**, who mulls over the purported counterexample of a human percipient in an entirely simulated environment. His elegant treatment of the celebrated "brain in the vat" predicament probes the position taken by the target article that intentionality is located in the relation between percipient and environment. The intent of the counterexample is to show that the subject–environment relation can alter radically, and yet the semantic (intentional) content remains the same. Churchland's thought experiment, however, is not fully enough specified to constitute a clear counterexample to the position taken by my article. It is not clear, first, what the semantic content is supposed to be. According to my account, at least, it would not be the subject's thinking "he's at a baseball game," but rather the configuration of info(t)-bearing events occasioned (under normal conditions) by the surface features of the baseball, the hot dog, or some such object or set of objects. In terms of Section 10, the content is the set of events O at the input of the anterior channel O-R (and not, we should carefully note, the actual objects that provide occasion for those events). Nor is it clear, second, whether the relation between subject and environment (although certainly different in the "brain in the vat" case) is different in ways that should affect the content. All that affects content, as far as the anterior channel is concerned, is the input O and the output R, and not the particular causal processes by which O and R happen to be connected. As long as the statistical characteristics of the channel O-R remain constant, the causal relation between its termini is irrelevant to the subject's content.

Now, what exactly is the counterexample? By stipulation, the posterior channel R-C remains unaltered, and so does the activity of the subject's efferent system. But for there to be a counterexample at all, the content O must also remain unchanged. Given the unaltered channel R-C, this entails that the channel O-R itself remains unaltered in its statistical characteristics. Because we are assuming that the efferent system produces control signals pertaining to O in exactly the same fashion as before, it follows that everything in the relation between subject and environment that affects the content O remains unchanged. All that is changed is the causal processes that provide the occasion of O; and, extensive as these may be, they are not essential to the content of the subject's perception. In brief, given the conditions of "perfect deception" that the thought experiment imposes, if the content is the same, then so are all the relevant environmental relationships, and no counterexample to my account has been produced. Think of it this way. Just as in Berkeley's theory of perception there is no detectable difference between a real physical object and a "perfect illusion," so on my account there is no detectable difference between the real perceptual environment and a "perfect computer simulation." The philosopher's notion of a completely undetectable deception removes the distinction between "deception" and "reality" beyond the pale of scientific significance.

How the explanation should proceed. Several commentators, while agreeing with the thrust of my expalnatory project, raised criticisms about the manner in which it was pursued. **Dennett**, for instance, complains that my account is lacking in detail, being only a "theory sketch" too much given to "handwaving." **Morphis**, by contrast, expresses confidence in my basic approach, but thinks it should be executed with a "broader and more adventurous hand." **Leyton** believes that there are resources in ecology and thermodynamics that my article ignores with baleful consequences, while **Powers** is mildly scornful of my failure to see that all the problems could be solved by control theory exclusively.

In response to **Dennett**, it should first be said that the reason I am not confusing the "specs" of a system with its design is that I am not at present in the business of building systems at all. (I once was – see Sayre, 1973 –

and that presumably is the reason he finds my account of noise reduction "relatively detailed.") Dennett's delightfully whimsical "Vigilator" illustrates once again his considerable skill at what he himself not long ago described as the exploration of "highly abstract constraints" of psychological theories, "without worrying about the mechanics and biochemistry of concrete 'realizations'" – an enterprise he hails as "the most promising and exciting work . . . being done today" in AI, psychology, and philosophy of mind (Dennett 1979, p. 255). There is something deliciously droll about being cited by the author of that job description for having produced only a "sketch" of a theory that leaves "the question of mechanism untouched."

But seriously, what I *am* trying to do in the target article – to summarize it once again – is to describe a functionally articulate set of processes that in upshot would accomplish a directionality upon objects, all of which can be understood as information-processing procedures in an explicit and precise sense of the term "information," and all of which incidentally could be implemented in both naturally and artifically contrived systems. The advance I purport to have achieved over Gibson (1966; 1979) (since **Dennett** asks) is with regard both to functional articulateness and to explicitness in sense of "information" involved. The superiority to be claimed over Marr (1982) (certainly not a matter of detail) is in the attempt I make to limit use of the term "information" to precisely defined senses, and in my taking seriously the problem of explaining how brain states can represent the properties of objects in the first place. Dennett notes that the concept of fidelity has been seen as a necessary component of representation by "everyone from Gibson to Marr to Fodor to Winograd," who have not made much of it because they take it for granted. I agree. My complaint is that they should not have taken it for granted because it is a very tricky concept. And if, as Dennett allows, my analysis in terms of high mutual info(t) "captures this [concept of fidelity] nicely," then this is certainly a step in the right direction – and just as certainly an illustration of the usefulness of MTC in psychological explanation.

A response in the same vein is due **Dennett**'s welcome remark that there is nothing to quarrel with in my use of the concepts of negentropic coupling and negentropic flexibility (pace **Leyton**). It may be, as he claims, that there is nothing new here except the language (although I doubt it); but even if this is so, the language has not been getting much play in CS, and I would like to think that my article shows reasons why it deserves more attention. I also note with pleasure that in Dennett's estimation if we had a "scaling" defined for negentropic coupling as we do for mutual information, we might have gained new powers of description and problem setting. The "scaling" in question is degree of success in the organism's perceptually guided ventures.

Finally, **Dennett** thinks my proposals "curiously undercut" my criticism of the computational approach. What I find curious is that he should think so. My criticism is that CS cannot rely upon the intentionality of computer states for an explanation of the intentionality of psychological states in humans, because computer intentionality derives from the human variety. I have attempted to provide an alternative model of intentionality – one not relying on the assumption of human intentionality in the first place – and the fact that intention turns out to be "relational" in this model is neither here nor there. It remains the case that we cannot rely on computer intentionality to explain the human variety, because the former still derives from the latter despite the outcome of my attempt.

What **Morphis** means by saying that my approach would be better pursued in a broader and more adventurous manner, I take it, is that intentionality (like pain) involves many interactions with other functions of the organism and can be fully understood only in a broader context than my account provides. I believe this is right. Although a considerably broader context is provided by Sayre (1969; 1976), a great deal more is required, and Morphis is to be commended for pointing this out.

I hope **Leyton** will not find me disingenuous in agreeing with most of his factual observations while failing to see why he thinks they are even slightly damaging to my account. The term "negentropic coupling" is explicitly introduced in the target article as a label for the relationship an organism enjoys with its host environment that enables it to assimilate negentropy in forms and amounts needed, and to rid itself of the resulting byproducts. Energy for work is one form of negentropy in question, and what Leyton calls an "energy sink" is what absorbs the byproducts figuring in my definition. So although I am using "negentropic coupling" in a somewhat broader sense than he is, his sense is certainly included in mine. Nor is it an objection to point out that there are some features of negentropic flexibility I do not mention (those having to do with trophic levels on food chains). Such features may be presumed irrelevant to the topic of the article. Leyton is right in pointing out that what he calls "nonspecific information" is important for understanding the defensive mechanisms of sentient organisms. (These matters are discussed at length in Sayre, 1976). But since the "information" involved is nonspecific, and hence not focused upon specific objects, it is of no particular concern for an account of intentionality. Finally, Leyton cites various types of perceptual phenomena not involved in ongoing manipulative tasks, the upshot of which is supposed to be that two "irreconcilable demands" fall upon my account: stabilization (which my account is claimed to emphasize) and destabilization (illustrated supposedly by cases in which animals reject present perceptions – destabilize them? – in search of others that are more rewarding). But there is no conflict here that goes beyond Leyton's terminology, because in the perceptual life of any healthy animal the "stabilization" of one percept is at the expense of the "destabilization" of its predecessors. In sum, there is no incompatibility of any sort between my account and the factors Leyton cites, and certainly nothing approaching logical inconsistency as he extravagantly charges. Of all the commentators Leyton seems to appreciate best the thermodynamic dimensions of my approach, and I look forward to opportunities for productive interchange once these misunderstandings are cleared away.

Powers asserts that discovering how signals introduced into the efferent system could possibly change according to external disturbances and constraints would "solve the problem of intentionality" and then opines that with a certain amount of "arm waving" I move things "within a millimeter of control theory" which is supposed to provide all the answers. As he is a control theorist, one might wonder, why doesn't he rejoice at the progress and take the final millistep himself? Then we would cordially congratulate each other for having solved a very difficult problem. But the reasons for his lack of cordiality soon appear. He thinks the problem has been solved long ago – by control theory, of course – and that its solution has been illustrated "hundreds of thousands of times." Presumably it is illustrated by any system whose behavior is driven by the difference between its "perception" of a particular object and a "reference signal" establishing the "needed" state of that object. Readers acquainted with early cybernetic literature may recall that systems of this description (goal-seeking mechanisms) were at one time popular as *illustrations of "purposive" systems*. Now "intentional" in one of its several senses is roughly synonymous with "purposive." This leads me to suspect that Powers assumes the target article is about intentional in the sense of purposive rather than in the sense of being about an object, which of course it is not. Perhaps my article is insufficiently clear on this matter, for both **MacKay** and **Perlis & Hall** seem at times to share the same misapprehension (MacKay in his talk of "meaning for the goal-pursuing system," and Perlis & Hall in their discussion of the visual mechanism that "intends" to follow a road). If so, I apologize to the commentators concerned. Reasons for being dissatisfied with goal-seeking mechanisms as models of intention in the sense of purpose, incidentally, may be found in Chapter 3 of Sayre (1969).

Whether MTC has a role. There are actually two related issues here: whether communication theory, however conceived, has a proper role to play in a project like mine; and, if so, whether in the strict form of MTC or in some more "relaxed" version. An interesting statistic is that, whereas three commentaries (**Brown, Earle & Lea, Gordon,** and **MacKay** – all and only the English commentators) express the opinion that no version of communication theory is relevant, and six (**Heil, Leyton, Marks, Perlis & Hall, Powers,** and **Turvey**) take no stand on the issue, the majority remaining all express confidence that communication theory is relevant in some form or another. And of the latter, all save **Dretske** – a notable exception – believe (if I represent them correctly) that MTC strictly conceived is the version from which help can be expected. This (to me) is most heartening, because I would have guessed that before quite recently the received view was that MTC has little or nothing to offer in the study of cognition.

The notable exception is **Dretske,** who probably has thought as much as anyone else in the world about the role of communication theory in cognitive studies. Why do we (he and I), he asks, both look in the same places but find different things? To be sure, we are not looking in exactly the same places, for I am looking in the direction of strict MTC and he is looking toward a distinctly "relaxed" version (see Sayre 1983; Dretske 1983b). So the question really is: Why, with this shared goal of developing an information-theoretic approach to cognition, do we persist in looking in different directions? Part of the answer, I am inclined to think, is that I want to stick strictly to MTC as a mathematical theory, because this is a necessary condition for using the theorems and principles of that theory in the study of cognition. For me, at least, these are important resources. Especially important are Shannon's 10th theorem (Section 8) and the formula defining mutual information (Sections 5, 8, 9, 10, and 11). ($2^{C_t} = 2^{C_1} + 2^{C_2}$ also figures briefly in Section 8.) The answer to **MacKay's** question – why "assume" the sum rather than the product in the right hand of the equation? – is simply that the equation as stated *is* a theorem, while the other form would be a falsehood. A proof may be found in the Solution Book, pp. 51–52, of Gallager (1968), problem 4.18, p. 525. Another part of the answer may lie with the fact that I envisage in the very long run a theory of cognition that is mathematical in roughly the sense of thermodynamics, while Dretske appears content with *explanations reflecting the categories of folk psychology*. I return to specific issues raised by Dretske's important commentary in later paragraphs.

Other possible resources. Methodological criticisms of a different sort concern the emphases of my proposals. **Brown et al.,** for instance, object to my bringing efferent processes into the account of intentionality; several other commentators stress the importance of internal "goal-criteria" (**MacKay**), "reference signals" (**Powers**), or "concepts" (**Perlis & Hall**), with reference to which (as Powers puts it) the efferent output of the intentional system controls its afferent input. I believe that Brown et al. are pretty much on the wrong side of this issue. The notion of "aboutness" they appear to favor, as a feature of specific representations at the visual input, abstracted from the functions these representations serve in the activity of the system overall, is a misconception fostered by the computational model. My account assumes that visual representations have an important (but not exclusive) function in the guidance of behavior (could anyone seriously think otherwise?) and gives this function due emphasis in its explication of perceptual intentionality. I am basically sympathetic with the suggestion that an account of intentionality should afford room for internal "goal" representations with reference to which the system regularly "evaluates" (**MacKay's** term) the course of its perceptual affairs – as long as thinking in these terms does not lull one into confusing intention in the sense of purpose with the intentionality of being about something (to mention that danger once again). In my account, the role played by established patterns drawn from Percept Storage (Section 9.3) comes close to the comparator functions stressed by these several commentators.

Brown et al. are lavish in their criticism, objecting (as best I can make out) (1) that my account contains more "free parameters" (functionally independent variables?) than what it purports to explain, (2) that it uses something like an "analysis by synthesis" model which has been tried and found failing, (3) that my account, as an alternative to Marr's, ignores his distinction among computational, algorithmic, and representational levels of explanation, and (4) that my account cannot be usefully extended. Regarding (1), although I agree that it would be a move in the wrong direction to concoct an explanans that left more factors unexplained than are present in the explanandum, I am at a loss as to how Brown et al. have found out how many "free parameters" are acutally present in the perceptual process. Regarding (2), I confess that my account derives from earlier models in Sayre (1965), which (acknowledging indebtedness to Licklider, 1952) used something like an "analysis by synthesis" approach; but the present approach to problems of intentionality has not been tried before and has not yet been proven a failure. Regarding (3), although I too admire the detail achieved in Marr's theory, I have little confidence in an account that attributes to neurophysiological mechanisms the kind of mathematical computations Marr performs in his calculations (neurophysiological networks don't compute like people do); and with my low confidence in this approach goes an attentuated interest in its threefold distinction among levels of explanation. Regarding (4), it may suffice to say that I cannot see why it is an objection to my account that I don't extend it to the intentionality of action (which they say psychologists have not found crucial anyway), and that the syllogism to which I do extend it is no longer considered paradigmatic of reason (I had not suggested otherwise). In upshot, the commentary of Brown et al. seems to illustrate how hard it is to produce an objective critique of views substantially different from those one has been brought up on.

I note with interest the other mathematical formalisms **Golden** mentions as possibly helpful in the study of intentionality, particularly the "special case" class of state vectors that have both thermodynamic and neurophysiological connections. While agreeing that other forms of mathematics will almost certainly prove useful in this study, I remain appreciative of Golden's estimation that my approach through MTC should result in new insights.

Marks's commentary seems best read as a recommendation that CS not overlook the broadly adaptive influences upon our preferential behavior. This seems legitimate. I find useful Marks's distinction between theories of syntactics, semantics, and pragmatics in cognitive studies. In these terms, my account attempts to lay the basis for a semantic theory by incorporating pragmatics in some general sense.

I would like to be able to reply at greater length to **Churchland**'s advocacy of a "network" theory of semantic content, and to his reading of my argument in Section 3 as a rejection of this approach. I hope this is a misreading, because I am in sympathy with a great deal of what Churchland says (in his commentary and in 1979) in behalf of this approach. Part of the disparity may be cleared up by noting that the semantic networks of AI, which I argue in Section 3 do not illuminate intentionality, are not in the same boat with Churchland's theory – at least if we give full credit to his observation that the intentional states of cognitive creatures enjoy nomic and statistical connections *not only with the environment* (my emphasis on his words) but also with an enormous range of other cognitive states. It seems to me by and large correct that the semantic content of our thoughts and language is determined primarily (although not exhaustively) by the role they play within our respective "cognitive economies." Churchland and I may differ in our views of how semantic content is introduced into these "cognitive economies" in the first place (I say through percepts, which are promoted to meanings by being freed from exclusive stimulus control), but there appears to be considerable overlap between his concept of a "cognitive economy" or netowrk and mine of a conceptual linkage (Section 11). (More views of mine on such matters may be found in Chapters 11 and 12 of Sayre, 1976). Moreover, I suspect we don't differ fundamentally even on the role of perception in such matters, because I agree entirely with his remark that the network "retains systematic causal connections with reality, connections, moreover, that carry information about reality" (Churchland 1979, p. 41).

The quantification problem. One requirement for applying any mathematical formalism to empirical processes is a specification of the factors that will stand as values of the variables in the formalism; in the case of MTC, the requirement is to specify the empirical events that will stand as members of the input and output sets of information channels. The tack taken by the target article is to conceive that portion of the overall perceptual cascade *O-C* stretching from the retina to the visual cortex (the channel *R-C;* Section 10) as consisting of nonindependent sets of neuronal events (firings, inhibitions, summations, and so forth; Section 9) and the portion stretching from object to retina (the channel *O-R*) consisting, on the *R* side, of sets of neuronal events in the retina and, on the *O* side, of physical events at the surface of the object – more exactly, of electromagnetic events resulting from the interaction of physical features of objective surfaces and incident light rays. Before my account could be applied experimentally, it would be necessary to achieve a great deal more specificity in identifying the physical events actually involved in the organism's perceptual processes, and at this point I have little more to contribute toward that end. Several commentators, however, quite properly raise questions in this regard, and they deserve the best response I can muster.

Golden, in a probing analysis, notes that MTC is based on assigning probabilities to specific events, and that before calculations of the sort my account postulates can actually take place it is necessary to identify relevant neurophysiological events and to characterize the channels these events constitute. He is right in pointing out that the target article says less than is needed on these and related issues for a full-fledged application in the experimental study of perception. What the article perhaps does not make adequately clear is that the account it presents, with its crude division into four info(t)-processing stages, is an attempt to get a fighting start on problems of just this sort. Before we can develop specific characterizations of the channels implicated in perception, we must have some general ideas about the types of information-processing task involved. This is what the target article has attempted to provide.

One aspect of the problem of specificity is what **Gordon** calls the question of quantifying the stimulus array. Allowing for ambiguities in the technical use of the term "stimulus," this might mean either identifying the info(t)-bearing events at *R* or identifying corresponding events at the source *O* of the perceptual channel. Gordon's concern is primarily with the latter, and he raises questions about the "uncertainty" associated with the view of a mountain from which the mist has just cleared, which looks somewhat like a human face, and so forth. Although, to be sure, an adequate theory of visual perception had better have something to say about large-scale views of distant mountains, Gordon's choice of examples makes it difficult to connect his questions with the target article's discussion of perceptual objects on a much more modest scale. (However, the issues raised by Gordon *are* treated in Chapter 9 of Sayre, 1976). Another thing about Gordon's remarks that makes responding difficult is that he seems to be talking about "uncertainty" in the sense roughly of subjective unexpectedness, whereas MTC (as has been stressed by all its major expositors, and as is made reasonably clear in Section 5) is concerned with uncertainty exclusively in the sense of *nonsubjective* improbability. A similar misapprehension underlies the comments by **Eskew**, whose main objection to my account is based on the "fundamental point" (questionably attributed to Searle, 1980a) that info(t) must be observer-relative. As carefully stated at the very beginning of Dretske (1981), info(t) is "an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes." Info(t)-processing tasks of the general sort ascribed to visual perception in the target article are entirely of this objective sort. These tasks are performed in the observer's nervous system but in no other sense is info(t) "in the observer" as Eskew suggests.

Nonetheless, **Eskew** is to be commended for having thought carefully about several technical issues raised by my account, particularly those regarding the importance of sequential or temporal features in the processes conjectured to take place at various junctures in the perceptual cascade. At this point I can only say that a more complete account of those processes would stress (1) the need for channels in the afferent system with relatively stable (time independent) conditional probabilities, (2) the sequential sampling of states of the Accumulator over real time, treating that component as a Markov source, and (3) scanning techniques for testing relative novelty of neuronal activity from sector to sector of the Retina in its real-time operation (perhaps roughly in the manner of an electron microscope, with temporal sequence imposed upon the data as a result of the scanning procedure).

A different type of concern about quantification is raised by **Daugman**, which leads him to suspect a "basic incoherence" in my application of MTC to perception. This dire consequence is supposed to follow from my attempting to provide a fidelity analysis of a channel between a three-dimensional (3D) object and a two-dimensional (2D) retina or cortex, which can't be done because the units are incommensurable. But this complaint is misguided on both counts, because (1) I explicitly reject (Section 9.2) the rather peculiar assumption that retinal (or cortical) representations are two-dimensional (the retina is hemispherical), and (2) if Daugman's strictures about dimensional commensurability were credible, it should be incoherent to talk about the fidelity of an image of a 3D scene (for example, a sporting event) conveyed across a one-dimensional telephone line, which patently it is not. Moreover, while I certainly agree with Daugman that some sort of "grain structure" would have to be imposed on the object or image for the measures of MTC to be applicable, I cannot imagine why he thinks this would result in making those measures "irrelevant." Also, although he is quite right in pointing out that the "input/output" distinction is irrelevant in calculating mutual information (this is shown by the basic theorem $I(A;B) = I(B;A)$), this poses no problem for my account whatever; the needed asymmetry is instead secured by the dynamics of the perceptual–behavioral control loop (Section 10). Finally, it may fairly be protested that my account attributes much more to intentionality than mere maximization of mutual information. The (admittedly rather "gadgety") stage-wise procedure of Section 9 is part of an effort to say what more there might be. Whatever the shortcomings of this breakdown into functions, however, and despite Daugman's charge to the contrary, it bears no resemblance to the abortive attempt to *explain* sleep-inducing power by merely applying a label *meaning* "sleep-inducing power."

The identity relationship. Those commentators who probe the nature of the relationship of identity between *O* and *C* (notably **Dretske**, **Ellerman**, and **Morphis**) go right to the heart of my account. When a representation at *C* is intentionally related to an objective structure *O*, there is an identity of info(t) at *C* and *O* (when, but not only when – there are countless identity relationships of this sort in nature that do not even approximate intentionality). The nature of the identity here is perhaps clearest in the case of a noiseless channel, where the probability of a given event occurring at the input is mathematically the same as that of an indication of that event at the output. In this limiting case, the entire amount of info(t) received at the output is identical to the info(t) entered at the input. In a channel with noise, on the other hand, not all the info(t) at the input is transmitted through to the output; but *some* is nonetheless, otherwise the input and the output sets would not constitute an information channel. Indeed, the terminology of identity is just another way of talking about mutual information, which any channel must possess in some degree to be an information channel in the first place. The amount of info(t) indicated identically at input and output, of course, varies from channel to channel, being equivalent to the overlap of what is indicated at the two termini separately (see Abramson, 1963, Sections 5–7, for details). My proposal, in briefest possible summary, is that when a representation *C* is established in the cortex that adequately serves the function of focusing the organism's behavior on a particular objective structure *O* at the other end of the perceptual channel *O*-*C*, then the identity of info(t) at the two termini of this channel constitutes the relationship of intentionality between *C* and *O*. To explain how this relationship between *C* and *O* is established in the perceptual data processing of an organism is, in **Dennett**'s apt phrase, "to turn the corner . . . from info(t) to info(s)." Unclear as my "corner-turning" may be to Dennett and others, this at least is what it's all about.

Accordingly, I must reject **Dretske**'s gloss in proposing that what I want to say is that, in the perception of a yellow flower, the "information that there is a yellow flower out there" is identically present at the cortex, in the retina, and "out there" as well. What I mean by "identity" in this connection has nothing to do with propositional content (*that* so and so) and everything to do with the sameness of info(t) present at these several junctures in the technical sense of MTC. Although propositional contents may be very much on the scene in higher level cognition, I do not think they have much to do with visual perception as such (seeing a yellow flower is not the same as seeing that the flower is yellow). **Ellerman** quotes my summary remark in Section 10 that the intentionality of perception is the identity of info(t) structure between *O* and *C* (overlooking the essential qualifications that immediately follow) and proceeds to construct an alleged counterexample to what he takes to be the "best possible case" illustrating that summary remark. The supposed "best possible case" is a noiseless and deterministic channel in which all and only 1s at the input are indicated by 1s at the output, and all and only 0s by 0s at the output. The counterexample is a channel with the same characteristics except that all and only 1s at the input are indicated by 0s at the output, and vice versa for input 0s. The reason this is supposed to be a counterexample is that, although the mutual information of the two channels is the same (which it is), the first channel is such that 1 at the output *means* 1 at the input *to an observer* who knows the channel characteristics, whereas with the second channel a 1 at the output *means* a 0 at the input. That is, the mutual information is the same, but the meanings differ; hence intentionality cannot be a matter of mutual information. This argument fails for two reasons (besides overlooking the qualifications that tell a good share of the story). First, the presence of an observer who interprets the output signals is entirely extraneous to my account and indeed could not be incorporated in it (without circularity) even if one tried. And second, as far as every channel feature germane to my account is concerned, Ellerman's two cases are indistinguishable, differing only in matters of notation that have nothing to do with channel specifications.

Morphis is right in detecting the essentially mathematical character of the identity relationship but sees no advantage in this account over variations of the "robot reply" discussed in Searle (1980a). **Ellerman** and **Heil** also identify my account as a version of the "robot reply." Although this account was not intended as a response to Searle's challenge, I would like to speak briefly to that issue.

The "robot reply." The argument of **Heil** against my account of perceptual intentionality proceeds by (1) taking it for granted that I propose to locate intentional content *inside* the organism, (2) construing my account as an embellishment of Searle's Chinese-room thought experiment (specifically, as a version of the "robot reply" to Searle's argument), (3) generalizing beyond the manipulation of symbols (like those of the Chinese alphabet) whose intentionality is derivative to any code system whatever, and (4) drawing the familiar conclusion that nowhere in all this is intentionality to be found. This rendition of my account has the fault of laying it open to criticism for not providing adequate answers to questions it never addresses. For one thing, my account of perceptual intentionality is not a version of the "robot reply" because I am not coming to the defense of "strong AI." Indeed, Section 3 should make it quite clear that I support Searle's side in the controversy. Yet I am urging that perceptual intentionality arises nonderivatively in a complex interaction between biological organism and perceptual environment – an interaction complex enough to encompass all the environmental, biological, and social factors Heil cites as germane to an adequate account. And because there is no apparent reason why the info(t) processes implicated in my account could not be performed by entirely mechanical as well as biological systems, it follows that an artificial system interacting in these complex ways with the world at large might also be endowed with perceptual intentionality in the same way as human percipients.

So why, with that admission, is my account not just another version of the "robot reply?" For one reason, because my account is of perceptual intentionality, and not of understanding (linguistic symbols or whatever), which is the focus of Searle's discussion. A second and more important reason is because my account does not locate intentionality *inside* the perceiving system in the first place. The representation possessing intentional content is inside the system, of course. But its object is typically outside the system; and if by "intentionality content" we mean the object, as I propose, then this content is not inside the system as **Heil** surmises. From a perspective like mine that stresses the real-time dynamic interaction between organism and environment as the locus of the intentional relationship, the thesis of "strong AI" seems so implausible that one wonders at the fuss Searle's argument has occasioned.

One thing my account has in common with the "robot reply," however, is that it features what certainly appears to be a causal interaction between organism and environment. **Heil, Morphis,** and **Ellerman** may have considered this sufficient warrant for grouping it with that position. **Churchland** also brings up the matter of causal interaction. There are issues here that deserve further discussion.

Causal interaction. In his brief but penetrating commentary, **Kelley** suggests that if my account is confirmed by neuroscience the result might be a comprehensive theory of how the nervous system gives rise to intentional states. He believes, however, that my "stronger claim" that intentionality is a relationship of high mutual information between a cortical representation and a set of objective circumstances runs afoul of a "classic problem" for causal theories of perception. In general, causal theories identify the object of a perception with its cause but have trouble separating the object proper from other causal antecedents. By way of solution, causal theorists typically lay down other conditions to pick out the intentional object; and my conditions, as Kelley observes, have to do with an identity of info(t) structure between *O* and *C*. But since *R* has a share in that identity also, why isn't the structure at *R* as much the intentional object as that at *O*? This is a good question.

Kelley even knows the answer I want to defend, which is roughly that the perceptual system "locks on to" and "tracks" the structure *O* in the objective environment. The fuller answer in Section 9 makes use of the concept of a perceptual–behavioral control loop between *C* and *O*, which enables the organism to direct its behavior with respect to those objective circumstances. Kelley's objection to this answer is that the perceptual process persumably does this by locking on to and tracking the retinal events *R* as well. But this is not the case at all. The mediating events at *R* are only part of a complex series of info(t) processes by which the perceptual system maintains its focus upon *O* wihtout exceeding the limits of its information-processing capacities. There is no sense at all, as far as my account allows, in which the perceptual–behaviorial control loop maintains a focus on R. Making sense of this would require the organism to be directing its behavior toward, or with respect to, events in its retina instead. Even if the organism were conducting eye surgery on itself, or trying to view its retina in a mirror, the perceptual object O would consist of events at the surface of the eye itself and not the neurophysiological events at R by which its behavior is being partially guided. Very likely my account requires further development to stand as a solution to all the "classic" problems, but this one at least seems not particularly dire.

Heil also raises an interesting problem in connection with the causal aspects of my account. Observing that the info(t)-processing channels in my account do not require any particular causal linkage (strictly, do not require any causal linkage at all, because informational nad not causal connections do the work of the theory), he believes this opens it up to familiar Twin Earth–type counterexamples. If Smith is looking at banana b on Earth, and his Doppelganger Smythe is undergoing exactly the same neural transformations on Twin Earth, why is not Smythe as well as Smith looking at b? Heil claims that the answer I want to give is unavailable. I want to say, in line with my previous response to **Kelley**, that Smith perceives b because b is the object with respect to which his behavior is directed through the operation of his perceptual–behavioral control loop, and that Smythe does not perceive b because his behavior is not so directed. This seems simple enough, but Heil says it misses the difficulty. As best I can make out, the difficulty is that the object of perception – Heil's "appropriate object" – must be specified independently of being identified with the object toward which behavior is directed. I suspect that what Heil has in mind here is the familiar requirement confronting any mind–brain identity thesis that claims identity between brain state B and mental state M: if the claim that B = M is to have empirical content (not be merely tautological), there must be some way of specifying either state independently of the other. But if there is a requirement of this sort facing my account, it is satisfied by the facts that (1) the neurophysiological processes are supposed (as Kelley saw) to be empirically specifiable, and (2) the mental component – the intentional direction of C upon O – is specified as a relationship between factors at opposite ends of the perceptual–behavioral control loop. If there is an additional requirement that O be specified independently of this control loop, I confess I cannot understand it. Perhaps Heil's sense that there is a problem here could be dispelled by realizing that the only identity to which my account is committed is the mathematical identity (as **Morphis** and **Ellerman** saw) between info(t) structures at C and O, which is not the sort of identity ("contingent") involved in mind–brain identity theses at all.

With these considerations in view, I trust that **Ellerman** will conclude that my account does not confuse meaning and cause as he seems to suspect, and that **Churchland** will find further reason to be assured that there are no causal presuppositions in my account that would produce tension with his "network" approach to cognitive phenomena.

Heil raises the further objection that my account does not give semantic content any role to play in the system, which I gather means that it does not rely on semantic content to explain anything else. This is correct. My account concerns the genesis of such content (as **Kelley** clearly saw), with only afterthoughts about the roles intentional content might play in higher cognitive functions. If the conjectures of Section 11 are even roughly on the right track, however, there is much for it to explain on these higher levels. I turn once again to these topics, and once again only briefly.

Postscript revisited. It would have been preferred by **Dretske** that I pay more attention to the higher level cognitive functions about which he says so much in his 1981 work, and he gently rebukes me for having prefaced what little I do say about them in Section 11 with the disclaimer "not intended for critical discussion." What this disclaimer means is simply that the conjectures of this section are not developed in enough detail to be defended as they stand, and that they are unaccompanied by any discussion of the many issues they raise. Dretske nonetheless chooses to raise some of them for me. I do not think him unfair for having done so, and I welcome the chance to say just a bit more on these topics (relying upon the reader to realize that the bit more I can say in a few paragraphs is not enough to remove the disclaimer). What I owe Dretske is (1) a somewhat fuller account of how the intentionality of perception differs from that of thought and belief, (2) some thoughts about how misrepresentation is possible on these different cognitive levels, and (3) a brief explanation of the "casualness" of the target article with regard to these issues.

(1) To begin with, the intentionality of perception as I portray it is a matter of being directed upon or about a specific object, and not of being an attitude toward a propositional content (because perception as I see it is not a propositional attitude). As **Dretske** notes, I certainly talk this (the former) way at times; indeed, I intended to talk that way all the time. Being of this character, perceptual intentionality involves the perceptual presence of the object to the perceiving organism and thus clearly differs from the intentionality of thoughts and beliefs about objects that are not perceptually present. Perceptual intentionality, by this account, consists in an adequately high level of mutual information between C and O under certain conditions having to do with the perceptual control of the organism's behavior toward O – where "adequately high" means adequate to maintain perceptual control under those conditions. No particular difficulty is posed for this account by Dretske's case of seeing a flower at a distance and in poor light. If one's behavioral project were merely to find some flowers to make up a bouquet for the dinner table, the relatively low level of mutual information these circumstances provide might well suffice – which it presumably would not in a percpetually more demanding project, like finding a coreopsis in a field of wild daisies. (Dretske is right, by the way, in pointing out that the kind of picking out in question is not a matter of distinguishing a particular object from all others of its kind; for him to perceive his cousin Clyde it is clearly not necessary that he notice details sufficient to distinguish Clyde from Clyde's twin brother. The picking out in question is a matter of distinguishing the particular object with which one currently is occupied from other objects in one's perceptual field.)

Now, if the percept "yellow" gets its intentionality by picking out perceptually present objects in this fashion, **Dretske** asks, how can meanings or concepts manage to be about yellow when nothing yellow is perceptually present? Because he thinks the answer I sketch in Section 11 is an evasion, I will try to sketch the outlines of an equivalent answer in somewhat different strokes. First, once a stable pattern *C* has been established in the percipient's cortex, the object *O* need not remain perceptually present for it to continue to enjoy a relation of high mutual information with *C*. Although *O* must of course be perceptually present to stand in a relation of perceptual intentionality with *C* – that is, it must be perceptually present in order to be perceived – there are no requirements of perceptual presence attached to the relation of mutual information itself. Although Smythe in **Heil**'s example is not perceiving Smith's banana, the events in Smythe's cortex (being identical to those in Smith's) are nonetheless characterized with the same mutual information with respect to events at the banana's surface. Since Smythe is on Twin Earth, he cannot see Smith's banana; but nothing in the imagined facts of the case prevents his speaking or thinking about it. Second, a well-established pattern C may retain its capacity for guiding behavior with respect to an object *O* even when that object is not perceptually present. When I am looking for my watch, for example, my perceptual field is constantly being tested for fit with a pattern representing that object in thought, and my behavior is guided by the anticipation of coming to perceive an object fulfilling the conditions of fitting that object – conditions to be spelled out in terms of mutual information. In looking for something, I am not perceiving it; yet the pattern that once picked out that object in perception may now pick it out as part of another cognitive function – in this case, thoughtfully directed (as distinct from random) searching. Third, when representations once active exclusively in perception become associated with other representations in meaning relationships – which happens when people begin to direct their cooperative ventures with vocal signals – then the cooperating individuals begin to form conceptual networks like those (I take it) stressed in **Churchland**'s commentary. For more about how I view these networks, the interested and patient reader may consult Chapters 11 and 12 of Sayre (1976).

(2) How, given this general view of representation, is *mis*representation possible? With trepidation, I offer the following for **Dretske** to consider. In the case of percepts, misrepresentation is a matter of bringing an established pattern to bear (out of Percept Storage; Section 9.3) that initially shows enough fit with structures currently present in the percipient's perceptual field to set up expectations regarding further encounters with these objective structures, which expectations are subsequently thwarted. When the initially promising fit is not sustained as perceptually based dealings with the object unfold, the initially promising pattern is proved a misfit – that is, a misrepresentation. Mistake in belief requires a different story. Since by my lights a belief is not a representation in the first place, a mistaken belief is not a misrepresentation. A belief, rather, is a state of mind involving readiness to use concepts in any one of multifarious ways, and a mistaken belief is a readiness to use concepts incorrectly. This indicates the direction in which I suspect a proper answer to the problem of misrepresentation is to be found, but it contributes little to the answer itself.

(3) How can I be so casual in addressing these issues, whereas for **Dretske** misrepresentation is such a central problem? For him, in fact, the problem is crucial, because it affects the very cogency of the account of informational content that stands at the heart of his analyses of knowledge and belief. For an internal structure (representation) to carry the information *that s is F* (Dretske 1983, p. 57, my emphasis) is for all tokens of that structure to carry that information, which at least appears to rule out the possibility that some of those tokens fail to carry that information as his account of false belief requires (see my commentary on Dretske, 1983a; Sayre 1983, p. 79). Dretske's account lands in this predicament, if I read it correctly, largely because of his conception of information content as propositional (*that* s if F). Because information in my account is not propositional (not even info(s), as far as perception is concerned), the problem of misrepresentation does not challenge the cogency of the account overall. It can safely be postponed to another occasion.

References

Abramson, N. (1963) *Information theory and coding*. McGraw-Hill. [DPE, tarKMS]

Ackley, D., Hinton, G. & Sejnowski, T. (1985) Boltzmann machines: Constraint satisfaction networks that learn. *Cognitive Science* 9:147–69. [RMG]

Alley, T. R. (in press) Organism–environment mutuality, epistemics, and the concept of an ecological niche. *Synthese*. [MTT]

Anderson, J. A. & Mozer, M. C. (1981) Categorization and selective neurons. In: *Parallel models of associative memory*, ed. G. E. Hinton & J. A. Anderson. Erlbaum. [RMG]

Anderson, J. A., Silverstein, J. W., Ritz, S. A. & Jones, R. S. (1977) Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review* 84:413–51. [RMG]

Ashby, R. (1956) *An introduction to cybernetics*. MIT Press. [MLey]

Bar-Hillel, J. (1964) *Language and information*. Addison-Wesley. [taKMS]

Barlow, H. J. (1985) The twelfth Bartlett memorial lecture: The role of single neurons in the psychology of perception. *Quarterly Journal of Experimental Psychology* 37A:121–45. [IEG]

Barr, A. & Feigenbaum, E. (1981) *The handbook of artificial intelligence*. Vol. 1. HeurisTech Press. [taKMS]

Beadle, G. & Beadle, M. (1966) *The language of life*. Doubleday. [taKMS]

Bennett, J. (1976) *Linguistic behaviour*. Cambridge University Press. [taKMS]

(1983) Cognitive ethology: Theory or poetry? *Behavioral and Brain Sciences* 6:356–58. [taKMS]

Ben-Zeev, A. (1984) The Kantian revolution in perception. *Journal for Theory of Social Behavior*. 14:69–84. [MTT]

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society Series B* 36:192–326. [RMG]

Brillouin, L. (1962) *Science and information theory*. Academic Press. [taKMS]

Broadbent, D. E. (1958) *Perception and communication*. Pergamon. [taKMS]

Carello, C., Turvey, M. T., Kugler, P. N. & Shaw, R. E. (1984) Inadequacies of the computer metaohor. In: *Handbook of cognitive neuroscience*, ed. M. S. Gazzaniga, Plenum. [MTT]

Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*. Cambridge University Press. [PMC, tarKMS]

Churchland, P. M. & Churchland, P. S. (1983) Content: Semantic and information-theoretic. *Behavioral and Brain Sciences* 6:67–68. [MM]

Cohen, P. & Feigenbaum, E. (1982) *The handbook of artificial intelligence*. Vol. 3. HeurisTech Press. [taKMS]

Collins, A. M. & Quillian, M. R. (1969) Retrieval time from semantic

memory. *Journal of Verbal Learning and Verbal Behavior* 8:240–47. [RB]
Dennett, D. C. (1969) *Content and consciousness*. Routledge & Kegan Paul. [MTT]
(1978) *Brainstorms*. Bradford Books. [MM, DP]
(1979) Current issues in the philosophy of mind. *American Philosophical Quarterly* 15:249–61. [rKMS]
(1983) Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences* 6:343–90. [DCD, MLeb, taKMS, MTT]
Dixon, N. F. (1981). *Preconscious processing*. Wiley. [RTE]
Dretske, F. I. (1981). *Knowledge and the flow of information*. MIT Press. [MM, taKMS]
(1983a) Précis of *Knowledge and the flow of information*. *Behavioral and Brain Sciences* 6:55–90. [FID, tarKMS]
(1983b) Why information? *Behavioral and Brain Sciences* 6:82–89. [DPE]
Elton, C. (1927) *Animal ecology*. Macmillan. [MLey]
Fodor, J. A. (1975) *The language of thought*. Thomas Y. Crowell. [DCD, taKMS]
(1980) Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences* 3:63–73. [DPE, RTE, taKMS]
(1981a) The mind-body problem. *Science American* 214 (January):114–23. [taKMS]
(1981b) *Representations*. MIT Press. [taKMS]
Gallager, R. G. (1968) *Information theory and reliable communication*. Wiley. [rKMS]
Garner, W. R. (1962) *Uncertainty and structure as psychological concepts*. Wiley. [taKMS]
Gibson, J. J. (1966) *The senses considered as perceptual systems*. Houghton Mifflin. [DCD, tarKMS, MTT]
(1979) *The ecological approach to visual perception*. Houghton Mifflin. [tarKMS, MTT]
Haber, R. N. (1983) Can information be objectivized? *Behavioral and Brain Sciences* 6:70–71. [MM]
Haugeland, J. (1978) The nature and plausibility of cognitivism. *Behavioral and Brain Sciences* 1:215–26. [taKMS]
(1981) *Mind design: Philosophy, psychology, artificial intelligence*. Bradford Books. [taKMS]
Hebb, D. O. (1949) *The organization of behavior*. Wiley. [taKMS]
Heil, J. (1981) Does cognitive psychology rest on a mistake? *Mind* 90:321–42. [JH]
(1983) *Perception and cognition*. University of California Press. [JH]
Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America* 79:2554–558. [RMC]
Hubel, D. H. & Wiesel, T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* 160:106–54. [taKMS]
Johnson-Laird, P. N. (1983) *Mental models*. Cambridge University Press. [RB]
Kandel, E. R. (1970) Nerve cells and behavior. *Scientific American* 223 (July):57–70. [taKMS]
Katz, S. (1983) R. L. Gregory and others: The wrong picture of the picture theory of perception. *Perception* 12:269–79. [RTE]
Kim, J. (1982) Psychophysical supervenience. *Philosophical Studies* 41:51–70. [JH]
Kish, G. B. (1955) Learning when the onset of illumination is used as reinforcing stimulus. *Journal of Comparative and Physiological Psychology* 48:261–64. [MLey]
Kolodner, J. L. (1984) *Retrieval and organizational strategies in conceptual memory: A computer model*. Erlbaum. [MLeb]
Krechevsky, I. (1937) Brain mechanisms and variability: 2. Variability where no learning is involved. *Journal of Comparative and Physiological Psychology* 23:139–63. [MLey]
Kugler, P. N., Kelso, J. A. S. & Turvey, M. T. (1982) On the control and coordination of naturally developing systems. In: *The development of movement control and coordination*, ed. J. A. S. Kelso & J. E. Clark. Wiley. [MTT]
Kugler, P. N. & Turvey, M. T. (in press) *Information, natural law and the self-assemby of rhythmic movements: Theoretical and experimental investigations*. Erlbaum. [MTT]
Kugler, P. N., Turvey, M. T., Carello, C. & Shaw, R. E. (1985) The physics of controlled collisions: A reverie about locomotion. In: *Persistence and change: Proceedings of the first international conference on event perception*, ed. W. H. Warren, Jr. & R. E. Shaw. Erlbaum. [MTT]
Lebowitz, M. (1980) Generalization and memory in an integrated understanding system. Technical Report 186, Yale University Department of Computer Science. [MLeb]
(1983a) Generalization from natural language text. *Cognitive Science* 7:1–40. [MLeb]
(1983b) RESEARCHER: An overview. Proceedings of the Third National Conference on Artificial Intelligence, Washington, D.C. pp. 232–35. [MLeb]
Lee, D. N. (1980) Visuo-motor coordination in space-time. In: *Tutorials in motor behavior*, ed. G. E. Stelmach & J. Requin. North-Holland. [MTT]
Lee, D. N., & Reddish, P. E. (1981) Plummeting gannets: A paradigm of ecological optics. *Nature* 293:293–94. [MTT]
Lee, D. N., Young, D. S., Reddish, P. E., Lough, S. & Clayton, T. M. H. (1984) Visual timing in hitting an accelerating ball. *Quarterly Journal of Experimental Psychology* 35A:333–46. [MTT]
Lettvin, J., Maturana, H., McCulloch, W. & Pitts, W. (1959) What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers* 47:1940–51. [taKMS]
Licklider, J. C. R. (1952) On the process of speech perception. *Journal of the Acoustical Society of America* 24:590–94. [RB]
Lindeman, R. L. (1942) The trophic dynamic aspects of ecology. *Ecology* 23:399–418. [MLey]
Luce, R. D. (1960) The theory of selective information and some of its behavioral applications. In: *Developments in mathematical psychology*, ed. R. D. Luce. Free Press. [taKMS]
Luenberger, D. G. (1979) *Introduction to dynamic systems: Theory, models and applications*. Wiley. [RMC]
Mace, W. M. (1977) James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. In: *Perceiving, acting and knowing: Toward an ecological psychology*, ed. R. Shaw & J. Brandsford. Erlbaum. [MTT]
McDermott, D. V. (1981) Artificial intelligence meets natural stupidity. In: *Mind design*, ed. J. Haugeland. MIT Press. [MLeb]
MacKay, D. M. (1950) The nomenclature of information theory. In: *Proceedings of the information theory symposium*. Reprinted as Appendix of MacKay, D. M. *Information, mechanism and meaning*. MIT Press, 1969. [taKMS]
(1951) Mindlike behaviour in artefacts. *British Journal for the Philosophy of Science* 2:105–21; [DMM]
(1956) Towards an information-flow model of human behaviour. *British Journal of Psychology* 47:30–43. [DMM]
(1962) Theoretical models of space perception. In: *Aspects of the theory of artificial intelligence*, ed. C. A. Muses. Plenum Press. [DMM]
(1963) Internal representation of the external world. Proceedings of the AGARD Symposium on Natural and Artificial Logic-Processor, Athens. (Mimeo) [DMM]
(1969) Information, mechanism and meaning. MIT Press. [DMM]
(1978) The dynamics of perception. In: *Cerebral correlates of conscious experience*, ed. P. A. Buser & A. Rougeul-Buser. Elsevier. [DMM]
(1981) What kind of neural image? *Freiburger Universitaetsblatter* 74:67–72. [DMM]
Marr, D. (1982) *Vision*. W. H. Freeman. [RB, DCD, RMC, tarKMS, MTT]
Mayhew, J. E. W. & Frisby, J. P. (1981) Psychophysical and computational studies towards a theory of human stereopsos. *Artificial Intelligence* 17:349–85. [RB]
Michaels, C. F. & Carello, C. (1981) *Direct perception*. Prentice-Hall. [MTT]
Minsky, M. (1968) Matter, mind, and models. In: *Semantic information processing*, ed. M. Minsky. MIT Press. [DP]
(1975) A framework for representing knowledge. In: *The psychology of computer vision*, ed. P. Winston. McGraw-Hill. [taKMS]
Morphis, M. R. (1980) The total process approach to pain and the problem of the "privacy" of pain. Ph.D. dissertation, Brandeis University. [MM]
Neisser, U. (1967) *Cognitive psychology*. Appleton-Century-Crofts. [RB]
Norman, D. A. (1981) Twelve issues for cognitive science. In: *Perspectives on cognitive science*, ed. D. A. Norman. Erlbaum. [MM]
Pattee, H. H. (1973) Physical problems of the origin of natural controls. In: *Biogenesis, evolution, homeostasis*, ed. A. Locker. Springer-Verlag. [MTT]
(1977)Dynamic and linguistic modes of complex systems. *International Journal of General Systems* 3:259–66. [MTT]
Perlis, D. (1985) Languages with self-reference. 1. *Artificial Intelligence* 25:301-22. [DP]
Piaget, J. (1954) *The construction of reality in the child*. Basic Books. [taKMS]
Poggio, T. & Torre, V. (1984) Ill-posed problems and regularization analysis in early vison. Artificial Intelligence Memo #773. A. I. Laboratory, Massachusetts Institute of Technology. [JGD]

Prigogine, I. (1980) *From being to becoming: Time and complexity in the physical sciences.* W. H. Freeman. [MTT]
Pylyshyn, Z. W. (1978) Computational models and empirical constraints. *Behavioral and Brain Sciences* 1:93–127. [taKMS]
(1980a) The "causal power" of machines. *Behavioral and Brain Sciences* 3:442–44. [taKMS]
(1980b). Computation and cognition: Issues in the foundation of cognitive science. *Behavioral and Brain Sciences* 3:111–32. [taKMS]
Reed, E. & Jones, R. (1982) *Reasons for realism: Selected essays of James J. Gibson.* Erlbaum. [MTT]
Sayre, K. M. (1965) *Recognition: A study in the philosophy of artificial intelligence.* University of Notre Dame Press. [rKMS]
(1969) *Consciousness: A philosophic study of minds and machines.* Random House. [tarKMS]
(1973) Machine recognition of handwritten words: A project report. *Pattern Recognition* 5:213–28. [tarKMS]
(1976) *Cybernetics and the philosophy of mind.* Routledge & Kegan Paul. [tarKMS]
(1979) The simulation of epistemic acts. In: *Philosophical perspectives in artificial intelligence,* ed. M. Ringle. Humanities Press. [taKMS]
(1983) Some untoward consequences of Dretske's "causal theory" of information. *Behavioral and Brain Sciences* 6:78–79. [FID, tarKMS]
Schank, R. C. (1972) Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology* 3:532–631. [MLeb]
Schank, R. C. & Abelson, R. (1977) *Scripts, plans, goals, and understanding.* Lawrence Erlbaum. [taKMS]
Schrödinger, E. (1967) *What is life?* Cambridge University Press. [taKMS]
Searle, J. R. (1980a) Minds, brains, and programs. *Behavioral and Brain Sciences* 3:417–24. [PMC, DPE, RTE, JH, MM, tarKMS]
(1980b) Intrinsic intentionality. *Behavioral and Brain Sciences* 3:450–56. [DPE]
(1980c) The intentionality of intention and action. *Cognitive Science* 4:47–70. [MLeb]
(1983). *Intentionality.* Cambridge University Press. [MTT]
Shannon, C. E. (1948) A mathematical theory of communication. *Bell System Technical Journal* 27 (July):379–423; (October):623–56. [JGD, taKMS]
Shannon, C. E. & Weaver, W. (1949) *The mathematical theory of communication.* University of Illinois Press. [taKMS]
Shaw, R. E. & Turvey, M. T. (1982) Coalitions as models for ecosystems. In: *Perceptual organization,* ed. M. Kubovy & J. Pomerantz. Erlbaum. [MTT]
Skinner, B. F. (1969) *Contingencies of reinforcement: A theoretical analysis.* Appleton-Century-Crofts. [taKMS]
(1974) *About behaviorism.* Knopf. [RTE]
Solomon, J., Carello, C. & Turvey, M. T. (1984) Flow fields: The optical support for skilled activities. In: *Cognitive sport psychology,* ed. W. F. Straub & J. M. Williams. Sport Science Associates. [MTT]
Stitch, S. (1984) *From folk psychology to cognitive science: The case against belief.* MIT Press. [DP]
Taylor, C. (1964) *The explanation of behaviour.* Routledge & Kegan Paul. [RB]
Tribus, M. (1961) *Thermostatics and thermodynamics.* Van Nostrand. [taKMS]
Turing, A. (1950) Computing machinery and intelligence. *Mind* 59:433–60. [MM]
Turvey, M. T. & Carello, C. (1985) The equation of information and meaning from the perspectives of situation semantics and Gibson's ecological realism. *Linguistics and Philosophy* 8:81–90. [MTT]
Turvey, M. T. & Carello, C. (in press) The ecological view of perceiving-acting: A pictorial essay. *Acta Psychologica.* [MTT]
Turvey, M. T. & Kugler P. N. (1984) A comment on equation information with symbol strings. *American Journal of Physiology,* 246:R895–96. [MTT]
Turvey, M. T., Shaw, R. E., Reed, E. S. & Mace, W. M. (1981) Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition* 9:237–304. [MTT]
Waddington, C. H. (1968) The basic ideas of biology. In: *Towards a theoretical biology,* ed. C. H. Waddington. Aldine. [MTT]
Wagner, H. (1982) Flow-field variables trigger landing in flies. *Nature* 297:147–48. [MTT]
Walls, G. L. (1942) *The vertebrate eye.* Hafner. [taKMS]
Whitrow, G. J. (1967) Entropy. In: *The encyclopedia of philosophy,* ed. P. Edwards. Macmillan & Free Press. [taKMS]
Winograd, T. (1972) *Understanding natural language.* Academic Press. [DCD, taKMS]
Winston, P. (1984) *Artificial intelligence.* 2d ed. Addison-Wesley. [taKMS]

XIth CONGRESS OF THE INTERNATIONAL PRIMATOLOGICAL SOCIETY
Göttingen (FR Germany) 20-25 July, 1986

The XIth Congress of IPS will be held from 20 to 25 July, 1986 at the Central Lecture Hall (ZHG) of the University of Göttingen, Nikolausbergerweg 9c, Göttingen.

Symposia and plenary sessions on the main topics of primatological research, including ecology and conservation, will be held.

The language of the Congress will be English.

For further informations, please contact:

German Primate Center (DPZ), Kellnerweg 4
D-3400 Göttingen, FR Germany
Telephone: (0551) 38510. Teletex: 5518110-DPZ