Foundations of Perceptual Theory
S.C. Masin (Editor)

THE ROLE OF COMPUTATIONAL COMPLEXITY IN PERCEPTUAL THEORY

John K. Tsotsos

Department of Computer Science
University of Toronto, Toronto, Canada

ABSTRACT

The validity of perceptual theories cannot be considered only in terms of how well the explanations fit experimental observations. Rather, it is argued that sufficient consideration must also be given to the physical realizability of the explanation. Experimental scientists attempt to explain their data and not just describe it, in essence, providing an *algorithm* whose behavior leads to the observed data. Thus, computational plausibility is not only an appropriate but a necessary consideration. One dimension of plausibility is satisfaction of the constraints imposed by the computational complexity of the problem, the resources available for the solution of the problem, and the specific algorithm proposed. It is shown that such constraints play critical roles in the explanations of perception, intelligent behavior, and evolution.

The foundation of many modern perceptual theories arises from the *computational hypothesis*: biological perception can be modeled computationally as an information processing task. However, many argue that computational theories cannot explain biological behavior and that a computational theory would at best be an analogy. Is there any other type of explanation? In physics, cosmology, or chemistry, explanations and theories are put forward and the only requirement for their validity is that they account for the experimental observations. Would a cosmologist be required to create a universe in order for his theories to be taken seri-

ously? Or a biologist life? A theory that accounts for more observations than another is a better theory. Theories whose predictions are falsified are modified or rejected. Further, computation plays a large role in modern theory formation in the above disciplines. Computer simulation in particular has been a very powerful tool in the physical sciences. Yet no cosmologist would claim that he is creating a universe when building a computer simulation, and no one would criticize that cosmologist for not doing so. Yet the resulting theories would be considered valid if actual observations were explained. So, modeling as used in this essay has no implications of creating artificial life; rather computation is used as the formalism for an explanatory theory of perception with predictive power.

A key difference here is that disciplines such as physics or cosmology do not appear to inherently involve information processing, whereas perception and intelligence in general do appear to involve information processing. It seems inconceivable that we will ever be able to construct a universe of the magnitude and complexity of the one we live in. But, it does seem possible to soon construct machines with the same number of processors and connections as the human brain.[1]

Given the computational hypothesis, I claim that there is one basic issue that constrains all theories, regardless of implementation, namely computational complexity (amount of time and processing hardware required to reach a solution). Unfortunately, it appears that all problems are too hard in their general form (require unrealizable amounts of time or hardware or both). Thus, a new style of complexity analysis is required that attempts to solve these *too hard* problems in the context of the available resource limits and performance requirements. The full generality of the problem will necessarily be sacrificed in order to achieve this. Thus,

- problem complexity
- resource limits
- performance specifications

form an important first set of constraints that must be satisfied. If one is concerned with brain models, then values relating to human brains and performance must be considered. If one is building a machine system, the type of analysis is the same, even though the results of the analysis will differ. Other design constraints include:

- cost to develop
- cost to replicate

---

[1] The most recent computer from Thinking Machines Corporation, for example, the CM-5, has a peak processing rate of about 10^{11} operations per second, while the human visual cortex may have about 10^{15} operations per second. This is not too far from the numbers of neurons and connections in the brain; in addition, each of the CM-5 processors seem to have much more power than a neuron.

- physical shape and size, or packaging
- weight
- power consumption
- temperature control
- communication requirements and restrictions.

How do all of these constraints interact? It does not seem sensible to approach design as an optimization task of any sort: optimizing along one design variable necessarily will affect the others. We need to seek satisfying solutions, that involve design trade-offs along all variables in order to yield a solution. The trade-offs may differ depending on the implementation medium; in particular, trade-offs for a silicon implementation may not be the same as those for a neural implementation.

This contribution focuses on the first of the above constraints: problem complexity. The combinatorial problems are very apparent, and in fact in most (if not all) natural problems, optimal solutions are computationally intractable in any implementation, machine or neural. A few examples are in order.

i) Vision (the first two examples will be elaborated below):
   - Visual Search with unknown targets using a passive sensor system is *NP*-Complete (Tsotsos, 1989, 1990a);
   - Visual Search with unknown targets using an active sensor system is *NP*-Complete (Tsotsos, 1992b);
   - Polyhedral Line-labeling is *NP*-Complete (Kirousis & Papadimitriou, 1988).

ii) Reasoning:
   - Finding the best explanation for a class of independent problems using probability theory (and several other forms of abduction) is *NP*-hard (Bylander, Allemang, Tanner, & Josephson, 1989);
   - Abductive reasoning for all but the simplest theories is *NP*-complete (Selman & Levesque, 1989a);
   - Many forms of default reasoning are *NP*-hard (Kautz & Selman, 1990; Selman & Kautz, 1990);
   - Many of the strategies for defeasible inheritance in taxonomic hierarchies are intractable (Selman & Levesque, 1989b).

iii) Neural Networks:
   - For directed Hopfield Nets, determining whether a stable configuration can be found is *NP*-Complete (Godbeer, 1987);
   - $PERF_{AOF_{ns}}$ is *NP*-Complete (the loading task for neural networks) (Judd, 1990).

This listing only scratches the surface of the literature on the topic; there are many more examples and they form quite broad and natural problem classes. It appears that any interesting intelligent problem has the characteristic that it is susceptible to combinatorial explosion. It is important to stress however that the examples given above do not by themselves *prove* that these problems or that cognition are computationally intractable. They simply constitute evidence that the computational issues are real and may place severe constraints on algorithms proposed for the problems of cognition.

A BRIEF INTRODUCTION TO COMPUTATIONAL COMPLEXITY

Computational complexity is studied to determine the intrinsic difficulty of mathematically posed problems that arise in many disciplines. See Garey & Johnson (1979) and Stockmeyer & Chandra (1979). Many of these problems involve combinatorial search, i.e., search through a finite but extremely large, structured set of possible solutions. Examples include the placement and interconnection of components on an integrated circuit chip, the scheduling of major league sports events, or bus routing. Any problem that involves combinatorial search may require huge search spaces to be examined; this is the well-known combinatorial explosion phenomenon. Complexity theory tries to discover the limitations and possibilities inherent in a problem rather than what usually occurs in practice. After all, the worst case does occur in practice as well. This approach to the problem of search diverges from that of the psychologist, physicist, or engineer. In the same way that the laws of thermodynamics provide theoretical limits on the utility and function of nuclear power plants, complexity theory provides theoretical limits on information processing systems. If biological vision can indeed be computationally modeled, then complexity theory is a natural tool for investigating the information processing characteristics of both computational and biological vision systems. If the results of these analyses provide deeper insights into the problem and yield verifiable predictions, this would constitute evidence in favor of the computational hypothesis.

Using complexity theory, one can ask for a given computational problem C, how well, or at what cost can it be solved? More specifically, the following questions can be posed:

(1) Are there efficient algorithms for C?

(2) Can lower bounds be found for the inherent complexity of C?

(3) Are there exact solutions for C?

(4) What algorithms yield approximate solutions for *C*?
(5) What is the worst-case complexity of *C*?
(6) What is the average complexity of *C*?

Before studying complexity one must define an appropriate complexity measure. Several measures are possible, but the common ones are related to the space requirements (numbers of memory or processor elements) and time requirements (how long it takes to execute) for solving a problem. Complexity measures in general deal with the cost of achieving solutions.

Complexity theory begins with a 1937 paper in which the British mathematician Alan Turing introduced his well-known Turing Machine, providing a formalization of the notion of an algorithmically computable function. He postulated that any algorithm could be executed by a machine with an infinitely long paper tape, divided into squares, a printer that writes and erases marks on the tape, and a scanner that senses whether or not a given square is marked. This imaginary device can be programmed to find the solution to a problem by executing a finite number of scanning and printing operations. What is remarkable about the Turing Machine is that in spite of its simplicity, it is not exceeded in problem solving ability by any other known computing device. If the Turing Machine is given enough time, it can in principle solve any problem that the most sophisticated computer can solve, regardless of serial/parallel distinctions or any other type of ingenious design. As a result, the fact that a problem can be solved by a Turing Machine has been accepted as a necessary and sufficient condition for the solvability of the problem. A similar thesis was also put forward by another mathematician, Alonzo Church (1936), and thus it is usually referred to as the Church/Turing Thesis: any problem for which we can find an algorithm that can be programmed in any programming language running on any computer, even if unbounded time and space are required, can be solved by a Turing Machine. Perception can in principle be solved;[2] it can thus be implemented on a Turing machine and its computational nature confirmed.

Turing proved that the problem of logical satisfiability—for a given arbitrary formula in predicate calculus, is there an assignment of truth values of its variables such that the formula is true?—cannot be decided by any algorithm in a finite number of steps. This provided the basis for

[2] A simple-minded, *in principle* solution to perception is: store all possible images of all objects and events one may ever encounter; then for each stimulus search that store until a match is found; link that match to an appropriate action by searching through all possible stimulus-action associations. This solution is guaranteed to be correct; however, it is impossible to ever construct a simulation of it or to realize it with neural hardware simply because far too many space and time resources are required. Recall however, that Turing Machines have infinite tape. If one considers doing this for say a day, by using two video cameras, one recording what the eye sees, and the other recording the agent's actions, the task no longer seems so formidable.

other similar proofs of intractability. Once one could prove problems were inherently intractable, it was natural to ask about the difficulty of an arbitrary problem and to rank problems in terms of difficulty.

In what sense are complexity results inherent to a particular problem? Certain intrinsic properties of the universe will always limit the size and speed of computers. Consider the following argument from Stockmeyer and Chandra (1988): The most powerful computer that could conceivably be built could not be larger than the known universe (less than 100 billion light-years in diameter), could not consist of hardware smaller than the proton (10^{-13} cm in diameter), and could not transmit information faster than the speed of light (3×10^8 m/s). Given these limitations, such a computer could consist of at most 10^{126} pieces of hardware. It can be proved that, regardless of the ingenuity of its design and the sophistication of its program, this ideal computer would take at least 20 billion years to solve certain mathematical problems that are known to be solvable in principle. Since the universe is probably less than 20 billion years old, it seems safe to say that such problems defy computer analysis. In a subsequent section a new example with biological importance will be introduced which further demonstrates this point.

Some Basic Definitions

The following are some basic definitions common in complexity theory (Garey & Johnson, 1979). A problem is a general question to be answered, usually possessing several parameters whose values are left unspecified. A problem is described by giving a general description of all of its parameters and a statement of what properties the answer, or solution, is required to satisfy. An instance of the problem is obtained by specifying particular values for all of the problem parameters. An algorithm is a general step-by-step procedure for achieving solutions to problems. To solve a problem means that an algorithm can be applied to any problem instance and is guaranteed to always produce a solution for that instance. An important issue here is whether or not a proposed algorithm is decidable (or solvable). Basically, the requirement for this is that there exists a Turing Machine which can compute yes or no for each element of the set A for the following question: if the set A is countably infinite,[3] and there is another set B which is a subset of A, is a given element of A contained in B? A proof of decidability is sufficient to guarantee that the a problem can be modeled computationally.

The time requirements of an algorithm are conveniently expressed in terms of a single variable, n, reflecting the amount of input data needed to

[3] A set is countable if there is a one-to-one and onto mapping from the natural numbers (integers beginning with 0) and the set. The set may be finite or infinite.

describe a problem instance. A time complexity function for an algorithm expresses its time requirements by giving, for each possible input length, an upper bound on the time needed to achieve a solution. If the number of operations required to solve a problem is an exponential function of n, then the problem has exponential time complexity. If the number of required operations can be represented by a polynomial function in n, then the problem has polynomial time complexity. Similarly, space complexity is defined as a function for an algorithm that expresses its space or memory requirements. Algorithmic complexity is the cost of a particular algorithm. This should be contrasted with problem complexity which is the minimal cost over all possible algorithms. The dominant kind of analysis is worst-case: at least one instance out of all possible instances has this complexity.

A worst-case analysis provides an upper-bound on the amount of computation that must be performed as a function of problem size. If one knows the maximum problem size, then the analysis places an upper bound on computation for the whole problem as well. Thus, one may then claim, given an appropriate implementation of the problem solution, that processors must run at a speed dependent on this maximum in order to ensure real-time performance for all inputs in the world. Worst-cases do not only occur for the largest possible problem size; rather, the worst-case time complexity function for a problem gives the worst-case number of computations for any problem size; this worst case may be required simply because of unfortunate ordering of computations (for example, a linear search through a list of items would take a worst-case number of comparisons if the item sought is the last one). Thus, worst-case situations in the real world may happen frequently for any given problem size. Many argue that worst-case analysis is inappropriate for perception because of one of the following reasons:

1) relying on worst-case analysis and drawing the link to biological vision implies that biological vision handles the worst-case scenarios;

2) biological vision systems are designed around average or perhaps best-case assumptions;

3) expected case analysis more correctly reflects the world that biological vision systems *see*.

Each of these criticisms will be addressed in turn.

1) This kind of inference is quite incorrect. As was shown in (Tsotsos, 1990a), it is impossible for the biological (or any other) visual system to handle worst-case scenarios. The whole argument exists only to prove that all worst-case scenarios cannot be handled by human vision in a bottom-up fashion and that the quest for general solutions is futile.

2) It is far from obvious what kind of assumptions (if any) went into the design of biological vision systems. Vision systems emerged as a result of a

complex interaction of many factors including a changing environment, random genetic mutations, and competitive behavior. It is probably the case that the best we will ever be able to do under such circumstances is to place an upper bound on the complexity of the problem, and this is all worst-case analysis will provide.

3) Analyses performed by other authors (Grimson, 1988, for example) based on expected or average cases, depend critically on having a well-circumscribed domain and an algorithm. Thus the complexity measures derived reflect algorithmic complexity and not problem complexity as is the goal of the present paper. Only under those conditions can average or expected case analyses be performed. In general, it is not possible to define what the average or expected input is for a vision system in the world. Furthermore, the result of the analysis will be valid only for the average input, and does not place a bound on the complexity of the vision process as a whole. This also would not provide any guidance in the determination of required processing power for real-time performance. See also Uhr (1990).

Critical ideas in complexity theory are that of complexity class and, related to it, reducibility. If a problem S is known to be efficiently transformed (or reduced) to a problem Q then the complexity of S cannot be much more than the complexity of Q. Efficiently reduced means that the algorithm that performs the transformation has polynomial complexity. The class P consists of all those problems that can be solved in polynomial time. If we accept the premise that a computational problem is not tractable unless there is a polynomial-time algorithm to solve it, then all tractable problems belong in P.

In addition to the class P of tractable problems, there is also a major class of presumably intractable problems. If a problem is in the class N, then there exists a polynomial $p(n)$ such that the problem can be solved by an algorithm having time complexity $O(2^{p(n)})$; the time complexity function is asymptotically (as n becomes large) dominated by the polynomial $p(n)$. A problem is NP-Complete if it is in the class NP, and it polynomially reduces to an already proven NP-Complete problem. These problems form an equivalence class. Clearly, there must have been a first NP-Complete problem. The first such problem was that of satisfiability (Cook's 1971 Theorem). There are hundreds of NP-Complete problems. If any NP-Complete problem can be solved in polynomial time, then they all can. Most doubt the possibility that non-exponential algorithms for these problems will ever be found, so proving a problem to be NP-complete is now regarded as strong evidence that the problem is intrinsically intractable. If an efficient algorithm can be found for any one (and hence all) NP-Complete problems, however, it would be a major intellectual breakthrough.

Implications of *NP*-Completeness

What can be done when confronted with an *NP*-Complete problem? A variety of approaches have been taken:

(1) Develop an algorithm that is fast enough for small problems, but that would take too long with larger problems. This approach is often used when the anticipated problems are small.

(2) Develop a fast algorithm that solves a special case of the problem, but does not solve the general problem. This approach is often used when the special case is of practical importance.

(3) Develop an algorithm that quickly solves a large proportion of the cases that come up in practice, but in the worst case may run for a long time. This approach is often used when the problems occurring in practice tend to have special features that can be exploited to speed up the computation.

(4) For an optimization problem, develop an algorithm which always runs quickly but produces an answer that is not necessarily optimal. Sometimes a worst-case bound can be obtained on how much the answer produced may differ from the optimum, so that a reasonably close answer is assured. This is an area of active research, with sub-optimal algorithms for a variety of important problems being developed and analyzed.

(5) Use natural parameters to guide the search for approximate algorithms. There are a number of ways a problem can be exponential. Consider the natural parameters of a problem rather than a constructed problem length and attempt to reduce the exponential effect of the largest valued parameters.

NP-Completeness effectively eliminates the possibility of developing a completely satisfactory algorithm. Once a problem is seen to be *NP*-Complete, it is appropriate to direct efforts toward a more achievable goal. In most cases, a direct understanding of the size of the problems of interest and the size of the processing machinery is of tremendous help in determining which are the appropriate approximations. One could hypothesize that the evolutionary process discovered these methods through millennia of experimentation.

ON ALGORITHMS: "GOOD" vs BIOLOGICALLY PLAUSIBLE

The notions of a good algorithm and an intractable problem was developed in the mid-to-late 1960's. A good algorithm is one whose time requirements can be expressed as a polynomial function of input length. An intractable problem is one whose time requirements are exponential functions of problem length, or in other words, a problem which cannot be

solved by any polynomial time algorithm for all instances. Note that the boundary between good and bad problems is not precise. A time complexity of n^{1000} is surely not very practical while one of $2^{0.001}$ is perfectly realizable. Yet empirical evidence seems to point to the fact that natural problems simply do not have such running times, and that the distinction is a useful one.

Biological plausibility of a given theory or algorithm is not the same notion as that of good algorithm, yet few bother to make the distinction. Usually, physical limitations, however real, do not enter into the discussion just like they do not enter the discussion in any theoretical complexity argument (for example, see Gopalkrishnan, Pamakrishnan, & Kanal, 1991). It is also true in complexity theory that algorithms with polynomial complexity are believed to be *good* while those with exponential complexity are *bad*; yet, there are an infinite number of values of exponents and variables that would lead to the exact reverse when an algorithm is physically realized. Consider simply the following pair of functions: $O(An^x)$ and $O(2^{xn/A})$. It is easy to see that there is an infinite space in which the polynomial function has actual value larger than the exponential depending on the values of the constants A and x. And of course, there is an infinite number of such function pairs that we may compare. Early complexity theorists of course understood this problem. Yet, they claimed that polynomial functions with bad behavior do not occur in practice and likewise exponential functions with good behavior also do not occur in practice. Thus, the search for polynomial and sub-polynomial complexity functions is the driving goal of theory.

But an important issue seems forgotten: if the practice of complexity analysis is to lead to tangible benefits then the theorems must lead to algorithms that must be physically realizable and the physical realization must in some way be better than others with respect to time or space efficiency. No matter what the time and space complexity functions, there is an infinite space of possible variable values or problem sizes which will not be practically realizable. The fact that all computers have finite memories is sufficient to guarantee this. One cannot in practice take infinite time to read or load an infinite Turing Machine tape. Engineering design specifications always impose constraints: the amount of memory may be limited by power consumption or cost; the number of processors is likewise constrained; real-time response places a hard constraint on time complexity and thus on problem size. These constraints cannot be ignored in any complexity discussion which may eventually be used to solve real problems. And the whole point of complexity theory is to formally provide insights on the relative difficulty of real problems. Yet, virtually all theoretical discussions do exactly this. The concern of this essay is on

what are the constraints whose satisfaction is required in order for a theory to be biologically plausible.

It is claimed that it is not sufficient for a perceptual theory to only explain a set of experimental observations; experiments typically can use no more than a minuscule subset of all possible stimuli. Broader considerations beyond experiment are needed. Biological plausibility of a perceptual theory will thus be characterized in three stages. First, a theory must of course be sufficient to explain the observations. Second, it is important to define the size of problem which the algorithm must be able to handle, and this follows:

- The algorithm that embodies the theory accepts up to the same number of input samples of the world per unit time as human sensory organs. It is a non-trivial task to determine exactly the quantitative nature of the input to the human sensory system. With respect to the visual system, there are two eyes; each has about 110-125 million rods and 6.3-6.8 million cones; each eye can discriminate over a luminance span of 10 billion to one; the spatial resolution of the system peaks at about 40 cycles/degree while the temporal resolution peaks at about 40 Hz but the two are not independent; finally, there are many inputs from other sensory and motor areas. See Dowling (1987) for further discussion.
- The implementation that realizes the algorithm exists in the real world and requires amounts of physical resources which exist.
- The output behavior of the implementation as a result of those stimuli is comparable both in quality, quantity and timing to human behavior. The behavioral literature on exactly what the quality, quantity, and timing of human behavior is to a variety of stimuli is immense, but far from complete. What is required however, are responses from the algorithm that agree qualitatively and quantitatively with human responses and that are generated with the same time delays as human responses.

The third stage of the definition requires that the functions for time and space complexity of any algorithm which we claim performs some information processing task in the brain only permit values of their variables which lead to brain-sized space requirements and behaviorally-confirmed time requirements. Issues of polynomial vs exponential do not enter the discussion of biological plausibility at all. In other words:

- solutions should require significantly fewer than about 10^9 processors operating in parallel, each able to perform one multiply-add operation over its input per millisecond;
- processor average fan-in and fan-out should be about 1000 overall; and

- solutions should not involve more than a few hundred sequential processing steps.

Any perceptual theory must satisfy the above characterization; and similarly, any theory of any other aspect of intelligent behavior would have a corresponding characterization of biological plausibility.

ON COMPUTATIONAL MODELING AND PERCEPTION

Complexity theory is as appropriate for analysis of visual search specifically and of perception in general as any other analysis tool currently used by biological experimentalists. Experimental scientists attempt to explain their data and not just describe it; it is no surprise that their explanations are typically well-thought-out and logically motivated, involving procedural steps or events. In this way, a proposed course of events is hypothesized to be responsible for the data observed. There is no appeal to non-determinism nor to oracles that guess the right answer nor to undefined, unjustified, or undreamed-of mechanisms that solve difficult components. In essence, experimental scientists attempt to provide an *algorithm* whose behavior leads to the observed data. Attempts at providing algorithmic explanations appeared even before the invention of the computer. For example, perception as hypotheses and unconscious inference theory (Helmholtz, 1963) is remarkably similar to the current reasoning paradigm in artificial intelligence, where reasoning is formalized as a logical process using formal mathematics.

The basic formal requirement for the computability of perception is that perception be formally decidable (see Davis, 1958, 1965, for in-depth discussions of decidability). If a problem can be formulated as a decision problem, that is, we wish to know of each element in a countably infinite set A, whether or not that element belongs to a certain set B which is a proper subset of A, then the problem is decidable if there exists a Turing Machine which computes yes or no for each element of A. This requires that perception, in general, be formulated as a decision problem. This formulation does not currently exist. Visual search, an important sub-problem however, can be formulated as a decision problem (Tsotsos, 1989) and is decidable (it is an instance of the Comparing Turing Machine defined in Yashuhara, 1971). More research is needed to try to formalize other subproblems of perception in the same way. If some aspect of perception is determined to be undecidable, this does not mean that all of perception is also undecidable nor that other aspects of perception cannot be modeled computationally. For example, one of the most famous undecidable problems is whether or not an arbitrary Diophantine equation has integral so-

lutions (Hilbert's 10th problem). This has theoretical interest, but more importantly, this does not mean that mathematics cannot be modeled computationally! Similarly, another famous undecidable problem is the halting problem for Turing Machines: it is undecidable whether a given Turing Machine will halt for a given initial specification of its tape. This too has important theoretical implications, but since Turing Machines form the foundation of computation, it certainly does not mean that computation cannot exist!

Appeals to non-algorithmic explanations cannot seriously be entertained because, by definition of algorithm, they would not give a step-by-step procedure for achieving a solution to a problem. Thus, the problem would remain unsolved except by appeals to inexplicable processes, and this does not lead us any closer to understanding perception. Since biological scientists provide algorithmic explanations, computational plausibility is not only an appropriate but a necessary consideration. One dimension of plausibility is satisfaction of the constraints imposed by the computational complexity of the problem, the resources available for the solution of the problem and the specific algorithm proposed.

Any computational paradigm is a candidate for use in constructing a biologically plausible model. Neural network approaches are not the only ones that are biologically plausible as is often believed. Neural networks are Turing-equivalent and they are subject to the same constraints of computational complexity and computational theory as any other implementation (see Judd, 1990, for further discussion and proofs of this statement). It is important to note that relaxation processes are specific solutions to search problems in large parameter spaces and nothing more. Neural networks use variations of such search procedures which in general may be termed optimization techniques. If optimization is the process by which real neurons perform some of their computation, it is subject to precisely the same considerations of computational complexity as any other search scheme.

Visual Search

Visual search is a common if not ubiquitous sub-task of vision, in both man and machine. A basic visual search task is defined as follows: given a target and a test image, is there an instance of the target in the test image (Rabbitt, 1978)? Typically, experiments measure the time taken to reach a correct response. Region growing, shape matching, structure from motion, the general alignment problem, connectionist recognition procedures, etc., are specialized versions of visual search in that the algorithms must determine which subset of pixels is the correct match to a given prototype or description. The basic visual search task is precisely what any model-

based computer vision system has as its goal: given a target or set of targets (models), is there an instance of a target in the test display? Even basic vision operations such as edge-finding are also in this category: given a model of an edge, is there an instance of this edge in the test image? It is difficult to imagine any vision system which does not involve similar operations. It is clear that these types of operations appear from the earliest levels of vision systems to the highest.

In Tsotsos (1989), a computational definition of the visual search task was presented, and the unbounded case was distinguished from the bounded case. Unbounded visual search refers to a search task where the target is not given explicitly in advance, and even if it can be given it is not used by the sensory apparatus to optimize search in any way. Bounded visual search on the other hand, is a search task where the target is known explicitly in advance and it is used to optimize the search process. An equivalence was drawn between unbounded search and bottom-up processes, and bounded search and task-directed visual processes.

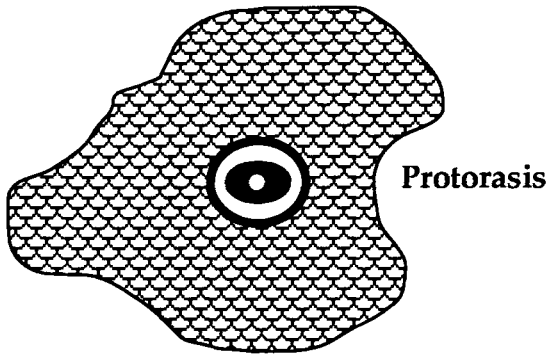
It was shown in Tsotsos (1989, 1992b) that unbounded visual search, regardless of whether the images are time-varying or the camera system is dynamically controlled (active), is *NP*-Complete. This is due solely to the fact that the subset of pixels in an image which corresponds to a target cannot be predicted in advance and all subsets must be considered in the worst case. The bounded problem, on the other hand, requires linear time for the search process. This qualitatively confirms all of the visual search data that has been experimentally discovered (say, by Treisman, 1988). These results are true for an active camera system as well (Tsotsos, 1992b). The four theorems proved in those papers show that in general, a bottom-up approach to perception (as suggested by Marr, 1982) is not only computationally intractable, but biologically implausible. Yet, task-directed approaches do have direct biological counterparts.

ON COMPUTATIONAL COMPLEXITY AND EVOLUTION

In the section introducing complexity theory, an example due to Stockmeyer and Chandra (1979) was given in order to demonstrate the concept of an intractable problem. A new example that has direct biological relevance is now presented to further support this notion. Consider for a moment the following simple-minded and straightforward evolutionary strategy. First, suppose that an organism named Protorasis was the very first organism with a visual system (Figure 1). Its visual system consists of a single eye, whose retina contains a single photoreceptor, which responds uniquely to only 10 shades of gray. Of those shades of gray, only 7 have

meaning to the organism, and are linked to some sort of action. Thus, the only visual processing requirements for Protorasis' brain are that those 7 models are somehow represented and that matching can be done (presumably by a simple network of neurons).

stimuli to which Protorasis' photoreceptor responds



models of meaningful stimuli linked to actions

Figure 1. A fanciful depiction of the very first organism with a visual system, Protorasis. The visual system consists of a single eye, whose retina contains a single photoreceptor which responds uniquely to only 10 shades of gray shown at the top. Of those shades, only 7 have meaning to the organism, and are linked to some sort of action.

This network could be as simple as that depicted in Figure 2, where 7 output neurons are completely connected to the 10 photoreceptor outputs with excitatory and inhibitory connections. Also, suppose that random mutations can cause one or more of the following to change: the number of eyes; the number of photoreceptors in an eye; the range of stimuli to which each photoreceptor can uniquely respond; the number of neurons (and their

connectivity) available for storing and matching models. Also assume that the perceptual and behavioral strategy for each subsequent organism was exactly the same as for Protorasis.

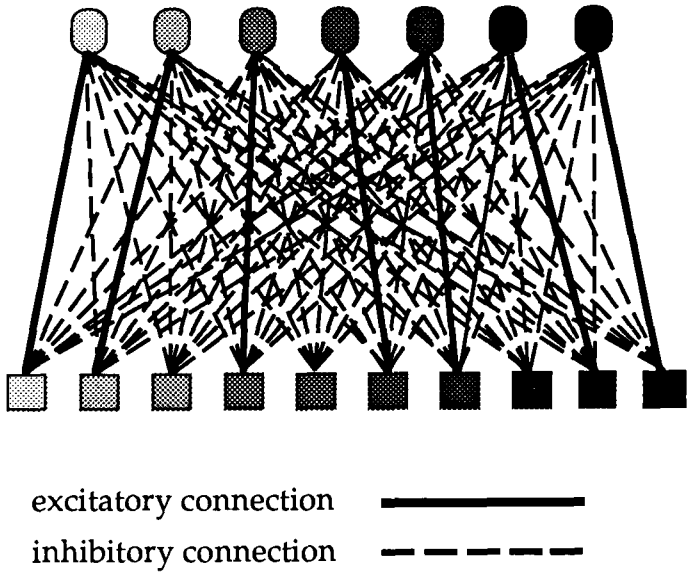


Figure 2. The network required to solve the task of linking a small number of perceptual stimuli to units whose response initiates action for Protorasis. The network has 10 input units, 7 output units, and 70 connections.

What constraints are there on subsequent visual systems so that they may still function as well as that of Protorasis? Let:

E represent the number of eyes (without any assumptions about whether they are convergent on the same portion of the scene or not);

P_i be the number of photoreceptors in eye i;

N be the number of models which may be stored and matched (coarsely speaking, this gives a measure of the amount of brain devoted to visual processing);

S be the number of unique stimuli to which each photoreceptor uniquely responds.

It is easy to characterize the resulting number of possible images and required number of units for matching in a quantitative manner. The number of possible images is given by

$$\prod_{i=1}^{E} S^{P_i}.$$

If each of these images has some significance to an organism, then

$$N = \prod_{i=1}^{E} S^{P_i};$$

otherwise N has some value less than this. In either case the number of connections in the model of Figure 2 is given by $S*N$.

Now suppose that Protorasis-2 was the result of some mutation that increased the size of the brain for model storage and matching from 7 to 700, and also increased the number of photoreceptors from 1 to 4. The number of possible images would be 10^4. Even though there was a comparatively much larger increase in brain size, there is no longer sufficient computation power to recognize more than 7% of the possibilities using the simple-minded strategy of Protorasis. Similarly, if Protorasis-3 resulted from an ability to detect many more shades of gray in the environment, say 100, and included an increase in brain size from 700 to 10,000, the problem is even more acute. Small changes in P or S lead to exponentially large increases in N. In this case, there are 10^8 potential images for the 10,000 storage units!

Perceptual power for an organism may be estimated using the quotient

$$\frac{N}{\prod_{i=1}^{E} S^{P_i}}.$$

This reflects the percentage of world events to which the organism can perceive and react. The larger this value is, the more powerful the perceptual capabilities of the organism are with respect to its sensory apparatus. This of course assumes a constant time recognition strategy as is assumed with the network of Figure 2.

These are just the static images. If τ is the time interval in seconds during which significant time-varying events may occur, and the visual system may sample every α seconds then the total number of possible image sequences would be given by

$$\left(\prod_{i=1}^{E} S^{P_i}\right)^{\tau/\alpha}.$$

It does not require much further analysis or discussion in order to see that this simple and straightforward strategy cannot possibly be the one evolution actually used. Our brains would be wildly larger than they are (recall the characterization of biologically plausible).

It is clear that not all possible images will have significance to any given organism. However, it may be postulated that greater perceptual power, via the ability to recognize a larger and richer set of images, would lead to better likelihood of survival for an organism. For example, a greater variety of food sources could be recognizable as would a greater variety of predators. So, an evolutionary goal could be to achieve large values of E, S, and P for a given value of N.

Formally, as mentioned earlier, the problem of visual search—finding a target in an image—requires exponential time in its worst case using a single processor (or exponential processors for a constant time solution), and the strategy described for Protorasis is similarly exponential in nature (in number of processors and connections). Could it be that through random mutations, evolution discovered the same principles of approximation and optimization that have been determined to be appropriate for dealing with *NP*-Complete problems?

The theory of computational complexity permits these conclusions; and further, to lay a theoretical foundation *proving* why evolution did not take the direct and simple path of Protorasis. Moreover, as described earlier, the theory gives guidelines on how to deal with such exponentially difficult problems, and the surprise is (or perhaps it is not that surprising after all!) that many of the suggestions have biological counterparts:

- Develop a fast algorithm that solves a special case of the problem, but does not solve the general problem: the observation that not all combinations of locations in an image need be considered in a matching process because objects and events are spatio-temporally localized permits spatio-temporally localized receptive fields to be used as an approximating measure. This special case is solved much faster (the exponential is reduced to a low order polynomial function), but the general problem is not solved since general location combinations are not considered (Tsotsos, 1988).
- Develop an algorithm that quickly solves a large proportion of the cases that come up in practice, but in the worst case may run for a long time: the observed serial bottleneck believed to be the reason for visual attention may be a manifestation of this. Simpler tasks can be solved quickly in parallel, while more complex tasks require serial, selective attention (Tsotsos, 1991).

• Use natural parameters to guide the search for approximate algorithms: hierarchical organization and hierarchical abstraction are well-accepted methods for reducing search and these methods reduce the search dramatically in perception (Tsotsos, 1988).

A model that incorporates the above approximations and abstractions leads to the possibility of Protorasis' recognition system performing its tasks with different time requirements. That is, different objects or stimuli may be recognized with different time costs. Thus, a different comparative measure of power is possible: the number of different perceptual events recognizable divided by the time required for their recognition:

$$\frac{N}{\sum_{j=1}^{N} T_j},$$

where model $j \in N$ requires time T_j. The larger this quotient, the more powerful the perceptual system, and this is independent of the details of the sensors themselves (number of eyes, photoreceptors, etc.). The organism which can recognize and act on a larger set of perceptual events, on average faster than another organism, has a competitive advantage over that other organism. This function naturally accounts for the fact that many events may be recognizable very quickly while others may take more time; and that a larger proportion of the former is better than a larger proportion of the latter.

Unfortunately, it is not obvious how these measures can be put into practice. It is probably not possible to ever know the value of N for any organism; however, for a given large set of objects or event stimuli, the corresponding times for recognition could be measured and thus different organisms compared. It would be illuminating to carry out such a comparative experiment.

ON COMPUTATIONAL MODELING AND BEHAVIORISM

The philosophy for realizing intelligent behaviors in machines as articulated by Brooks, his colleagues and others, has received a great deal of attention (Brooks, 1991a, 1991b). Brooks believes that machines constructed out of simple modules with simple communication will exhibit intelligent behavior as an emergent property; the behavior is not directed by a single homunculus nor is it explicitly specified in the machine in any way. These principles are the cornerstones of the *subsumption* architecture Brooks proposed in 1986 for intelligent control. A simple description of the

subsumption idea includes: control layers define a total order on a robot's behaviors; the dominance of layers follow a hypothesized evolutionary sequence; each layer may spy on layers at lower levels and inject signals into them. It is claimed that the structure is scalable to human-like behavior and Brooks argues strongly against:

- the sense-model-plan-act framework for robot control;
- the representation of intermediate or hierarchical computations;
- the explicit representation of goals; and,
- CAD-like models of the world.

He goes on to claim that perception is connected to action, and further that his approach can be extended to cover the whole story, both with regards to building intelligent systems and to understanding human intelligence. As proof of his position he offers compelling evidence: many mobile robots that seem to have robust and interesting performance.

Brooks seems to be re-kindling the torch of old behaviorism, a philosophy appearing about 1913 in the psychology community (Watson, 1919). Behaviorism stood for one basic belief: humans are biological machines and as such do not consciously act, do not have their actions determined by thoughts, feelings, intentions, or mental processes. Human behavior is a product of conditioning: humans react to stimuli. Behaviorism is not popular currently in psychology nor in cognitive neuroscience.

Similarly, arguments against Brooks' position are not new. For example, Kirsh (1991) focuses on one of Brooks' claims, that intelligent behavior is concept-free. Kirsh claims that concepts are necessary for some types of behavior and also can make computational processes simpler. He argues for the need of representation in a theory of perception simply because vision is complex and must be sometimes solved in general ways.

But Brooks is not alone in his belief that some sort of behaviorist theory is the most appropriate. Ramachandran's (1990) utilitarian theory is remarkably similar. Ramachandran rejects previous well-known theories of perception (Helmholtz's perception as unconscious inference, Gibson's direct perception, Marr's natural computation) and proposes rather that perception does not involve intelligent reasoning, nor resonance with the world, nor the creation of internal representations. Rather, perception is a *bag of tricks*. Through millions of years of evolution, the visual system has evolved numerous short-cuts, rules-of-thumb, and heuristics each one adopted only because it works and not because of any other appeal. Ramachandran is particularly critical of computational theories. Although he does make some valid points, he has developed a perspective on perception that can be labeled as a behaviorist approach just like Brooks, and thus is subject to the same criticism, as will be outlined next.

Behaviorists seem haunted by one of their claims, namely that their paradigm and the solutions formulated within it will scale up to problems which are human-like in their size. This is particularly true of Brooks, who claims a solution to intelligence in general. The arguments Brooks presents on scaling are inadequate (Brooks, 1991a, 1991b). Although Brooks mentions the issue, these arguments never appear in any concrete and direct fashion. Ramachandran seems unaware of this issue.

It can be proved that strict behaviorism (that is, not deviating in any way from the published principles and dogma, specifically, that no explicit targets are permitted) is not supported by current biological evidence and may require time to execute that is given by an exponential function in the image size in the worst-case (Tsotsos, 1992a). It does not matter what kind of computational medium is used for the implementation (recall the Church/Turing Thesis); the exponential worst-case behavior depends solely on the inability of a behaviorist system to know where the stimuli that trigger tricks or behaviors are found. Parallelism does not help; if the search is conducted in parallel, an unrealizable number of processors (given by an exponential function of image size) will be needed, again something which is not biologically plausible. It does not matter if the visual system is passive or active, the same conclusions are reached. It is this search action which is inherent in the behaviorist or utilitarian view but which is never explicitly addressed that results in the rejection of these theories. Small signals (such as simple voltages, or sonar blips) would not lead to the same problem; thus the success of the current implementations. The alternative is to employ a satisfying set of approximations and optimizations (Tsotsos, 1988, 1990, 1992a, 1992b) that tie the behaviors or tricks together.

CONCLUSION

In this essay I argued for the need to consider issues of realizability within biologically plausible limits for any theory that is proposed as an explanation for perception, or intelligence in general. The theoretical foundations for realizability can be laid within the framework of computational complexity theory. Further, that theory provides guidelines for how to deal with problems that appear to be unrealizable. In previous papers, it was shown that a small number of unifying approximations and optimizations are sufficient for reducing the potential combinatorial explosion and satisfying the definition of biological plausibility outlined earlier (Tsotsos, 1988, 1990, 1992a, 1992b).

Now, this kind of argument is not new: Uhr (1980) and Feldman and Ballard (1982) among others, have attempted to make similar arguments. Each drew their own conclusions: Uhr that pyramid structures were needed; Feldman and Ballard argued for massive parallelism. However, none of the previous authors tied such back-of-the-envelope calculations directly to a formal theory and none put all the elements together to show that they satisfy biological plausibility.

The results force a change to Marr's (1982) view of computational vision, namely, that in principle solutions are not necessarily realizable and thus are not necessarily acceptable. A necessary condition on their validity is that they must also satisfy the complexity constraints of the problem and the resources allocated to its solution. Similarly, the results force a change to the behaviorist approach to intelligence.

One final point: we cannot assume that evolution finds optimal solutions in the same sense that complexity theory seeks. Evolution finds satisfying solutions and it is those solutions which perceptual theorists are attempting to find. It would be an uninteresting conclusion if complexity theory applied only to artificial computation problems and not natural ones. Thus, this essay argued for a new style of complexity analysis, that attempts to balance problem complexity, available resources for its solution and required performance time in the context of the computational modeling of biological perception.

Acknowledgment. Parts of this paper were written while the author was visiting the Multimedia Systems Institute of Crete at the Technical University of Crete, with the kind support of Prof. Stavros Christodoulakis. The author is the CP-Unitel Fellow of the Canadian Institute for Advanced Research. This research was funded by the Information Technology Research Center, one of the Province of Ontario Centers of Excellence, the Institute for Robotics and Intelligent Systems, a Network of Centers of Excellence of the Government of Canada, and the Natural Sciences and Engineering Research Council of Canada.

DISCUSSION

V. S. Ramachandran (*Neurosciences Program, University of California at San Diego, La Jolla, CA*): In this essay Tsotsos raises several interesting issues concerning the computational approach to vision. It seems to me that he and I see eye to eye on many issues but not on all. In this "reply" I will not comment on the more formal aspects of his theory but will confine myself, instead, to some of the meta-theoretical questions that he ad-

dresses. It is not clear to me what he means by the term "behaviorist." I should point out as the answer that my criticisms were directed mainly against the Marr school of Computational Vision rather than computational vision in general (Ramachandran, 1985b, 1990; Churchland, Ramachandran, & Sejnowski, 1993). In this essay I shall try to summarize some of these ideas.

David Marr's ideas created nothing short of a revolution in our understanding of human vision comparable to the Chomskyian revolution in linguistics. The major strength of his approach to vision is that it allows a much more precise and rigorous formulation of perceptual problems than what one could achieve by doing psychophysics or physiology. Unfortunately, there are also several major pitfalls associated with his approach and I shall take this opportunity to spell them out briefly.

Levels of Analysis. Any complex information processing systems—including the human visual system—can be understood at several distinct "levels"—e.g., the level of the "computational problem," the level of algorithm (a sequence of steps) and finally, least important (in Marr's scheme), the actual neural hardware that is used to implement the algorithm. To ensure progress in understanding vision it is important not to get "confused" between these levels, especially since the logical structure of arguments at each level is quite independent of the other levels.

This argument may be valid for some simple machines, but when we are talking about complex biological systems, I would like to submit that the only sure way to progress, in fact, is to deliberately get confused between these levels—deliberately make what orthodox philosophers might call "category mistakes." There is now a wealth of experimental evidence which suggests that our perceptual experience of the world is powerfully constrained by the actual neural machinery, i.e., the "hardware" that mediates perception (e.g., see Ramachandran, 1985b, 1992; Ramachandran & Gregory, 1978, 1991; Ramachandran, Rogers-Ramachandran, Stewart, & Pons, 1992; Ramachandran, Stewart, & Rogers-Ramachandran, 1992). And, in general, I think no important discovery in science has ever been made by respecting the distinctions between levels. For example, consider Mendelian inheritance. You can't think of two more different levels than the behavior of pea plants and the structures of molecules, and yet it is by bridging these two that the science of Biology was born.

Identifying the computational problem. According to Marr, the single most important step in understanding human vision is to provide a precise mathematical formulation of the "computational problem" confronting the organism. In doing this it is best to start from first principles (e.g., by considering "natural constraints") and to avoid being confused by results obtained by psychophysical and physiological methods.

Unfortunately it is not always obvious what the so-called computational problem is in any given situation. (Try answering the simple question: what is the goal of color vision?) Indeed the real visual system often seems to subdivide any given problem into many "sub-problems" many of which would be difficult to discern unless you do experiments and acquire a certain familiarity with the phenomenology of human vision. It should come as no surprise, therefore, that most of the computational problems that *AI* researchers are currently preoccupied with were in fact identified by psychophysicists (e.g., the stereo-correspondence problem by Julesz and Wheatstone; the structure from motion by Wallach, the aperture problem by Wallach, and the motion correspondence problem by Ternus). They certainly weren't deduced from "first principles."

Modularity. According to Marr's doctrine of modularity, "early vision" processes, such as stereo, motion correspondence, shape-from-shading, structure from motion, etc., are mediated by several autonomous modules. These modules remain largely insulated from each other and convey the results of their computations to higher visual centers for subsequent processing. Vision, in this scheme, is a strictly bottom-up affair.

It may indeed be useful to treat vision as modular at least as first approximation but there is now a great deal of evidence that the modules must interact with each other significantly even at the very earliest stages of visual processing. We have shown, for example, that both motion correspondence (Ramachandran, 1985a; Ramachandran & Anstis, 1986) and stereopsis (Ramachandran, 1986; Nakayama, Shimojo, & Ramachandran, 1990) can be strongly influenced by image segmentation based on implied occlusion. More remarkably, we find that a jumping sound source superimposed on a dynamic noise display will cause the noise to "jump" along with the sound—an example of cross-modal motion capture (Ramachandran, Intrilagator, & Cavanagh, unpublished manuscript).

A simple version of cross modal motion capture can be produced by using a single dot blinking on and off adjacent to a white square. Subjects viewing this display usually do not see any motion—they just see a spot blinking on and off. We then added an auditory stimulus presented by earphones. Simultaneous with the blinking on of the light, a tone is sounded in the left ear; simultaneous with the blinking off, a tone is sounded in the right ear. Subjects see the single dot move to the right behind the occluder. In effect, the sound "pulls" the dot in the direction the sound moves (Ramachandran, Intrilagator, & Cavanagh, unpublished manuscript). This is convincing evidence for some form of "heterarchy," and against a pure, straight through, noninteractive hierarchy. (A weak subjective motion effect can be achieved when the blinking of the light is accompanied by somatosensory left-right vibration stimulation to the hands.)

It comes as no surprise that visual and auditory information is integrated at some stage in neural processing. After all, we see dogs barking and drummers drumming. What is surprising about these results is that the auditory stimulus has an effect on a visual process (motion correspondence) that Pure Vision orthodoxy considers "early."

Segmentation. This is a special case of the modularity argument. According to this view, certain elementary visual functions such as stereopsis, motion, color, etc. are mediated relatively early in visual processing by specialized modules, whereas segmentation of the visual scene into separate objects is assumed to be a more complex process that can actually use the output of these early vision modules. Since the modules perform their functions prior to image segmentation, the argument goes, one can successfully model them and study them experimentally without worrying about segmentation (Marr, 1981). Contrary to this view, out evidence suggests that image segmentation can profoundly influence a number of early visual processes such as stereopsis (Ramachandran, 1986; Nakayama, Shimojo, & Ramachandran, 1990), structure from motion (Ramachandran, Cobb, & Rogers-Ramachandran, 1988), motion correspondence (Ramachandran, 1985a; Ramachandran, Rao, & Vidyasagar, 1973), and shape-from-shading (Ramachandran, 1988). The implication is that the early vision modules are not autonomous—they interact significantly with each other and with segmentation. Any program of research on perception must take these facts into account.

Consider stereopsis, the matching of slightly dissimilar images from the two eyes to recover stereoscopic depth. Julesz stereograms (Julesz, 1971) are often cited by Marr and his colleagues to illustrate the view that stereopsis is a prime example of modularity—of an early visual process that is relatively autonomous and insulated from other visual processes such as segmentation. The stereogram depicted in Ramachandran (1986, Figure 7) flatly contradicts this view. We created this stereogram using two illusory squares by introducing small horizontal disparities between the vertical edges of the cut sectors. The disks themselves were at zero disparity in relation to the surrounding frame.

If the top pair (conveying crossed disparities) is stereoscopically fused, one sees a striped square standing well in front of a background consisting of black circles on a striped mat. If the bottom pair (uncrossed disparities) is fused, one sees four holes in the striped opaque foreground mat, and through the holes, well behind the striped mat, one sees the four corners of a partially occluded striped square on a black background. These are especially surprising results, because the stripes of the perceived foreground and the perceived background are, by definition, at zero disparity. The only disparity that exists on which the brain can base stereo depth per-

ception comes from the edges of the pacmen. Notice that in this display, the illusory contours must emerge after stereoscopic fusion and yet these contours can in turn influence the matching of finer elements in the display. Assuming that perceiving subjective contours is a "later" effect requiring global integration, and that finding stereo correspondences for depth is an "earlier" effect, then this result appears to be an example of "later" influencing—in fact enabling—"earlier." The emergence of qualitatively different percepts (illusory square in front of disks, versus illusory square behind portholes) cannot be accounted for by any existing stereo algorithms. Most of these algorithms would simply predict a reversal of in sign of perceived "depth" if the disparities are reversed (Ramachandran 1986; Nakayama, Shimojo, & Ramachandran, 1990).

Absence of "top-down" influences. According to Marr, the computations of early vision modules are unaffected by high level object knowledge and semantics. The segmentation of Gregory's "Dalmatian dog" according to this view, occurs not because we know it is a dog and use this knowledge to segment the image but because of certain hidden cues intrinsic to the image, e.g., collinear edges that generate illusory contours around the edge of the dog.

But if this is strictly true, why does a hollow mask (viewed from the inside) look convex rather than hollow? Are Helmholtz (1963) and Gregory (1970) incorrect in assuming that the reason faces look convex is because we know them to be faces? This is an important issue, for if they are right then it would be a striking example of the role of "top-down" influences in vision and would imply that even semantic knowledge can influence the processing of early vision modules such as those concerned with shape from shading and stereopsis.

But does this depth reversal effect really have anything to do with faces? Is it possible, for example, that the reversal of the hollow mask results simply from a generic assumption that objects are usually convex? Or does high-level semantic knowledge also play a role? To find out, Richard Gregory, Kerrie Maddock, and I presented subjects with two adjacent masks, one of which is right side up, the other is upside down. Upside down faces are often poorly recognized, and in any case, upright faces are what we normally encounter. In the experiment, subjects walked slowly backwards away from the pair of stimuli, starting at 0.5 m, moving to 5.0 m. At a distance of about 0.5 m, subjects see both masks as depth inverted (concave). At about 1 m, subjects usually see the upright mask as convex; the upside down mask, however, they continue to see as concave until they are at a distance of 1.5-2.0 m. Because the stimuli are identical except for orientation, this experiment illustrates that "later" process (face categorization) has an effect on an "earlier" process (shading and stereopsis).

Hierarchical Processing. Marr's scheme implies that vision is largely a "bottom-up" process with a one-way flow of information from the sense organs to the motor output. Although generating an appropriate motor output is the ultimate goal of vision there has, until now, been no evidence that the motor programs themselves (that are used to generate the output) can influence the early stages of perception. This idea has been dubbed "dead vision" by Ballard (1989) to contrast it with what Brooks (1986) has called "active vision."

It is remarkable that this myopic view of vision has held sway for so by especially given the flatly contradictory evidence from physiology—the existence of massive backprojections from so-called "higher" to lower visual areas. It is a well known, but often glossed over fact, for example, that there are three times as many fibers coming back from *V1* to the *LGN* than vice versa—even though the textbooks usually mislead as by showing only forward projecting arrows. It is usually assumed tacitly that these back projections may simply be involved in some aspect of overall gain control but that they may not be crucially involved in the actual computations that lead to perception.

The fact that what we see depends not only on the input but also on what you intend to do with the information (i.e., the type of behavior you wish to generate) receives support from a new series of experiments that we have been doing on patients with squint (Ramachandran, Cobb, & Valente, 1992).

Exotropia is a form of squint in which both eyes are used when fixating on small objects close by (e.g., a foot from the nose) but when looking at distant objects, the "squinting" eye deviates outward by as much as 40° to 60°. Curiously, the patient does not experience double vision—the deviating eye's image is usually assumed to be "suppressed." It is not clear, however, at what stage in visual processing the suppression occurs.

Surprisingly, it is claimed by orthoptists that in a small subset of these patients, "fusion" occurs not only during inspection of near objects, but also when the squinting eye deviates (see Duke-Elder, 1949, for a review). This phenomenon, called "anomalous retinal correspondence" or *ARC*, has not always been taken seriously, perhaps because it was assumed that *ARC* implies a rather improbable lability of binocular receptive fields. Clinicians and physiologists raised in the Hubel-Wiesel tradition usually take it as Gospel that (1) binocular connections are established in area 17 in early infancy and that (2) binocular "fusion" is based exclusively on anatomical correspondence of inputs in area 17. For instance, if a squint is surgically induced in a kitten or an infant monkey, area 17 displays a complete loss of binocular cells (and two populations of monocular cells) but the maps of the two eyes never change. No apparent compensation such as

"anomalous correspondence" has been observed in area 17 and this has given rise to the conviction that it is highly improbable that an *ARC* phenomenon truly exists.

To explore the possibility that there might be more to the *ARC* reports, Ramachandran, Cobb, and Valente (1992) recently studied two patients who had intermittent exotropia. Ramachandran's two patients appeared to "fuse" images both during near vision and during far vision—when the left eye deviated outward—a condition called "intermittent exotropia with anomalous correspondence."

To determine whether these patients do indeed have two (or more) separate binocular "maps" of the world, Ramachandran, Cobb, and Valente (1992) devised an experimental procedure that tested the binocular alignment of after-images; the after-image for the right eye being generated independently of the after-image for the left eye. Here is the procedure: (1) The subject (with squint) was asked to shut one eye and to fixate on the bottom of a vertical slit-shaped window mounted on a flashgun. A flash was delivered to generate a vivid monocular afterimage of the slit. He was then asked to shut this eye and view the top of the slit with the other eye (and a second flash was delivered). (2) The subject opened both eyes and viewed a dark screen, which provided a uniform background for the two afterimages.

The results were as follows: (A) The subject (with squint) reported that he saw afterimages of the two slits that were perfectly lined up with each other, so long as he was converging within about arm's length. (B) On the other hand, if he relaxed vergence and looked at a distant wall (such that the left eye deviated), the upper slit (from the anomalous eye) vividly appeared to move continuously outwards so that the two slits eventually became misaligned by several degrees. They then repeated this experiment on two normal control subjects and found that no misalignment of the slits occurred for any ordinary vergence of conjugate eye movements. Nor could misalignments of the slits be produced by passively displacing one eyeball to mimic exotropia in the normal individuals.

Ramachandran, Rogers-Ramachandran, Stewart, and Pons (1992) have dubbed this phenomenon "dynamic anomalous correspondence." The phenomenon itself is not new but these authors have been able to establish its existence clearly and have pointed out a number of implications that appear not to have been recognized by the Neuroscience and *AI* community.

1. Binocular correspondence can change continuously in "real time" in a single individual depending on the degree of exotropia. Hence, binocular correspondence (and "fusion") cannot be based exclusively on the anatomical convergence of inputs in area 17. The relative displacement observed between the two afterimages also implies that the "local sign" of retinal

points (and therefore binocular correspondence) must be continuously updated as the eye deviates outwards.

2. Since the two slits would always be "lined up" as far as area 17 is concerned, the observed misalignment implies that feedback (or feedforward) signals from the deviating eye must somehow be extracted separately for each eye and must then influence the egocentric location of points selectively for that eye alone. This is a somewhat surprising result, for it implies that "remapping" of egocentric space must be done very early—before the "eye of origin" label is lost—i.e., before the cells become completely binocular. Since most cells beyond area 18 (e.g., *MT* or *V4*) are symmetrically binocular we may conclude that the correction must involve interaction between reafference signals and the output of cells as early as 17 or 18.

It is quite remarkable that a complete remapping of perceptual space in *x-y* coordinants can occur selectively for one eye's image simply in the interest of preserving binocular correspondence. It would be interesting to see if this remapping process can be achieved by algorithms of the type proposed by Zipser and Anderson (1988) for parietal neurons or by "shifter-circuits" of the kind proposed by Van Essen and Anderson (1990).

Identifying "natural constraints." An important idea put forth by Marr is the notion of "natural constraints." Marr points out (as did J. J. Gibson, R. L. Gregory, and Helmholtz) that the evolving visual system did not have to cope with problems of arbitrary complexity (i.e., not like solving a problem in number theory, for example). The system can capitalize, instead, on certain statistical regularities in the natural world—regularities based on the physics of matter, and these properties can be used to impose constraints on solutions to perceptual problems.

But how do you go about identifying these constraints? It would be wonderful if they could be deduced from first principles, of course, but you really can't do this because you never know which particular constraints a given creature is exploiting unless you watch what it is doing. For example, at some abstract level both bats and humans have the same problem—avoiding obstacles and grasping edible objects (either with the mouth or with the hands)—but bats use echolocation and humans use a variety of depth cues—primarily visual—such as stereo, motion parallax, convergence, etc. And there is no way you could have guessed that bats used echolocation unless you did bat psychophysics—e.g., jamming their sonar to see what happens to their behavior.

In this context, Tsotsos also makes a valid point. His work suggests that in addition to natural constraints (imposed by the environment), there is also another important source of constraints, namely, that arising from computability and resources available for computation. This is just as real

and in some ways more fundamental from a theoretical perspective since such constraints would apply regardless of sensory modality or domain of application.

Finally, it is surely obvious that natural constraints, by themselves, do not impose a unique solution to perceptual problems; there are usually far too many theoretically plausible solutions and the only way to find out which particular one is used is by doing psychophysics and physiology.

Conclusion. It would seem, therefore, that many of Marr's ideas are fundamentally flawed (e.g., the notion that segmentation does not constrain early vision), whereas others may be only partially true (e.g., the idea that early vision is relatively immune from top-down influences). I hasten to add, however, that this long list of criticisms should not in any way be seen as detracting from the originality and importance of Marr's contribution. Marr was a brilliant scholar and had he been alive today he would almost certainly have cheerfully acknowledged these shortcomings. I do hope, however, this commentary will stimulate at least some of his colleagues to incorporate our findings into a new and more viable theory of human vision.

Acknowledgment. I thank the ONR and AFOSR for funding this commentary and Francis Crick, Patricia Churchland, Terry Sejnowski, Dan Plummer, Steve Cobb, and Richard Gregory for stimulating discussions.

Tsotsos: Since computational behaviorists cite Ramachandran's "collection of tricks" metaphor as biological evidence for their models, I included his work in my broad view of "behaviorism" in order to strengthen my arguments by extending their domain of applicability to include a biological model. Unfortunately, it is clear that the term behaviorism has no single interpretation. Ramachandran's utilitarian view was never intended as a formal model nor should it be classified as behaviorist. What does matter however, is how Ramachandran's collection of tricks is to actually work so that human perception is the result. This is a goal of the computational behaviorists. Are all tricks always active, always looking at the whole visual field to see if they should react? It can be easily shown that this is an intractable solution. Does each simply look at specific parts of the visual field in order to reduce the amount of computation? This would not lead to the flexibility of function that human visual systems possess. The only tractable yet flexible solution is that each is activated only by an appropriate stimulus; thus an integration strategy is needed. The computational behaviorists provide one, but do so in an ultimately intractable and biologically implausible manner. Ramachandran's contribution is to illuminate some of the interactions which must be

explained. Together, we both stress the fact that theorists and modelers alike must respect computational and biological realities.

The physical constraints I list at the beginning of my paper are all orthogonal dimensions in design to those discussed by Marr (1982). According to Marr, the computational level of a theory addresses the questions: What is the goal of the computation? Why is it appropriate? and, What is the logic of the strategy by which it can be carried out? Marr called solutions at this level "in principle" solutions. At the representational and algorithmic level one asks: How can this computational theory be implemented? What is the representation for the input and output? What is the algorithm for the transformation? And, finally, at the implementational level one asks: How can the representation and algorithm be realized physically? Complexity considerations (problem complexity, resources, performance specifications) span these three levels and are not just implementational details as Marr implies. If the task to be performed or the algorithm to be implemented is tractable, then perhaps efficiency is only an implementational detail. However, if the task is an intractable one, as vision in its most general form seems to be, complexity satisfaction is not simply a detail to contend with during implementation, just as discretization and sampling effects or numerical stability are not simply implementational details. Complexity satisfaction is a major constraint on the possible solutions of the problem. It can distinguish between solutions that are realizable and those that are not. Ascertaining how much computation can be performed will strongly constrain which computations are chosen to actually solve the problem. It is this class of "natural constraints" which I propose play important, but until now, ignored roles in perceptual theories. It is not the case as Ramachandran states that we should purposely confuse the levels. Rather, the levels are intimately related, and they are related by other orthogonal design dimensions.

Finally, I would like to strengthen Ramachandran's argument against the independent modules view proposed by Marr. When Marr proposed the independent modules view, he was working on a hypothesis that, in the mid-to-late-1970's, reflected current best knowledge of neurobiology. John Allman and Jon Kaas had recently discovered area *MT* in the owl monkey which seemed to be concerned exclusively with motion computations (Allman & Kaas, 1971). Semir Zeki had reported observations on area *V4* (Zeki, 1977), and it appeared as if the role of *V4* was to process color independently of motion. Since these two areas had such unique and seemingly independent properties, a good hypothesis to test would be whether or not the independence applied throughout the visual cortex. This would also be good for computational modelers; we could work on solving simpler and smaller sub-problems, and then only worry about their

integration into a whole rather than have to deal with the many interactions among functionalities. This was a very sensible thing to propose at the time, and David Marr left his mark on the field for realizing this. Evidence accumulated since then, however, paints a very different picture of the visual cortex and a serious look at the current neurobiology leads to strong contradiction. The paper by Felleman and Van Essen (1991), for example, if anything else, is a crystal clear demonstration that no area of the visual cortex is without massive input from many other areas, most of the pathways are both bottom-up as well as top-down, and further that we are quite in the dark about the details of what each of the visual areas is computing. Even the independent *P* and *M* pathways distinction has fallen by the wayside (Maunsell, 1992; Martin, 1992). The view recently proposed by Oliver Braddick on the computations underlying the perception of motion is even more problematic (Braddick, 1992). He cites evidence that leads him to believe that the computations are composed of many interacting computational loops and re-entrant processing streams. No independent modules here! The hypothesis has been refuted with respect to biological visual systems and those who continue to follow that perspective are out of date.

REFERENCES

ALLMAN, J. M., & KAAS, J. H. (1971). A Representation of the visual field in the caudal third of the middle temporal temporal gyrus of the owl monkey. *Brain Research*, **31**, 85-105.

BALLARD, D. (1989). Reference frames for animate vision. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1635-1641). Palo Alto, CA: Morgan Kaufmann. [VSR]

BRADDICK, O. (1992). Visual Perception: Motion may be seen but not used. *Current Biology*, **2**, 597-599.

BROOKS, R. (1986). A Layered Intelligent Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation*, **2**, 14-23.

BROOKS, R., (1991a). Intelligence without representation. *Artificial Intelligence*, **47**, 139-159.

BROOKS, R. (1991b). Intelligence without reason. In *Proceedings of the Twelve International Joint Conference on Artificial Intelligence* (pp. 569-595). Palo Alto, CA: Morgan Kaufmann.

BYLANDER, T., ALLEMANG, D., TANNER, M., JOSEPHSON, J. (1989). Some results Concerning the Computational Complexity of Abduction. In *Proceedings of the First International Conference on Principles of*

Knowledge Representation and Reasoning (pp. 44-54). Palo Alto, CA: Morgan Kaufmann.

CHURCH, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, **58**, 345-363.

CHURCHLAND, P., RAMACHANDRAN, V. S., & SEJNOWSKI, T. (1993). *A critique of pure vision*. Cambridge, MA: MIT Press. [VSR]

COOK, S. (1971). The complexity of theorem-proving procedures. In *Third Annual Symposium on Theory of Computing* (pp. 151-158). New York: Association for Computing Machinery.

DAVIS, M. (1958). *Computability and Unsolvability*. New York: McGraw-Hill.

DAVIS, M., (1965). *The Undecidable*. New York: Hewlett Raven Press.

DOWLING, J. (1987). *The Retina: An Approachable Part of the Brain*. Cambridge, MA: Harvard University Press.

DUKE-ELDER, S. W., (1949). *Textbook of ophtalmology* (vol. 4). St. Louis, MS: Mosby. [VSR]

FELDMAN, J., & BALLARD, D. (1982). Connectionist models and their properties. *Cognitive Science*, **6**, 205-254.

FELLEMAN, D., & VAN ESSEN, D. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, 1:1, 1-47.

GAREY, M., & JOHNSON, D. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: Freeman.

GODBEER, G. (1987). *The Computational Complexity of the Stable Configuration Problem for Connectionist Models* (Technical Report No. 208-88). Toronto: University of Toronto, Department of Computer Science.

GOPALKRISHNAN, P., PAMAKRISHNAN, I., & KANAL, L. (1991). Approximate algorithms for the knapsack problem on parallel computers. *Information and Computation*, **91**, 155-171.

GREGORY, R. L. (1970). *The Intelligent Eye*. New York: McGraw-Hill. [VSR]

GRIMSON, E. (1988). The Combinatorics of Object Recognition in Cluttered Environments using Constrained Search. In *Proceedings of the Second International Conference on Computer Vision* (pp. 218-227). Silver Spring, MD: IEEE Coputer Society Press.

HELMHOLTZ, H. v. (1963). *Handbook of Physiological Optics*. New York: Dover. (Translated by J. P. C. Southall. Originally published in 1867).

JUDD, J. S., (1990). *Neural network design and the complexity of learning*. Cambridge, MA: MIT Press.

JULESZ, B. (1971). *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press. [VSR]

KAUTZ, H., & SELMAN, B. (1990). Hard Problems for Simple Default Logics. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning* (pp. 189-197). Palo Alto, CA: Morgan Kaufmann.

KIROUSIS, L., & PAPADIMITRIOU, C. (1988). The complexity of recognizing polyhedral scenes. *Journal of Computer and System Sciences*, **37**, 14-38.

KIRSH, D. (1991). Today the earwig, tomorrow man? *Artificial Intelligence*, **47**, 161-184.

MARR, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

MARTIN, K. (1992). Visual Cortex: Parallel pathways converge. *Current Biology*, **2**, 555-557.

MAUNSELL, J. (1992). Functional visual streams. *Current Opinion in Neurobiology*, **2**, 506-510.

NAKAYAMA, K., SHIMOJO, S., & RAMACHANDRAN, V. S. (1990). Transparency: Relation and depth, subjective contours, luminance and neon spreading. *Perception*, **19**, 497-513. [VSR]

RABBITT, P. (1978). Sorting, categorization, and visual search. In E. Carterette & M. Freidman (Eds.), *Handbook of Perception: Perceptual Processing* (vol. 9, pp. 85-136). New York: Academic Press.

RAMACHANDRAN, V. S. (1985a). Apparent motion of subjective surfaces. *Perception*, **14**, 127-134. [VSR]

RAMACHANDRAN, V. S. (1985b). Guest editorial: The neurobiology of perception. *Perception*, **14**, 97-105. [VSR]

RAMACHANDRAN, V. S. (1986). Illusory contours capture stereopsis and apparent motion. *Perception & Psychophysics*, **39**, 361-373. [VSR]

RAMACHANDRAN, V. S. (1988). Perception of depth from shading. *Scientific American*, **269**(8), 76-83. [VSR]

RAMACHANDRAN, V. S. (1989). *Visual perception in people and machines*. Presidential lecture given at the annual meeting of the Society for Neuroscience, Phoenix, AZ. [VSR]

RAMACHANDRAN, V. S. (1990). Interactions between motion, depth, color, and form: The Utilitarian theory of perception. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 346-360). New York: Cambridge University Press.

RAMACHANDRAN, V. S. (1992). Blind spots. *Scientific American*, **266**(5), 86-91. [VSR]

RAMACHANDRAN, V. S., & ANSTIS, S. M. (1986). Perception of apparent motion. *Scientific American*, **254**(6), 102-109. [VSR]

RAMACHANDRAN, V. S., COBB, S., & ROGERS-RAMACHANDRAN, D. (1988). Recovering 3-D structure from motion: Some new constraints. *Perception & Psychophysics*, **44**, 390-393. [VSR]

RAMACHANDRAN, V., COBB, S., & VALENTE, C. (1992). Dynamic anomalous retinal correspondence: A problem for theories of binocular vision. Unpublished manuscript. [VSR]

RAMACHANDRAN, V. S., & GREGORY, R. L. (1978). Does colour provide an input to human motion perception? *Nature*, **275**, 55-56. [VSR]

RAMACHANDRAN, V. S., & GREGORY, R. L. (1991). Perceptual filling in of artificially induced scoromas in human vision. *Nature*, **350**, 699-702. [VSR]

RAMACHANDRAN, V., INTRILAGATOR, J., & CAVANAGH, P. (unpublished manuscript). Moving sound sources generate motion capture. [VSR]

RAMACHANDRAN, V. S., RAO, V. M., & VIDYASAGAR, T. (1973). Apparent motion with subjective contours. *Vision Research*, **13**, 1399-1401. [VSR]

RAMACHANDRAN, V. S., ROGERS-RAMACHANDRAN, D., STEWART, M., & PONS, T. (1992). Perceptual correlates of massive cortical reorganization. *Science*, **258**, 1159-1160. [VSR]

RAMACHANDRAN, V. S., STEWART, M., & ROGERS-RAMACHANDRAN, D. (1992). Perceptual correlates of massive cortical reorganization. *Neuroreport*, **3**, 583-586. [VSR]

SELMAN, B., & KAUTZ, H. (1990). Model-Preference Default Theories. *Artificial Intelligence*, **45**, 287-322.

SELMAN, B., & LEVESQUE, H. (1989a). Abductive and Default Reasoning: A Computational Core. In *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 343-348). Palo Alto, CA: Morgan Kaufmann.

SELMAN, B., & LEVESQUE, H. (1989b). The Tractability of Path-Based Inheritance. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1140-1145). Palo Alto, CA: Morgan Kaufmann.

STOCKMEYER, L., & CHANDRA, A. (1988). Intrinsically difficult problems. *Scientific American Trends in Computing* (vol. 1, pp. 88-97). New York: Scientific American.

TREISMAN, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, **40**(A-2), 201-237.

TSOTSOS, J. K. (1988). A 'complexity level' analysis of immediate vision. *International Journal of Computer Vision*, **1**, 303-320.

TSOTSOS, J. K. (1989). The Complexity of Perceptual Search Tasks. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1571-1577). Palo Alto, CA: Morgan Kaufmann.

TSOTSOS, J. K. (1990a). A Complexity Level Analysis of Vision. *Behavioral and Brain Sciences*, **13**, 423-455.

TSOTSOS, J. K. (1990b). A Little Complexity Analysis Goes a Long Way. *Behavioral and Brain Sciences*, **13**, 459-469.

TSOTSOS, J. K. (1991a). Is Complexity Analysis Appropriate for Analyzing Biological Systems? *Behavioral and Brain Sciences*, **14**, 770-773.

TSOTSOS, J. K. (1991b). *Localizing Stimuli in a Sensory Field Using an Inhibitory Attentional Beam* (Technical Report RBCV-TR-91-37). Toronto: University of Toronto, Department of Computer Science.

TSOTSOS, J. K. (1992a). *Behaviorist intelligence and the scaling problem* (Technical Report RBCV-TR-92-42). Toronto: University of Toronto, Department of Computer Science.

TSOTSOS, J. K. (1992b). On the relative complexity of active vs passive visual search. *International Journal of Computer Vision*, **7**, 127-141.

TURING, A. (1937). On computable numbers with an application to the Entscheidungs problem. *Proceedings of the London Mathematical Society*, **2**, 230-265.

UHR, L. (1980). Psychological motivation and underlying concepts. In S. Tanimoto & A. Klinger (Eds.), *Structured computer vision* (pp. 1-30). New York: Academic Press.

VAN ESSEN, D., & ANDERSON, C. H. (1990). Reference frames and dynamic remapping processes in vision. In E. L. Schwartz (Ed.), *Computational Neuroscience* (pp. 278-294). Cambridge, MA: MIT Press. [VSR]

WATSON, J. B. (1919). *Psychology from the Standpoint of a Behaviorist*. Philadelphia: Lippincott.

YASHUHARA, A., (1971). *Recursive Function Theory and Logic*. New York: Academic Press.

ZEKI, S. (1977). Colour coding in the superior temporal sulcus of the rhesus monkey visual cortex. *Philosophical Transactions of the Royal Society of London*, **B197**, 195-223.

ZIPSER, D., & ANDERSON, R. (1988). A back propagation network that simulates response properties of a subset of a posterior parietal neurons. *Nature*, **331**, 676-684. [VSR]