Computational Processes in Human Vision:
An Interdisciplinary Perspective

Edited by Zenon W. Pylyshyn

The Canadian Institute for
Advanced Research Series in
Artificial Intelligence

Ablex Publishing Corp.
Norwood, NJ

1988

pp. 3-26

1
Playing Twenty Questions with Nature*

Whitman Richards
Aaron Bobick

Massachusetts Institute of Technology

The Twenty Questions game played by children has an impressive reputation: in this game, participants rapidly guess an arbitrarily selected object with rather few, well-chosen questions. This same strategy can be used to drive the perceptual process, likewise beginning the search with the intent of deciding whether the object is "animal, vegetable, or mineral." For a perceptual system, however, several simple questions are required even to make this first judgment as to the Kingdom in which the object belongs. Nevertheless, the answers to these first simple questions, or their modular outputs, provide a rich data base which can serve to classify objects or events in much more detail than one might expect, thanks to constraints and laws imposed upon natural processes and things. The questions, then, suggest a useful set of primitive modules for initializing perception.

* Support for this work is provided by NSF and AFOSR under a combined grant for studies in Natural Computation, grant 79-23110-MCS, AFOSR 86-0139 and 83-12240-IST. Comments and criticisms by Shimon Ullman, Donald Hoffman, and John Rubin were greatly appreciated. Technical support was kindly provided by William Gilson. This paper is adapted from MIT A.I. Memo 660, Dec. 1982, entitled "How to Play Twenty Questions with Nature and Win." This title originated in an article by Alan Newell (1973), who pointed out the frustration of posing certain lines of questions for research on information processing, such as serial vs. parallel, peripheral vs. central, conscious vs. unconscious. I believe the Twenty Questions approach presents a worthwhile alternative for such research.

THE NAME OF THE GAME

Perceiving systems are subject to a massive bombardment of signals from the external world. Sometimes these signals are completely unexpected or unpredictable, such as when you hear a novel sound, or when I show you a postcard. Yet from this deluge of unforeseen data, the sound or scene is understood. How is this possible?

One strategy for interpreting unexpected scenes or sounds is to build a very general perceiver—a perceiver built from a hierarchy of modules of increasing complexity and scope. "Points" are aggregated into "blobs," "lines," or "edges." These elementary features then become the basis for more complex representations of shapes and regions, and their relations to one another, which are finally interpreted as "objects." Such a view is the one currently accepted by most. It seems to me a very unsatisfactory one. First the goals of the system are not really specified clearly. Presumably they include object recognition and manipulation. Yet to date no one has been able to offer a general definition of "object" that is precise enough to embody in a computer vision system. We can define special objects for which we have models, such as planes, trees, houses, or people, but not "object" in general. This difficulty currently casts doubts upon our ability to build a general purpose perceiver and raises questions about whether such a system indeed exists. Yet there is no doubt that special purpose devices can be built that match inputs to models. Here, the Twenty Questions Game can be applied profitably.

FUNDAMENTAL HYPOTHESIS (NATURAL MODES)

Before presenting one strategy for building a very general, yet special purpose perceiver, we make an important claim about the world we perceive. The claim is that the structure and events in the world are not arbitrary or random. Rather they can be seen as clustered in a multidimensional space. As a result of natural selection and environmental pressures, nature does not adopt all possible solutions to the problems it encounters (Stebbins & Ayala, 1985; Mayr, 1984). Fish and whales, although biologically quite different species, look similar because this particular body design is quite efficient for locomotion through fluids. Animals are not asymmetric and arbitrary, but are symmetric. Even chaotic systems have modal behaviors (Levi, 1986). Our fundamental hypothesis about the world is thus the following:

> *Principle of Natural Modes:* Structure in the world is not arbitrary and object properties are clustered about modes along dimensions important to the interaction between objects and their environment.

Such modal behavior seems necessary if a perceiver is to be able to categorize correctly structures and events in the world (Marr, 1970; Bobick & Richards, 1986). Of course the scheme used by the perceiver to perform the categorization has not been specified. Nor have the modes which are appropriate (these may differ for different perceivers). Yet it is this structure of the world which will allow us to play the Twenty Questions Game profitably. The basic idea is to choose questions that are keyed to identify the natural modes of the world of interest. Correlations between the answers then permit a "natural" categorization of the event or structure.

FROM TEMPLATES TO QUESTIONS

In simple worlds such as many industrial settings and laboratories, the modalness is quite apparent because the range of objects and views is quite limited and well-defined. An open-end wrench appears on the conveyor belt, or a red "cube" is placed on a table. Because the "object" of interest is known in advance and controlled, simple template matching usually suffices to solve these tasks. This is a very primitive form of the game we propose, where the perceiver's questions are tailored to the world. Examples can also be found in natural environments: the blowfly feeds when its receptors identify the ring structure of a sugar, and rejects the hydrocarbon chains of alcohols (except Inositol, which is an unnatural ring alcohol [Hodgson, 1961]). Other examples are the hungry fledgling gull that responds immediately to the looming red spot on its parents' beak; the mating call of the cricket (or bee), which is so precisely engineered that a simple pattern of pulses can be tailored to reflect even subtle species differences. Such examples are countless (Tinbergen, 1951; Wilson, 1971). In each case, an important primitive goal, such as feeding or the reproduction of the species, is achieved successfully in a very direct and reflexive manner only because the environment is highly structured.

Our strategy is to capitalize on this structure of the world, to build perceptual modules that as directly as possible identify whether a mode is reflected in the sensory data. To begin, we consider the three most obvious natural modes in our environment: structures that are either animal, vegetable, or mineral. That these are indeed separate kingdoms receives much biological support (see Figure 1). Although there are at least two other independent modes (fungi and slime), we consider these not relevant to our categorization goals.

Consider now the classical children's game of Twenty Questions, where the goal is to identify an object. By "object" we mean an entity constructed from properties that exhibit natural mode behavior. (Objects composed of random structures and properties thus lie outside our domain.) Our first questions attempt to identify whether the object is animal, vegetable, or mineral. Subsequent questions attempt to determine the size, shape, or mass, or the sounds "it" might make, how "it" moves, or perhaps its function. The final questions then become very specific and detailed. If we are clever and shrewd in our choices, we rapidly converge to the object. A perceptual system we propose is designed along similar lines. Imagine that for our first set of questions we identify a dozen or two—let's say twenty—very general but independent attributes of "things," where an "attribute" is a particular modal property present in the world. We simply ascertain whether each attribute is present or not. Then 2^{20} or roughly a million different types of events could be crudely categorized (Webster's Dictionary only lists 60,000 words total.) Certainly, such assertions all computed in parallel would form a useful way of initializing the perceptual process, providing an initial description of the events or contents of a scene. Can we indeed find such questions that are powerful and general, yet are simple enough to be computed from the sense data? Let's play a slightly modified version of the Twenty Questions Game to explore its power.

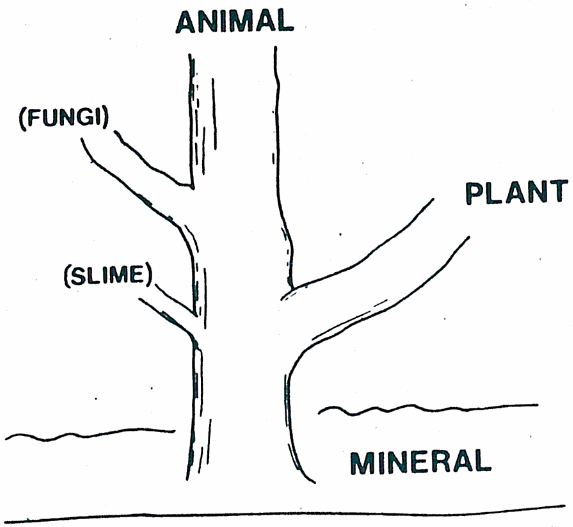

Figure 1. Tree of Life, showing the animal and plant kingdoms. Recently, biologists have added Fungi, Protozoa, and Slime as separate branches. (Woese, 1981).

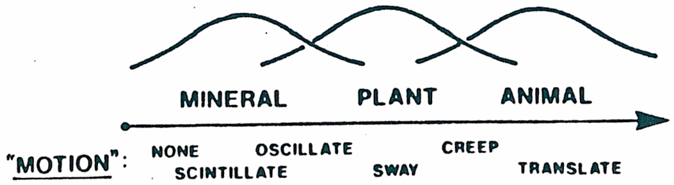

Figure 2. Qualitative types of motion or mobility associated with different kinds of living or inanimate objects, crudely ordered along the "Tree of Life" dimension shown in Figure 1.

PLAYING THE GAME

Imagine that an "object" has just entered our field of view, emitting some distinctive sounds. Our task is to identify as quickly as possible the general nature of the object. Loosely speaking, we would like to distinguish a man from a cat or a bird, but monkeys and men or clouds and smoke may be confused.[1] The principal rule of the game is that all our "questions" must be ones for which the answers can plausibly be computed from the sense data.

In the classical Twenty Questions game, our first question was, "Is it animal (or vegetable or mineral)?" How can we answer this question from the sense data? In fact, there are many ways to determine whether the "event" arose from an animal, vegetable, or mineral. For example, animals translate, rocks or plants do not (Figure 2). Animal sounds are different from the sounds of minerals (running water or falling rocks) or of the wind through the trees. Plants and animals have different shapes or colors; they "feel" different. Many of these attributes can be computed from the sense data using foreseeable technology.

[1] To specify rigorously the precision required of the Twenty Questions game is an important issue, but one which requires a clearer statement of the objectives and goals of the inquirer.

Surprisingly, the answers to the first set of questions posed to determine whether the event is animal, vegetable, or mineral tell us much more than just which of these three categories the event falls into. Consider Game 1 (shown in Appendix I). Our first question—"Is it moving?"—gave the answer *translation*, implying *animal*. The second question yielded the answer "legs"—confirming the *animal* interpretation. Yet the answer to the third question—that the emitted acoustic frequencies are broadband, rather than narrow-band as expected for an animal—causes us to question whether the "event" indeed arises from an animal. In this particular game, which is a transcript of one actually played, eight more questions are required to pinpoint the object. By playing such games, we see the power of an appropriate set of questions. Although the answers are restricted to a choice of triples[2] the collection of such answers is sufficient to narrow down an object or event much more precisely than just whether it is animal, vegetable, or mineral. The animal-vegetable-mineral distinction merely serves as a useful dimension along which values of various properties or attributes can be represented. In some sense, it is a dimension of "stuff" or "behavior." Mineral "stuff," plant "stuff," and animal "stuff" each represent different branches of the Tree of Life (Figure 1). We will see later that these fundamentally different properties will be useful descriptors of attributes outside their kingdom of origin. The utility of the animal-vegetable-mineral dimension for "stuff" thus goes far beyond what is implied by our first game.

CRITERIA FOR TWENTY QUESTIONS

Table 1 summarizes some useful preliminary questions that address various properties of natural things.[3] The first column is the attribute measured or extracted from the raw sense data. The next three columns indicate the initial three output states of the question box or module. The animal-vegetable-mineral categories serve to guide the choice of the type of output assertion to be computed. The final

[2] In practice, a default response may be necessary on occasion. Thus each question requires 2-bits for the answers. More answer categories may be counter-productive if one wishes to create an indexable representation for memory that can be efficiently accessed (Dirlam, 1972).

[3] The list makes no distinction between "shape," "stuff," and "structure," although the strategies for computing these properties are clearly quite different. For example, see Rubin & Richards, 1982; Hoffman & Richards, 1982. Chemical attributes are not included because localization for scene segmentation is usually difficult.

Table 1. Example Questions and the Three General Categories of Their Answers.

	AUDIO-VISUAL			
ATTRIBUTE (Question)	MINERAL	PLANT	ANIMAL	(REF)
acoustic frequency	none or broadband (lo)	broadband (hi)	narrowband	1
acoustic modulation	none	pseudo-sine	interrupted	2
frequency change	no	no	yes	3
motion	none	sway	lateral	4
support	no "leg"	one "leg"	several "legs"	5
symmetry	irregular	3-D (one axis)	mirror (bilateral)	6
axis	none	vertical	horizontal?	7
"texture"	irregular (2-D wideband)	fractal	1-D parallel (hair)	8
"color"	yellow, brown, blue	green, red	agouti	9
	TACTILE			
ATTRIBUTE	MINERAL	PLANT	ANIMAL	(REF)
heat emmission/ absorption	cold	neutral	warm	10
texture	rough	rough and smooth	soft, smooth	11
hardness	rigid	crunchy, crisp	soft, elastic	
			hairy, feathers	12
movement	none	passive (bend)	active (wriggles)	13
adhesion/ viscosity	none (dry or wet)	sticky	oily	14

(fourth) column gives a reference in Appendix II as to how feasible it is to compute these outputs, using current or foreseeable technology.

Our preliminary choice of questions has been guided by several considerations. The first, already mentioned, is the computational feasibility. A second is the degree to which an attribute can encode a useful modal property of a "thing." Here we rely upon our intuitions about which properties are likely to exhibit highly modal behavior. For example, the sound an object makes reflects something about the structure of the source. If the sound is narrowband, then the source must have a tuned resonant cavity, which neither plants nor minerals have. All candidate objects from these two kingdoms can then be rejected—a rather strong assertion (see Rubin & Richards, 1982). Furthermore, because the size of the cavity determines the fundamental frequency of the sound, some indication of the source size can be inferred from the pitch. An elephant "roars" because it has a large resonant cavity whereas the mouse "squeaks" because of necessity it must have a small cavity. The sounds an animal can emit thus depend critically upon its size and therefore encodes size, if one wishes to examine this attribute in more detail. We see immediately that the simple question "What is the *pitch* of the source?" not only may tell us whether the object is animal, plant, or mineral, but also provides some information about its size. Translatory visual motion information can be similarly utilized to indicate animal size, as shown in Figure 3. Such questions about the pitch of a sound or the rate of motion are ideal questions because their refinement provides still more useful and quantitative information about the object. This ability to refine a question to extract more information was a third factor which influenced our selection of questions.

Other selection criteria for our Twenty Questions relate to the perceiver-object relation. Obviously one would like an object representation to be independent of the viewer's position or the particular disposition of the object. Yet most of our sense data seem to depend critically upon our particular view. For example, image intensities on the retina are seriously confounded with the orientation and reflectivity of the surfaces that reflect the light; or auditory intensities will depend upon the source distance and the intermediate absorbing and reflecting media. Is it at all reasonable, then, to hope to find descriptive attributes of objects and "things" that are insensitive to our particular viewpoint or position? Of the five basic physical variables—charge, mass, length, time, and temperature—only time is independent of the observer's position and the medium in which he exists. The best examples of viewer-independent attributes of an event or "thing" will thus be those where the temporal pattern encodes the property. The sparkle of water, the scintillating pattern of fluttering leaves on a tree, the gait of an animal, the chirp of a cricket—all are important characteristics of the "object" whose pattern remains the same regardless of where the perceiver is located. The dynamical environment is thus a critical ingredient of the Twenty Questions game.

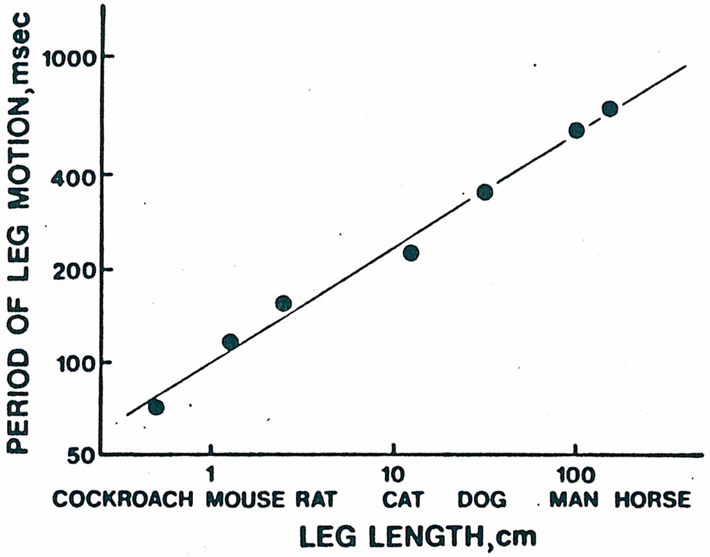

Figure 3. The rate at which the legs move encodes leg length and hence animal size, as shown by the high correlation between size and gait. (Adapted from McMahon, 1975). This desirable property of the motion Question we call "conveyance".

However, it should be stressed that all Questions, including those based on the temporal dimension, are designed to deliver answers about the external world, and do not just report the state of the sensorium. *Has "legs"* does not just mean that there are two or more roughly parallel elongated blobs in the image. The latter are features. (We view feature as an image-based data structure, whereas attributes or properties are assertions about an event or structure in the world external to the observer.) The inference process posed by the Twenty Questions is thus exceedingly difficult, and one which we believe must rely on the modal character of the natural world. (See Bennett, Hoffman, & Prakash, 1988.) However, the positive and very optimistic point we are making is that, given this modal character, only a few well-chosen modules can serve as a basis for a successful perceiver.

To summarize, we have five major criteria for our choice of questions:

1. *Computational Validity*—The representation of the attribute must be easy and reliable to compute.
2. *Conveyance*—One should be able to refine the attribute to yield a more metrical measure of an object property.
3. *Modality*—Different attributes or questions should be capturing different modal qualities of the "events" or "things."
4. *Viewer Independence*—Representations of attributes should be insensitive to the particular relations between the perceiver and the "object", i.e., to object distance, scale, or disposition.
5. *Configuration Independence*—The attributes should be independent of the particular state or configuration of the object.

Computational Validity

Given the above criteria, how do we know when they have been satisfied? Particularly difficult in this regard is the "modality" of the set of questions, to be addressed shortly, and their computational validity. The best evidence for the ability to answer one of the Twenty Questions is an example of a machine system that will deliver the correct answer. The references in Appendix II document the feasibility of designing sensors or information processors that can answer the question posed.

In the audio-visual realm, narrow-band sensors that measure the frequency of the acoustic spectrum have been available for many years (Flanagan, 1972). The measurement of acoustic frequency and intensity changes is thus readily accomplished for isolated sound sources. Not so easily achieved, however, is the isolation of a sound source, although this is a task performed reliably by the most simple natural binaural system (Howard & Templeton, 1966; Knudsen & Konishi, 1979). As long as the environment does not have more than one or two competing sources, the source direction or isolation can be found fairly reliably using either signal onset times or intensity differences, or both (Altes, 1978; Searle, Davis, & Colburn, 1980). Additional work needs to be done in this area, however, for source isolation (and direction) is a critical computation that must precede many of the acoustic questions, especially if one wishes to determine details about the physical properties of the source (i.e., is it metallic, wood, or rustling leaves?), or the nature of animal sounds (Klatt, 1977).

Similarly, for vision, a rather powerful input representation is also required before the Twenty Questions game can proceed with reasonable success. Although lateral motion or scintillation or sway can be computed crudely for a region using only primitive intensity information (Thompson & Barnard, 1981; Ullman, 1981), the exact shape of the region cannot yet be found reliably (Horn & Schunck, 1981; Hildreth, 1982). "Edge"-finding algorithms are still quite primitive, and confuse many types of intensity changes such as surface markings, shadows, or occluding edges. For vision, the most useful data base for the Twenty Questions game would be Marr's primal sketch (Marr, 1976; Marr & Hildreth, 1980), which is still unavailable and poses many quite difficult computational problems. Furthermore, it is still a feature-based representation and does not make assertions about properties in the world. Thus, although questions such as "number of supports" or "symmetry type" seem feasible in the long term (Brady & Asada, 1984; Richards, Dawson, & Whittington, 1986), as yet we do not have a sufficiently powerful "primal sketch" to permit these questions to be answered reliably.

More tractable are questions about the surface properties such as its roughness or composition, although obstacles also occur here. Many sensors are available to measure the spectral composition of reflected light, but we must remember that a reliable determination of the spectral reflectance of a surface also requires knowledge of the source illumination. Fortunately, this is rather constant in natural environments, and our crude color question is computationally feasible (Judd & Wysecki, 1975; Myrabo, Lillesaeter, & Hoimyr, 1982). Remote measures for surface roughness or quality, on the other hand, are still rather primitive and far from robust, although several recent studies, particularly in the remote sensing area, show promise of providing practical applications (Moon & Spencer, 1980; Milana, 1981). Tactile sensing, on the other hand, appears quite tractable, with several impressive recent advances in detecting surface properties (Hillis, 1982; Raibert & Tanner, 1982).

In sum, it is still uncertain the extent to which the technology of the near future can give reliable answers to all the posed questions. Those that concern "shape" appear particularly difficult, whereas those that address the "stuff," composition, or size of the object seem more tractable. The challenge is obvious.[4]

[4] In many cases the property-based questions can not be entirely decoupled from the shape descriptors, at least for vision. For example, many grouping tasks for connecting isolated contour segments may require that a property tag be attached to the contour descriptor (such as its codon type). This requirement complicates the integrated structure of the set of Twenty Questions, but does not obviate the need for them.

Modality

We have criteria and constraints on the types of questions we should ask, but we still have not found a rule or procedure that tests whether our questions are independent in the sense that they capture different modal properties in the world. At best, we have suggested that the behaviors or properties cf objects within each of the three kingdoms will differ, yet this is clearly not the case in practice. Very often a property, such as a hard "shell" (rock), or soft "feathers" (grass) may appear in more than one kingdom.

The problem of orthogonality and modalness is further complicated by the wide scale of sizes over which objects and events may exist—from the amoeba to the dinosaur; from the blade of grass to the giant Sequoia, or from the tiny grain of sand or speck of dust to Mount Everest. This enormous range of scales has led to the application of different natural laws to solve similar problems. The amoeba locomotes one way, the elephant another; the speck of dust behaves differently from a massive stone when subject to the wind or forces of nature. At any one scale, however, where size and mass are comparable, the behaviors are similar, at least to the degree that the "stuff" is the same. As the "stuff" differs, then the behavior will differ. Hence, the nature of the "stuff" becomes a dimension along which different behaviors or attributes may be categorized at any one scale. The log placed on water acts differently from stone because its "stuff" differs. The animal-plant-mineral distinction is thus basically a crude dimension to a property list. To the extent that the properties are independent, the questions will be independent. We appeal to the process of natural selection to converge upon an optimal set of questions that captures these different properties.

SUCCESSES AND FAILURES

The strategies and remaining problems encountered with the Twenty Question approach become more apparent as the game is played. Ideally, one would like to have available a massive dictionary against which the game could be played on a computer. In this way, the "top-down" and "bottom-up" inferences might be made more explicit, while at the same time, the evolution of the best questions (and their priorities) could be examined. In lieu of this, Appendix I presents two sample games to show what inferences may be (or are) drawn from successive questions when the game is played serially. (Of course, any biological implementation may elect to ask many of the questions in parallel.)[5]

Several problems become immediately apparent when one plays the game. For example, often one can be badly misled by the first or second question. If the answer to "motion" is "none," obviously one cannot immediately infer that the thing is not an "animal," for it may be a stationary animal, lying down. Similarly an animal in such a state will seem to have "no legs" and will emit no sounds. Clearly our deductions will be way off in this case. Have we therefore missed the mark?

Once again, we must consider the rather primitive goals of the Twenty Questions game: namely, to provide a crude classification of "things," often as they bear upon our survival. Certainly if the animal is not moving, then its immediate threat as a predator is less than if it is looming toward us. Given the alternatives, ones attention is focused upon the most active events in the environment.

Finally, the dimensions and attributes of our Twenty Questions have been driven by the natural, biological environment. The man-made world is quite different. In some sense, its qualities, although largely made of mineral "stuff," may extend the mineral-plant-animal dimension further to the right. Automobiles or planes translate more swiftly; their bodies are more resilient and "metallic." Yet what natural animals possess these same qualities? If there are none, then our original Twenty Questions strategy can still be applied successfully even in the world of man-made objects.

APPENDIX I: EXAMPLE GAMES[*]

"I'm thinking of an object. It is in its natural habitat (which is the same as yours), and is behaving in its most natural way. What is the object?

"The only questions you are allowed to ask and receive answers to are those which could be used by a rather simple sensory device, that is, one which is feasible to build today. For simplicity, the de-

[5] We must be careful about comparing the performance of a serial Twenty Questions Game (Siegler, 1977) with that obtained with parallel questioning. In the former, the earlier questions influence the context applied to succeeding questions whereas answers obtained in parallel share the same context. Here we have totally neglected the control problems in playing the Twenty Questions game—but for some discussion of this problem, see Bobick & Richards, 1986.

[*] NOTE: The original A.I. Memo also included a set of questions that determined the habitat.

vice will have only three outputs (plus a default if no firm answer is possible).

"Each of the output states indicates a different quality of the "thing" or dimension relevant to your question. For example, if you ask "Is it moving?," the relevant dimension is whether it behaves like an animal, plant or mineral, in which case it will either translate, sway, or not move at all.

"There are three main dimensions that you may use to frame your questions. One characterizes the basic biological structure from mineral to plant to animal. The second dimension pertains to the habitat or environment, ranging from arctic to temperate to tropic. A third dimension captures a different aspect of the location of the "thing" in the environment, namely, is it in the air, or the ground, or subterranean—below ground or under water."

Game 1

Habitat: (previously determined to be temperate environment, green rolling hills. Elevation of "thing" is on the ground.)

	ANIMAL	PLANT	MINERAL
Q1: Is it moving?	*translates*	sway	no
Q2: How many supports?	2, 4 or > 4	1	0
Implication: Confirms animal—has four "legs".			
Q3: What acoustic frequencies are emitted?	narrowband	*broad*	*broad*
Implication: Disconfirms animal. "Thing" makes low frequency, broad-band sounds, moves and has 4 legs. Must be big. Elephant or cow?			
Q4: Acoustic source	*point*	extended	extended
Implication: Confirms "animal" or isolated object.			
Q5: Visually symmetric?	*mirror*	3D	irregular
Q6: What major axis?	*horizontal*	vertical	none
Implication: Still seems to be some kind of large animal with horizontal major axis..			
Q7: Modulation of acoustic intensity?	interrupted	pseudo-sine	*none*
Implication: A large, moving animal with horizontal major axis that continually emits a steady broadband sound.			
Q8: Color?	agouti	green, red	yellow, brown, *blue*
Implication: "Animal" is blue. This is unlikely.			
Q9: Texture?	1-D parallel	fractal	irregular

Answer: None of the above. (Note that with two bits for answers, we have room for the default category.)

Q10: Hardness?	soft, elastic	crunchy, crisp	*rigid*

Implication: Large animal with horizontal axis that moves on ground and emits a steady sound; surface is blue and hard like a "mineral," but the texture is not hairy or irregular. Car?

Q11: (Scale of dimension) What is rate of leg motion?

Answer: Zero

Implication: Object has no legs, but moves (on wheels?). Confirms a car.

Game 2

	ANIMAL	PLANT	MINERAL
Q1: Is it moving?	*translates*	sway	*no*
Q2: What acoustic frequencies are emitted?	narrow-band	broad	broad
		(None of the above.)	
Q3: How many supports?	2, 4 or > 4	1	*0*

Implication: Animal at rest or a mineral.

Q4: Visually symmetric?	*mirror*	3-D	irregular
Q5: Texture?	fine	*smooth*	rough

Implication: Neither an animal nor mineral.

Q6: Hardness?	soft	crunchy	*rigid*
Q7: Color?	brown	green, red	*yellow-white-blue*

Implication: Hard, whitish-blue, mirror symmetric object with a smooth surface that lies flat on ground without support and makes (is making) no sound. (A round, white, smooth rock?)

Q8: What is its elevation?

Answer: Subterranean

Q9: What is its immediate environment?	solid	*soft*	*liquid*

Implication: Object is in moist soil and partially submerged under water. (As if in a pond or lake or ocean?) Oyster, clam or snail?

Q10: What is its approximate size?
Answer: Slightly smaller than a man's hand.

Confirms oyster or clam.

APPENDIX II

1. *Acoustic Frequency*. Comb filtering has been used for several years to separate sound sources (Shields, 1970; Flanagan, 1972; Zwicker & Terhardt, 1979). Unless many broadband sources are active simultaneously at S.P.L.'s comparable to the narrow-band sources, this question can be answered with available technology (Klatt, 1977). As initially formulated (Richards, 1980), the question simply addresses whether the source is broadband or not (such as wind through the trees, rushing water, or an animal cry). Much more useful, but also much more difficult, would be to extract the physical properties of the source—i.e., its acoustic "color": Is it metallic, wood striking wood, or a footfall?
2. *Acoustic Modulation*. Tracking a sound source to determine its modulation characteristics (Atal, 1972) also requires localization (as may Question #1). For narrow-band, harmonic sources with different spectral signatures, such localization is possible provided there are only a few competing sources (Altes, 1978). Again, as in Question #1, work should be undertaken to understand how the "textural" properties of the source can be extracted from the modulations. For example, is the source "harsh" or grating, or like clacking sticks, or "suave" and "smooth," or "roaring" like a brook or lion.
3. *Frequency Change*. Here again, as in Questions #1 and #2, localization is helpful but not as necessary because only Animals are generally capable of producing sounds of variable frequency. Simple ⅓ octave filtering should allow the detection of frequency change (Flanagan, 1972; Klatt, 1977.)
4. *Motion*. The motion of an "object" can be both visual and auditory. Clearly the detection of auditory movement requires localization (Altes, 1978; Searle *et al.*, 1980), and may be difficult. Visual motion detection has progressed enormously over the past 10 years, and can be detected with simple systems provided the background is stationary (Horn & Schunck, 1981; Thompson & Barnard 1981; Ullman, 1981; Hildreth, 1982). More work is still required, however, to use motion to segregate a visual scene, especially if sway or scintillation is to be disambiguated from translation or rotation.

5. *Support*. Although a powerful question, to estimate the numbers of "legs" supporting a region is quite complicated. First, the ground plane must be determined (see Question #20); secondly the candidate "support" must be recognized (e.g., leg or trunk); and finally a region should be identified as being supported although it may have a different color or texture. In the case of stationary supports, the local parallelism of the vertical occluding edges of the support may serve as a basis for determining the supporting member (Stevens, 1980). What to do in the case of animal motion, however? Also, shrubs clearly may have many "supports." The computational validity of this attribute is questionable, therefore, although a strong assertion would be quite useful.
6. *Symmetry*. Given that the occluding contour can be determined from an image, then mirror symmetry can be answered from available technology (Kanade, 1981; Hoffman & Richards, 1982). To determine 3D symmetry also requires a depth map, which is computable if binocular vision is available (Grimson, 1981). The difficult part of this question, therefore, is extracting the occluding contours, which at present can be done only on restricted classes of images (Davis & Rosenfeld, 1981; Binford, 1981; Richards, Nishihara & Dawson, 1982).
7. *Axis*. Again, as in Question #6, the orientation of a region can be answered rather easily (Ballard & Brown, 1982, Witkin, 1981) provided either that the occluding contour can be found, or the approximate areal extent of the region can be determined, such as by its spectral or textural qualities.
8. *"Texture"*. The intent of this question is to determine whether the surface property of the region is typical of rocks or metals, grass or shrubs, or animal skin, hair, or feathers. Schemes for disambiguating such surface properties have only recently been considered (Horn, 1977; Milana, 1981; Moon & Spencer, 1980; Cook & Torrance, 1982; Rubin & Richards, 1982) This is an area ripe for research.
9. *"Color"*. The value of spectral information in assessing food quality (Francis & Clydedale, 1975), printing inks or pho-

tographic reproductions (Judd & Wÿsecki, 1975) and in remote sensing (Chance & Lemaster, 1977; Lintz & Simonett, 1976; Myrabo et al., 1982) have provided a variety of practical tools.

10. *Heat Emission/Absorption*. The determination of surface temperature relative to one's own body temperature is a simple sensory ability if contact is used (Hertzfeld, 1962). Of course remote sensing is also possible here, as performed in surveillance or Landsat imagery (Lintz & Simonett, 1976; Barbe, 1979; Trivedi, Wyatt, & Anderson, 1982).
11. *Texture*. Passive touch sensing is coming close to obtaining the resolution required to determine surface roughness, as well as the texture pattern of the surface. At present, grid resolutions of 16 × 16 per cm^2 have been obtained (Hillis, 1982; Raibert & Tanner, 1982).
12. *Hardness*. The measurement of hardness of a point on a surface is a routine metallurgical technique and is trivial (Cox & Baron, 1955; O'Neill, 1962). The difficult task is to devise a skin-like sensor for the rigidity using force-feedback and the pattern of deformation. Recent progress in touch-sensing suggests that such sensors may be forthcoming in a few years, with possible applications for testing food ripeness (Kato, Kudo, & Ichimaru, 1977; Harmon, 1982).
13. *Movement*. The Hillis (1982) touch sensor could, in principal, be redesigned to measure whether a grasped object is wriggling or breathing. Whitney (1979) and Harmon (1982) also provide reviews describing the spectrum of compliant sensors now available.
14. *Adhesion/Viscosity*. Although a variety of rheometers are available to measure the viscocity and flow of fluids and gases (Van Wager, 1963), I do not know of a skin-like sensor that measures "stickiness" or "oiliness." Again, compliant sensors in this area, although perhaps relatively straightforward compared to remote sensing, will probably await commercial needs.

REFERENCES

Agin, G. (1972). Representational description of curved objects. Stanford AI Project Memo AIM-173, Stanford University.

Altes, R. A. (1978). Angle estimation and binaural processing in animal echolocation. *Journal of Acoustical Society America, 63*, 155–183.

Atal, B. S. (1972). Automatic speaker recognition based on pitch contours. *Journal of Accoustical Society of America, 52*, 1687–1697.

Ballard, D. H., & Brown, C. M. (1982). *Computer vision*. New Jersey: Prentice-Hall.

Barbe, D. F. (1979). Smart sensors. *Proc. Soc. Photo-Opt. Instrum. Eng., 178*.

Bennett, B., Hoffman, D., & Prakash, K. (1988). *Observer Theory*. Cambridge, MA: MIT Press.

Binford, T. (1981). Inferring surfaces from images. *Artificial Intelligence, 17*, 205–244.

Bobick, A., & Richards, W. (1986). Classifying objects from visual information. MIT AI Memo 879. Cambridge, MA: Massachusetts Institute of Technology.

Brady, M., & Asada, H. (1984). Smooth local symmetries and their implementation. *International Journal of Robotics, 3*, 36–61.

Chance, J. E., & LeMaster, E. W. (1977). Suits reflectance models for wheat and cotton: theoretical and experimental tests. *Applied Optics, 16*, 407–412.

Cook, R. L., & Torrance, K. E. (1982). A reflectance model for computer graphics. *ACM Transactions on Graphics, 1*, 7–24.

Cox, C. P., & Baron, M. (1955). A variability study of firmness in cheese using the ball-compressor test. *Journal of Diary Research, 22*, 386–390.

Davis, L., & Rosenfeld, A. (1981). Computing processes for low level usage: A survey. *Artificial Intelligence, 17*, 245–263.

Dirlam, D. K. (1972). Most efficient chunk sizes. *Cognitive Psychology, 3*, 355–359.

Flanagan, J. L. (1972). *Speech analysis: Synthesis and perception*. Berlin: Springer-Verlag.

Francis, F. J., & Clydedale, F. M. (1975). *Food colorimetry: Theory and applications*. Westport, CN: AVI Publishing.

Grimson, W. E. L. (1981). *From images to surfaces*. Cambridge, MA: MIT Press.

Harmon, L. (1982). Automated tactile sensing. *International Journal of Robotics Research, 1*(2):3–32.

Hertzfeld, C. M. (1962). *Temperature, its measurement and control in Science and Technology. Vol. 3*. Reinhold, N.Y.

Hildreth, E. (1982). The integration of motion information along contours. IEEE Proceedings of a Conference of Computer Vision Representation and Control, September, pp. 83–91.

Hillis, W. D. (1982). A high resolution image touch sensor. *International Journal of Robotics Research, 1*(2), 33–44.

Hodgson, E. S. (1961). Taste receptors. *Sci. Amer., 204*, 135–144.

Hoffman, D. D., & Richards, W. A. (1982). Representing smooth plane curves for recognition: Implications for figure-ground reversal. Proceedings of the National Conference on Artificial Intelligence, August 18–20, and MIT AI Memo 630, Cambridge, MA: Massachusetts Institute of Technology.

Horn, B. K. P. (1977). Image intensity understanding. *Artificial Intelligence, 8*, 201–231, and MIT AI Memo 335, Cambridge, MA: Massachusetts Institute of Technology.

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence, 17*, 185–203.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London: Wiley.

Judd, D. B., & Wysecki, G. (1975). *Color in business and industry*. New York: Wiley.

Kanade, T. (1981). Recovery of the three-dimensional shape of an object from a simple view. *Artificial Intelligence, 17*, 409–460.

Kato, I., Kudo, Y., & Ichimaru, I. (1977). Artificial softness sensing—An automatic apparatus for measuring viscoelasticity. *Mechanism and Machine Theory, 12*, 11–26.

Klatt, D. (1977). Review of the ARPA speech understanding project. *Journal of Acoustical Society of America, 62*, 1345–1367.

Knudsen, E. I., & Konishi, M. (1979). Mechanisms of sound localization in the barn owl. *Journal of Comparative Physiology, 133*, 13–21.

Levi, B. (1986). New global formalism describes paths to turbulence. *Physics Today, 39*(4), 17–18.

Lintz, J., & Simonett, D. S. (1976). *Remote Sensing of the Environment*. Reading, MA: Addison-Wesley.

Marr, D. (1970). A theory of cerebral neocortex. *Proceedings of the Royal Society of London B, 176*, 161–234.

Marr, D. (1976). Early processing of visual information. *Phil. Trans. R. Soc. Lond. B., 275*, 483–524.

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

Marr, D., & Hildreth, E. (1980). A theory of edge detection. *Proceedings of the Royal Society of London, B, 207*, 187–217.

Mayr, E. (1984). Species concepts and their applications. In E. Sober (Ed.), *Conceptual issues in evolutionary biology: An anthology* pp. 531–541. Cambridge, MA: MIT Press.

McMahon, T. A. (1975). Using body size to understand the structural design of animals: Quadrupedal locomotion. *Journal of Applied Physiology*, 1975, 619–627.

Milana, E. (1981). Apparatus for testing surface roughness. *Patent Nos. 4, 290, 698*.

Moon, P., & Spencer, D. E. (1980). An empirical representation of reflection from rough surfaces. *Journal of IES, 9*, 88–91.

Myrabo, H. K., Lillesaeter, O., & Hoimyr, T. (1982). Portable field spectrometer for reflectance measurements. *Applied Optics, 21*, 2855–2858.

Newell, A. (1973). You can't play 20 Questions with nature and win: Projective comments on the papers of this symposium. In William Chase (Ed.), *Visual information processing*. New York: Academic Press.

O'Neill, H. (1962). *Hardness measurement of metals and alloys*. London: Chapman and Hall.

Raibert, M., & Sutherland, I. (1983). Machines that walk. *Scientific American, 248*, 44–53.

Raibert, M. H., & Tanner, J. E. (1982). Design and implementation of a VLSI tactile sensing computer. *International Journal of Robotics Research, 1*, 3–18.

Richards, W., Dawson, B., & Whittington, D. (1986). Encoding contour shape by curvature extrema. *Jrl. Opt. Soc. Am. A*, Series 2, Vol. 3.

Richards, W., Nishihara, H. K., & Dawson, B. (1982). Cartoon: a biologically motivated edge detection algorithm. AI Memo 668, Mass. Inst. of Tech. Artificial Intelligence Laboratory, Cambridge, MA.

Richards, W. (1980). Natural computation: Filling a perceptual void. Presented at the 10th Annual Conference on Modelling and Simulation. April 25–27, 1979, University of Pittsburgh. Proceedings, N. G. Vogt & M. H. Mickle (eds.), 10:193–200.

Rubin, J. M., & Richards, W. (1982). Color vision and image intensities: When are changes material? *Biological Cybernetics, 45*, 215–226.

Schunck, B. G., & Horn, B. K. P. (1981). Constraints on optical flow computation. *Proceedings of IEEE Conference on Pattern Recognition and Image Processing*, August, 1981, 205–210.

Searle, C. L., Davis, M. F., & Colburn, H. S. (1980). Model for auditory localization. *Journal of Acoustical Society of America, 60*, 1164–1175.

Shields, V. C. (1970). *Separation of Added Signals by Digital Comb Filtering*. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Siegler, R. S. (1977). The twenty-question game as a form of problem solving. *Child Development, 48*, 395–403.

Stebbins, G. L., & Ayala, F. J. (1985). Evolution of Darwinsim. *Scientific American, 253*(1), 74.

Stevens, K. (1978). Computation of locally parallel structure. *Biological Cybernetics, 29*, 19–28.

Stevens, K. (1980). Surface perception by local analysis of texture and contour. MIT AI Technical Report 512, Cambridge, MA: Massachusetts Institute of Technology.

Thompson, W. B., & Barnard, S. T. (1981). Lower-level estimation and interpretation of visual motion. *IEEE Computer*, 14:20–28.

Tinbergen, N. (1951). *The study of instinct*. Oxford: Clarendon Press.

Trivedi, M. M., Wyatt, C. L., & Anderson, D. R. (1982). A multispectral approach to remote detection of deer. *Photogrammetric Engineering and Remote Sensing, 48*, 1879–1889.

Ullman, S. (1981). Analysis of visual motion by biological and computer systems. *IEEE Computer Magazine, 14*(8), 57–69.

Van Wager, J. R. (1963). *Viscosity and flow measurement; A laboratory handbook of rheology*. New York: Interscience.

Whitney, D. E. (1979). Discrete parts assembly automation—An overview. *Trans. ASAT, 101*, 8–15.

Wilson, E. O. (1971). *The insect societies*. Cambridge, MA: Belknap Press.

Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence, 17*, 17–45.

Woese, C. R. (1981). Archaebacteria. *Scientific American, 244*, 98–122.

Zwicker, E., & Terhardt, E. (1979). Automatic speech recognition using psychoacoustic methods. *Journal of Acoustical Society of America, 65*, 487–498.

2
Aspects of Visual Texture Discrimination*

Rick Gurnsey

Department of Psychology
&
Department of Computer Science
University of Western Ontario

Roger A. Browse

Department of Psychology
&
Department of Computer and Information Science
Queen's University

On the simplest view, textural segmentation may be characterized as involving (a) the measurement of certain image properties and (b) detection of differences in these properties between neighbouring regions. Apparently following this simple view, questions have been asked about which image properties permit effortless discrimination of textures by humans. We show results which indicate that several other factors must be taken into account in a satisfactory theory of textural segmentation. As might be expected, the probability of discriminating two textures depends on how long subjects have to examine the textural display. Contrary to what might be expected, however, texture discrimination is not symmetric; the probability of discriminating two textures, in many cases, depends on which forms the foreground (disparate) region and which forms the background. It would seem then, that discrimination cannot be based entirely on local differences between textured regions; a difference signal should be indifferent to the "sense" of the boundary. Furthermore, the ability to discriminate two textures depends on the amount of practice that subjects have had with the materials and procedure. In general, with naive subjects, the probability of discriminating two textures and the probability of a discrimi-

* This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (A2427). We are grateful to Dr. Brian Butler and Marion Rodrigues for their comments.