MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 660

December 1982

How To Play Twenty Questions With Nature and Win*

Whitman Richards

Abstract. The 20 Questions Game played by children has an impressive record of rapidly guessing 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 judgement as to the Kingdom 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.

Acknowledgement. This report describes research done at the Department of Psychology and the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for this work is provided by NSF and AFOSR under a combined grant for studies in Natural Computation, grant 79–23110–MCS. Comments and criticisms by Shimon Ullman, Donald Hoffman, John Rubin and Aaron Bobick were greatly appreciated. Technical support was kindly provided by William Gilson; artwork was the creation of Julie Sandell and K. van Buskirk.

*This title is adapted from 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 my 20 Questions present a worthwhile alternative for such research.

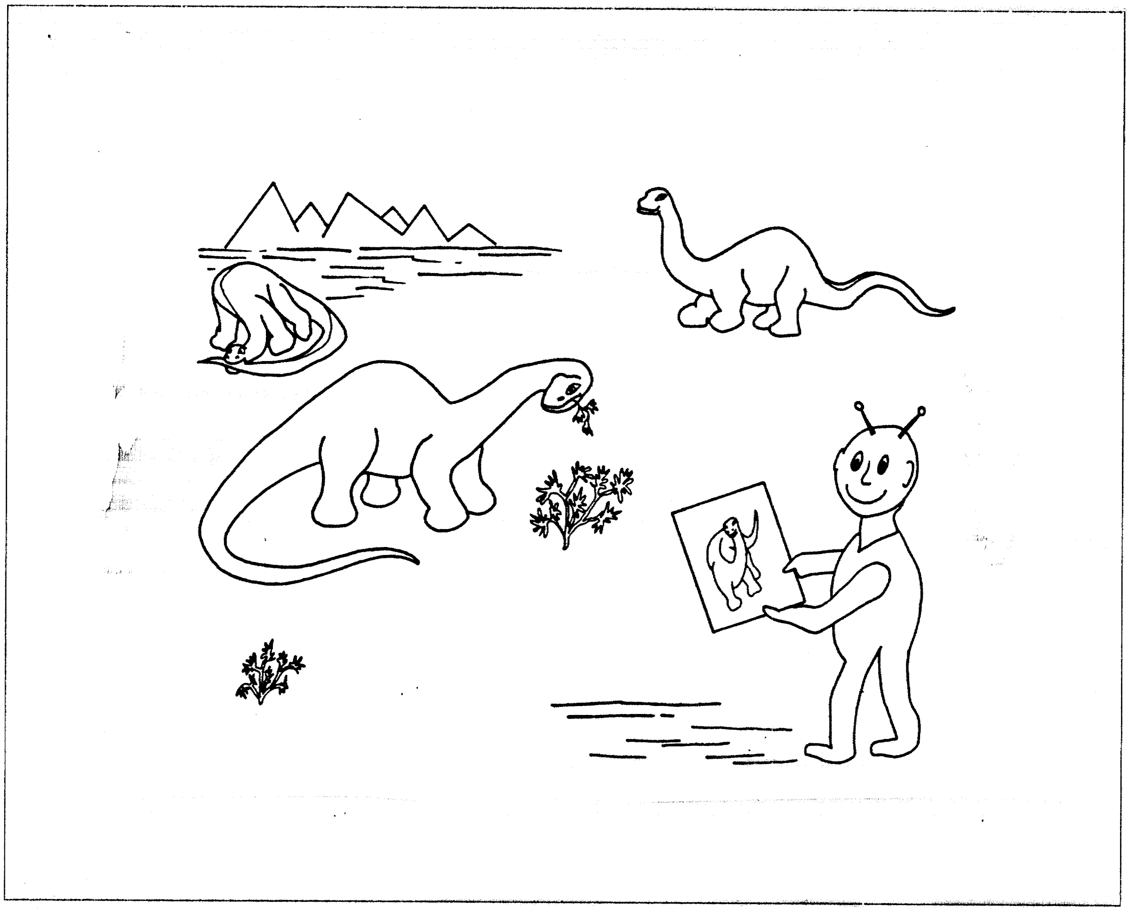

Figure 1 "These can't be dinosaurs. None of them match this picture!"

1.0 The Name of the Game

Perceiving systems are subject to a massive bombardment of signals from the external world. From this deluge of data, useful bits and pieces of information are abstracted from which intelligent decisions can be made. These information abstraction processes cannot be completely arbitrary. They clearly depend upon the goals of the system, its environment, and often upon certain expectations.

In simple environments such as many industrial settings and laboratories, the goal of the perceiver is usually quite limited and well-defined: find the "red" cube on the table, or the open-end wrench on the conveyor belt. Because the "object" of interest is known and expected in advance, simple "template" matching often suffices to solve these tasks. Examples of template-matching 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]!). Or 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 sucessfully in a very direct and reflexive manner only because the environment is limited or well controlled.

Yet how can such a simple template-matching strategy serve a more sophisticated being, who lives in a complex, changing environment? Here, surprises may often be the rule. When we look out a window, walk into an unfamiliar building, or simply view a novel picture or postcard, we have no difficulty in grasping the meaning or context of the scene. The greater our perceptual repertoire, the larger is the spectrum of the unexpected and the variety of "things" that must be recognized and dealt with, often out of the immediate context or frame of mind. Simple template matching to prestored models then becomes impossible, for there are just too many possibilities. Even for one simple item — let's say a dinosaur — the possible views and configurations is usually an infinity in itself (Fig. 1). Without some method of initializing the perceptual system, it must founder as a perceptron will (Minsky and Papert, 1965). What is needed at the outset are some low level representations or assertions that are powerful enough to capture the essence of the "event" or "thing", yet are readily and routinely computable from the raw sense data. These primitive, low level assertions will constitute the answers to our 20 Questions. What inquiries then should we ask? Under what conditions can we expect such a set of questions to provide a useful set of answers?

2.0 From Templates to Questions

In the case of lower animals which react to certain stimuli in essentially a reflexive manner, the system is preprogrammed to recognize a simple pattern. The presence of this pattern is almost guaranteed to represent an "event" or "thing" of importance to the animal. The pattern is thus an attribute uniquely associated with the event of interest, given the expected context. The red spot on the beak of the gull suffices for the fledgling gull because from its nest it will almost never encounter other instances of looming red spots — such as traffic lights or red balloons. In this case a simple template-matching strategy works well because of the controlled context. A simple question suffices to make reliable assertions about a complex event, namely that a parent has arrived, presumably with food.

The situation becomes considerably more complicated, however, for a general purpose perceptual system that must respond intelligently to a wide range of events in a variety of contexts. We cannot hope to find attributes or features unique for each event of interest and for each possible contextual situation. How then can we even hope to find simple questions that will have the same power as the red spot on the gull's beak? The proposed solution is to choose the questions carefully so they inquire about the more general properties of all things regardless of context.

Consider the classical children's game of 20 Questions, where the goal is to identify an object. The first questions usually attempt to identify the general class of the object. Is it 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. Why can't a perceptual system be 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". 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 20 questions game to explore its power.

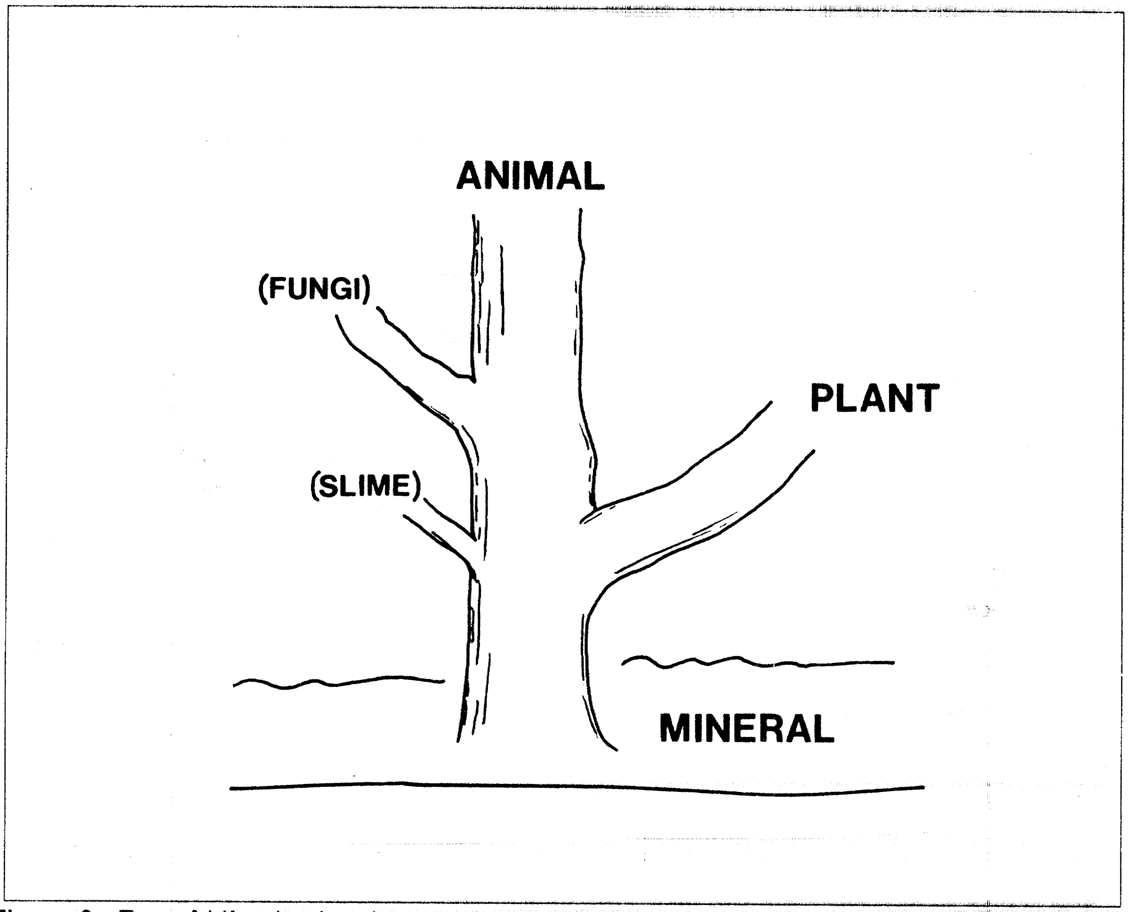

Figure 2 Tree of Life, showing the animal and plant kingdoms (more recently biologists have added Fungi, Protozoa and Slime as separate branches — Woese, 1981).

3.0 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 20 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 (Fig. 3). 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.

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 4 "legs" — confirming the animal interpretation. Yet the answer to the third

[1]To specify rigorously the precision required of the 20 Questions Game is an important issue, but one which requires a clearer statement of the objectives and goals of the inquirer.

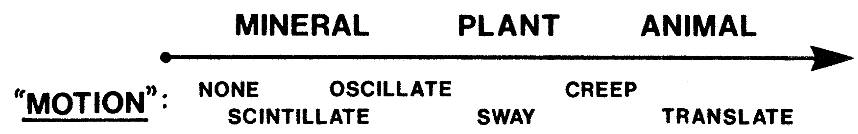

Figure 3 Qualitative types of motion or mobility associated with different kinds of living or inanimate objects, crudely ordered along the "Tree of Life" dimension shown in Fig. 2.

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 (Fig. 2). We will see later that these fundamentally different properties will be useful descriptors of features 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.

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

[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 and Richards, 1982; Hoffman and Richards, 1982. Chemical attributes are not included because localization for scene segmentation is usually difficult.

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 emission/ 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

Figure 4 Table I. Example Questions and the three general categories of their answers.

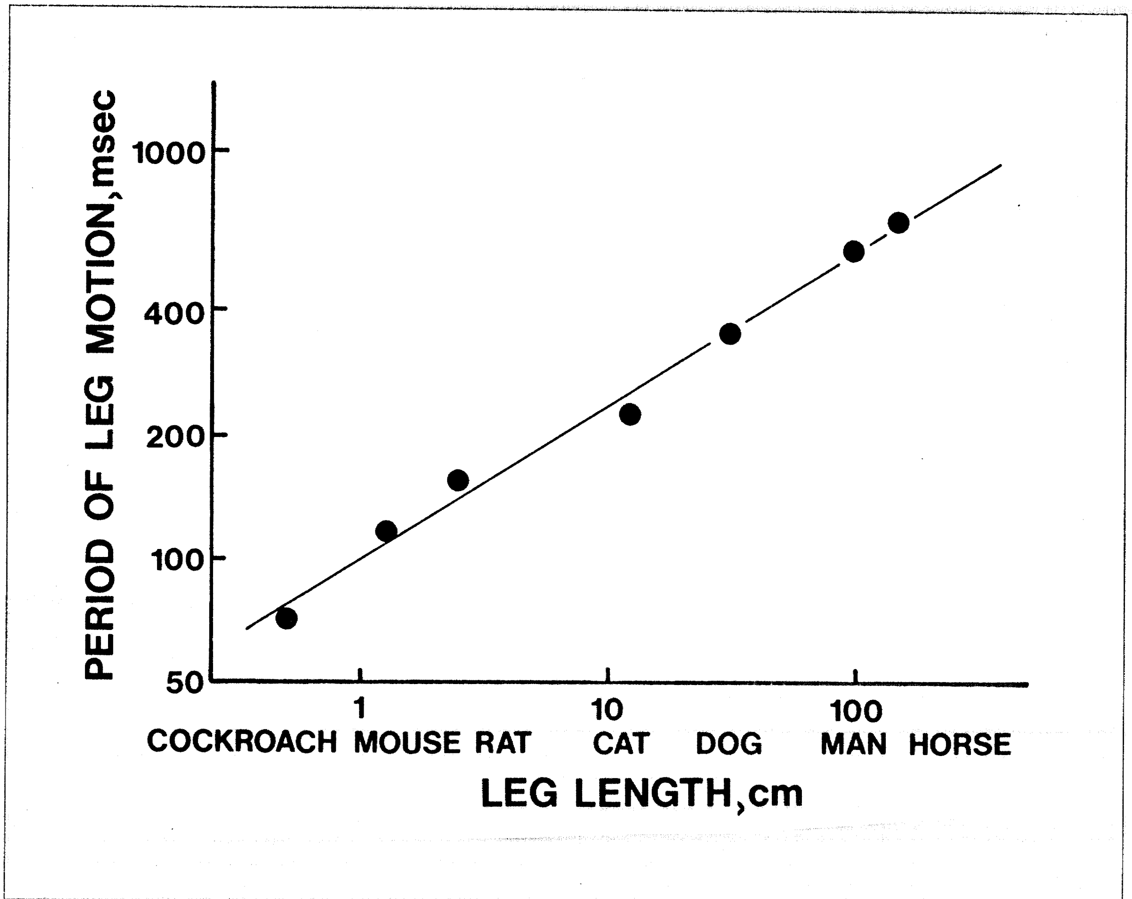

Figure 5 The rate at which the legs move encodes leg length and hence animal size, as shown by the high correlation between size and gait. (Adapted from McMahon, 1975).

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 property of a "thing", such as its size, shape or mass. Often, the very nature of the world or the goals of living things provide strong constraints upon attributes and the information they convey. For example, the sound an object makes reflects something about the structure of the source. If the sound is narrow-band, 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 and 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 encode its size. 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 Fig. 5. Such questions about the pitch of a sound or the rate of motion are ideal questions because the answer encodes a very relevant yet general property of the event.

However, many attributes of "things" are clearly not suitable for our 20 Questions Game. One of the most obvious is 3D shape. To create a 3D model that provides a canonical description of a "thing" is an extremely difficult computational problem (Marr and Nishihara, 1979; Bajcsy and Badler, 1982). Too many restrictive assumptions and intermediate constructs are required for such canonical representations. To play the Twenty

Questions Game safely — to the extent that one can bet one's life on the answers — it is necessary to make accurate inferences as directly as possible from the available sense data. This is the lesson of the innate-releasing mechanisms or "templates" of the more primitive animals.

Yet the basic idea of representing a "thing" by a canonical description is critical. Such a representation will be independent of the viewer's position or the particular disposition of the object, and hence will be a property of the "thing" itself. An important selection criteria for any one of our 20 Questions is thus that it be independent of the perceiver-object relation. Yet most of our immediate 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. When such temporally-varying patterns are emitted, whether they be visual, auditory, or tactile, they generally remain the same regardless of distance or disposition. (This is why most communications schemes encode information in a temporal pattern.) 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 dynamic environment is thus a critical ingredient of the 20 Questions Game.

In sum, we now have four major criteria for our choice of questions:

(i) *Computational Validity* - The representation of the attribute must be easy and reliable to compute.

(ii) *Conveyance* - The attribute should encode a general property of object (such as size, mass, etc).

(iii) *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.

(iv) *Orthogonality* - Different attributes or questions should be capturing independent qualities of the "events" or "things".

4.1 Computational Validity

Given the above criteria, how do we know when they have been satisfied? Particularly difficult in this regard is the orthogonality of the set of questions, to be addressed shortly, and their computational validity. The best evidence for the ability to answer one of the 20 Questions is an example of a machine system that will deliver the correct answer. The references in the last column of Tables I and II document the feasibility of designing sensors or information processors that can answer the question posed.

In several cases, where simple physical variables such as temperature, humidity, or hardness or soil composition are to be measured, many sensors are currently available. In fact, technology has become so advanced that many physical properties of a surface can be measured at a distance, rather than by "touching" as required by many of our questions. The most obvious example is surface temperature using infra-red detectors.

However, humidity and surface roughness may be added to this list (Sabins, 1978; Milana, 1981).

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 and Templeton, 1966; Knudsen and 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 et al., 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 it is desired 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 20 Question 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 and Barnard, 1981; Ullman, 1981), the exact shape of the region cannot yet be found reliably (Horn and 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 20 Question Game would be Marr's primal sketch (Marr, 1976; Marr and Hildreth, 1980), which is still unavailable and poses many quite difficult computational problems. In the meantime, there is some merit to focusing on the recovery of occluding edges, but this cannot be done reliably without creating a sparse, rather disconnected representation of edges that must be linked or grouped (Richards, et al., 1982). Thus, although questions such as "number of supports" or "symmetry type" seem feasible in the near term (Hoffman and Richards, 1982), 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 and Wysecki, 1975; Myrabo et al., 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 and 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 and 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 20 Questions, but does not obviate the need for them.

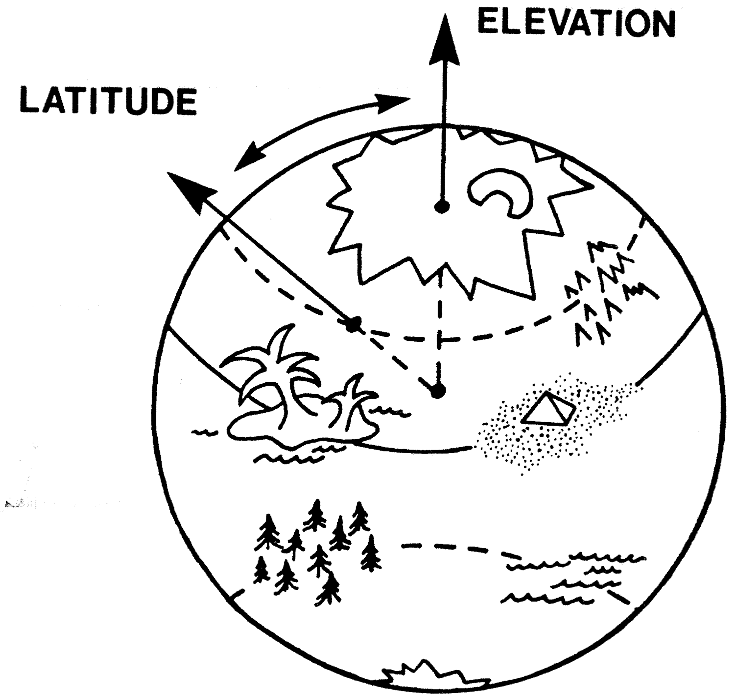

Figure 6 Habitat is basically a description of the environments offered on the planet Earth. There are two major dimensions: latitude and elevation.

4.2 Orthogonality

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 and orthogonal. At best, we have suggested that the behaviors or properties of 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 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.

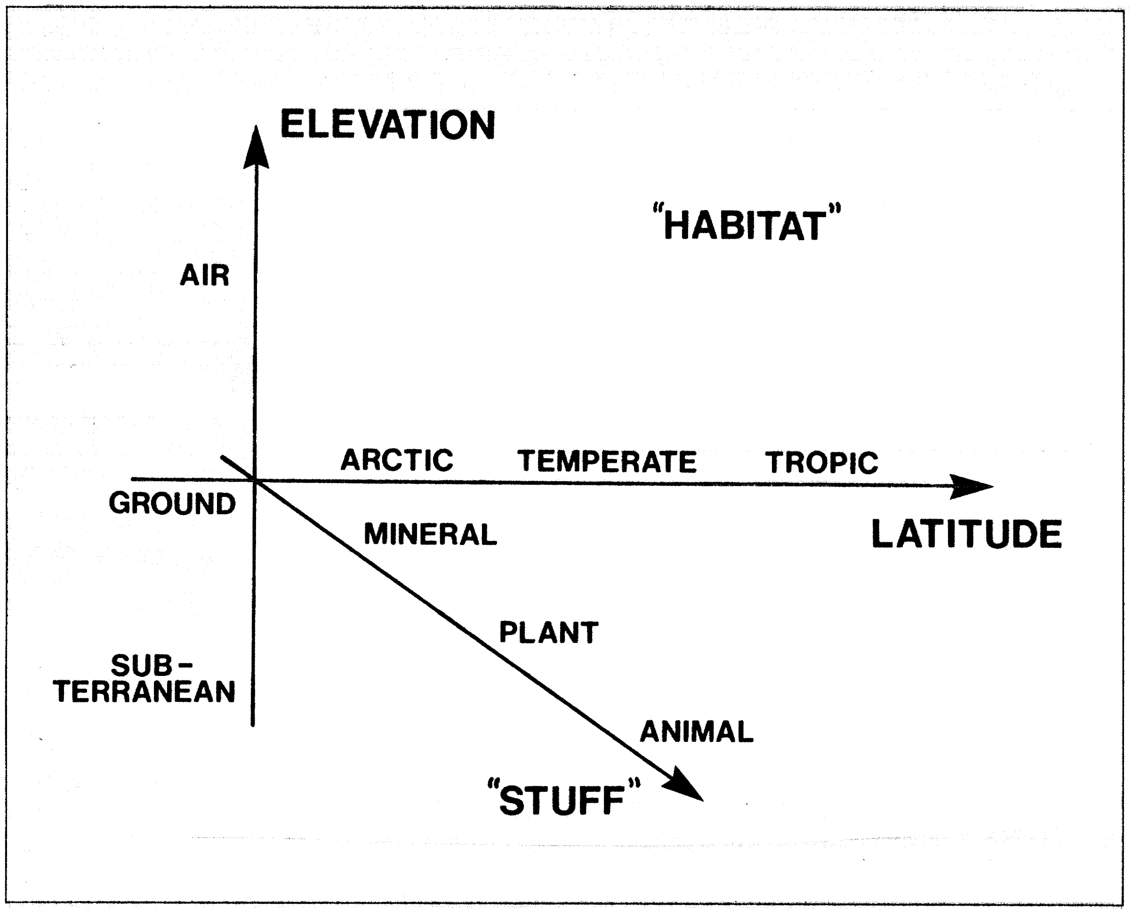

Figure 7 The major dimensions of the 20 Questions.

5.0 Habitat

One important reason why template–matching succeeds in simple creatures like the blowfly or the fledgling gull is that their environment is simple and highly constrained, just like it might be in a laboratory or industrial setting. At present, our 20 Question scheme ignores this advantage. Yet clearly if the game is played and we deduce the "thing" is a large, four–legged animal, we are not going to guess "*CAMEL*" if we are in the arctic, nor "*POLAR BEAR*" if we find ourselves in the Sahara. Our conclusion about the "thing" is thus heavily influenced by the habitat in which we find ourselves. An independent set of questions is therefore needed to set the context.

HABITAT is basically a description of where on the planet Earth the perceiver is located. Normally a position on the earth is described by the three dimensions of latitude, longitude and elevation. However, because the environment does not change substantially with longitude, *HABITAT* will have only two dimensions: latitude and elevation (Fig. 6).

The scheme adopted to characterize the environment or setting is shown in Fig. 7. The arrow coming out of the page is our previous property or "stuff" axis. The first *HABITAT* dimension is shown on the vertical axis, which characterizes the effect of elevation above (or below) the Earth. Is the "thing" in the air, on the ground, or subterranean — either below ground or under water? Although only three categories are shown for this dimension, finer discriminations can obviously be made. Points on this dimension should be relatively easy to assign for any perceiver equipped with a vestibular system, or who can "see" the horizon.

The second axis depicting the effect of latitude is more complicated, but has been

HABITAT

ATTRIBUTE (Question)	ARTIC	TEMPERATE	TROPIC	(REF)
temperature	cold	warm	hot	15
humidity	dry	moderate	wet	16
terrain	flat	hilly	mountain	17
ground cover	white	yellow/brown	green	18
earth	solid	soft	liquid	19
ELEVATION	air	ground	subterrain	20

Figure 8 Table 2. "Habitat." Further Questions whose answers set the overall context of the scene.

tentatively divided into *ARTIC*, *TEMPERATE* and *TROPIC*. However, the environmental attributes that are really of interest are such things as *TEMPERATURE* (Low, Medium, High), *HUMIDITY* (dry, moderate, wet), *GROUND COVER* (white, yellow-brown, green); or *TERRAIN* (flat, hilly, mountain). Note that although the parameters assigned to each attribute generally map one-to-one onto the *ARTIC-TEMPERATE-TROPIC* dimension, this does not mean that the environment is restricted to these choices. Figure 8 summarizes the kinds of questions needed to establish the HABITAT context, completing our initial twenty questions.

6.0 Successes and Failures

The strategies and remaining problems encountered with the 20 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 (or are) drawn from successive questions when the game is played serially. (Of course, any biological implementation would probably elect to ask the questions in parallel.)[5]

Several problems become immediately apparent when playing 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

[5]We must be careful about comparing the performance of a serial 20 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.

may be a stationary animal, lying down. Similiarly 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 20 Question 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, one's attention is focused upon the most active events in the environment.

It is also clear that the relative priorities of the questions must change depending upon our immediate needs. Although dynamic events probably always come first, if we need *FOOD*, we must engage in an active search with "symmetric red (orange) object, above the ground" taking a high priority. The control structure of the 20 Questions is thus an interesting and important issue in its own.

Finally, the dimensions and attributes of our 20 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", 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 20 Questions strategy can still be applied successfully even in the world of man-made objects.

7.0 Levels of Perception

The thrust of the 20 Questions Game is to provide a crude and quick rough categorization to an event or "thing". What precedes the 20 Questions representation? What lies beyond?

7.1 The Input Representation

Playing the 20 Questions Game requires more than merely asking the questions. For example, each question must be properly formulated from intensity-based primitives, such as the spatial, temporal and spectral derivatives of intensity. These primitives form one kind of input representation roughly corresponding to Marr's primal sketch for vision (Marr, 1982) as previously discussed. Still another representation is needed, however, before the 20 Questions can be posed.

Implicit in our game is that the questions are all addressed to one region in 3-space. But a region is a collection of locations that are somehow bound together. How are these common locations determined? A second representation is needed to make this information explicit.

Finding locations that belong together depends upon what we mean by an "object". Intuitively, an object is an entity that occupies space and consists of roughly the same "stuff". These notions lead one to postulate the following constraints:

(i) *Uniqueness*: Only one object (event) can occupy any spatial location at any given instant of time. Thus any events associated with a given location (in 3-space) at any instant must be associated with the object.

(ii) *Common Fate*: If a property associated with one location is also associated with a neighboring location, the locations belong to the same object, especially if a continuous property path is present between them.

Those familiar with the Marr-Poggio stereo algorithms (1976,1979), will recognize that the above are simply a generalization of the uniqueness and continuity constraints required for developing a successful stereo matcher. This makes sense, for the strategy of matching the different images presented to the two eyes is to match events having the same origin.

Given the above, we can now specify another important goal for early information processing, namely to make explicit which locations belong together, both within and between the senses (vision, audition, tactile). In vision this requires finding the occluding edges of an object, together with its component parts suffering common fate, using image-based primitives (i.e., a "primal sketch"). In audition, we need to know which frequency bands come from the same direction, or which are similarly modulated in time. The full "common-fate" representation then must bring these auditory and visual (also tactile) correlations together so that the 20 Questions can be asked and answered for this region of common events.

The goal of the 20 Questions is then to assign properties, structure, and possibly primitive actions to the "object" identified by the lower-level "common-fate" representation. Clearly the two types of representations are not entirely independent, for some of the tools needed for the construction of "common-fate" assertions can also be used to construct the 20 Questions. For example, the very low-level raw "primal sketch" could serve both representations. This possibility and its implications become clearer if the Schneider "two visual system" proposal (1969) is invoked to place the 20 Question, property-based representation in the primary sensory cortex, whereas the "common-fate" representation could be resident subcortically in the colliculus. This dual processing has the obvious advantage that the computationally complex tasks, such as making property shape assertions, can be accomplished using relatively isolated hardware modules, loosely coupled, while in parallel the grouping tasks that require linking of property lists can proceed elsewhere.

7.2 Output Representations

We have been playing the 20 Question Game in order to provide a crude categorization of an "event" or "thing" in a scene, presumably as a first stage in recognition. Clearly, additional effort is required to create a precise, detailed and useful object description. For example, we may have been fortunate enough to ascertain that the "thing" is a *PRIMATE* in motion, but is it a monkey or man, and if the latter, who is it? To obtain answers to such questions, still another, more detailed representation of the "event" must be constructed. Does this imply that we must invoke still another set of different 20 Questions for the next level of analysis (Harmon, 1973)? Or can our original set be used once again, but more locally?

So far we have sought only a representation useful for the very early stages of recognition. Beyond this may be the need to know what an object is doing — i.e., its actions and intentions. Are the same 20 Questions useful here? If not, then the power of the game is greatly weakened. And what if we desire to manipulate an object, or to show how it is built (i.e., the prints or anatomical sketches) to aid in modifications or repairs? As our representational goals change, certainly at the very least the relevance of any one of the 20 Questions must also change. The hope would be that the answers to these few questions would still provide a substrate upon which a rich variety of higher-level representations could be constructed. How is this possible without introducing a never ending hierarchy or sets of additional questions?

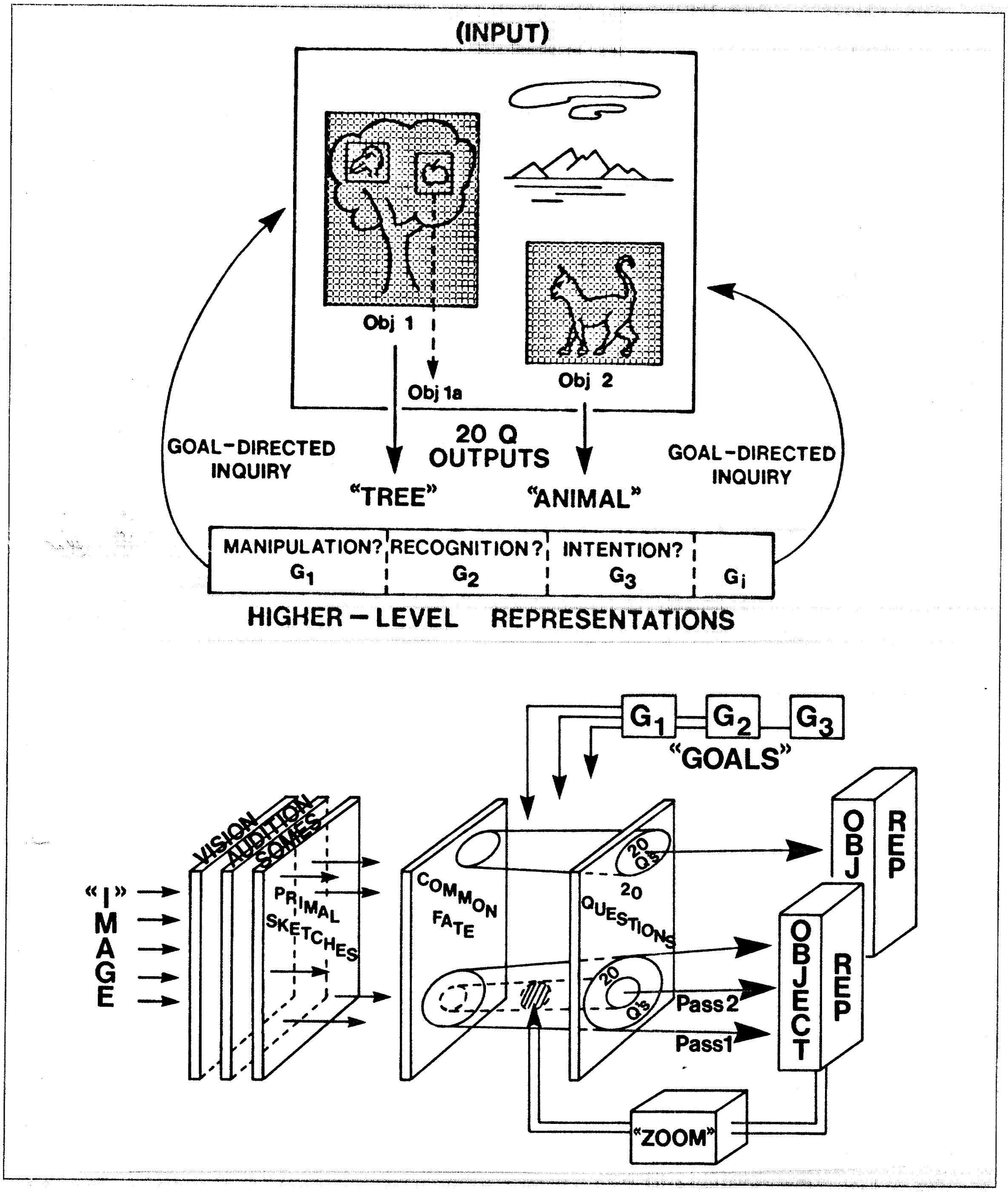

Figure 9 Schematic of 20 Question Game, showing its Control Structure.

7.3 Ullman's Routines

Faced with the implausibility of a hierarchy of 20 Questions (or representations), Ullman (1982) has suggested a control structure might be devised that reexamines the original 20 Questions outputs in greater detail, perhaps even digging deeper into the incoming sense data itself. Exactly which questions and primitive elements are selected will depend upon the immediate objective. (We have already noted that a similar control structure is needed to set priorities for the 20 Questions.) Thus, once a representational goal has been chosen

Figure 10 The "All-Natural" girl: her lips are a tulip bud; her cheeks are peaches; her eyes are sparkling water; her eyebrows are bushy shrubs; her hair is straw. The sweater is wool, of course.

(recognition, manipulation, modification, repair, etc.), a weighted sequence of probes might be invoked to test the data for the additional information needed to complete the desired representation. For example, at this point the spatial relations between various property or shape descriptors could be made explicit. This scheme has the advantage of allowing a wide wide variety of representations to be constructed depending upon the immediate need, without adding an entirely new hierarchy of 20 Questions.

8.0 From Wholes to Parts

The simple Twenty Question Game thus serves as a springboard that allows a perceiver to begin to interpret a novel scene, seen completely afresh, without context. The output of the 20 Questions is a baseline from which more detailed explorations on the data are made. These explorations can take many directions. Most obvious are routines that zoom down on the "thing" and attempt to examine its parts — like the fruit on a tree, or the face of a man.

Figure 9 illustrates this point. The input is a scene containing a tree, a mountain and an animal. Let us say the 20 Questions have identified a category *TREE* as indicated by the large box of questions superimposed on the tree. Similarly, the *ANIMAL* has been categorized in parallel. Let us now set as our goal to determine the intentional relations between the object *TREE* and the object *ANIMAL*. Our present 20 Question answers are clearly grossly inadequate. We need more information about just what the animal is doing, as well as details about the tree. Is there a bird there? We are thus forced to change the scale of our 20 Question inquiries and to note other "significant" features of each of these two regions of the image. Can we still ask the same questions over once again to advantage? The idea behind the 20 Questions is that we can, because it is basically a property list which should apply at any scale.

Consider that the first pass through the 20 Questions yielded the object *ANIMAL* (cat-like). An intentional representation requires that we obtain more information about the actions, posture, or behavior of the animal. Although we might already know it is moving toward the tree, is it making friendly or aggressive noises, is its tail stiff or its teeth bared and its claws extended? The original set of 20 Questions can answer these inquiries, for aggressive noises are distinctly different from friendly ones, a stiff tail appears like a rigid rod, bared teeth might look like a row of white "rocks" or shells, extended claws might map into a defoliated shrub. The exact nature of the property of the part is disambiguated by the context *ANIMAL* (just as *HABITAT* reduces the object possibilities at a coarser level). Intentions can be inferred from the properties of the parts, given the context.

The montage of Fig. 10 is another example of how the same set of 20 Questions can be applied to advantage at different scales. Here the facial features are all textures taken from the Vegetable or Mineral Kindgoms: a tulip forms the lips; the cheeks are peaches; her eyes are sparkling water, etc. Recall how many poets apply similar descriptions! Of course the context "PRIMATE" rules out these "unnatural" origins in practice, so the 20 Question answers are usually unambiguous, especially if reinforced by the presence of the appropriate spatial relations.

In sum, the 20 Question strategy is to apply the same questions in sequence at a number of spatial scales to the scene. To accomplish this requires a rather flexible control structure for manipulating the questions, plus a lower-level, parallel representation that initially decides which locations share "common-fate" and hence share the same pool of answers. (This latter representation must be part of the initializing routine). Beyond the 20 Questions is not an additional hierarchy of further sets of hard-wired questions — this seems implausible. More likely is that the same set of questions is reapplied at a finer scale.

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, i.e., one which is feasible to build today. For simplicity, the device 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" the 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 envionment, ranging from artic 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? *Implication*: It's an animal in motion.	*translates*	sway	no
Q2: How many supports? *Implication*: Confirms animal — has four "legs".	*2,4 or >4*	1	0
Q3: What acoustic frequencies are emitted? *Implication:* Disconfirms animal. "Thing" makes low frequency, broad-band sounds, moves and has 4 legs. Must be big. Elephant or cow?	narrowband	*broad*	*broad*
Q4: Acoustic Source *Implication:* Confirms "animal" or isolated object.	*point*	extended	extended

	ANIMAL	PLANT	MINERAL
Q5: Visually symmetric?	*mirror*	3D	irregular
Q6: What is 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 broad-band 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 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?	narrowband	broad (None of the above)	broad
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

	ARTIC	TEMPERATE	TROPIC
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

Documentation of devices that can provide answers to each of the twenty questions.

1. **Acoustic Frequency.** Comb filtering has been used for several years to separate sound sources (Shields, 1970; Flanagan, 1972; Zwicker et al., 1979). Unless many broad-band sources are active simultaneously at S.P.L.'s comparable to the narrowband sources, this question can be answered with available technology (Klatt, 1977). As initially formulated (Richards, 1980), the question simply addresses whether the source is broad-band 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 1/3 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 ten years, and can be detected with simple systems provided the background is stationary (Horn and Schunck, 1981; Thompson, 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 and 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 and Rosenfeld, 1981; Binford, 1981; Richards et al., 1982).
7. **Axis.** Again, as in Question #6, the orientation of a region can be answered rather easily (Ballard and Brown, 1982) 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 and Spencer, 1980; ; Cook and Torrance, 1982; Rubin and Richards, 1982). This is an area ripe for research.
9. **"Color".** The value of spectral information in assessing food quality (Francis and Clydedale, 1975), printing inks or photographic reproductions (Judd and Wysecki, 1975) and in remote sensing (Chance and Lemaster, 1977; Lintz and Simonett, 1976; Sabins, 1978; Myrato 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 temprature is a simple sensory ability if contact is used (Herzfeld, 1962). Of course remote sensing is also possible here, as performed in surveillance or Landsat imagery (Lintz and Simonett, 1976; Barbe, 1979; Trivedi et al., 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 x 16 per cm^2 have been obtained (Hillis, 1982; Raibert and Tanner, 1982).
12. **Hardness.** The measurement of hardness of a point on a surface is a routine metallurgical technique and is trivial (Cox and Baron, 1955; O'Neill, 1967). 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 (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.
15. **Temperature.** In contrast to Question #10, which measured local temperature, this question addresses the global temperature of the environment. Again, however, many methods are currently available (Herzfeld, 1962).
16. **Humidity.** This property of the environment is routinely measured (Wexler, 1965), (although perhaps not in quite the same manner as in human beings).
17. **Terrain.** Whether the terrain is flat, rolling or mountainous is a problem in remote sensing. Although surface topography can be determined from Landsat images (Lintz and Simonett, 1976) a completely automated scheme is not yet fully developed, although Grimson's (1981) and Witkin's (1981) algorithms come close.
18. **Ground Cover.** Thanks to the Landsat program to assess food crops, remote sensing techniques in this area are quite sophisticated (Lintz and Simonett, 1976; Sabins, 1978).
19. **Earth.** Is the terrain made of rock, soil or sand, or is it mud or marsh? This question is the global analog of #12 (and perhaps #14) which addressed hardness (and viscosity). Although soil mechanics has been studied for some time (Tumikis, 1962), what we desire here is a global sensor of the terrain (or sea) in which we move. Certainly force sensors can assess the hardness of the ground we step on (Whitney, 1979), whereas others might measure the drag as we move through mud. This is all provided that legged motion is possible in the forseeable future (Raibert and Sutherland, 1983).

20. **Elevation.** Altitude above or below ground or sea level is a rather trivial measurement of pressure (Benedict, 1969). The exact elevation of a viewed object, however, is difficult, usually requiring a reference plane such as the ground or ocean surface. However, for a viewed object, we initially seek to know only whether the object is on the ground (i.e., supported by it), above the ground, or below it — a computationally feasible question given we know our own viewing position (Stevens, 1980).

REFERENCES

Agin, G., Representational description of curved objects. *Stanford AI Project Memo AIM–173, Stanford University, 1972.*

Altes, R.A. Angle estimation and binaural processing in animal echolocation. *J. Acoust. Soc. Amer.,* 1978, *63,* 155–183.

Atal, B.S. Automatic speaker recognition based on pitch contours. *J. Acoust. Soc. Amer.,* 1972, *52,*1687–1697.

Bajcsy, R. and Badler, N., *Representation of Three–Dimensional Objects.* New York: Springer–Verlag, 1982.

Ballard, D.H. and Brown, C.M. *Computer Vision.* New Jersey: Prentice–Hall, 1982.

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

Benedict, R.P., *Fundamentals of Temperature, Pressure, and Flow Measurements,* New York: Wiley, 1969.

Binford, T., Inferring surfaces from images. *Art. Intell.,* 1981, *17,* 205–244.

Boden, M., *Artificial Intelligence and Natural Man.* New York: Basic Books, 1975??

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

Cook, R.L. and Torrance, K.E., A reflectance model for computer graphics. *ACM Trans. on Graphics,* 1982, *1,* 7–24.

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

Davis, L. and Rosenfeld, A., Computing processes for low level usage: a survey. *Art. Intell.,* 1981, *17,* 245-263.

Dirlam, D.K., Most efficient chunk sizes. *Cogn. Psych.,* 1972, *3,* 355–359.

Flanagan, J.L., *Speech Analysis: Synthesis and Perception.* Berlin: Springer–Verlag, 1972.

Francis, F.J. and Clydedale, F.M., *Food Colorimetry: Theory and Applications.* Westport, CT: AVI Publishing, 1975.

Grimson, W.E.L., *From Images to Surfaces.* Cambridge, MA: MIT Press, 1981.

Harmon, L.D., Automated tactile sensing. *Int. J. Robotics Res.,* 1982, *1 (2),* 3–32.

Harmon, L., The recognition of faces. *Sci. Amer.,* 1973, *229,* 70–82.

Herzfeld, C.M., *Temperature, Its Measurement and Control in Science and Industry, vol. 3.* New York: Reinhold, NY, 1962.

Hildreth, E., The integration of motion information along contours. *IEEE Proceedings of a Conference of Computer Vison Representation and Control, Sept., 83–91, 1982.*

Hillis, W.D., A high resolution image touch sensor, *Int. J. Robotics Res.,* 1982, *1 (2),* 33–44.

Hoffman, D.D. and Richards, W.A., Representing smooth plane curves for recognition: Implications for figure–ground reversal. *Proceedings of the National Conference on Artificial Intelligence, August 18–20, 1982,* and *MIT AI Memo 630.*

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

Horn, B.K.P., Image intensity understanding. *Art. Intell.,* 1977, *8,* 201–231, and *MIT AI Memo 335.*

Horn, B.K.P. and Schunck, B.G., Determining optical flow. *Art. Intell.,* 1981, *17,* 185–203.

Howard, I.P. and Templeton, W.B., *Human Spatial Orientation.* London: Wiley, 1966.

Judd, D.B. and Wysecki, G., *Color in Business and Industry.* New York: Wiley, 1975.

Kanade, T., Recovery of the three-dimensional shape of an object from a simple view. *Art. Intell.,* 1981, *17,* 409–460.

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

Klatt, D., Review of the ARPA speech understanding project. *J. Acoust. Soc. Amer.,* 1977, *62,* 1345–1367.

Knudsen, E.I. and Konishi, M., Mechanisms of sound localization in the barn owl. *J. Comp. Physiol.,* 1979, *133,* 13–21.

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

McMahon, T.A., Using body size to understand the structural design of animals: quadrupedal locomotion. *J. Appl. Physiol.,* 1975, *1975,* 619–627.

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

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

Marr, D. and Hildreth, E., A theory of edge detection. *Proc. Roy. Soc. Lond. (B),* 1980, *207,* 187–217.

Marr, D. and Nishihara, H.K., Representation and recognition of the spatial organization of three-dimensional shapes. Proc. Roy. Soc. Lond. (B), 1979, *200*, 269–294.

Marr, D. and Poggio, T., Cooperative computation of stereo disparity. *Science,* 1976, *194,* 283–288.

Marr, D. and Poggio, T., A computational theory of human stereo vision, *Proc. Roy. Soc. Lond. (B),* 1979, *204,* 301–328.

Mervis, C.B. and Rosch, E., Categorization of natural objects. In M.R. Rosenwerg and L.W. Porter (Eds.), *Annual Review of Psychology,* 1981, *32,* 89–115.

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

Minsky M. and Papert, S., *Perceptrons.* Cambridge, MA: MIT Press, 1965.

Moon, P. and Spencer, D.E., An empirical representation of reflection from rough surfaces. *J. IES,* 1980, *9,* 88–91.

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

Newell, A., 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, 1973.

O'Neill, H., *Hardness Measurement of Metals and Alloys.* London: Chapman and Hall, 1962.

Raibert, M. and Sutherland, I., Machines that walk. *Sci. Amer.,* 1983, *248,* 44–53.

Raibert, M.H. and Tanner, J.E., Design and implementation of a VLSI tactile sensing computer. *Int. J. Robotics Res.,* 1982, *1,* 3–18.

Richards, W.A., Nishihara, H.K. and Dawson, B., *Cartoon: A biologically motivated edge detection algorithm. MIT AI Memo 668,* 1982.

Richards, W.A., 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 and M.H. Mickle (eds.), 1980, *10,* 193–200.

Rubin, J.M. and Richards, W., Color vision and image intensities: When are changes material? *Biol. Cybern.,* 1982, *45,* 215–226.

Sabins, F.F., *Remote Sensing: Principals and Interpretation.* San Francisco: Freeman, 1978.

Schneider, G.E., Two visual systems. *Science,* 1969, *163*, 895–902.

Schuanevddt, R.W., Durso, F.T. and Mikherji, B.R., Semantic distance effects in categorization tasks. *J. Exp. Psychol.,* 1982, in press.

Schunck, B.G. and Horn, B.K.P., 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., and Colburn, H.S., Model for auditory localization. *J. Acoust. Soc. Amer.,* 1980, *60,* 1164–1175.

Shields, V.C., Separation of added signals by digital comb filtering. *Masters Thesis, MIT, Dept. of Electical Engineering,* 1970.

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

Stevens, K. Surface perception by local analysis of texture and contour. *MIT AI Technical Report 512,* 1980.

Stevens, K., Computation of locally parallel structure. *Biol. Cybern.,* 1978, *29,* 19–28.

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

Tinbergen, N., *The Study of Instinct.* Oxford: Clarendon Press, 1951.

Trivedi, M.M., Wyatt, C.L. and Anderson, D.R., A multispectral approach to remote detection of deer. *Photogram. Engr. and Remote Sensing,* 1982, *48,* 1879–1889.

Tumikis, A.R., *Soil Mechanics.* Princeton, NJ: Van Nostrand, 1962.

Tversky, A., Features of similarity. *Psychology Review,* 1977, *84,* 327–352.

Ullman, S., Visual routines, *MIT Cognitive Science Paper No. 20, 1982.*

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

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

Wexler, A., *Humidity and moisture: Measurement and control in science and industry.* New York: Reinhold, 1965.

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

Wilson, E.O., *The Insect Societies.* Cambridge, MA: Belknap Press, 1971.

Witkin, A.P., Recovering surface shape and orientation from texture. *Art. Intell.,* 1981, *17,* 17–45.

Woese, C.R., Archaebacteria. *Sci. Amer.,* 1981, *244,* 98–122.

Zwicker, E. and Terhardt, E., Automatic speech recognition using psychoacoustic models. *J. Acoust. Soc. Amer.,* 1979, *65,* 487–498.