Thoughts on the specific nerve energy

Andrei Gorea

Laboratoire de Psychologie Expérimentale Associé au C.N.R.S.,
Université René Descartes, 28 rue Serpente, 75006 Paris, France

The *specific nerve energy* (*SNE*) doctrine is derived from the common sense observation that there must be something specific about sensory processing merely because, everybody would agree, sensations are specific*. There is also common sense beneath the credo that objects "give off images of themselves, which are carried to the mind by the nerves" (see Boring, 1942, p. 69). At the time when they were spelled out, these two propositions were equally untestable, and for that reason regarded as axiomatic.

Let me stress from the very beginning the fact that a strict definition of what is actually meant by either "specific nerve energy" or "giving off images of itself" has never been provided. In fact, Johannes Müller himself phrases the essence of the former as a mirror image of the latter:

"*A sensation is not the conveyance to consciousness of a quality or a state of an external object, but rather the conveyance to consciousness of a quality or state of our nerves, brought about by an external cause.*" [translated by Bela Julesz from the *Handbuch der Physiologie des Menschen*, 4th edition, Coblenz, Verlag von J. Hoelscher, 1844.]

Although equally reasonable from the "common sense" standpoint, by the end of the 19th century, the *SNE* approach had completely overthrown the "giving off images..." idea. Indeed, all of the evidence provided by anatomy, cytology, electrophysiology and neuropsychology transformed the *SNE* doctrine into a matter of investigation, "proved" correct its consequences and allowed the transfer of its axiomatic status into the realm of current knowledge.

The sensory scientist and more particularly the modern psychophysicist might be aware that the implicit and continuous use of this "current knowledge" in everyday model/theory building is not as obvious as it might appear. It is my claim that such a feeling is not only due to the general difficulty of applying "current knowledge" to specific investigation, but essentially to the fact that the *SNE* doctrine remains a doctrine in that it has never been satisfactorily tested. Paraphrasing Boring, not all the men who discarded this problem as trivial understood it (Boring, 1942, p. 68).

*In latin, sensation is "qualis", i.e. quality. "Specific sensation" may thus be looked at as tautological.

1 THE *SNE* AS A DEDUCTIVE APPROACH

1.1 A Syllogism

A. The basic concepts of the *SNE* doctrine are rooted in the following two observations: "The same stimulus acting on different nerves gives rise to different qualities [or perceptual states]; different stimuli acting on the same nerve give rise to the same quality. It is the nerve, not the stimulating object, that matters." (Müller, *passim* Boring, p. 71).

B. The *SNE* doctrine cannot be related to specific *receptors* (where a neural pathway begins) since stimulation beyond the receptor site still gives rise to specific perceptual states.

C. *SNE* must then be related to the site where a pathway **ends**, the *sensory center*, or to the neural pathway as a whole (including the "sensory center").

D. Modern and contemporary neuroscience - as initiated by Mountcastle (1957) and extensively developed by Hubel and Wiesel (1977) - basically consists of new experimental and theoretical instances of Müller's point of view.

Points A to C develop the logic beneath the *SNE* doctrine and point D provides experimental support. Altogether, points A to D lead to the apparently obvious conclusion that "... if you cross-connect (say) the optic and the auditory nerves, you could see tones and hear colors" (du Bois-Reymond, *passim* Boring, p. 78).

Whether logical or intuitive, the above statement is based on the *belief* that "sight is not hearing because the optic fibers are projected on the occipital lobe and the auditory upon the temporal lobes" (*op. cit.*, p. 78). It is crucial to note that, despite point D, this premise is no more testable than the fact that we do *see* colors and *hear* tones, which is a pure matter of convention. In fact points A to D do not tell us what a specific sensation or "perceptual state" is, or what a nervous "ending" (or "site") is meant to stand for.

1.2 Where does deduction stop?

Let us consider the following:

E. If you cross-connect, during a *critical period*, the optic and the auditory nerves you would still see colors and hear tones. There is no *a priori* reason to reject this proposition. In the absence of any available test, it is as axiomatic as its reversed formulation. (However, accepting it will bring the *SNE* concept down in pieces. Or, will it?)

F. Accepting propositions A,B,C and E as true entails that the specificity of sensations does not lie either in the stimulus, or in the receptors, or in the pathways, or in the "sensory center", whatever their precise definition might be. Where does it lie, then?

I see two possible answers to this question. It either lies nowhere: there is no such thing as *specific* sensations or perceptual states. Or, it is related to a specific *conjunction* of the physical characteristics of the stimulus, their sensory receptors, the pathways and the "sensory centers", altogether.

The first alternative presents the advantage of eliminating all reference to the *SNE* concept. However, since it also presents the disadvantage of bringing back the sensory sciences to the deep blur of untestable philosophical propositions, I shall not discuss it here.

The spirit of the second alternative is definitely more biological. If sensation is the *stimulus-brain conjunction*, none of the enumerated "links" is "specific" with respect to sensation. On the other hand, at least in the early stages of development, none of these links should be missing without specificity of sensation being lost. All links are, however, specific in their own right. The photoreceptors respond to light, while the ciliary cells respond to vibration, etc. The optic and auditory tracts are spatially distinct and neurons at different stages are selective to distinct aspects of the stimulus in different physical domains. Depending on the location within the processing hierarchy and on the extent of the critical period, their selectivity is more or less experience dependent.

The inconvenience of the above perspective is that, while offering a general paradigm for conceptualising our general knowledge and deductions, it does not lead to specific models and theories and even less to specific experiments in the field of sensory research. The concept of *SNE* remains inherent to this formulation but it is no longer a matter of test. How could one test the (self-evident and circular) proposition that sensation is "labelled" by the interactions among all the elements, from stimulus to brain, intervening in that particular behavior?

The *SNE* doctrine appears to be rather diffuse in our minds. If identified with the concept of *stimulus-brain conjunction*, it provides an implicit frame for our thinking and, under this acceptation, it is more like a paradigm (Kuhn, 1962). If taken under its more localized version, it is more like a tool. As a tool, however, it requires strict definitions of what we mean by concepts such as "perceptual states" (or specific sensations) and neural "sites" (*viz.* "nervous endings" or "sensory centers").

It is in the rapidly changing ways of using the *SNE* concept, especially as a tool but also as a paradigm, that one may realize the extent to which the choice of the stimuli, the setup of our experiments, the interpretation of our results and the building of our models are pervaded by tacit assumptions the source of which can be, in most cases, traced down to the *SNE* doctrine. I shall try to make this point clear by briefly discussing what I think have been the main ideas within our field of research during the last three decades and their dependency on distinct definitions of the perceptual state-neural site duo.

The reader should be aware that, in this brief historical discussion, I have taken a (psychophysically) biased perspective and chosen to cite only a very limited number of authors. The likelihood of having omitted some basic developments in vision research is thus quite high.

2 PERCEPTUAL STATES AND NEURAL SITES IN THE LAST THREE DECADES OF VISION RESEARCH

2.1 Stimulus-specific (feature) detectors

2.1.1 General considerations. The main idea behind the feature- detector approach - at least as it started in the early fifties - is that there exists a set (to be specified) of distinct, canonical visual mechanisms (sensory centers) whose sensitivity profiles are such that they respond selectively to a set of distinct, canonical visual stimuli (features). It is implicit in this formulation that a given sensation or perceptual state is directly related to the activation of a given mechanism (or *homunculus*). As such, the feature-detector approach represents an orthodox implementation of the *SNE* doctrine.

The immediate problem with such an approach is that mechanisms and stimuli are defined with respect to one another. The actual existence of a canonical mechanism cannot be proven unless we specify the corresponding canonical stimulus and, reciprocally, the specification of a canonical stimulus requires knowledge about the corresponding canonical mechanism. Thus, in the same vein as the paradigmatic *stimulus-brain conjunction* approach, the feature- detector approach apparently misses from the very start its ultimate objective of providing the basis for a one-to-one relationship between the activation of a specific neural site and the experience of a specific sensation (or perceptual state). How was it actually used?

In 1953, Barlow discovered cells in the frog retina responding selectively to *small* objects *moving*, within a restricted velocity range, across their receptive fields. He proposed that *small* + *movement* = *fly* (which is to be eaten; paraphrasing Barlow, 1953, p. 86) and that the cells he described are thus *fly detectors*. This was an *a posteriori* interpretation. Its merit consisted in its biological *meaningfulness* (for frogs). It soon became a matter of dogma to assume (implicitly or explicitly) that all features (i.e. canonical stimuli) and feature-detectors are biologically meaningful. While meaningfulness would be a superb guide for our experimental and theoretical work, it is rarely (if ever) specified *a priori* (see also 2.5.1).

The initial suggestion of Hubel and Wiesel (1959) that orientation is a meaningful attribute of the visual image and that "oriented" units are meaningful relays in visual processing was an *a posteriori* assessment too. Visual scientists were ready, however, to accept meaningfulness in early vision as a key concept in visual research and proceeded for more than a decade as if the assessment of the meaningfulness of the mechanisms they were about to study was an *a priori* endeavour. This assumption started the golden age of the feature-detector approach.

2.1.2 Feature-detectors and hierarchical processing. The meaningful sensory center may be peripheral (for the frog) or, at least in principle, as central as one desires (for more sophisticated species; Barlow, 1972). Correlatively, perceptual states may be as elementary as *movement* and as complex as *small* + *movement* = *fly*. Of course, they can be *faces* but also *grandmothers*. As everybody in the field must have realized and acknowledged from the very start of the feature-detector tradition, *grandmothers* and, say, *yellow* are equivalent perceptual states, although grandmothers are frequently colored, old and typically friendly. If this is so, another major problem with the feature-detector approach is its effort to match an apparently hierarchical structure of meaningful sensory centers (a very influencial view since Hubel and Wiesel, 1968) with an apparently unstructured domain of perceptual states.

On the one hand, this convergence scheme (the higher the processing stage, the higher the complexity of the processed visual attribute) is difficult to conciliate with the idea that distinct detectors are the substrate of distinct perceptual states or, at least that distinct perceptual states are *always* related to the activation of distinct mechanisms. How could this be if the neural substrate is just a relay within a hierarchical chain of neural transformations? On the other hand, the ultimate implication of the convergence scheme, namely the existence of a higher up *master homunculus*, is identical with that of the *SNE* doctrine.

During the golden years of the feature-detector approach, few debated the inherent ambiguity of the stimulus/mechanism definition or the meaningfulness of orientation (rather than curvature, for example) and orientation-selective (rather than curvature-selective) units for visual behavior. Instead, "flies" and "fly"-detectors multiplied as mushrooms and pervaded the classical frontiers of early vision.

Similar considerations apply to the psychophysical feature- detector approach where the existence of bar and edge detectors was questioned on grounds that a different stimulus description (*viz.* in the frequency domain) may be more meaningful and certainly more general. The psychophysicists put then more emphasis on an alternative approach based on the idea of *filtering*. However, the basic assumption that filters, as well as feature-detectors, are *labelled* such that they can be directly related to a specific perceptual state, was not abandoned (see 2.1.4). Two reasons for this were that no other paradigm was immediately available and that the labelling concept was extremely fruitful in naming a whole set of *specific perceptual states* and *detectors* on solid (electrophysiological and psychophysical) experimental grounds. Given the above considerations, the extent to which these specific detectors, their visual functions and the underlying experimental evidence are beyond any doubt remains a matter of debate.

2.1.3 Feature-detectors and parallel processing. The view of a strictly hierarchical processing of the visual image coexisted practically from the very beginning with the view that visual information is initially blown into a number of primitives processed in parallel up to some higher (unknown) associative areas (for recent reviews see DeYoe and Van Essen, 1988; Livingstone and Hubel, 1988; Zeki and Shipp, 1988). Within this context, vision research emphasized the idea of "super-flies" related to concepts such as space and time, form and motion, chromatic and achromatic dimensions each of which is presumably processed within parallel pathways (e.g. Livingstone and Hubel, 1988).

The first challenge of the hierarchical view may have been the discovery by Enroth-Cugell and Robson (1966) of the X and Y ganglion cells in the cat retina. X and Y cells were shown to differ in many respects of which their distinct spatial and temporal processing characteristics were of main interest. In a relatively short lapse of time, psychophysical research managed to impose the idea that shape (i.e. spatial information) and what was indifferently referred to as flicker or motion (i.e. temporal and spatio-temporal information) were processed by more or less independent mechanisms. The issue soon became ambiguous both neurophysiologically and psychophysically. The X/Y distinction at higher processing levels became controversial and the status of motion perception (which is inherently spatio-temporal) raised theoretical problems concerning the separability of space and time (see Burt, 1987).

While space and time may be conceptually (and experimentally) difficult to relate to distinct perceptual states, the perception of color, form, motion and depth may be easily regarded as orthogonal and studied independently. This conceptual and experimental facility, I would guess, led to the reinforcement of the generalized parallel processing idea which partly overshadowed the hierarchical processing one. The initial feature-detectors which, in principle, could be selective to any specific combination of visual attributes (like "yellow submarine") were replaced by specific pathways dealing with specific attributes at all levels of complexity. Saying that two pathways are distinct is to say that they carry specific (perceptual) information and thus specific nerve energies.

There are two main objections to this approach. The first is experimental and relates to the increasing number of cross- connections between presumably distinct pathways and to the difficulty of demonstrating their exclusive selectivity to a given stimulus dimension (e.g. DeYoe and Van Essen, 1988; Zeki and Shipp, 1988). The second is theoretical and relates to the integration of attributes processed independently within an unique and meaningful visual object. This integration problem, repeatedly addressed by both neurophysiologists (e.g. Zeki and Shipp, 1988) and experimental psychologists (e.g. Treisman and Gelade, 1980) is far from being solved.

It is interesting to note that there is an integration problem only when one rejects the possibility that the cross-talk among distinct pathways can be captured within the activity of a unique (meaningful) mechanism. Indeed, there is no such a problem if one is ready to accept the existence of a feature-detector (of unspecified complexity) selective within a multidimensional physical space. Such a feature-detector is in fact a neurophysiological replica of the stimulus and, as such, it can be directly associated with a perceptual state (of unspecified complexity). Hence, there is an integration problem only outside of the conceptual frame determined by *SNE* doctrine.

2.1.4 "Identification" and the labelling argument. Recently, Watson and Robson (1981) performed the following experiment. They randomly presented during one of two temporal intervals one out of two spatial frequency patches whose contrast covered the whole threshold range. Observers were asked to specify the interval which contained the stimulus (detection) and to identify the stimulus (identification). They measured the detection/identification performances as a function of contrast for a number of stimulus pairs and found that when the two stimuli in a pair were sufficiently disparate (in spatial frequency), the detection and identification functions of contrast overlapped. Since the system is capable of identifying the stimulus any time it detects it and since, at threshold, the probability of activating more than one (optimal) detector is very low, it follows that this detector *must* be *labelled*. Thus, *all* detectors must be labelled. What "labelled" was meant to specify is unclear, although everybody would probably agree that the "labelling" idea is directly related to that of a specific perceptual state and of a specific nerve energy. If so, is the above described experiment a proof of, or just another way of restating the *SNE* doctrine?

The relationship between identification and detection has been discussed by Helmholtz and its modelling has been shown since to depend on factors such as the underlying detection theory, the linearity of the detection process, the independence of the detectors, etc. (Graham, 1989). It is clear, for example, that the interpretation of the above experiment is critically dependent on the assumption that threshold performances are determined by the activation of an optimal detector. One may, however, doubt whether this assumption will ever be a matter of formal testing. Moreover, the optimal detector is specified psychophysically and formal proofs of its neurophysiological site are missing (see, however, Newsome in this volume). Thus, the application of the *SNE* concept to early vision remains a matter of consensus.

2.2 Linear filters and feature-detectors - Mechanisms

2.2.1 Traditional approach. Linear filters (Campbell and Robson, 1968; Sachs, Nachmias and Robson, 1971) and feature-detectors have been and still are hostile friends. From the *SNE* point of view, they are equivalent to a large extent. The basic and perhaps only difference between them is that, in principle, the filter approach requires a limited number of filters to account for a much larger number of perceptual states. The underlying idea (which can be traced back to Young) is that a perceptual state is related to some specific pattern of activation of a limited set of low-level units. Current understanding of color perception and of discrimination/ identification visual performances in early vision (see para 2.1.4) is heavily dependent on this principle.

The manipulation of the filter concept eventually led to the specification of theoretically *optimal detectors* which, in turn, permitted the specification of the *appropriate stimulus* (Watson, Barlow and Robson, 1983) to be used in the process of testing (psychophysically or electrophysiologically) the existence of the optimal detectors... such as spatial frequency or directionally tuned filters displaying a more or less pronounced even or odd spatial symmetry, with a more or less Gaussian spatial sensitivity weighting function, etc.

This increasingly sophisticated engineering approach also raised problems related to the independence of spatial and temporal processing (see para. 2.1.3), to biological noise and its correlation across distinct detectors and, more generally, to the linear *vs.* nonlinear processing of visual information (Graham, 1989). As Bela Julesz pointed out a while ago, the fact that visual processing is strongly nonlinear necessarily leads us back to the feature-detector approach since nonlinearities are features (or bugs or flies).

2.2.2 Pyramids. It has been proposed (Marr, Ullman and Poggio, 1979) that early vision may be modelled as a parallel, multiple-scale filtering process. Since the representation of physical information is isomorphical to the related percept at any scale of the "pyramid", a perceptually "popping-out" feature is a "popping-out" neuron (or group of neurons) at at least one of these filtering levels.

This "pyramidal" scheme (see Part 1 of this volume) is an obvious extension of the filter approach and it was initially developed to provide higher efficiency coding (but not decoding) primitives (kernels, wavelets, 2-D Gabors, etc.) and algorithms in the luminance domain. It cannot thus account for more than second-order, black-and-white phenomena such as texture discrimination, "pop-out" effects and the like.

In principle (but not yet in practice), the pyramidal approach could be applied at *all perceptual domains*, whether at the same early vision processing stage (such as the chromatic domain), or at a higher processing stage (such as, say, the domain within which we account for shape-from-motion phenomena). At the same processing level, a large population of units would share the same multidimensional tuning space, while others would be more or less one-dimensionally biased. At different processing stages, primitives would differ qualitatively so that the higher the processing level, the more elaborated the coding primitive. Moreover, the multiple-scale processes (at all complexity levels) may be made interactive and the perceptual states may be related to the state of the pyramid(s) as a whole rather than to the activity of some of its (their) layers (see para. 2.5).

If implemented, this architecture of interactive "pyramids-on-pyramids" may develop unexpected behaviors. It dilutes any specific meaning of the perceptual state concept. It also leads to a major problem: What is the scaling metric for higher level pyramids?

2.2.3 Textons and statistics. Bela Julesz never hesitated to identify the texton and the feature-detector concepts (Julesz, 1981). They are both just different names for low-level visual primitives (i.e. "atoms of perception") and, in principle, may be extended to any visual entity independently of its complexity (the "grand-mother" detector). The problem, of course, is of defining what a visual entity is. For the texton theory, *crossings* and *terminators* were important perceptual "atoms". Which brings us to (in this particular case, binary) *statistics*.

The texton story started with the idea that, from a Fourier point of view, two stimuli (textures) with identical power spectra can be discriminated only on the basis of their *spatial phase* characteristics. Black-and-white stimuli with identical power spectra are also identical in terms of their second order statistics (they are *iso-dipoles*) but they are not necessarily identical in terms of their higher order statistics. In the Fourier domain, statistics of an order higher than 2 may always be related to the phase spectrum of the stimulus.

Julesz and col. showed that the iso-dipole texture-pairs they initially used, were not "instantaneously" discriminated and concluded that what they had already coined as the *preattentive* visual system was not sensitive to spatial phase. Within this theoretical framework, discrimination based on spatial phase (or higher-order statistics) requires *scrutiny* (i.e. some kind of ill-defined mind's eye search process). Later on, Julesz and col. found a few iso-dipole texture-pairs readily discriminable. Since, according to them, the first set of experiments had shown that spatial phase information could not be processed without scrutiny, they concluded that discrimination of texture-pairs which do not share 3rd or higher order statistics must be based on the analysis of very *local* and distinct patterns which they called *textons*.

The first logical step having led to the texton concept was not sufficiently validated. First, the phase distortion in the texture- pair was not quantified and it was probably quite small. Besides, whether attentive or preattentive, phase discrimination as such is quite poor to start with. It is hence an error to conclude on the basis of the above experiments that phase information is not processed by early vision. Second, phase information may be regarded as the relevant parameter only to the extent that one has in mind a *global* Fourier analysis.

Since the description level at which these textons could be characterized was not obvious, Julesz and col. "scrutinized" the texture-pairs producing high and low discrimination performances and pinpointed some specific shapes which they called "blobs", "terminators", "crossings", "connections"... In the last few years, a series of papers has demonstrated, however, that all typical and apparently atypical cases of texture discrimination can be accounted for by the parallel processing of the image by a population of local linear filters at different spatial scales. The texton no longer had any reason to exist.

The idea that the visual system might compute and subsequently discriminate visual stimuli on the basis of their n^{th}-order statistics was new and insightful. Conceptually, it can be regarded as one of the first attempts to get rid of the *SNE* doctrine. Statistics must be computed over a large number of units which do not need to be labelled with respect to the dimension along which discrimination takes place and whose correlative perceptual states thus become irrelevant. The perceptual state is related to the statistics themselves.

While computation by the visual system of n^{th}-order statistics did not receive experimental support, it definitely prefigurated the *connectionist* philosophy (see para. 2.5), as well as recent electrophysiological research demonstrating resonant activity in neural populations (see Gray in this volume).

2.3 Matching as a perceptual state

Research in stereopsis (Julesz, 1960, 1971) and motion perception (Reichardt, 1961) led in the early sixties, to the formulation of the concept of *matching* as a direct substrate of perceptual states. The underlying idea was that a given sensation is characterized by the extent to which the activities of a given (rather than of any other) pool of neurons are matched (or cross-correlated) in space (for stereopsis) or in space and time (for motion). Out of the very large number of possible binary matchings, only those which are *globally coherent* (or concordant) are finally selected through global interactions.

This formulation only apparently solves the dilemma introduced by the *SNE* doctrine: relating perceptual states (as well as states of mind) to matching states in the brain does not require, in principle, the use of labelled *primitives (see para 2.5)*. Posing that a given perceptual state depends on the matched activity within a neural population does not exclude that it also depends on the particular neurons involved in the matching process. On the other hand, identical neural populations may give rise to very different perceptual states.

Depth perception is in all respects distinct from motion perception. The underlying matching processes, as modelled, are of a very different kind. But so are the neurons subserving each of the two perceptual states. In contrast, motion and texture perception may be related to very similar matching processes across similar or identical cell populations (Gorea and Papathomas, 1990). The remaining two combinations are also possible. Is thus the specificity of sensations related to the process (of matching) or to its neurophysiological substrate? Whether blunt or dull, this question has no obvious answer. Definitely no more than "Where are the nervous sites of our perceptions?" Hence, the use of the matching concept as the neurophysiological counterpart of a perceptual state is not entirely independent of the *SNE* doctrine.

2.4 The computational approach

In order to build a machine that "sees", one is facing conceptual problems analogous to those encountered in the process of unveiling the nature of biological vision (see Ullman in this volume). Marr's work (1982) is exemplary in having intimately combined these two domains of research.

One of Marr's conceptual contributions to the study of biological vision was to let the neural "matching" process be guided by **real world constraints**. A second contribution was to reverse the perspective of the current theoretical inquiry. Instead of asking "What is it [the visual system] doing?", he asked "What is it supposed to do?" The underlying idea is that descriptions of function should provide information about (neural) substrate. The analysis of the "natural stimulus" and of the "biologically plausible functions" of a system vis-à-vis this "natural stimulus" may stand as a revival of Gibson's (1966, 1979) philosophy and as the ultimate concept behind the *computational* approach.

On the one hand, one may say that Marr's approach stressed the *natural* stimulus-end of the visual process. Vision (but also any other sense and for that matter, experience) is constrained. More than anything, the job of the vision scientist is to realize, inspect, understand and determine how the system reacts to and takes advantage of those constraints.

Constraints are physical in the sense that the physical arrangement in space and time of the visible matter determines the nature of *visual assumptions* concerning the visual meaningfulness of that physical matter (see the chapters by Anstis and by Cavanagh in this volume). This makes the irrefutable point of materiality, namely that vision is the stimulus-visual brain conjunction.

On the other hand, constraints are biological in the sense that *we see what we need*. This makes an ambiguous point. It may have been intended to mean that the system "needs" specific information concerning some vital functions of ours like moving within a sophisticated environment. But it may also mean that the system "needs" something sufficiently well specified to be experimentally evaluated by some master biologist. The point here is that it is equally likely that *we need what we see*.

The distinction between *seeing what we need* and *needing what we see* is crucial when elaborating the concept of *perceptual state*. In the first case, perceptual states are given *a priori*. In the second case, they are physically determined. It is thus the second alternative which leads explicitly to the specification of perceptual states in terms of physical dimensions. But it is also the alternative which objects to the interest of asking "What is the system supposed to do?". This contradiction in the premises of Marr's thought is, of course, inherent to the nature/nurture dilemma. The consequence of which is that the specification of perceptual states remains a paradoxical matter.

It is the inherent implementation of Marr's approach which, despite its formal rigour, brings us back to the *SNE* concept. Whether explicitly or implicitly accepted, processing stages and parallel processes have *meaning*. They are labelled. Knowledge about "out there" is provided *directly* at/within those processing stages and parallel pathways. Of course, in a strictly computational sense, knowledge is a purely decisional matter, but one may argue that sensing and interpreting is also a decisional matter.

The specification of the physical and biological constraints of visual behavior is not necessarily an objective matter. Certainly, the visual system of some diving birds has adapted so as to automatically correct for refraction errors. But is there any *objective* reason explaining why our visual systems did not evolve to process infrared light? Why is our retina inhomogenous? And then, why don't we fly? Etc.

2.5 The connectionist approach - Networks

There is (almost) nothing new under the sun. Most fashionable nowadays, networks are "matching" devices. In addition, the connectionist approach leans heavily on the idea that meaningfulness of neural processing is intrinsically related to the interaction of processing units both with the outer world and among themselves. This implies that the units themselves and their interconnections (networks) are (or must have been at some point) memory devices. The idea that memory is a distributed process may be traced back to William James and is unanimously accepted nowdays.

What is new about the connectionist approach is that it might offer a possible solution to the problem of "high-level-vision". The solution is conceptual and, to the extent that it can be *simulated*, it is objective. The notorious problem with this approach is that it is (notoriously) untestable.

"High-level-vision" is certainly something ill-defined. Being ill-defined may have hidden advantages. For example the advantage of insinuating that the concept of perceptual state is itself ill- defined.

Most will agree that there is more to vision than orientations, disparities, movement detectors and so (see Barlow in this volume). The connectionist approach has just started to face problems such as shape and size constancy, 3-D recovering from 2-D representations (also addressed by the computational approach), etc. Of course, most will also agree that there is more to vision than shape and size constancy... The question is how much more. The question is, *How do we define the scope and represent the complexity of what vision is supposed to account for in our behavior?*

Going beyond early vision is a dominant preoccupation today (see Cavanagh in this volume) and the connectionist approach is simultaneously a consequence of this preoccupation and a means to study (simulate) behaviors related to it. From the standpoint of the present discussion, the connectionist approach appears to dilute the problem of both sensory centers and perceptual states. Accounting for complex visual behaviors such as watching a yellow submarine or visualizing a tempest in terms of specific sensory centers and perceptual states is definitely an uneasy task. In what sense would these two behaviors be qualitatively different?

The *SNE* doctrine is tautological with the concept of specific-neural-sites-distinct-perceptual-states. In that respect, the connectionist approach may be the alternative solution. The "states" of a network, which are difficult to qualify as qualitatively different, are perceptual states. An untestable solution... Unless, contrary to traditional modelling and experimentation, simulation is to be accepted as scientific proof.

3 CONCLUSION

Things (and thoughts) can be indefinitely more confusing. Consider this. When you listen to a complex tone, you may, especially if you are well trained, pick up some of its components. Recent papers suggest that this kind of selectivity indicates that the specific underlying filters do have "direct access to perception" (e.g. Welch, 1989). Would those scientists agree on the reciprocal, *viz.* that "direct access to perception" necessarily implies the existence of specific filters, mechanisms and what more? Probably not, if you consider that "direct access to perception" of a yellow submarine does not imply the existence of a yellow submarine specific detector...

It is consensually accepted that *access to perception* refers to a sensorial (visual) entity. It is generally implied that if the neural substrate of a sensorial entity is itself a neural entity (namely that it may be spatially localized in the cortical space) specific for analyzing a given physical (or otherwise conceptual) dimension of the stimulus, that neural entity has direct access to perception.

The unanimously shared conviction that we do have direct access to Gabor-patches (as visual primitives), to oriented edges (by the virtue of zero-crossings), to red (but also to yellow) etc., is puzzling. How is that anatomically possible? If cells in V4 code *color as seen* (i.e. respect color constancy - Zeki, 1980), what about visual behavior accounted for by the activity of CGL, color-opponent cells? Through what path do the latter access perception?
What shall we think about the perceptual status of a feature-detector, if the only evidence we have about its materiality is obtained *via* stimulation with an *exclusive* class of stimuli, whether defined along a physical or otherwise conceptual dimension?

Orientation and spatial frequency specific detectors exist "beyond any doubt" and their stimulation is positively assumed to account for the capacity of "identifying" our own orientation- and frequency-related sensations. However, all evidence is against the slightest capacity of visually identifying the harmonic components of a square-wave grating.

Suppose that we have a metric for ordinating faces. Are we sure that selective adaptation, masking, subthreshold summation and the like experiments with faces *would not* provide results equivalent to those obtained with sinusoidal gratings? What would our conclusions be?

*

None of the insights provided by the theoretical (and conceptual) approaches of these last decades has been proven definitely wrong. What we know about vision today is what all of them taught us.

We (think we) know that our visual system is built up of (spatial and spatio-temporal) orientation detectors, face and hand detectors, and also of more or less narrowly tuned (chromatic, but also color, spatial and temporal frequency, disparity, etc.) filters and of more or less specific (X-Y, magno-parvo, luminance-chrominance, etc.) pathways... We also have the firm conviction that all these detectors and filters and pathways, all of which *must* have perceptual meaning and thus direct access to perception, interact within rather huge networks whose states also have perceptual meaning, presumably at a higher complexity level...

A few might think they even know that perceptual meaning is a perfectly useless concept. Like the ether, say. But, a "unifying" theory of vision making the economy of this concept has not as yet been proposed. The *SNE* doctrine is the doctrine of perceptual meaning. As such, it could never be formulated as a question to be answered experimentally. It is a state of mind.

The *SNE*'s paradigmatic nature may be looked at in a different way. If we mix all the required ingredients: feature-detectors, one-and multidimensional filters, matching devices, pyramids-on-pyramids, parallel pathways and distributed processing *plus* a rich ecological visual environment (in Gibson's (1979) sense) and if we let it be, this artificial system *must* develop a perceptually meaningful behavior identical to that of our visual brain. It seems to me that this unescapable conclusion is rooted in the philosophy according to which understanding the visual brain cannot go beyond this isomorphical, but also circulary, explanation (see chapters by Klein and by Tyler in this volume).

Visual behavior is meaningful. Meaningfulness does not require consciousness. The purpose of studying visual behavior is to uncover the neural substrate of visual meaningfulness as defined at a given moment. Sooner or later, this is achieved either when the neural substrate (or a model of it) appears to match the meaningful behavior as defined, or when we manage to redefine meaningfulness such that it matches a given substrate. The problem of the appropriate stimulus matching a sensory entity as *experienced* is nowadays as intact as it has ever been.

ACKNOWLEDGEMENT. I am particularly grateful to Bela Julesz, Patrick Cavanagh, Christopher Tyler, Stanley Klein, Horace Barlow, Shimon Ullman and Maggie Shiffrar for their constructive comments on earlier versions of this paper.

REFERENCES

Barlow H.B. (1953) Summation and inhibition in the frog's retina, *J. Physiol (London)* **119**, 69-88.

Barlow H.B. (1972) Single units and sensations: a neuron doctrine for perceptual psychology? *Perception* **1**, 371-394.

Boring E.G. (1942) *Sensation and perception in the history of experimental psychology*, New York, D. Appleton-Century Company.

Burt P.J. (1987) The interdependence of temporal and spatial information in early vision, In *Vision, brain and cooperative computation* (Eds. M.A. Arbib & A.R. Hanson), Cambridge, MIT Press.

Campbell F.W. & Robson J.G. (1968) Application of Fourier analysis to the visibility of gratings, *J. Physiol (London)* **197**, 551-566.

DeYoe E.A. & Van Essen D.C. (1988) Concurrent processing streams in monkey visual cortex, *Trends Neurosci.* **11**, 219-226.

Enroth-Cugell C. & Robson J.G. (1966) The contrast sensitivity of retinal ganglion cells of the cat, *J. Physiol (London)* **187**, 517-552.
Gibson J.J. (1966) *The senses considered as perceptual systems*, Boston, Houghton Mifflin.
Gibson J.J. (1979) *The ecological approach to visual perception*, Boston, Houghton Mifflin.
Gorea A. & Papathomas T.V. (1990) Texture segregation by chromatic and achromatic visual pathways: an analogy with motion perception, *J. Opt. Soc. Am. A* **7**
Graham N.V.S. (1989) *Visual pattern analyzers*, New York, Oxford University Press.
Hubel D. & Wiesel T.N. (1959) Receptive fields of single neurones in the cat's striate cortex, *J. Physiol. (London)* **148**, 574-591.
Hubel D. & Wiesel T.N. (1968) Receptive fields and functional architecture of monkey's striate cortex, *J. Physiol. (London)* **195**, 215-243.
Hubel D. & Wiesel T.N. (1977) Functional architecture of macaque monkey visual cortex, *Proc. R. Soc. London* B. **198**, 1-59.
Julsez B. (1960) Binocular depth perception of computer-generated patterns, *Bell Syst. Tech. Jour.* **39**, 1125-1162.
Julesz B. (1971) *Foundations of cyclopean perception*, Chicago, University of Chicago Press.
Julesz B. (1981) Textons, the elements of texture perception and their interaction, *Nature* **290**, 91-97.
Kuhn T.S. (1962) *The structure of scientific revolutions*, The University of Chicago, 1st edition; 1970, 2nd edition.
Livingstone M. & Hubel D. (1988) Segregation of form, color, movement and depth: anatomy, physiology and perception, *Nature* **240**, 740-749.
Marr D. (1982) *Vision*, San Francisco, Freeman & Co.
Marr D., Ullman S. & Poggio T. (1979) Bandpass channels, zero- crossings, and early visual information processing, *J. Opt. Soc. Am.* **69**, 914-916.
Müller J. (1844) *Handbuch der Physiologie des Menschen*, 4th edition, Coblenz, Verlag von J. Hoelscher.
Reichardt W. (1961) Autocorrelation, a principle of evaluation of sensory information by the central nervous system, In *Sensory coding* (Ed. W.A. Rosenbluth), New York, John Wiley.
Sachs M.B. Nachmias J. & Robson J.G. (1971) Spatial-frequency channels in human vision, *J. Opt. Soc. Am.* **61**, 1176-1186.
Treisman A. & Gelade G. (1980) A feature integration theory of attention, *Cognitive Psychol.* **12**, 97-136.
Watson A.B. & Robson J.G. (1981) Discrimination at threshold: labelled detectors in human vision, *Vision Res.* **21**, 1115-1122.
Watson A.B., Barlow H.B. & Robson J.G. (1983) What does the eye see best? *Nature* **302**, 419-422.
Welch L. (1989) The perception of moving plaids reveals two motion-processing stages, *Nature* **337**, 734-736.
Zeki S. (1980) The representation of colours in the cerebral cortex, *Nature* **284**, 412-418.
Zeki S. & Shipp S. (1988) The functional logic of cortical connections, *Nature* **335**, 311-317.