

---

# Solving the "Real" Mysteries of Visual Perception: The World as an Outside Memory

J. KEVIN O'REGAN *Groupe Regard,  
Laboratoire de Psychologie Expérimentale, CNRS, EHESS, EPHE,  
Université René Descartes, Paris*

**Abstract** Visual science is currently a highly active domain, with much progress being made in fields such as colour vision, stereo vision, perception of brightness and contrast, visual illusions, etc. But the "real" mystery of visual perception remains comparatively unfathomed, or at least relegated to philosophical status: Why it is that we can see so well with what is apparently such a badly constructed visual apparatus?

In this paper I will discuss several defects of vision and the classical theories of how they are overcome. I will criticize these theories and suggest an alternative approach, in which the outside world is considered as a kind of external memory store which can be accessed instantaneously by casting one's eyes (or one's attention) to some location. The feeling of the presence and extreme richness of the visual world is, under this view, a kind of illusion, created by the immediate availability of the information in this external store.

**Résumé** À l'heure actuelle, de nombreux travaux sont exécutés dans le domaine de la perception visuelle et des progrès considérables sont réalisés en ce qui concerne la vision des couleurs, la vision stéréoscopique, la perception de l'intensité et des contrastes, les illusions visuelles, etc. En comparaison de ces progrès toutefois, le «véritable» mystère de la perception visuelle demeure insondé ou relégué au rang philosophique: comment se fait-il que nous puissions si bien voir avec un appareil visuel qui semble si piètrement conçu?

Dans le présent document, je me pencherai sur plusieurs défauts de la vision et sur la façon de les surmonter selon les théories classiques. Je ferai une critique de ces théories et proposerai une autre approche selon laquelle le monde extérieur constitue en quelque sorte une mémoire externe à laquelle il est possible d'accéder instantanément en fixant les yeux (ou l'attention) sur un endroit précis. Dans cette perspective, l'apparence de la présence du monde visuel et son extrême richesse sont une sorte d'illusion créée par la possibilité d'accéder sur-le-champ à l'information contenue dans cette mémoire externe.

Figure 1a is a diagram of the eye of a horseshoe crab. It is constructed in a logical way, with the photosensitive layer directly facing the incoming light. In contrast, the human eye, like that of other vertebrates (Figure 1b), is constructed in a curiously inverted manner: Before reaching the photosensitive rods and cones, the light must first traverse not only a dense tangle of neural matter formed by the axons and layers of neurons that serve the first stages of visual computation, but also a vast web of blood vessels that irrigate the retina (Figure 1c). Both of these obscure the photosensitive layer and would be expected to impede vision. An additional defect of the human retina is related to the fact that the axons and blood vessels come together into a sort of cable that leaves the ocular globe at a place which is about 10-13 degrees on the nasal side of the retina, where there can be no photosensitive cells. The resulting "blind spot" is surprisingly large, subtending a visual angle of about 3-5 degrees, which corresponds to the region obscured by a lemon held at arm's length. Other apparent defects of the retina are its severe nonuniformity. There is no region where cones are arranged with uniform spacing. Rather, as eccentricity increases, the inter-cone distance increases rapidly and strongly all the way across the retina. Indeed, this is true *even within the fovea*, since cone separation increases at the same rate within the fovea, across the macula and up to about 14 degrees into periphery. Thus, contrary to conventional wisdom, even the fovea is not a region of uniform acuity. In addition to the strong gradient in cone spacing across the retina, a further apparent defect of the retina derives from the increasing numbers of rods present beyond about 3-5 degrees from center, and the thinning of the yellowish macular pigment, both making colour vision strongly non-homogeneous. Optical aberrations off the optical axis, and the two diopter difference in lens power for red and blue light also degrade the quality of the image. Finally, saccadic eye movements create calamitous smearing and displacement of the retinal image; fixation accuracy during normal activity such as walking may be far from perfect, with retinal image slip possibly attaining 4 degrees per second (Steinman & Collewijn, 1980).

And yet, despite all these defects, vision seems perfect to us: the world does not seem of different resolution or colour at different eccentricities, and there is no obvious hole in each eye's field of view corresponding to the position of the blind spot. We are not generally aware of colour fringes or other optical aberrations off the optical axis. The smearing and displacement of the retinal image caused by saccades and fixational instability is not usually noticed.

Explanations for these phenomena are generally not considered in textbooks on vision, and visual scientists tend to avoid them. Yet it seems to me that they are the "real" mysteries of visual perception. Even though classic visual phenomena like the illusions and effects displayed in science museums and the specific domains currently discussed by visual scientists, such as colour

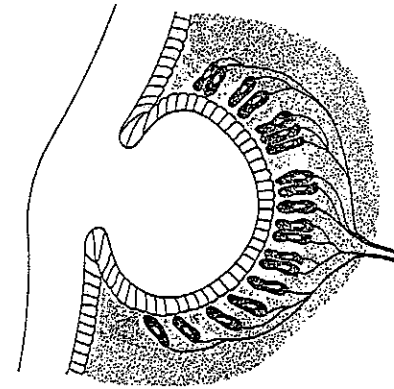


Fig. 1a An ocellus or eyespot. The light-gathering vesicle illustrated is one of many kinds found in lower organisms. This one is found in the horseshoe crab. It is tiny and the drawing has been much enlarged. (Figure and caption from Gibson, 1966, Fig. 9.9, with permission).

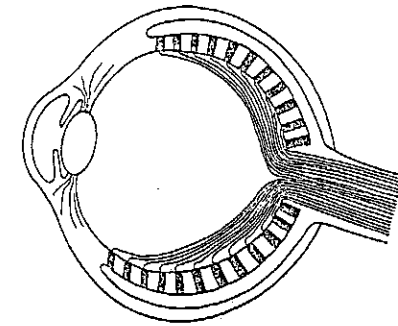


Fig. 1b The chambered eye of a vertebrate. This organ also has an image-forming lens but the retina is *inside out* as compared to the mollusc eye. That is, the nerve fibers from the receptive units are gathered together in front of the retina, not behind it, and emerge through a hole in the retina. The vertebrate optic nerve thus constitutes a flexible cable, an arrangement that permits the eyeball to be freely mobile within its bony orbit. (Figure and caption from Gibson, 1966, Fig. 9.8, with permission).

vision, stereopsis, movement perception, contrast sensitivity, etc., are important and interesting, they are in a way just the tip of the iceberg in the task of understanding vision. The deeper mystery of why we can see so well with such a terrible visual apparatus remains comparatively unfathomed.

In the present paper I will start by considering the classic explanations for two specific instances of the "real" mysteries: our lack of awareness of the blind spot, and our lack of awareness of the perturbations caused by eye movements. The explanations will involve "compensatory mechanisms" that implicitly assume the existence of an internal representation like a kind of panoramic "internal screen" or "scale model" which has metric properties like the outside world. I shall present problems with this idea, and suggest an alternative view in which the outside world is considered a form of ever-present external memory that can be sampled at leisure via eye movements. There is no need for an internal representation that is a faithful metric-preserving replica of the outside world inside the head. In a second part of this paper I shall raise the related question of how objects are recognized independently of the position on the retina on which they fall.

The ideas I shall put forward are closely related to those propounded at different times by Helmholtz (1925), Hebb (1949), Gibson (1950, 1966), MacKay (1967, 1973) and more recently by Turvey (1977), Hochberg (1984),

and Haber (1983), among others. Everything I shall say has probably already been proposed in one way or another by some of these authors. In the context of the contemporary debate about whether perception is "indirect", in the empiricist, (Berkeley, Helmholtz) tradition, or "direct", in the phenomenologists' (Mach, Hering) tradition (see Epstein, 1977, and Hochberg, 1988, for histories of this distinction), these authors might not want to be put together into the same bag. However, in my opinion, and as noted by Hochberg (1984), the distinction between the indirect and the direct theories may disappear, depending on how the theories are fleshed out, and both contribute to the truth. Moreover, the point I wish to make here concerns not the question of indirect versus direct perception, but the question of what visual perception is, or, put in another way, the question of what it means to "feel like we are seeing". I shall claim, and this is consistent with the views of the above authors, that many problems in perception evaporate if we adopt the view that the brain need make no internal representation or replica or "icon" of the outside world, because it is continuously available "out there". The visual environment functions as a sort of "outside memory store", and processing of what it contains can be done without first passing through some intermediate representation or what Turvey (1977) calls 'epistemic mediator'. Even if my viewpoint is not original, the recent flurry of experiments on "trans-saccadic fusion", plus the incredulity I have received with regard to the translation (in?)-variance experiment (Nazir & O'Regan, 1990) to be described below, lead me to believe that the viewpoint is worth bringing again to the attention of the community of workers involved in studying reading and scene perception. I merely hope that my own rendering will serve to make more amenable a view that seems to have been neglected.

COMPENSATORY MECHANISMS AND THE "INTERNAL SCREEN"  
*Compensating for the blind spot*

In the classic textbook explanation of why we do not see the blind spot, it is assumed that the brain "fills in" the missing information by some kind of interpolation scheme that perceptually inserts material into the region of the blind spot based on what is in its immediate vicinity. This provides an explanation of why it is that neither homogeneous or uniformly textured regions appear to have a hole in the place where the blind spot is situated. As far as I know, no serious testing of this idea has ever been done (but see an interesting study by Ramachandran & Gregory, 1991), although it is mentioned in virtually every textbook that discusses the blind spot. The notion of "filling in" is also often invoked in studies on brightness perception or contour illusions, where it is sometimes suggested that colour "flood-fills"

<sup>1</sup> In addition, Hochberg (1984) has pointed out that over the years the Helmholtz theory has become caricatured and elevated to the status of a straw man. The same is true of Gibson's theory, as witnessed by Ullman's (1980) critique.

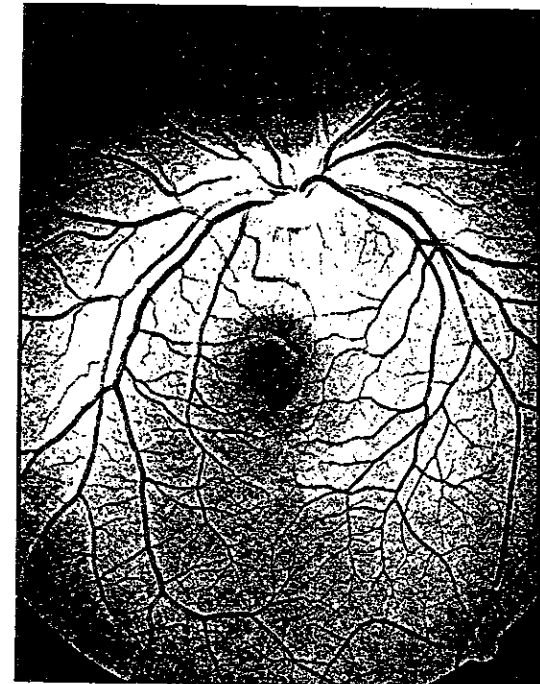


Fig. 1c Angiography of the ocular fundus of a human subject, taken with a green filter. The fovea is the dark central region, and the blind spot the white region on the left. The vast web of veins and arteries leaving the blind spot is visible. The arteries have a central white streak, the veins are uniformly dark. Also just visible are faint white striations converging on the blind spot in arcs from above and below. These are the axons from ganglion cells all over the retina converging on the blind spot. Like the blood vessels, these fibres lie on the inner side of the retina, so the light must pass through them before impinging on the photosensitive layer. Photograph provided by J.F. Le Gargasson, Service de Biophysique du Prof. Grall, Hôpital Lariboisière, Paris.

regions delimited by contours. (e.g. Gerrits & Vendrik, 1970; Grossberg & Mingolla, 1985; Paradiso & Nakayama, 1991).

Note now that although it is not generally explicitly mentioned, the notion of "filling in" implicitly assumes the idea that what a viewer has the subjective impression of "seeing" is something like a photograph, that is, something that has metric properties like those of our visual environment. The function of the interpolation scheme is to fill in the missing parts of this metric representation<sup>2</sup>.

<sup>2</sup> Admittedly the "filling-in" could fill in something into a non-metric representation. But I believe that most people who have thought about "filling in" do not consider this possibility. Certainly when the concept is used in the literature on brightness perception, some kind of cortical map having metric properties is assumed to be completed by a filling-in process (e.g. Grossberg & Mingolla, 1985).

### Compensating for eye movements

The idea of a metric-preserving internal representation like a photograph is also implicit in the mechanisms postulated to compensate for the defects caused by eye movements.

Eye movements interfere with visual perception in two ways: They *smear* the retinal image and they *displace* it. Smearing arises because the retina has an integration time of about one tenth of a second (c.f. Coltheart, 1980), so when the image sweeps across the retina during the 20-50 ms duration of a saccade, all the visual information accumulated over the time just before, during, and just after the saccade, will essentially be averaged or smeared together. The effect can be simulated with the eyes stationary by shifting the image or by flashing a luminous grey field during the estimated saccadic duration. Why is it that this "grey-out", which happens three to five times per second all the waking day, is not noticed?

To account for this, Volkmann, Schick and Riggs (1968; also Holt, 1903) suggested the existence of a "saccadic suppression" mechanism, which acts something like a faucet: When the brain sends the command for an eye movement, it turns off the faucet which allows visual information to enter, thereby locking out the expected smear. It now appears that a significant portion of the saccadic suppression mechanism might stem from retinal masking factors (e.g. Burr, 1980; Campbell & Wurtz, 1978; Yakimoff, Mitrani & Mateef, 1974; see E. Matin, 1974, for a review of saccadic suppression).

*Displacement* of the retinal image is the second type of perturbation caused by saccades: Elements of the image which impinge on one retinal location before the saccade, end up being at different locations when the eye comes to rest after the saccade. A similar displacement of the image can be obtained artificially by pressing on the side of the eye with the finger: When this is done rapidly, a shift of the world is perceived. Why is it that this shift is easy to see, but that when it occurs via an eye saccade, it is not noticed? How is it that we can accurately locate objects in our visual field despite the fact that their positions are continuously being shifted around? How can we fuse together information from successive fixations to give us the subjective impression of a seamless visual environment?

To deal with these problems another compensatory mechanism is usually postulated: the "extraretinal signal" (Matin, Matin & Pearce, 1969). This is a signal which indicates the extent of the saccade which is made, and which can be used to shift the internal representation of the environment in a way that compensates for the actual shift caused by the saccade. There is some debate in the literature concerning the origin of the extraretinal signal: Does it have its source in proprioceptive afference from the extraocular muscles, indicating the actually occurring movement of the eyes? Or does it come from an "efference copy" of the efferent command that gives rise to the saccade (for reviews on these notions see MacKay, 1973; Matin, 1972, 1986; Shebilske,

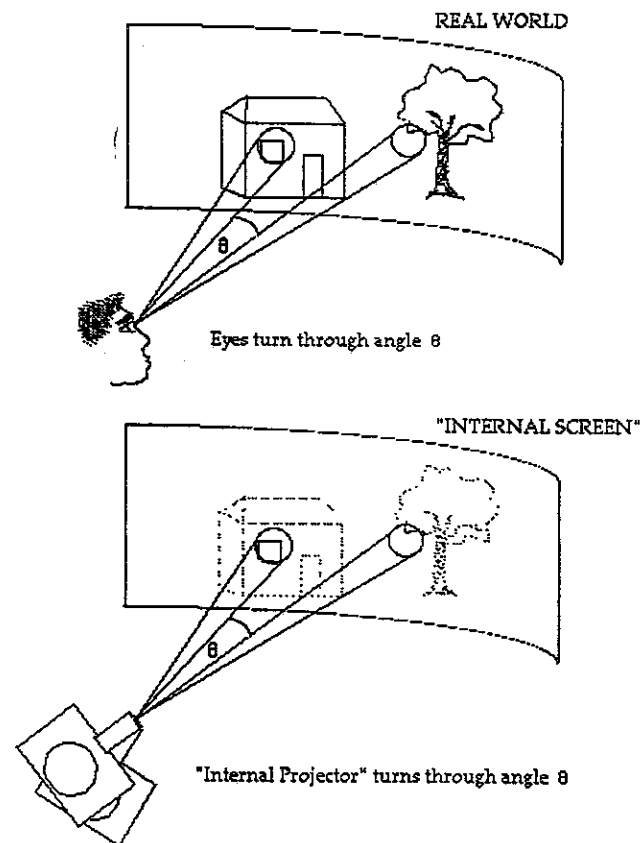


Figure 2. Mode of operation of the "internal screen". When the eye moves through an angle  $\theta$ , the internal projector must move through the same angle  $\theta$  before projecting the new information onto the screen.

1977)? However, despite questions as to its origin, few authors doubt that some signal indicating the extent of saccades is used to compensate for the image shift that they provoke, thereby guaranteeing a seamless visual percept and the ability to accurately locate objects in our environment.

As was the case for the compensatory "filling-in" mechanism postulated to explain why we don't see the blind spot, the idea that mechanisms like "saccadic suppression" and the "extraretinal signal" are needed to compensate for eye movements all implicitly assume that what we "see" has something like photographic quality, like a kind of internal panoramic "screen" (Figure 2) or "integrative visual buffer" (McConkie and Rayner, 1976) or a little 3D model that preserves the metric properties of the outside world. Incoming visual information is continuously being "projected" onto this screen or model, building up the internal representation as the eye scans around the visual environment (Irwin, 1992, has called this the "spatiotopic fusion"

hypothesis; O'Regan and Lévy-Schoen, 1983, referred to "trans-saccadic fusion"). During each stop of the eye, the "filling-in" process compensates for holes and other inadequacies in the projected image. At each eye movement, the internal "projector" is simultaneously moved through a certain angle, given by the "extraretinal signal", corresponding to the amplitude and direction of the saccade that is made. In that way the new incoming information is inserted onto the screen or model in the correct place. During the eye movement the "projector" is turned off so that the resulting smear is not registered: this is "saccadic suppression".

#### PROBLEMS WITH A METRIC-PRESERVING "INTERNAL SCREEN"

The idea of an internal screen or 3D model appears rather caricatural, and has never been explicitly mentioned in the literature on saccadic suppression, on the extraretinal signal or on the blind spot (although cf. Feldman, 1985). But it is nevertheless implicitly present, though probably not in any well worked-out manner, in the minds of researchers, particularly in the case of "filling in" and of the "extraretinal signal": The filling-in operation is rather like what an artist does when he touches up a painting, and this is a metric preserving operation; similarly, the extraretinal signal is an algebraic correction signal which shifts a coordinate system representing the outside world: the idea again implicitly involves the notion of a metric. Both ideas are also supported by the existence, shown by neuroanatomists, of "cortical maps" in the visual pathways that approximately preserve retinal topography. However, several problems arise with the notion of internal screen when it is taken seriously, and when one attempts to imagine how it might be implemented biologically. Some of the most obvious problems will be presented below. Turvey (1977) and Haber (1983) have discussed the issue of the "internal screen" in greater detail. Irwin (1992) has also reached the conclusion that the notion of internal screen must be discarded, and has devoted a series of articles to the task of determining what it should be replaced with.

A first problem with the notion of internal screen comes from the fact that depth information must somehow be coded in the internal screen—so internal "scale model" is a better concept than internal "screen". But it is not obvious how a mechanism would be designed that inserts information onto the scale model depending on the degree of eye convergence and accommodation; further, how would different degrees of focus arising from the different depths be taken into account and combined at a single point?

A similar problem resides in the fact that the internal screen notion requires a mechanism which allows information from successive fixations to be fused together at a single location in the internal screen, despite the fact that the information from the successive fixations may have widely different resolutions and colour quality, depending on which parts of the retina they stem from.

Another problem concerns the accuracy of the extraretinal signal. If it is not perfectly accurate, then errors will gradually build up and the estimated location of objects will be incorrect. This problem might be overcome by some (kind of) recalibration scheme based on the overlap from successive views, rather in the way satellite photographs are aligned. But again, the resolution and colour information from successive views may be very different, and it is not obvious how they can be combined together. A final problem is that not only eye movements, but also head and body movements modify what can be seen, and these should also be taken into account in determining the motion of the "projector".

In addition to the above theoretical problems, a number of recent empirical studies have attempted to determine the exact metrical properties of the internal representation. Perhaps the first such study was Lévy-Schoen and O'Regan (1979) and O'Regan and Lévy-Schoen (1983). At the time, we were convinced that the apparent stability of the visual world implied the existence of an internal metric-preserving representation that accumulates information over successive fixations made in the visual field. To test this idea of "trans-saccadic fusion", we constructed stimulus pairs for which each member of a pair consisted of apparently random lines, but which when superimposed formed a recognizable word (Figure 3). We presented one member of a pair just before the saccade, the other just after the saccade, but both in the same physical location in space. We predicted that even though the two stimuli impinged on different retinal locations, they should appear perceptually as being superimposed in the internal "screen".

In a variety of conditions of stimulus durations and delays between the two stimuli, we never observed the expected fusion. In additional unpublished experiments, we also attempted to favour fusion by drawing an identical frame around each stimulus. Because the frame was common to both pre- and post-saccadic stimulus, we thought it might provide a means for the visual system to correctly align them. However, again, we never found any fusion. Further work by other authors using similar paradigms (Bridgeman & Mayer, 1983; Irwin, Yantis & Jonides, 1983; Rayner & Pollatsek, 1983) has also showed no fusion, and it seems to be the present consensus that the notion of an internal metric-preserving "screen" must be severely questioned (two studies are still in favour of it: Hayhoe, Lachter & Feldman, in press, and Wolf, Hauske & Lupp, 1980; but Irwin, Zacks & Brown, 1990, attempted and were unable to replicate this last study; Irwin, 1992, gives an exhaustive critique of studies on trans-saccadic fusion).

#### DISCARDING THE INTERNAL SCREEN

The important idea underlying the hypothesis of internal screen or model is that what we have the subjective impression of "seeing" is not what is on the retina, but given by the accumulated composite picture of what is on the

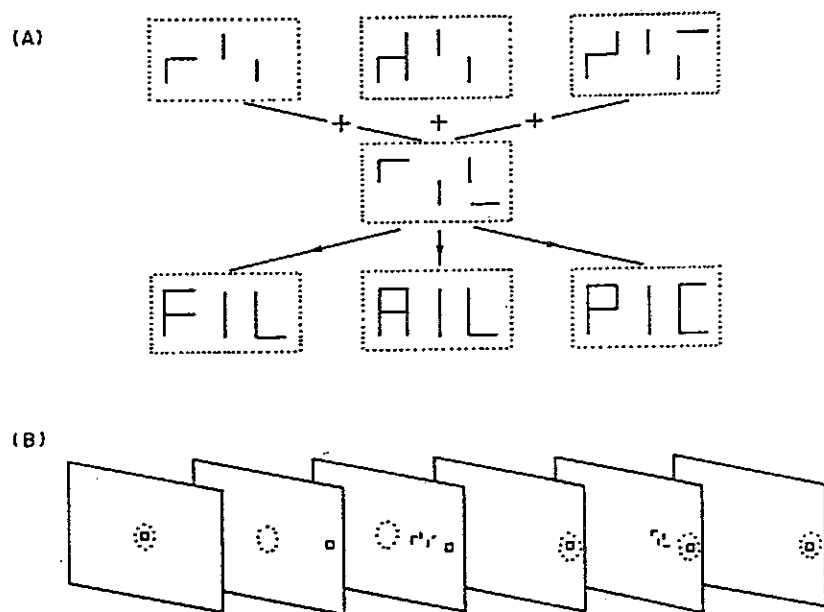


Fig. 3 Stimuli used in the experiment on trans-saccadic fusion (O'Regan & Lévy-Schoen, 1983). (a) There were three possible stimuli, each consisting of two halves. The half that occurred second was always the same. The two halves, when superimposed, formed a three letter word. The word subtended  $2.9^\circ$  horizontally and  $1.7^\circ$  vertically. (b) The sequence of events for an individual trial. The dotted circle shows the behaviour of the eye in a critical trial where one stimulus half appears before the saccade, and one after. The eye is fixating a fixation point in the center of the screen. A target point appears  $8.2^\circ$  to the right or to the left. During the eye's latency period, the first stimulus half appears for 1 ms midway between initial fixation point and fixation target. The eye makes a saccade and arrives at the fixation target. The second stimulus half appears for 1 ms. Its moment of occurrence is always 50 ms after the first stimulus half. (Figure and caption from O'Regan & Lévy-Schoen (1983), by permission). The result of the experiment showed that there is no "trans-saccadic fusion", in which visual information gathered before the saccade is integrated with visual information gathered after the saccade.

screen. However, as has been argued in different ways by Gibson, MacKay, Turvey and Haber, among others, it may be that the notion of internal screen is not necessary. Some arguments similar to theirs are given below.

In cinema viewing, even though the camera cuts continually from one viewpoint to another, viewers have no difficulty apprehending the spatial arrangement of the set. It seems that viewers do not attempt to build up a coherent metric replica of the set, but are satisfied with what might be termed a "semantic" representation of it, containing a number of statements such as: X is talking to Y, they are standing on the beach facing the waves, etc., which are coherent with the viewer's prior knowledge about beach scenes. Viewers appear not to need to know exactly the displacement of the camera, nor do

they appear to calculate the camera's displacements from the visual information they are provided with. Rather, what they may be doing is simply to attempt to qualitatively interpret each shot within the context of their prior knowledge of the set. Any knowledge about position in the scene will be represented in rather approximate terms: 'a little bit left of', 'on the far right', 'several paces behind', etc.

If this can be achieved in cinema viewing, why not in normal circumstances with eye movements? It could be that eye movements interfere with vision no more than camera cuts interfere with cinema viewing. Viewers simply take the incoming information as it comes, and do not attempt to integrate it into a precise, metric-preserving internal representation, but only into a kind of non-metric, schematic mental framework, or structural description. Furthermore, because in normal vision people have active control over their own exploratory eye- or body- movements, making sense of what comes before their eyes is probably facilitated by the fact that the things that do come before the eyes have been actively sought out. There is no need to *compensate* for eye movements, since they are the very *means* by which information is obtained. Of course when scenes change in a way which is out of control of the viewer, as is the case in cinema viewing, conventions of film cutting must be followed so as not to confuse the viewer. In fact, these conventions may provide information about the nature of our mental representations of the visual world. Hochberg's (1968) and Gibson's (1979) interest in cinema viewing appears to stem from this idea. D'Ydewalle (in press) has provided a useful summary of cinema cutting techniques.

#### A TACTILE ANALOGY: THE WORLD AS AN EXTERNAL MEMORY

Another argument against the need for an internal metric representation comes from an analogy with the tactile sense suggested by MacKay (1967, 1973). Suppose I close my eyes and take a bottle in my hand; suppose also that my fingers are spread apart so there are spaces between them. Consider the following tactile analogy of the problem of the blind spot: Why do I not feel holes in the bottle where the spaces are between my fingers? When asked within the tactile modality, the answer seems trivial to us: Why *should* I feel holes there? My tactile perception of the bottle is provided by my *exploration* of it with my fingers, that is by the sequence of changes in sensation that are provoked by this exploration, and by the relation between the changes that occur and my knowledge about what bottles are like. Thus, since I suspect that what I have is a bottle, even though I am not currently touching the neck of the bottle, I expect that if I move my hand up, the feeling I will experience will be one of diminishing diameter, and that ultimately I will encounter the cap or cork of the bottle. But the fact that I don't know yet whether it's a cap or a cork doesn't alter the fact that I am aware of a complete bottle. In fact even if I lift my hand off the bottle, I am still aware of the presence of the

bottle right near me. In summary, the tactile feeling of "perceiving the bottle" is actually a kind of cycle, or at least potential cycle: An action with the hand, causing a change in sensation, being used to modify or confirm an interpretation, being used to guide further action. Said in another way, "perception" is *getting to know* or *verifying* the sensations caused by possible actions. Note that no coherent global metric-preserving representation or model of the bottle is postulated in this account of bottle-perception: There is no need for an internal replica of the bottle since the bottle is continuously "out there", and any question that requires metric knowledge can be resolved by sampling the sensation present on the hand. For example, the base of the bottle is large compared to the hand, the neck is small. If I desire further details about some part of the bottle, I move my hand there. But until I actually wonder about them, I am unaware of their not being present in my consciousness, so I feel no lack. Similarly, I only perceive a lack in my memory of what I had for breakfast when I ask myself a question which I can't answer, as for example which way the jam jar was facing. "Remembering" requires an active interrogation of the past. I'm not right now remembering my grandmother, unless I actually *do it*. Thus, in the example of feeling the bottle, the bottle amounts to an outside memory store that can be interrogated or explored, and the feeling of "perceiving" comes from the exploratory activity *itself*.

If this view of perception is now applied to the visual modality, we would say that we experience the impression of "seeing the bottle" when through some physical action (eye or body movement) or mental (attentional?) interrogation of the outside memory constituted by the visual field, we obtain sensations that are compatible with the presence of a bottle. The "percept" of the bottle is an **action**, namely the visual or mental exploration of the bottle<sup>3</sup>. It is *not* simply the passive sensation we get from the retina or some iconic derivative of the information upon it. Rather, this sensation is being used to supplement a mental schema we have about the results of the possible actions that we can undertake with our eyes (or heads or bodies). We do not see a hole in the bottle where the blind spot is, nor do we see its color or surface quality as less clear in the regions we are not directly fixating, because our feeling of "seeing" comes not from what is on the retina, but from the result of using the retina as a tool for probing the environment. A tool, as for example a ruler, can be used to probe the environment, but not to probe itself: You can measure the length of an object with a ruler, but you can't check whether the ruler itself has changed length!

The notion of the outside memory store may be what Gibson (1950; 1966; 1979) calls the "ambient optic array", and what Turvey (1977) calls the

<sup>3</sup> Hebb's (1949) theory has similarities to this. Interestingly, Hochberg (1984) notes that a similar view was also expressed in Helmholtz's classic theory of perception, where what is perceived are *expected contingencies*, not what is on the retina or some iconic derivative of it.

"ordinal image" in opposition to the "anatomical image". Note that the idea that perceiving amounts to using the retina as a tool to interrogate this outside store leads to two kinds of predictions. Since "seeing" involves both *interrogation* of the visual field, and also *apprehension* or *integration* or *comprehension* within the current mental framework, one would predict that a person would fail to see something either (a) if he or she does not interrogate or wonder<sup>4</sup> about the appropriate aspect of the visual field or (b) if he or she is unable to integrate the obtained sensations into his or her mental framework. In particular, even if you are directing your eyes directly at something, unless (a) you are (at least unconsciously) wondering about it, and (b) you are able to apprehend it, you will not have the impression of "seeing" it<sup>5</sup>. This is compatible with the rather troubling result of Haines (1991), who found that pilots landing an airplane using a "head up display" in a flight simulator (in which the instruments panel is displayed superimposed on the windshield) would often not see a perfectly visible airplane parked in the middle of the runway (an almost inconceivable occurrence), and would blithely drive right through it. Neisser & Becklen (1975), studying how people view videos of two simultaneous, superimposed, action sequences, also concluded that we only "see" what we attend to.

The present view of what "seeing" is should be distinguished from a radical Gibsonian viewpoint, in which internal representations play no role. The idea that the outside world is an external memory store does not imply that no processing of the information in that store is done. On the contrary, I believe that what we have the subjective impression of "seeing" is precisely those aspects of the content of that store which we choose to process or to integrate into our mental framework by virtue of the appropriate cognitive operations.

<sup>4</sup> The notion of selective attention could be invoked here instead of 'interrogation' or 'wondering about'. However the idea of attention sometimes carries with it the idea of sensitizing or activating a metric-preserving (or at least retinotopically organized) cortical map. I want to avoid this notion, since it is a tempting step to take to then imagine that what we have the impression of "seeing" is what is activated on such a map. While retinotopic cortical maps undoubtedly exist, what we "see" is not what is on these maps, but rather what schematic (semantic) representations we construct on the basis of what the maps signal.

<sup>5</sup> It is amusing to reflect on the Wittgensteinian nature of this idea: In proposition 7 of the *Tractatus Logico-Philosophicus* Wittgenstein (1961) said: "What we cannot speak about we must pass over in silence" ("Wörter man nicht sprechen kann, darüber muss man schweigen"). This seems to correspond to our point (b), above, for which we could say here something like: (b) "What we cannot see must remain hidden from us!" (this sounds more impressive in German: "Was man nicht erkennen kann, das bleibt uns verborgen!"). The equivalent of (a) would be: "We cannot see what we don't look at! (or perhaps better: we only see what we notice)" ("Wir sehen nicht was wir nicht anschauen."). For memory we might say: (b) "We recollect only what we *can*" ("Das was man sich nicht erinnern kann, das bleibt vergessen") and (a) "We recollect only what we *try to*!" ("Was man sich nicht in Erinnerung ruft, das bleibt vergessen"). All these statements are only *seemingly* tautological!

## VISUAL VERSUS TACTILE PERCEPTION

A point also needs to be made about the difference between the impressions of "perceiving" via the tactile and via the visual sense. When I feel the bottle with my tactile sense, I cannot say I really feel the **whole** bottle; it would be more accurate to say that I am **aware** of the whole bottle, even though I can currently only **feel** a part of it. On the other hand, in the visual modality, perception is an intensely rich sensation of **total** external presence, and I have the impression I can perceive the whole bottle even when, on closer scrutiny, I realize that the exact shape and colour of its cap are not clearly visible to me because I am fixating elsewhere in the bottle.

Why is there this difference between the subjective wholeness of vision (giving the impression of seeing a whole scene) and the paucity of tactile perception (giving the impression of feeling only a part of an object, even though one is *aware* of the whole)?

If I stand on the edge of a cliff, but with my back to it, I have, as noted by Gibson (1979) an intense awareness of the presence of the cliff, although it is currently not in my visual field. This awareness does not have a precise metric quality, but it strongly influences my potential future actions. A similar awareness of objects in front of one comes when one closes one's eyes. I conjecture that the feeling of "seeing" consists of three parts: the first part consists precisely of this non-metric awareness of the presence of objects in front of one; the second part is the awareness of the possibility of interrogating parts of the environment with the retina as a tool; the third part is a global sensation of "lots of stuff" being on the retina. It is this latter quality that gives the feeling of "wholeness" to vision. Whereas with touch, the size of the zone used to sample the environment is small (finger/hand), with vision, it is enormous. It is as though we had an enormous hand that we could apply to the whole field in front of us<sup>6</sup>. Since in vision we are used to having such enriching sensations over a very wide field of view, tactile perception feels unsatisfactory to us, and does not convey to us the same feeling of outside reality that vision does. But I conjecture that congenitally blind persons, since they have never experienced such a wide field of enriching sensation, do not feel any lack in the wholeness of their tactile world, and in fact "perceive" the world as being just as "whole" and "present" as we do. Blind people are not groping around in the world like we sighted people grope for an unseen object in our pocket. They perceive the world as being thoroughly as "present" as we do<sup>7</sup>.

6 This analogy was suggested by Donald MacKay in a popular lecture I attended in about 1965.

7 In fact, the skin of the body covers a very large surface. We generally pay little attention to the contact our bodies make with our clothes, chairs, the floor, etc., because we are not using our whole bodies to probe the outside world. Possibly blind persons make more use of their bodies as "seeing" organs. This may be related to the notion of "Whole-Body-Seer" referred to by the author of a recent account of what it is like to become blind (Hull, 1991).

Another interesting difference with tactile perception is also a consequence of the very wide field of view afforded by the retina. When I move my hand as I explore the bottle, the position I move it to will be determined primarily by my (internal) knowledge about the bottle. But with vision, the continual presence of stimulation all over the retina provides "signals" which can be used to direct eye movements. Some years ago a person crossed the Atlantic in a kayak. He stored all his food in the front and back of his kayak, each meal attached to a labelled piece of string that he could pull out when he required it. In the same way, the poor quality visual sensations in peripheral vision may serve as signals that allow eye movements to obtain better quality information. But note that no metric representation of the arrangement of food in the kayak was necessary, just a mass of mingled strings providing a connection to each meal. In the same way, it may be that saccades to objects in peripheral vision are not adjusted according to a global metric: perhaps the movement needed to get the eye to a given position is learnt separately for each position. Deubel (1987) has done an interesting adaptation experiment suggesting that learnt saccade amplitude generalizes only over small lobe-like zones.

Why is the view of "seeing" expressed in the above paragraphs so strange to some researchers? I think that because the sensation provided by our retinas is so easily and unconsciously available, be it by an eye movement and/or a mental effort, researchers fall into the trap of thinking that what we see is what is on the retina, or some kind of internal icon. In addition, there is a large cultural heritage of graphical representation (maps, drawings, paintings, photographs, diagrams, film and video) which biases us into thinking that our representations of reality have a similar iconic quality. The neuroanatomy of "cortical maps" is a further biasing factor. But in fact, the impression that we see everything in front of us with the metric quality of a photograph is actually an illusion created by the fact that if we ask ourselves whether we see anything in particular, we can interrogate the external environment via the retinal sensations, possibly after an eye movement, and obtain information about it. But if we do not ask ourselves about some aspect of our environment, then we do not see it. It is the act of looking that makes things visible<sup>8</sup>. As Sherlock Holmes would have remarked: It is not sufficient to have something in front of your eyes to see it! "Seeing" is the action of interrogating the environment by altering retinal sensations, and by integrating these sensations into one's cognitive framework<sup>9</sup>. Anything that is not interrogated or that falls outside the cognitive framework is not seen.

8 Is this a perceptual version of solipsism? Perception is an *illusion* created by the desire to look...!

9 This may be part of the reason why under stabilized vision we see in a fragmentary way. Of course there may be additional mechanisms of habituation—but then these may be present precisely because seeing only requires changes.



## EYE CONTINGENT DISPLAY CHANGE EXPERIMENTS

In recent years, because of the possibility of online computer control of experiments, it has become possible to change in real time what is visible on a computer display as a function of the eye movements that an observer makes. One finding involves the use of text written in aLEtErNaTiNg CaSe. Conditions are set up in such a way that every time the eye makes a saccade, the particular letters that were in one case, change to the other case. Interestingly, the subjective impression one has in reading such continuously changing text is that no change at all is taking place (McConkie, 1979). It is possible to detect that case changes are occurring, but only by making the conscious effort of remembering the case of a letter in a word, reading on, and then coming back to the word to check if the case has changed.

More recently it has become possible to do similar manipulations on high quality colour images such as street or household scenes (McConkie, 1990). It is observed that surprisingly obvious and large objects in a picture, such as cars, lamp-posts and windows can be shifted, removed, or changed in colour during eye saccades, without this being noticed.

At first sight these are surprising findings. However, considered in the light of the present conception of what it is to "see", it becomes apparent that the results are just what is to be expected. "Seeing" the printed page or a picture is not passively contemplating an iconic representation of that page or picture. On the contrary, it involves continuously noting and inserting into one's cognitive framework, the interpretations of the sensory changes that are brought about by shifts of the eye. Eye-movement contingent display changes will only be noticed if the expectations generated before a saccade are sufficiently precise to be contradicted by the changes that the saccade produces. Thus, if I'm not (at least unconsciously) asking myself any particular question about a street scene, and am only checking whether it really is a street scene, then if a car appears or disappears on the road, this might well go unnoticed even though it is large and perfectly visible. (An exception to this would be changes that attract attention irrespective of the (unconscious) interrogation being made, e.g. by creating a flash or some gross perturbation of the picture's overall luminance. Such changes would be noticed.) Another example: If I am reading, since what I am trying to "see" is words, not letters, I don't notice that in the first four sentences of this paragraph some of the "g"'s used have a closed lower loop (g), and some have an open lower loop (g)...!). On the other hANd, it is more likEly one would see the odd letTers in the present sentence, siNce they creAtE a greater visual perturbAtion<sup>10</sup>.

<sup>10</sup> It is unclear exactly what constitutes a "greater" perturbation. I doubt that any purely visual definition of salience would work: I suspect that even a (normally) perfectly obvious flash occurring *during* a fixation might not be noticed if the observer is being very attentive to some aspect of a display (as in playing a difficult video game).

## "INVARIANCE" TO GEOMETRIC TRANSFORMATIONS

Another problem in vision which is implicitly related to the nature of our internal representation of the visual environment is the problem of invariance to geometric transformations: How is it possible to recognize an object independently of the size, position and orientation of its retinal projection? Many visual scientists and workers in artificial vision have considered that this problem in vision is an aspect of what has been called the "inverse optics problem": How does the brain reconstruct the correct three-dimensional representation of objects from the information available in the two retinal images? It seems clear that if vision is seen in this way, that is to say as consisting of a problem of "reconstruction", then an underlying assumption must be that the purpose of the first stages of image recognition is to create a kind of metric-preserving representation similar to the 3D scale model discussed above. It then makes sense to wonder what kind of transformation operators the visual system might possess that enable it to give the same outputs to a figure which occupies different retinal positions or that has been rotated or changed in size.

Various solutions to this problem have been used in the literature on artificial vision. A highly memory-intensive method is what might be called the "brute force memory" method, in which each different view of an object is stored as a separate template, and no transformation algorithm at all is used. A slightly less memory-intensive technique would be to store only a subset of all possible views of an object, and use an interpolation scheme to match those views which have no stored template. Both these methods neglect the operator nature of geometric transformations, and so have the disadvantage that the ability to recognize one object from all viewpoints does not generalize to another object: for each new object, all viewpoints must be learned anew. An alternative technique that does not suffer from this problem consists of storing a representation of the object in a canonical form, and using a global transformation operator to shift, rotate or change its size until it coincides with the canonical form (e.g. Marr & Nishihara, 1978). This method is less memory intensive, but requires more computation. Another method used in artificial vision consists in transforming the image into a representation that itself is independent of the image's size, orientation, etc. (Burkhardt & Muller, 1980; Cavanagh 1985; Reitboek & Altmann, 1984; Schwarz, 1981). For example, a log-polar transformation converts size changes into shifts in the transformed representation. This can then be further transformed using a Fourier transform, which is shift invariant, to render the final transform independent of size. Autocorrelation is another method that has been suggested (Gerrissen, 1982; Kröse, 1985; Uttal, 1975).

Which of these methods, if any, does the human visual system use? The particular linear or logarithmic non-homogeneity in receptor spacing possessed by the retina has been taken as evidence that the visual system may be using

a log-polar transform to obtain size invariance (Cavanagh 1985; Schwarz, 1981). But what little behavioural data there is suggests that such a transform is not used, since, contrary to what it would predict, recognition of a learned pattern may suffer a decrement when it is tested in a different size (Bundesen & Larsen, 1975; Kolers, Duchnick & Sundstroem, 1985; but see Cooper, Biederman & Hummel, 1992).

As concerns invariance to orientation, a large literature on "mental rotation" starting with Shepard and Metzler (1971) and Cooper and Shepard (1973) shows that the time taken to compare a figure to a rotated version of itself is a linear function of the angle of rotation. This has been taken to suggest that humans use a global rotation operator to rotate the figure until a match is obtained. However the evidence now appears less clear cut, because in other paradigms and using other types of stimuli, there are cases when rotation of the stimulus either has no effect on recognition or an inconsistent effect, and the size of the effects depends on the complexity and familiarity of the stimuli and on the degree of practice (see Jolicoeur, Snow & Murray, 1987; Tarr & Pinker, 1989).

The empirical evidence with regard to position changes is sparse. This is surprising, since translation invariance is probably the first problem that must be solved in an artificial image recognition system, and because the problem is even more critical for human vision owing to the inhomogeneity of the retina: Figure 4 is taken from Hebb (1949), and illustrates the dramatic changes in cortical representation of a square that occur when the fixation point is changed within the square.

A first point to note is that eye movements provide a possible mechanism to effect translations of the retinal image, and these might be used to move the image into a canonical position for recognition. Nevertheless, once the object to be recognized falls on a region with sufficient acuity, few people would doubt that it can then be recognized no matter what the exact position is on the retina on which it impinges. However in the few cases in which this assertion has been tested, it turns out that there is in fact a strong dependence of recognition on position fixated (but see Cooper, Biederman, & Hummel, 1992). For example, we have observed that the probability of being able to recognize a word depends strongly on where the eye is fixated in it (O'Regan, 1990, Fig. 9; Nazir, O'Regan & Jacobs, 1991). The time taken to recognize a word also depends strongly on the position within the word that the eye starts fixating (O'Regan, Lévy-Schoen, Pynte & Brugailière, 1984; Vitu, O'Regan & Mittau, 1990; O'Regan & Jacobs, 1992); and this is true even for words as short as four and five letters. A related finding is that of Kahn & Foster (1981) and Foster & Kahn (1985), who showed that discrimination accuracy for dot patterns diminishes as a function of inter-pattern distance, in a way that cannot be accounted for in terms of acuity.

As was the case for size and rotation changes, these studies show that

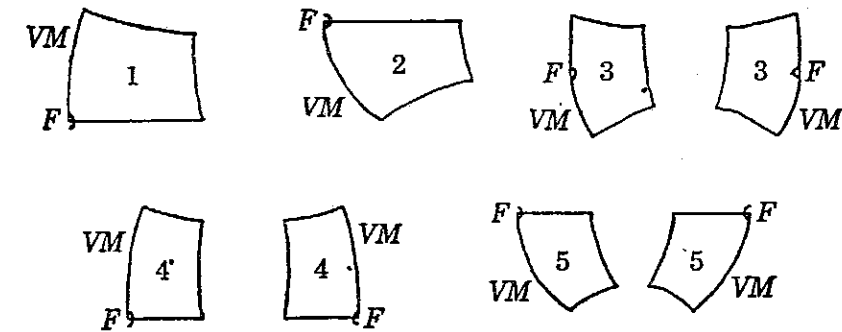


Fig. 4 Diagramming roughly the changes in cortical projection of a square when the fixation point only is changed: based on the data of Polyak (1941) and his Figure 100, for a square subtending a visual angle of  $18^{\circ} 20'$  (the size of the "central area" of the retina). 1, fixation on the upper right corner of the square, which thus falls in the lower left visual field and produces an excitation in the upper right cortex only; 2, fixation on the lower right corner; 3, bilateral projection with fixation on the center of the square; 4, bilateral fixation on the midpoint of the top line of the square; 5, fixation on midpoint of bottom line. F, projection of fixation point; VM, vertical meridian. (Figure and caption from Hebb, 1949)

human vision suffers a penalty in recognition performance when a word is translated to a new position. Part of the reason for this penalty may be that words have distinctive parts which have to be resolved to be recognized, so that when these parts fall on regions of the retina that have lesser acuity, difficulties arise. Note however that recognition is nevertheless generally possible, so some attributes of the stimulus are available with sufficient resolution to allow recognition; some form of translation invariance is therefore present. What mechanism underlies this invariance? In particular, is there some kind of global transformation operator, that can be applied to any translated pattern, or is a brute force memory method used in which each new pattern must be learnt in all possible translated positions?

We attempted to answer the question in an experiment set up to teach people a completely new and unfamiliar pattern (see Fig. 5), but in such a way that it impinged only on a single retinal location (Nazir & O'Regan, 1990). After learning, in a subsequent test phase, we then presented the pattern at other retinal locations. If a global transformation operator is used, then the new pattern should still be recognizable in the new retinal positions, but if brute force memory is used, then recognition should be impossible. The results of the experiment showed that subjects had difficulty performing the distinction at the new location. The first few times a subject saw the target stimuli in a new retinal location, his or her reaction was often one of astonishment: "I've never seen *that* before!" After a few presentations of the small set of stimuli however, subjects were able to make the correspondence with the discrimination they were performing at the initial retinal location and

so deduce which was the target and which were the non-targets, and performance improved. The other interesting aspect of the results was that the results were rather variable: Depending on the stimuli, translation to another retinal location could be either easy or hard; different subjects also had rather different patterns of results depending on the particular stimuli and particular retinal locations being translated to and from.

The result of this experiment surprises many workers in vision, who would have expected perfect translation invariance if acuity is sufficient to do the task. However it seems to me that it is surprising only in the context of the theories of invariance as proposed by engineers, for whom it is important to completely reconstruct the whole metric structure of an object. But it is not surprising if we admit the possibility that no reconstruction is necessary because the image is continuously available "out there": in that case the task of vision is to extract just a sufficient number of cues from this external memory store so that objects can be discriminated from each other and so that manipulation of objects and locomotion are possible. For discriminating patterns therefore, only a small battery of simple components or features may suffice in most cases, and providing these have been learnt at many retinal positions and in many sizes and orientations, then most new patterns can be classified by using these features, and by noting in what approximate spatial relationships they lie. In a task like our translation-invariance experiment described above, when the dot pattern is learnt at the training position, people attempt to extract a few descriptors that allow the patterns to be distinguished. Examples might be "large blob at top right"; or "vertical line near middle"; or "darker at top than at bottom". The notions of "blob", "line" and "darkness" as well as the ability to approximately spatially locate such components within the global configuration of the stimulus, may or may not have been learnt at many retinal locations throughout the long training period of early life. This idea was suggested by Hebb (1949, p. 47-48). An alternative might be that the brain is innately wired to have spatial invariance to a set of features such as these. In any case therefore, when the stimuli are presented in a new retinal location, to the extent that the particular features chosen to recognize the pattern are features that happen to be translatable to the new retinal location, and to the extent that the spatial relations between the features can also be sufficiently accurately reproduced in the new retinal location, the stimulus will be more or less accurately identified in the new location. This explains why the results of our experiment were not all-or-none, and why, depending on the stimuli and on the subjects, different degrees of translatability were observed.

The idea that patterns or scenes are recognized by extracting a small set of descriptors and their spatial inter-relations is of course an old idea: two recent influential promoters are Foster (1984) and Biederman (1987). Humphreys &

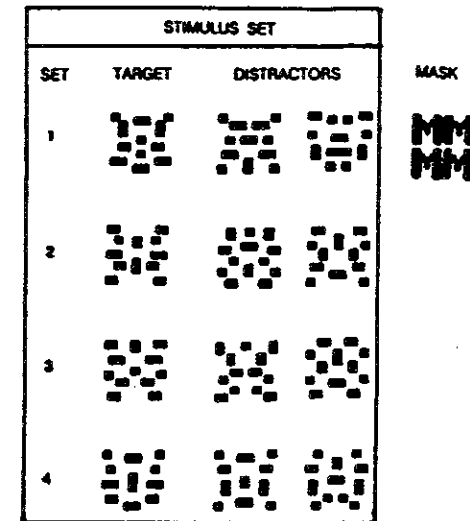
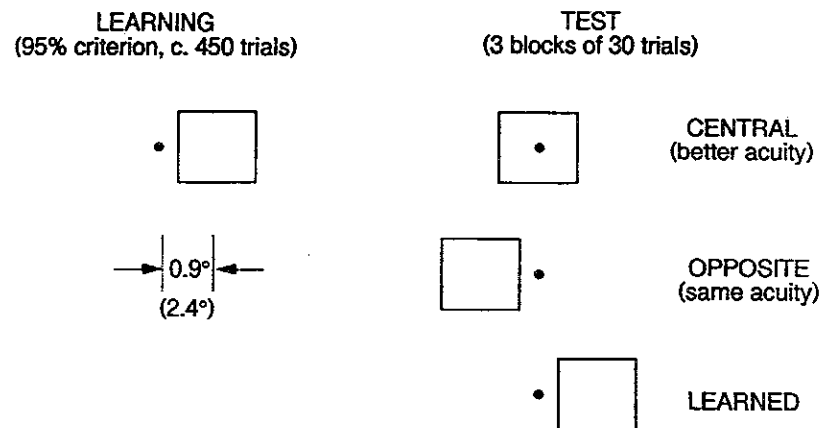


Fig. 5a Dot stimuli used in the translation (in)variance experiments of Nazir and O'Regan (1991). There were four sets of stimuli, each set having one target and two distractors. The stimuli were slightly less than  $1^\circ$  square. In a further experiment, not shown here, stimuli consisting of thin horizontal (or vertical) lines of small squares of different shades of grey were also used.

Bruce (1989) give an excellent survey of current theories. What I have added here is the suggestion that "seeing" does not involve simultaneously perceiving all the features present in an object, but only a very small number, just sufficient to accomplish the task in hand. The subjective impression we have of seeing **whole** objects arises first because the retinal stimulation is very rich and so provides the impression that "a lot of stuff is out there", and second because if at any moment we want to know what exactly any of that "stuff" is, we just use our retinas and eye movements to find out. These ideas qualitatively explain the pattern of results in the translation invariance experiment, in particular the variability between subjects and between patterns. The idea can also be used in a similar way to understand the variability in the results of mental rotation experiments as a function of practise, familiarity and stimulus complexity (Jolicoeur et al., 1987). A related finding is the fact, demonstrated by Thompson's (1980) striking "Margaret Thatcher" illusion, that though a familiar face may be recognized when it is upside down, recognition of the face's particular *expression* (smiling, frowning), may be inaccurate. This shows that recognition did not proceed by global transformation of the whole face. Young, Hellawell and Hay (1987) have suggested that face recognition proceeds by the combination of local features and (global) configurational information. Similarly, the text below ("READING UPSIDE DOWN") seems pretty much correct until you turn it over:

## Task: learn to distinguish target from 2 distractors



- Expt 1: "stabilized" vision, 2.4°  
 Expt 2: 150 ms exposure, 2.4°  
 Expt 3: 150 ms exposure, 0.9°  
 Expt 4: line stimulus at 0.5° above & below

Fig. 5b In the learning phase, subjects were trained on a single stimulus set (for example set 1), presented at a single retinal location, the 'LEARNT' location (for example 0.9° to the right, as shown here, or 2.4° in other experiments). Their task was to learn to distinguish the target from the distractors. When a 95% accuracy criterion had been reached, which took about 450 trials, the test phase began. This involved two short blocks of tests, in which the stimuli were presented at two retinal locations other than those at which they had been learnt (the 'OPPOSITE' and 'central' locations), and a third block again at the 'learnt' location. Two methods were used to ensure that the stimuli impinged on the desired retinal locations. In one experiment, computer-controlled real-time eye movement monitoring allowed the stimulus to be replaced by a mask as soon as a saccade was detected. In three other experiments, a short, 150 ms presentation duration followed by a mask was used, ensuring that no eye movements could occur. The results of the experiments showed a significant deficit in recognition at previously unlearnt retinal locations.

## READING UPSIDE-DOWN

The idea of visual perception involving component extraction is also compatible with Ivo Kohler's (1951) findings, according to which after training with spectacles that transform the visual world in various ways (inverting, reflecting), subjects re-establish normal upright perception in a *fragmentary* way, with aspects of the environment being corrected, and others not. An example given by Kohler is that of a person who, after adaptation to left-right inverting spectacles, saw cars as driving on the correct side of the road, but perceived their licence plate numbers as being written in mirror-writing.

It is interesting to note an important difference between the translation invariance experiment we did, and a picture-priming experiment by Cooper, Biederman and Hummel (1992), in which good evidence for translation invariance was found. The reason for the difference is presumably that in Biederman's experiment the objects used were easily decomposable into the subparts that Biederman calls "geons", and that these may be highly familiar components that have been seen in many locations on the retina (or else they are innately "wired" as translation-invariant). In our experiment however, no such obvious components were present, and subjects had to use ad hoc methods to define aspects of the dot patterns that could be used to differentiate them. This will have rendered translation to new locations more precarious. It is interesting to note that in defining the stimuli for our experiment, we experimented with a number of possibilities. We found that very simple stimuli, like lines of different orientation, could easily be translated. More surprising, very complex stimuli, with a large number of closely spaced dots, were also easy to translate. The reason appears to have been that for any two complex stimuli, it will always be easy to find some simple blob or alignment of dots that can be used to distinguish them, and this simple feature will most likely be translatable. Only when the stimuli are neither very simple, nor very complex, will it be hard to find simple translatable features that can distinguish them.

## Conclusion

Most people are familiar with facts of visual perception such as the Poggen-dorf, Zöllner, and Ponzo illusions, the illusion of dizziness, the Moon illusion, afterimages and aftereffects such as the McCulloch effect and the waterfall and other movement illusions, brightness and contour illusions like the Cornsweet-Crane illusion and the Kanisza triangle—since all these and other what I would call "minor" mysteries are the normal fare of science museums and textbooks. Whole domains of study related to contrast sensitivity, movement perception, colour vision, stereopsis, pattern recognition, etc. are the every-day interest of specialists in visual science. But all these phenomena are eclipsed by what I call the "real" mystery of visual perception: how can it be that we see so well with what an engineer would consider a very badly constructed visual system? Why do we not notice optical aberrations, differences in resolution, defects in retinal structure, and the smear and displacement caused by eye movements? Why does the visual world seem so rich and so perfect to us?

The answer to these questions, I have claimed here, is that they need not be posed at all. Like the concept of the "ether" in physics at the beginning of the century, the questions evaporate if we abandon the idea that "seeing" involves passively contemplating an internal representation of the world that has metric properties like a photograph or scale model. Instead I believe that

seeing constitutes an active process of probing the external environment as though it were a continuously available *external memory*. This allows one to understand why, despite the poor quality of the visual apparatus, we have the subjective impression of great richness and "presence" of the visual world: But this richness and presence are actually an illusion, created by the fact that if we so much as faintly ask ourselves some question about the environment, an answer is immediately provided by the sensory information on the retina, possibly rendered available by an eye movement.

I wish to warmly thank the following people, who commented on the manuscript, or provided useful information or stimulating discussions on related questions: François Bresson, André Bullinger, Irving Biederman, Peter de Graef, Glyn Humphreys, Dave Irwin, Pierre Jacob, Arthur Jacobs, Alan Kennedy, Ken Knoblauch, Ariane Lévy-Schoen, George McConkie, John Morton, Tatjana Nazir, Jacques Ninio, Joël Pynte, Jim Todd.

#### References

- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
- Bridgeman, B., & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society*, 21, 285-286.
- Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology, Human Perception and Performance*, 3, 214-220.
- Burkhardt, H., & Müller, X. (1980). On invariant sets of a certain class of fast translation-invariant transforms. *IEEE Transactions ASSP*, 28, 517-523.
- Burr, D. (1980). Motion smear. *Nature*, 284, 164-165.
- Campbell, F.W., & Wurtz, R.H. (1978). Saccadic omission: Why we do not see a grey-out during a saccadic eye movement. *Vision Research*, 18, 1297-1303.
- Cavanagh P. (1985). Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In D. Rose & V. Dobson (Eds.), *Models of the visual cortex* (pp. 85-95). New York: Wiley.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, 27, 183-228.
- Cooper, E., Biederman, I., & Hummel, J. (1992). Metric Invariance in Object Recognition: A Review and Further Evidence. *Canadian Journal of Psychology*, 46, 191-214.
- Cooper, L.A., & Shepard, R.N. (1973). Chronometric studies of the rotation of mental images. In W.G. Chase (Ed.), *Visual information processing* (pp. 75-176). New York: Academic Press.
- Deubel, H. (1987). Adaptivity in gain and direction in oblique saccades. In J.K. O'Regan & A. Lévy-Schoen (Eds.) *Eye Movements: From Physiology to Cognition* (pp. 181-190). Amsterdam: North Holland.

- Epstein, W. (1977). Historical introduction to the constancies. In W. Epstein (Ed.) *Stability and constancy in visual perception: Mechanisms and processes* (pp. 1-22). New York: Wiley.
- Feldman, J. (1985) Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Sciences*, 8, 265-289.
- Foster, D.H. (1984). Local and global computational factors in visual pattern recognition. In P.C. Dodwell & T. Caelli (Eds.), *Figural Synthesis* (pp. 83-115). Hillsdale, NJ: Erlbaum.
- Foster, D.H., & Kahn, J.I. (1985). Internal representations and operations in the visual comparison of transformed patterns: Side effects of pattern point-inversion, positional symmetry, and separation. *Biological Cybernetics*, 51, 305-312.
- Gerrits, H.J.M., & Vendrik, A.J.H. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. *Experimental Brain Research*, 11, 411-430.
- Gerrissen, J.F. (1982). Theory and model of the human global analysis of visual structure. *IEEE transactions on Systems, Man & Cybernetics*, 12, 805-817.
- Gibson, J.J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- 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.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures and neon color spreading. *Psychological Review*, 92, 173-211.
- Haber, R.N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences*, 6, 1-54.
- Haines, R. (1991). A breakdown in simultaneous information processing. In Stark, L., & Obrecht, G. (Eds.), *14th international Symposium on Presbyopia* (pp. 171-176). New York: Plenum.
- Hayhoe, M., Lachter, J., & Feldman, J. (in press). Integration of form across saccadic eye movements. *Perception*.
- Helmholtz, H. von (1925). *Physiological optics* (Vol. 3); (J.P.C. Southall, Trans.). Rochester, NY: Optical Society of America. (Original work published 1909).
- Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.
- Hochberg, J. (1968). In the mind's eye. In R.N. Haber (Ed.) *Contemporary theory and research in visual perception*. Holt, Rinehart & Winston, 309-331.
- Hochberg, J. (1984). Form perception: Experience and explanations. In P.C. Dodwell & T. Caelli (Eds.), *Figural synthesis* (pp. 1-30). Hillsdale, NJ: Erlbaum.
- Hochberg, J. (1988). Visual Perception. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey & R.D. Luce, *Stevens' handbook of sensory physiology* (pp. 195-276). New York: Wiley.

- Holt, E.B. (1903). Eye movement and central anaesthesia. *Harvard Psychological Studies*, 1, 3-45.
- Hull, J.M. (1991). *Touching the rock: An experience of blindness*. Pantheon.
- Humphreys, G.W., & Bruce, V. (1989). *Visual cognition: Computational, experimental and neuropsychological perspectives*. Hove, UK: Erlbaum.
- Irwin, D.E. (1992). Perceiving an integrated visual world. In D.E. Meyer and S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience—A silver jubilee* (pp. 121-142). Cambridge, MA: MIT Press.
- Irwin, D.E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics*, 34, 49-57.
- Irwin, D.E., Zacks, J.L., & Brown, J.S. (1990) Visual memory and the perception of a stable visual environment. *Perception & Psychophysics*, 47, 35-46.
- Jolicoeur, P., Snow, D., & Murray, J. (1987) The time to identify dioriented letters: Effects of practice and font. *Canadian Journal of Psychology*, 41, 303-316.
- Kahn, J.I., & Foster, D.H. (1981). Visual comparison of rotated and reflected random-dot patterns as a function of their positional symmetry and separation in the field. *Quarterly Journal of Experimental Psychology*, 33A, 155-166.
- Kohler, I. (1951). Über Aufbau und Wandlungen der Wahrnehmungswelt. *Österreichische Akademie der Wissenschaften, Sitzungsberichte, philosophisch-historische Klasse*, 227, 1-118.
- Kolers, P.A., Duchnick, R.L., & Sundstroem, G. (1985). Size in visual processing of faces and words. *Journal of Experimental Psychology, Human Perception and Performance*, 11, 726-751.
- Kröse, B.J.A. (1985) A structure description of visual information. *Pattern Recognition Letters*, 3, 41-50.
- Lévy-Schoen, A., & O'Regan, J.K. (1979). Comment voit-on en bougeant les yeux? Expériences sur l'intégration des images rétiniennes successives (Résumé). *Psychologie Française*, 25, 76-77.
- McConkie, G., & Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer & R.B. Ruddell (Eds.), *Theoretical models and processes of reading* (pp. 137-162). Newark: International Reading Association.
- McConkie, G. (1979). On the role and control of eye movements in reading. In P.A. Kolers, M.E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (pp. 37-48). New York: Plenum.
- McConkie, G.W. (1990). Where vision and cognition meet. Paper presented at the *H.F.S.P. Workshop on Object and Scene Perception*, Leuven, Belgium.
- MacKay, D.M. (1967). Ways of looking at perception. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 25-43). Cambridge, MA: MIT Press.
- MacKay, D.M. (1973). Visual stability and voluntary eye movements. In R. Jung (Ed.), *Handbook of sensory physiology, Vol. VIII/3A* (pp. 307-331). Berlin: Springer.

- MacKay, D.M. (1985). The significance of 'feature sensitivity'. In D. Rose & V.G. Dobson (Eds.), *Models of the visual cortex* (pp. 47-53). New York: Wiley.
- Marr, D., & Nishihara, H.K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, B*, 200, 269-294.
- Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81, 899-917.
- Matin, L. (1972). Eye movements and perceived visual direction. In D. Jameson & L.M. Hurvich, *Handbook of Sensory Physiology, Vol. VIII/4, Visual Psychophysics* (pp. 331-380). Berlin: Springer.
- Matin, L. (1986). Visual localization and eye movements. In K. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of perception and human performance, Vol 1* (pp. 20-1—2-45). New York: Wiley.
- Matin, L., Matin, E., & Pearce, D.G. (1969). Visual perception of direction when voluntary saccades occur. I. Relation of visual direction of a fixation target extinguished before a saccade to a flash presented during the saccade. *Perception & Psychophysics*, 5, 65-79.
- Nazir, T.A., & O'Regan, J.K. (1990). Some results on translation invariance in the human visual system. *Spatial Vision*, 5, 81-100.
- Nazir, T.A., O'Regan, J.K., & Jacobs, A.M. (1991). On words and their letters. *Bulletin of the Psychonomics Society*, 29, 171-174.
- Neisser, U., & Becklen, R. (1975) Selective looking: Attending to visually specified events. *Cognitive Psychology*, 7, 480-494.
- O'Regan, J.K. (1990). Eye movements and reading. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 395-453). Amsterdam: Elsevier.
- O'Regan, J.K., & Jacobs, A.M. (1992). The optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology, Human Perception and Performance*, 18, 1, 185-197.
- O'Regan, J.K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, 23, 765-769.
- O'Regan, J.K., Lévy-Schoen, A., Pynte, J., & Brugailière, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology, Human Perception & Performance*, 10, 2, 250-257.
- Paradiso, M.A., & Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research*, 31, 1221-1236.
- Polyak, S.L. (1941). *The retina*. Chicago: University of Chicago Press.
- Ramachandran, V.S. & Gregory, R.L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350, 6320, 699-702.
- Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception & Psychophysics*, 34, 39-48.

- Reitboek, H.J. & Altmann, J. (1984) A model for size- and rotation-invariant pattern processing in the visual system. *Biological Cybernetics*, 51, 113-121.
- Shebilske, W. (1977). Visuomotor coordination in visual direction and position constancies, In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes* (pp. 23-70). New York: Wiley.
- Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 3, 701-703.
- Schwarz, E.L. (1981). Cortical anatomy, size invariance and spatial frequency analysis. *Perception*, 10, 455-468.
- Steinman, R.M., & Collewijn, H. (1980). Binocular retinal image motion during active head rotation. *Vision Research*, 20, 415-429.
- Tarr, M.J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21, 233-282.
- Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception*, 9, 483-484.
- Turvey, M.T. (1977). Contrasting orientations to the theory of visual information processing. *Psychological Review*, 84, 67-88.
- Ullman, S. (1980). Against direct perception. *Behavioral and Brain Sciences*, 3, 373-415.
- Uttal, W.R. (1975). *An autocorrelation theory of visual form detection*. Hillsdale, NJ: Erlbaum.
- Volkman, F., Schick, A.M.L., & Riggs, L.A. (1968). Time course of visual inhibition during voluntary saccades. *Journal of the Optical Society of America*, 58, 562-569.
- Vitu, F., O'Regan, J.K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics*, 47, 583-600.
- Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus* (transl. D.F., Pears B.F. McGuinness). London: Routledge.
- Wolf, W., Hauske, G., & Lupp, U. (1980). Interaction of pre- and postsaccadic patterns having the same coordinates in space. *Vision Research*, 20, 117-125.
- Yakimoff, N., Mitrani, L., & Mateef, St. (1974). Saccadic suppression as visual masking effect. *Agressologie*, 15, 387-394.
- Young, A.W., Hellowell, D., & Hay, D.C. (1987). Configurational information in face perception. *Perception*, 16, 747-759.