Journal of Electronic Imaging 10(1), 123–151 (January 2001).

Representation of human vision in the brain: How does human perception recognize images?

Lawrence W. Stark
Claudio M. Privitera
Huiyang Yang
Michela Azzariti
Yeuk Fai Ho
Ted Blackmon
Dimitri Chernyak
Neurology and Telerobotic Units
School of Optometry
University of California at Berkeley
Berkeley, California 94720
E-mail: claudio@scan.berkeley.edu

Abstract. *The repetitive scanpath eye movement, EM, sequence enabled an approach to the representation of visual images in the human brain. We supposed that there were several levels of binding—semantic or symbolic binding; structural binding for the spatial locations of the regions-of-interest; and sequential binding for the dynamic execution program that yields the sequence of EMs. The scanpath sequences enable experimental evaluation of these various bindings that appear to play independent roles and are likely located in different parts of the modular cortex. EMs play an essential role in top-down control of the flow of visual information. The scanpath theory proposes that an internal spatial-cognitive model controls perception and the active looking EMs. Evidence supporting the scanpath theory includes experiments with ambiguous figures, visual imagery, and dynamic scenes. It is further explicated in a top-down computer vision tracking scheme for telerobots using design elements from the scanpath procedures. We also introduce procedures—calibration of EMs, identification of regions-of-interest, and analysis and comparison programs—for studying scanpaths. Although philosophers have long speculated that we see in our mind's eye, yet until the scanpath theory, no strong scientific evidence was available to support these conjectures.* © *2001 SPIE and IS&T.* [DOI: 10.1117/1.1329895]

1 Introduction

We introduce in this section the structure and the content of the paper.

We present in Sec. 2 the *scanpath theory* that outlines how a top-down spatial-cognitive model can control active eye movements, EMs, and visual perception. Philosophers had long speculated about perception and how we see in our *mind's eye*, but little scientific evidence was adduced until the scanpath sequence of eye movements enabled an approach to the problem. The scanpath sequence, itself, consists of alternating saccadic EMs and fixations that enable the active-looking paradigm. The controlling top-down model has its roots in philosophy, where for two thousand years, thoughtful scholars have considered how could it be possible for non-iconic inner awareness or conscious perception to confirm such a perception from the outside world. Our suggestion is perhaps the model can succeed by using iconic matching to physical signals arriving at the brain via peripheral nerves and sensory organs.

The earliest experiments showed the sequential and repetitive character of the scanpath and its idiosyncratic nature with respect to the person viewing and the picture or scene viewed. These experiments suggested the reality of the scanpath EM sequence for several kinds of static pictures. Most scenes are dynamic. Thus a test of the scanpath theory would be to ask whether EMs, while looking at such dynamic scenes, could be similarly characterized as a scanpath sequence. Ambiguous and fragmented figures shift in their visual perceptions; so do the scanpaths traced over the constant physical picture. Thus they appear to be generated from an internal model or schema rather than being controlled by external visual world signals impinging upon the brain. Quite a few laboratories have confirmed these results (see references below); also, one can go back to precursor static experiments like those of Yarbus to see evidence of the repetitive sequences we call the *scanpath*.

Like any theory the scanpath theory is *testable in experiments*; it can be disproven in that way and can persist if not disproven. Of course, no scientific theory can ever be *proven*. Experiments presented range from verbal descriptions to qualitative comparisons to quantitative measures available to statistical analysis. Section 3 documents how two scanpaths can be quantitatively compared as to their loci of fixations using a locus similarity index, Sp; Sp uses a Euclidean distance measure or a binary measure dependent upon a usual clustering of fixations about a point-of-interest, POI. Even more interesting is to compare sequenc-

Paper 021101 received Aug. 31, 1999; revised manuscript received Aug. 14, 2000; accepted Oct. 3, 2000.
1017-9909/2001/$15.00 © 2001 SPIE and IS&T.

ing of these fixations using a sequential similarity index, Ss, that measures how similar two strings of fixations are. Ss uses a string editing algorithm. Random similarities are used so as to enable statistical tests of the results. Methodological results are detailed sufficiently to satisfy the reader as to the care given to various aspects of these experiments from EM data acquisition, to fixation identification, to analysis of results and presentation as *parsing diagrams*.

In Secs. 2 and 3 key figures are presented related to the experiments on scanpath, ambiguous figures, and visual imagery studies. The above experiments and methods have already been published but we have included in Secs. 2 and 3 in this review some experimental figures and careful description of the quantitative methodology used to test the scanpath theory and its prediction.

More recently, *visual imagery* has been employed to show that the similarity of scanpaths when visualizing a picture were similar, qualitatively and quantitatively, to the scanpaths of a subject when actually viewing the imaged picture. This is, first of all, strong evidence for the top-down scanpath theory of vision, since there is no external world available to satisfy the naïve bottom-up concept that the external world enters the brain and controls visual perception. In addition, the visual imagery experiments enable us to explore further, in a quantitative way, the interactions between different modules for symbolic, spatial, sequential, and motor readout control of the scanpath and perception. Several of these visual imagery *grid* experiments are presented to document these new approaches. In Sec. 4, we have grouped the visual imagery grid experiments and detailed the experimental approach carefully so the reader can judge the quality of our laboratory studies. We have also explained the summary tables of these experiments and the parsing diagrams that had been explicitly designed to show the statistical significance of our results not only with respect to a random bottom *anchor* but also within different experimental variables. The scanpath experimental results as tested in the laboratories are related to finding behavioral data consistent with the scanpath theory, testing and confirming it.

Section 5 is about the brain and perception and attempts to show how new neurological and neurophysiological research concerning the modular cortex is consistent with and explanatory of the various mechanisms that must exist for the scanpath theory to *work*. In addition, these neurological results in combination with the scanpath theory helped to clarify the neuroanatomy and functioning of the brain. We have all seen circuit diagrams showing how every part of the brain is connected with any other part of the brain—a *reductio in absurdum*. The scanpath theory forces functional connection between different parts of the modular cortex in order to proceed from the top-down, TD, noniconic visual memory schema to a matching of the bottom-up, BU, iconic signal pathways in the visual cortex.

Simulation of a top-down computer-vision model in Sec. 6 enables one to have an explicit set of algorithmic mechanisms that gives body to the scanpath theory. A model is not an experimental proof of the theory, but is rather a vehicle for testing ideas in an explicit manner. This approach was a major motivation for the TD robotic vision system that our laboratory has been working with for some years and the model thus explicitly illustrates the feasibility of our TD scanpath approach.

In summary, this review presents the scanpath theory for top-down visual perception, and describes a number of experiments testing the theory that further enhance our understanding of this theory and its scientific support. We further discuss the philosophical and physiological underpinnings of the theory, describe a computer-vision simulation, and discuss the nature of the internal spatial-cognitive representation of visual perception.

2 Scanpath Theory: TD Vision and Scanpath Sequential EMs

We begin our introduction to the scanpath theory with a brief description of BU and TD vision. Then we describe early experiments on the scanpath as a sequence of EMs. These early experiments support the scanpath theory by showing that the repetitive and idiosyncratic sequences are present as predicted with viewing of ambiguous figures and during visual imagery. We then introduce visual imagery grid pictures as an experimental approach to memory representation and binding.

2.1 *Vision*

Human vision is complex. The essential problem is how to match bottom-up, BU, confirmatory signals coming both from the wide peripheral visual field, with only low resolution but with high sensitivity for moving objects, and from multiple high-resolution glimpses by the centrally located fovea, a small, circa one-degree region. These foveal regions-of-interest, ROIs, are sequentially visited by a string of fixations, shifted by a string of saccades, rapid eye movement, EM, jumps, and are simultaneously matched by top-down, TD, symbolic, spatial, and sequential representations or bindings of the hypothesized image.

When the retinal field is mapped onto the visual cortex, there is a considerable geometrical magnification of the signals coming from the fovea, and a consequent reduction of signals coming from the periphery. The log-polar distortion (Fig. 1) is a rather good depiction of the geometry of the visual image mapped onto the visual cortex.[1] When the high-resolution fovea is fixated on a particular part of the picture, such as the sailboat at the edge of the beach, that ROI is magnified on the visual cortex. Contrariwise, those parts of the image lying on the periphery of the retina are minified, so that only color and textural segmentation of large areas can be appreciated at the low resolution of the periphery. Two such foveal and peripheral representations (Fig. 1) provide an indication of the kind of BU information coming into the visual brain.

2.2 *Eye Movements, EMs, and Scanpath Theory*

The earliest experiments were observations of repetitive sequences of EMs while a subject looks at a picture. These led Noton and Stark[2–4] to the experimental definition of the scanpath as an idiosyncratic alternation of glimpses (called fixations or foveations) and rapid jumps of eye position (called saccades) to various ROIs, in the viewed scene. Other laboratories have confirmed the repetitive and idiosyncratic of the scanpath sequence of EMs. In addition there are many important studies not explicitly showing sequences of EMs but showing the location of fixations and

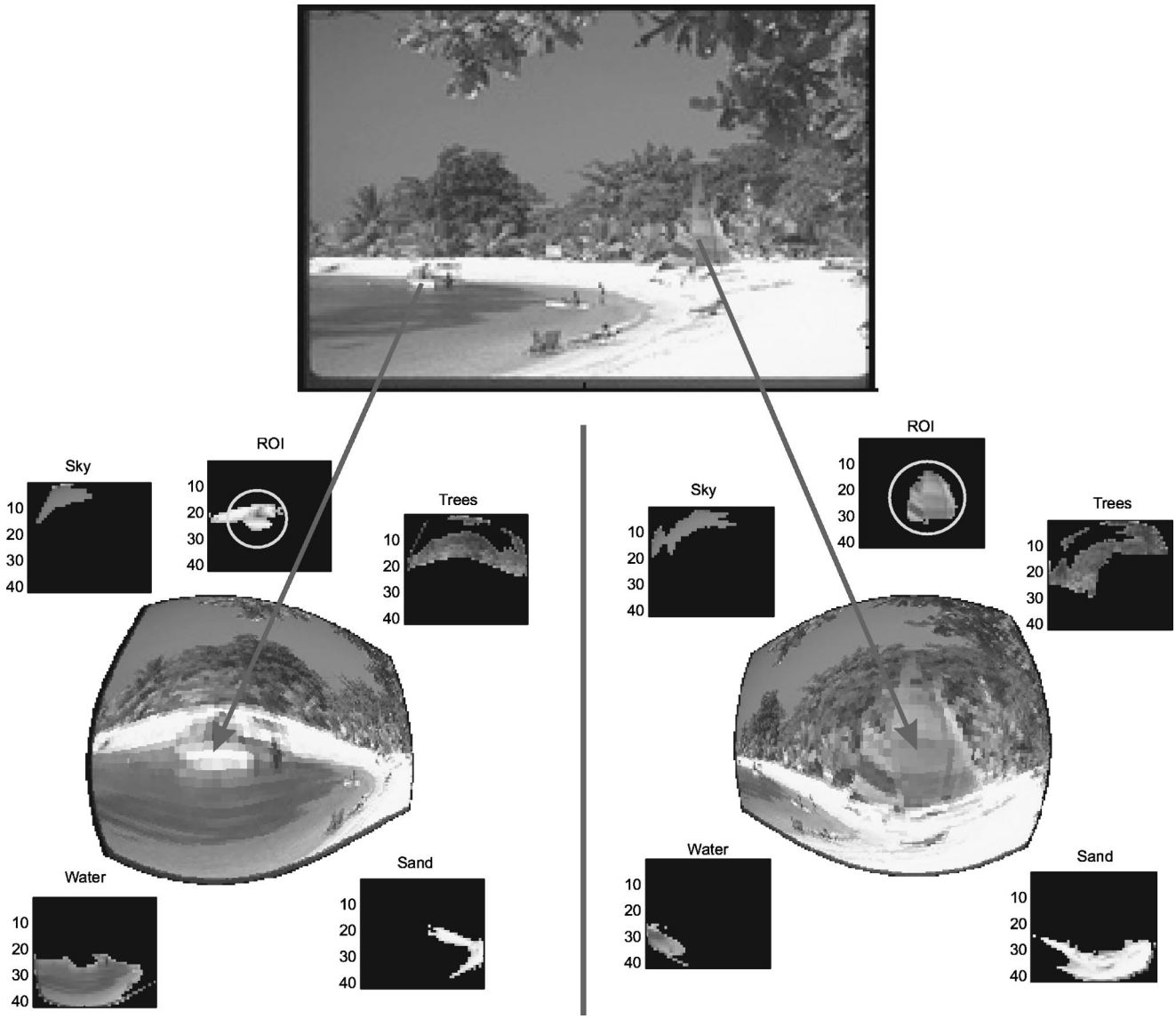


Fig. 1 Log-polar distortions of a picture. Two fixations (left and right panels below) on original picture (upper) show log-polar distortions with high cortical magnification (irregular shapes, lower left and lower right) of successive foveal ROIs (circles), as well as minification of peripheral regions likely captured as textured or colored segments (surrounding small squares, lower).

saccades in a static pattern overlaying a picture. A prime example is the important work of Alfred Yarbus. After the scanpath was established experimentally in 1971, and the scanpath theory put forward, it is easy to go back to an important predecessor such as Yarbus and see his EM data as important support for scanpath and the scanpath theory in his fundamental papers.[5–15]

EMs and attention shifts are very closely linked. It is only in an unusual laboratory situation that the two can be putatively separated; psychologists generally study attention shifts without measuring EMs, and neurologists study EMs without measuring attention shifts.[16–18]

The scanpath theory proposed that an internal spatial-cognitive model controls both perception and the active-looking EMs of the scanpath sequence[2–4] and further evidence for this came from new quantitative methods, experiments with ambiguous figures (Fig. 2),[19,20] and more recently from experiments on visual imagery[21–23] and from MRI studies on cooperating human subjects (see Sec. 5.4).[24] Further, it has been known for some time that the implicit or explicit task-setting in which the subject is immersed can strongly modify the scanpath.[15] Thus as a subject continually looks at the scene she may change her point of view, think of different task-goals and modify the scanpath.

Examples of two such EM scanpaths (Fig. 2) are shown for the classical ambiguous figure, *Two Faces or a Vase* (Fig. 2, left column). Depending upon the TD internal cognitive model, the subjects *sees* one or another of these two interpretations.[19,20] Some control over the current interpretation can be induced by *priming* the subject with a nonambiguous distortion of the ambiguous figure (Fig. 2, right column). Not only does the subject report on the interpretation, but her EMs show quite different patterns (left column), easily noted to be appropriate for the comparison of two faces in one case (upper row) or viewing the vase in

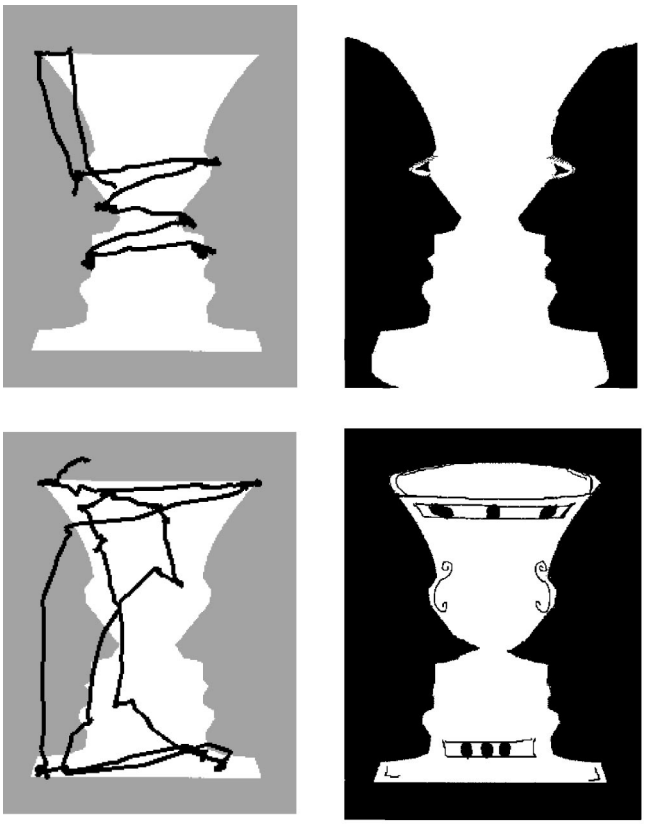


Fig. 2 EMs while looking at an ambiguous figure: The Ellis experiment. Identical ambiguous figures of vase (lower left) and two faces (upper left). EMs superimposed on ambiguous figures as they were actually seen following exposure to priming stimuli (right).

the other case (lower row). Of course, the actual picture viewed after the priming was the same in both cases. This evidence from Ellis and Stark[19] supports the scanpath theory (Fig. 2 provided by Privitera and Weinberger, 1998).

Of great interest in the context of an ongoing experiment in our research group is to study the nature of scanpath EMs when looking at dynamic scenes. Animations of constructed graphical scenarios (Fig. 3, upper panels) showed a set of moving cars on intersecting roads.[25] Subjects were clearly interested in the possibility of collisions. Their internal models evidently developed dynamical internal representations that guided their EMs to follow these objects of interest. Snapshots at five-second intervals, of a fifteen-second presentation of a reduced dynamic graphic display, are shown with time along the oblique axis (Fig. 3, lower panels). EMs were recorded throughout, and the heavy dots represent the location of EM fixations and thus, of the fovea during the five seconds preceding each of the snapshots. After analysis of the EMs into fixations and smooth pursuits the foveations and smooth pursuit tracking episodes are numbered in sequence (Fig. 3, middle panels, right).

Smooth pursuits are EMs that continually track a moving object; in so doing, they place the fovea on top of the moving target.[26] Thus, a smooth pursuit may appear to the perceiving brain as would a fixation on a stationary target. Of course, the brain is informed by efferent copy of the EM commands of the motion of the eyes as well.[27] We treat these smooth pursuits in a similar fashion to fixations for the purposes of the scanpath theory and apply the quantitative measures we use to analyze the scanpath (see below).

The string of such glimpses, both fixations and smooth pursuits, is shown in a more abstract form (Fig. 3, middle panels, right) with string labels, *FABCBE*. A noniconic representation of an operational model capable of generating such a scanpath (Fig. 3, middle panels, left) includes both fixations (lettered boxes) and commands to EMs to shift the fovea (circles with arrows). Glimpses and EM commands alternate. These sequences are not deterministic, but rather probabilistic. The solid lines are for the particular experiment illustrated, while the dashed lines represent other experimental scanpaths measured during repeat studies. This repetitive nature of the scanpath suggests that an internal spatial cognitive model is continually rechecking the scene with the same important points of interest in mind.

2.3 *Visual Imagery Scanpath with Grid Picture*

A simple grid with a pattern of alphabetical symbols is viewed by the subject for two seven-second periods (Fig. 4, upper row, left two grids). The EMs show somewhat repetitive scanpaths for these seven-second looking periods (middle row, left two grids). Next a blank grid is displayed (upper row, third grid). When asked to engage in visual imagery[23,24,28,29] and to imagine the previous pattern, the subject makes a scanpath (middle row, third grid) very similar to those made when looking at the patterns.[21,22,30] However, at this time there is no external pattern, only the subject's memory, that is, the internal cognitive spatial model, to guide the EMs. Then the subject is asked to draw the pattern from memory (right grids); this serves as a useful operational instruction to force the subject to localize carefully the components of the pattern to be remembered. Finally, a finite-state automata is derived from the experimental data (bottom row, middle) whose frequency of transitions is indicated by number and strength of connecting arrows; these could also be placed as coefficients in a Markov matrix (bottom row, right).[21,22] Figure 4 has been modified from Stark, Choi, and Yu, 1996.[31] The above scanpath experimental approaches have had as their goal documenting behavioral data consistent with the scanpath theory, testing and confirming it.

2.4 *Memory Representation and Binding*

How is the TD internal model distributed and operationally assembled? The concept of binding speaks to the assigning of values for the model and its execution by various parts of the brain.[32,33] We assume that there are several levels of *binding*—symbolic or semantic binding for naming of the model, spatial binding for the structural locations of the ROIs,[34–36] and sequential binding for the dynamic execution program that yields the sequence of EMs. The EM scanpath approach is complementary to studies being vigorously pursued in other laboratories—on attention shifts,[37] without recording concomitant EMs. Cognitive vision scientists now largely accept TD perceptual processes but only in the laboratory attention-shift paradigm; they still consider EMs to be controlled by the external visual scene, a most naive view. Further evidence for TD perceptual memory binding has come from studies on functional mag-

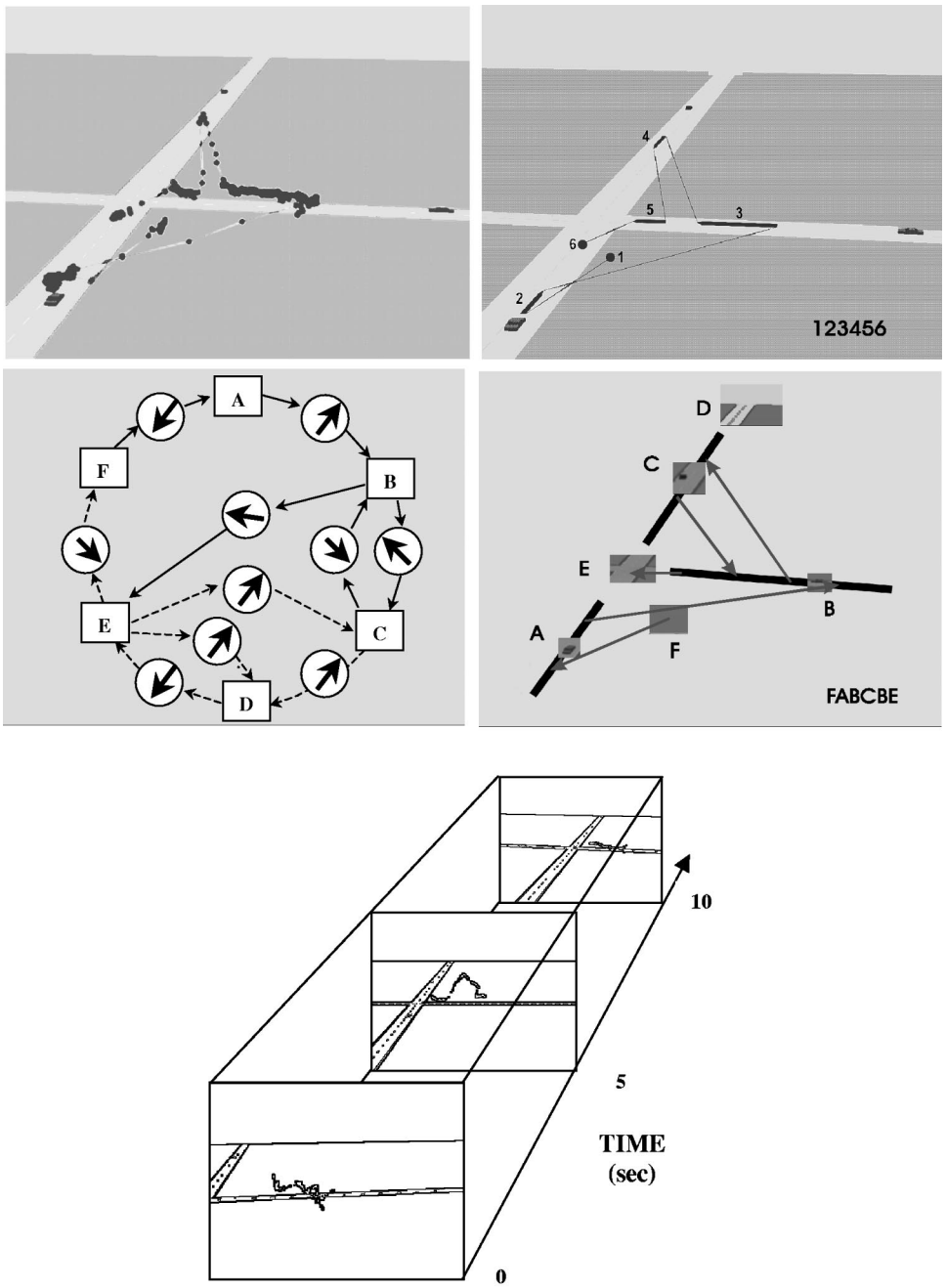


Fig. 3 Scanpath theory and dynamic display with EMs. EM positions (every 50 ms) during dynamic display shown as a connected sequence (upper left) while the dynamical ROIs visited form a connected sequence (upper right and middle right). By number or letter identification of smooth pursuit or static fixations, this sequential string of visited ROIs could be defined. A noniconic model of alternating perceptual ROIs (lettered squares) and saccadic EMs (circles with arrows) is shown by solid arrows for the experiment presented. This is the *feature ring* of the scanpath theory. On other presentations of the stimulus, other ROIs and sequences could be formed (dashed arrows). Animation of dynamical scenarios (lower panels, illustrated as snapshots every five seconds time proceeding from lower left to upper right). EM positions (black circles) taken every 50 ms are integrated over the preceding 5 s, and are superimposed onto snapshot images; they represent the basic data captured for this experiment.

netic resonance imaging, fMRI (see Secs. 5.4 and 5.5), and positron-emission tomography, PET, in man, and on the neuroanatomy and neurophysiology of animals. We thus try to use this current neurological information to localize these different aspects of the spatial-cognitive model in the brain. One aim of the experiments below is to attempt to dissect out different forms of binding and to test their respective contributions to the experimental scanpaths.

3 Metrics and Analyses

Like any theory the scanpath theory is testable in experiments; it can be disproven in that way and can persist if not

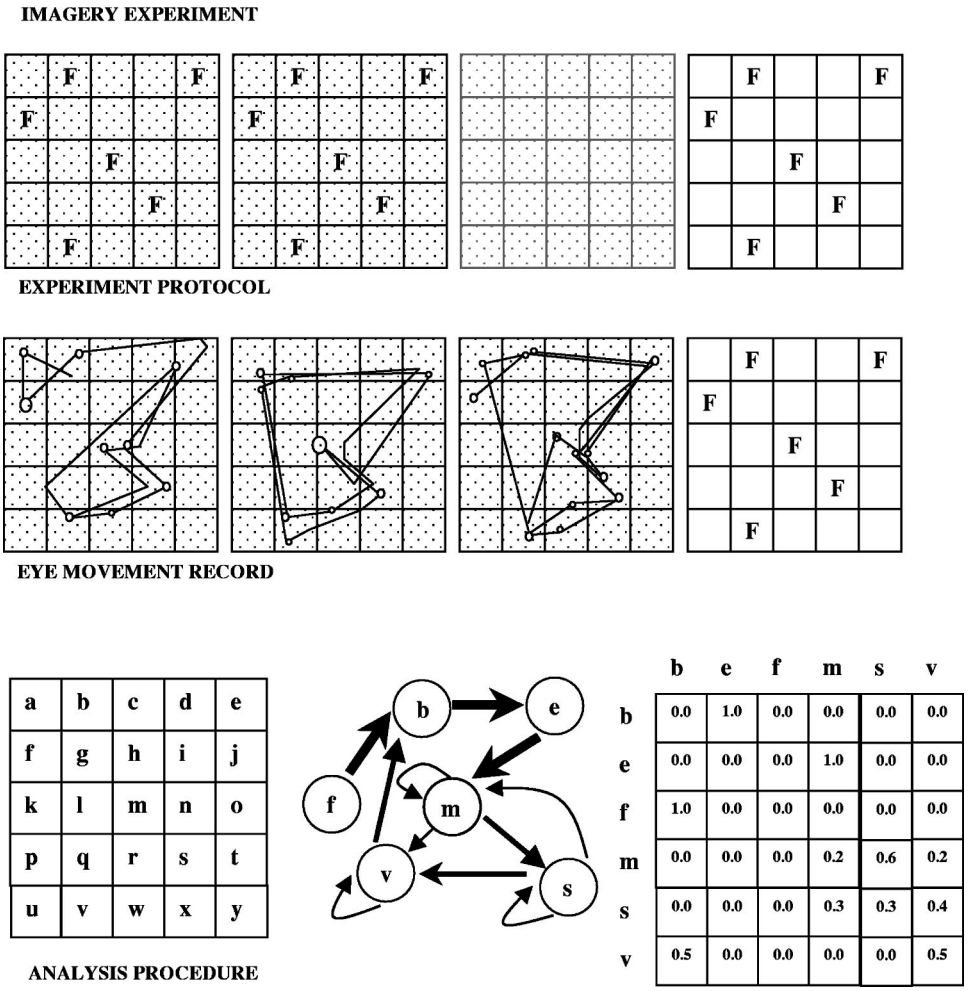


Fig. 4 EMs while engaged in visual imagery: The Brandt experiment. Scanpath EM sequence was almost the same for the second looking presentation (middle row, second grid) as for the first visual imagery presentation (middle row, third grid). During the visual imagery presentation, no information about the location of the alphabetic symbols, Fs, was available; thus, the remembered representational model must have controlled the scanpath in a TD fashion. Quantitative metrics could be obtained in the analysis procedure (lower row) by creating a finite state automata (middle) for generating the scanpath; then transition probability coefficients could be arranged in a Markov matrix (right) for later statistical analysis.

disproven. Of course, no scientific theory can ever be *proven*. Experiments presented range from verbal descriptions to qualitative comparisons to quantitative measures available to statistical analysis.

Two scanpaths can be compared as to their loci of fixations using a locus similarity index, Sp; Sp uses a Euclidean distance measure or a binary measure dependent upon a usual clustering of fixations about a point-of-interest, POI. Even more interesting is to compare sequencing of these fixations using a sequential similarity index, Ss, that measures how similar are two strings of fixations. Ss uses a string editing algorithm. Random similarities are used so as to enable statistical tests of the results.

The computational and mathematical platform utilized in our different experiments is introduced in this section. Methodological results are detailed sufficiently to satisfy the reader as to the care given to various aspects of these experiments from EM data acquisition, to fixation identification, to analysis of results and presentation as *parsing diagrams*.

3.1 *Stimuli and Computational Resources*

As anticipated in the previous section, several sets of stimuli were used—static and dynamic pictures for the *looking experiments* and also a series of grid patterns for the *visual imagery experiments*. A complex of computer workstations were interconnected in our Berkeley laboratory Internet; these included an Indigo SGI for display, the SGI and a PC-586 for collecting EM data, the PC-586 alone for the choice experiments, CH (see below), experimental display and data collection, and for analysis either of these workstations. Each of the computers was dedicated to running a part of the complex experiment. Software generated the protocol, displayed the stimuli, recorded the EMs, analyzed the collected data, compared (vector) sequences of fixations for analysis, and displayed the intermediate and final results. Software ranged from special lab C-programs to Matlab toolboxes for certain functionalities.

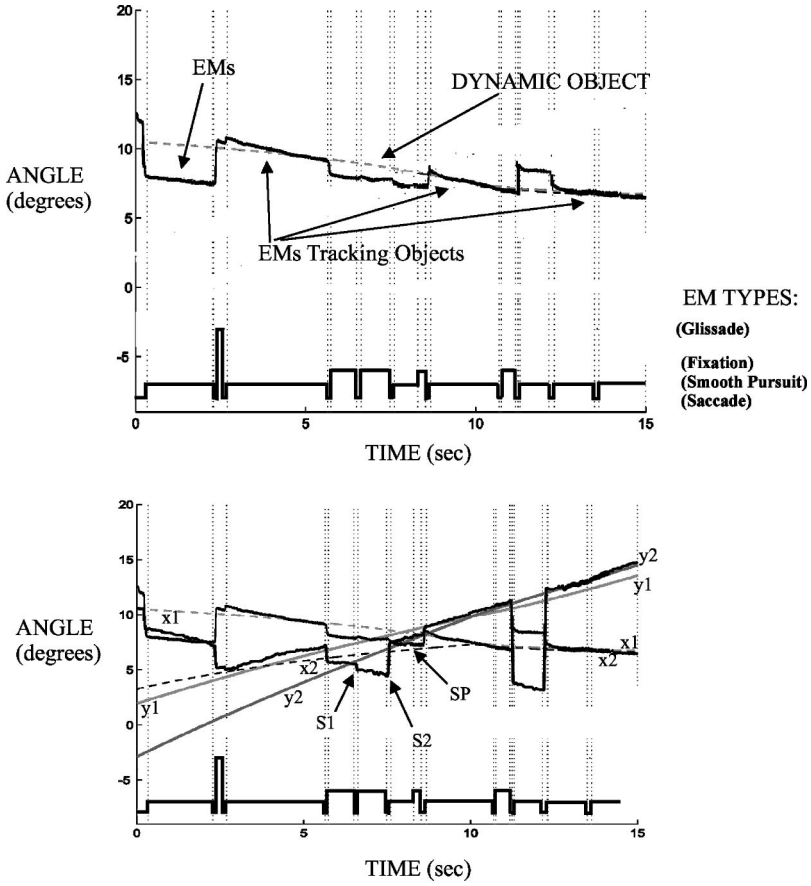


Fig. 5 EM trajectories and classification. The upper panel is an abbreviated version of the lower panel, it provides a simplified approach to demonstrate how the EM recordings and analyses proceed: some readers will not be so interested in the full depiction in the lower panel. Trajectories of EMs displayed as functions of horizontal angles (bold irregular lines) and time. *X* (dashed lines) and *y* (solid lines) components of the two dynamic objects are represented by regular lines. Note saccades, S, and smooth pursuits, SP, that show up clearly; these and other types of movements could be identified (lower steps and labels) and analyzed (see text).

3.2 *EM Recording and Analysis*

EM experiments must be carried out carefully. Computer controlled experiments presented pictures and carefully measured eye movements using video cameras.[38] An infrared source light was projected towards the eye of the subject, generating a bright Purkinje reflection that was easy to track. An infrared video camera was focused on one of the subject's eyes and the video image of the eye was digitized by mean of a framegrabber into the eye-tracking server. The subject was instructed to watch the visual stimuli on the screen of the client stimulus controller, which was socket-connected to the eye-tracking server. The subject was seated in front of the screen with head secured onto a slit-lamp chin-rest structure. Prior to the display of the visual stimuli, the subject was instructed to fixate on a sequence of nine different calibration points shown in random order on the screen. This calibration procedure established a mapping between the location of the Purkinje reflection in the video image and the direction of the subject's gaze on the screen. In this way, any drifts or nonlinearities could be removed.

An EM classification algorithm finally separated the raw EMs into fixations, rapid saccadic EMs, and smooth pursuits (Figs. 3 and 5) so that further analyses could be done on the loci and sequences of fixations considered as a string of loci of glimpses, the vector of active looking.

The performance of the algorithm is shown in Fig. 5 for a dynamic stimulus experiment. Figure 5 is truncated into two parts; the upper panel is a simplified representation that contains only the x coordinates of the oblique trajectory of one of the two moving objects and of the raw EM data. The full figure in the lower panel shows both objects viewed during the period of the figure and the corresponding EM x and y coordinates as functions of time.

The x and the y coordinates of the two dynamic objects, cars moving in the visual scenario, are represented by dashed and solid lines, respectively (car one, $\langle x1\, y1\rangle$ and car two $\langle x2\, y2\rangle$). (Note the change of gray level intensity of those lines as representing the two cars, Fig. 5, lower panel.) Bold irregular lines indicate the x and y coordinates of the EMs as functions of time. All are expressed in degrees of the visual angle (ordinate).

The EM classification algorithm extracts from this information those sections of the EM data corresponding to

i. smooth pursuits, SPs, while the eye tracked a moving object (see three arrows, Fig. 5, upper panel);
ii. saccades, S (note instances of saccades in the bottom descriptive sequential trace indicated as downward pulses, Fig. 5), when the eye rapidly jumped from one loci to another;
iii. fixations, F, while the eye was fixating on a still object in the input field of view (half-upward wide pulses); and even,
iv. various artifacts such as glissades, G (indicated by larger upward pulses).

For example, the presence of two saccades (S1 and S2, lower panel) is clearly indicated by the *y* component of the EMs in the full lower panel representation and thus validates the automatic computer analysis program; they were oblique saccades and their *x* components (upper panel) were small.

One of the most interesting aspects of the classification program was its ability to resolve conflicts between the independent parts of the program and finally get to an optimal classification of each temporal portion of the EM trajectory.

Dynamic smooth pursuit EMs were clearly generated by the observer tracking moving targets. These smooth pursuit EMs play a large role in scanpaths while subjects are observing dynamic scenarios and have a very interesting characteristic—they maintain the fovea in position over the moving object as long as this is possible and as long as the moving object is the one to which the top-down spatial cognitive model continues to pay attention. Given that the sensory system *sees* the moving object fixed on the fovea, we were able to modify the scanpath theory to include smooth pursuit as a sort of *superfixation*. In this way the scanpath sequence consists of saccades alternating with either fixations or superfixations, that is, dynamic smooth pursuit EMs.

Again in the full lower panel representation, we see, for example, a significant SP with a significant velocity in the *y* component; this again lends credence to our automated computer analysis. Finally, we would like to emphasize that we also do human manual checking of the computerized analyses in order to ensure that the data we put forward is reliable. We feel that with the video camera pixel resolution and lens magnification that we usually employ, a level of precision to about 1/4 to 1/3 of a degree is quite feasible. Even the effects of free head movements on accuracy and precision obtainable under favorable laboratory conditions with a video camera and a mobile HMD display remain within our precision limits.[39]

The programs in our laboratory have a long history dating back to 1959, starting at Yale University, continuing through our stay at MIT, the University of Illinois, and including our long residence (32 years) at Berkeley. Although it is impossible to list all the persons contributing to these programs, we must mention Robert Payne, Earl van Horn, Alan Sandberg, John Semmlow, Christian Freksa, especially for the scanpath analysis,[40] A. Terry Bahill, Michael Clark, An Nguyen, Yun Choi, Yong Yu, and most recently, Y. F. Ho.

Besides recording EMs, we also introduce another experimental technique, called choice, CH. Here, the subject is asked to click a mouse cursor onto various ROIs in the picture, or grid pattern, displayed on the computer screen. When presented with an empty grid, the subject must depend only upon memory bindings.

3.3 *Analyses of the Vectors of Looking*

Following the analysis of the EMs, the series of fixations is defined as a string. These strings represent actions of a finite-state automata (Fig. 3, middle panels). We compare each pair of strings, to see the number of letters they have in common; this matches the locations of the fixations and gives an Sp similarity index for the similarity of loci. We further compare each pair of strings as to the order of the string letters, using a string-editing algorithm (additional explanation is provided in the Appendix and Fig. 20). This provides Ss, the similarity index for sequence strings.[20,31,41–46]

The meaning of the two metrics Sp and Ss is exemplified in Fig. 6 (upper row) where two simplified scanpaths are compared. Two different scanpaths (left) with no locational similarity and no sequential similarity yield Sp=0 and Ss=0. When the two scanpaths have exactly the same locational similarity but no sequential similarity (middle) we have Sp=1 and Ss=0. Finally, two scanpaths with exact similarity yield Sp=1 and Ss=1 (right).

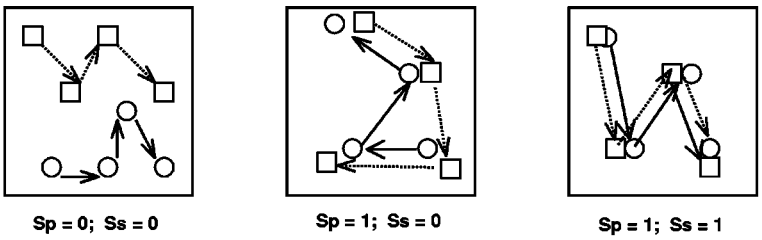
Pairwise comparisons of all scanpaths were averaged and assembled in Y-matrices. A Y-matrix is an ordered array of the string similarities, either Sp or Ss, to enable further sorting and averaging. Two simplified Y-matrices (Fig. 6, middle row, corresponding to repetitive looking of two different subjects and two different pictures) show such arrays and their row and column labels (Subject 1, S1, and Subject 2, S2, are two of the many subjects; Picture 1, P1, and Picture 2, P2, are two different scenarios). A complete Y-matrix for all the different subjects and input scenarios would be, of course, too large to display and thus to read. Parsing diagrams (Fig. 6, lower panels) with averages of similarity coefficients collected from the arrays of the Y-matrices, are a compact and intuitive alternative way to look at the data: R, for repetitive scanpaths, same subject looking at the same picture at different times; local=L, different subjects, same picture; idiosyncratic=I, same subject, different pictures; global=G, different subjects, different pictures; random=Ra, random strings compared.

3.4 *Analyses of Dynamic Scanpath Experiment*

To illustrate further how we use our analytic methods, we describe the Ss parsing diagrams for the dynamic scanpath experiment (Fig. 6, bottom panel).[32] Numbers in parentheses are standard deviations; bolded values represent significant differences (at a p value <0.01) from the Ra, random values of 0.16 ($p<0.01$). ANOVA analysis provided tests of significance, and arrows represent statistically significant differences with respect to the G, global value; this was considered a *bottom anchor*.

An important distinction is that between repetitive similarity R (Fig. 6, upper left box) and random similarity, Ra. When using different dynamic scenarios with the same general background, the same subject with the same stimulus

INDICES



Y-MATRICES

Sp	Subject 1		Subject 2	
	Picture 1	Pict 2	Picture 1	Pict 2
S1 P1	0.65 R	0.38 I	0.54 L	0.18 G
S1 P2		0.60 R	0.31 G	0.47 L
S2 P1			0.69 R	0.33 I
S2 P2				0.58 R

Ss	Subject 1		Subject 2	
	Picture 1	Pict 2	Picture 1	Pict 2
S1 P1	0.40 R	0.24 I	0.31 L	0.08 G
S1 P2		0.39 R	0.13 G	0.19 L
S2 P1			0.43 R	0.21 I
S2 P2				0.24 R

Ss PARSING DIAGRAMS

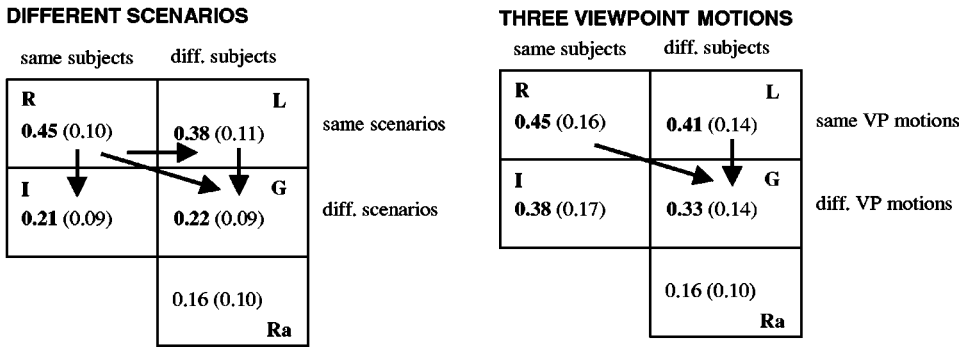


Fig. 6 Simplified, or toy, diagrams illustrating metrics for comparing scanpaths; and parsing diagram for the dynamic scanpath experiment. Quantitative methodology diagrammed to show similarity indices, Sp and Ss (upper panel). These pairwise comparisons are organized into Y-matrices (middle panel) and then indices segregated, averaged, and placed into parsing diagrams (lower panel). Note statistical tests of significance indicated by bolding, as well as arrows. These two parsing diagrams summarize the experimental base from a dynamic scanpath study.

showed a repetition value R of 0.45. This indicated that 45% of the sequences were congruent, and should be compared with the randomly expected Ss, Ra value of 0.16. However, when scanpaths for the same subject looking at different scenarios were compared (Fig. 6, bottom row, left parsing diagram), yielding the Ss-idiosyncratic value, the sequences were only 21% similar. This quantitative comparison documents that the scanpath theory generating the sequential EMs developed quite different sequences for different scenarios, that is, for different patterns of motion of the automobiles in the graphical scene.[25,47] We could conclude that different scenarios were viewed by different scanpath sequences.

A different result was obtained when different viewpoint motions were compared for the same scenario (Fig. 6, bottom row, right parsing diagram). The three viewpoint motions were panning, zooming, and static. Panning, or horizontal scanning, and zooming, or near/far approach of a camera, are standard movie filmic maneuvers; static means the camera point-of-view is at rest. The scenario remained the same, and the idiosyncratic similarity index, I, was 0.38. We could conclude that viewpoint motion did not make the scanpath sequence different for different motions. Statistical analysis, ANOVAs, supported these conclusions. For both sets of experiments, the similarities for different subjects, L, are almost the same as the R values. This may be due to the fact that the sequential motions of the different cars capture the attention of different subjects in a similar way, perhaps in a bottom-up fashion, or in cortical area MT.[25,48–52]

3.5 *Protocols and Subjects*

As indicated above, we could use a second method of *read-out–choice*, CH, clicking on a mouse cursor position, instead of the measurement of EM fixations. Of course, we

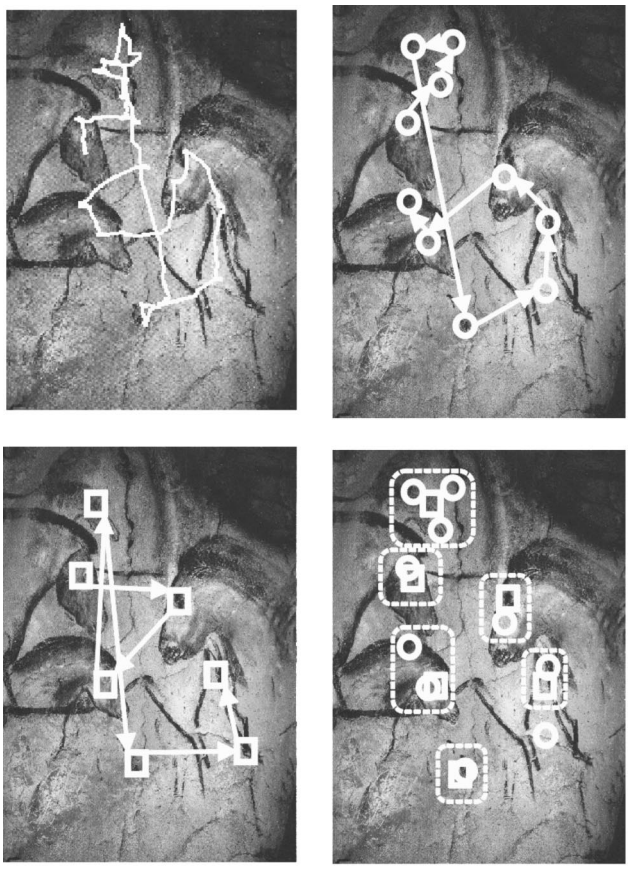


Fig. 7 EMs compared with choice, CH. Linearized EMs (upper left) were analyzed into fixations and saccades (upper right) while the subject looked at a cave painting of horses. Loci chosen by mouse clicks (lower left) could then be compared (lower right) with EM fixations.

studied the similarities and differences between the usual scanpath experiments, classical read-out method one, allowing subjects to gaze freely at the picture stimuli (e.g., Fig. 7), and the new second method of read-out, CH, by asking subjects to move a cursor over the stimuli pictures and click deliberately on ROIs. (see Sec. 4.1)

Another experimental task was to indicate the remembered patterns in the visual imagery experiment. We presented subjects with grids containing patterns of alphabetic symbols, letters, and asked them to image the pattern. The subjects then moved a cursor over blank grids and clicked deliberately on imagined or remembered loci. In this way, the CH method provided for an objective read-out of the structurally and sequentially bound memory traces. Further, we compared similarities and differences between this *cursor-CH* method and a third method of readout, that is, utilizing a locomotory pattern. In this third method, *walking*, WK, we asked subjects starting from a fixed initial position to walk over a blank grid marked on the floor and stop sequentially on those grid squares that represented remembered loci of the alphabetical letters.

Subjects were students visiting our laboratory who participated without pay; according to the rules of the Berkeley Committee for the Protection of Human Subjects they could terminate the experiment at will if they experienced any discomfort. They received oral, written, and also *operational* instructions, viewed a few preliminary pictures or grids, and usually were able to complete an experiment in less than twenty minutes. Operational instructions enforced a pattern of behavior by requiring the subjects to carry out procedures that served as additional reenforcement.

4 Visual Imagery Experiments

Visual imagery experiments, first of all, provide strong evidence for the top-down scanpath theory of vision, since there is no external world available to satisfy the naïve bottom-up concept that the external world enters the brain and controls visual perception. By embedding these experiments in a *grid* format we are able to explore in more detail and quantitatively, perceptual influences on the scanpath, in particular, the interactions between different modules for symbolic, spatial, sequential, and motor readout control of the scanpath and perception.

4.1 *Structural and Sequential Binding*

Our experimental results on binding, explained in detail below, compare the memory similarities between different read-out modes. Two different protocols compare EMs vs choice, CH, while looking at a set of pictures (Figs. 7 and 8), and choice, CH, vs walking, WK, while remembering a set of grid patterns (Figs. 9 and 10). Different *read-out* motor behaviors, indicating remembered patterns, were analyzed in the same way, and with the same methodology. Pairwise comparisons between scanpaths were carried out with each read-out mode (see Figs. 8 and 10, left and middle parsing diagrams), and then, between all pairs of one-mode against the other one-mode (see Figs. 8 and 10, right parsing diagrams).

By studying the within-mode similarities against the across-mode similarities we can assign quantitative numbers to the relative strengths of inherent and of read-out binding. In addition, we enriched the experimental protocols by examining the phenomenon of *consolidation*, meaning the memory coherence within repeated response patterns that may be stronger than the memory persistence from stimulus to response.

4.1.1 *Choice, CH, compared with EMs; picture viewing*

The alteration of the scanpath protocol, substituting mouse-cursor location and clicking for measured EM fixations, has many experimental advantages.[32] However, it had to be carefully evaluated by comparisons between EM fixations and choice, CH, loci in a variety of studies. In developing and expanding the protocol, we had subjects look at a number of pictures and then we studied their EMs. The methodological procedures to move from raw EMs to identified fixations were necessary; for example, in Fig. 7, linearized EMs of a subject looking at a cave painting of a horse (upper left) were transformed by a fixation algorithm to a sequential string of fixations (upper right, circles), with connecting vectors representing saccades and their sequence (upper right, arrows).

Next, sequential strings of CHs, produced by mouse cursor and clicks (lower left, squares) are also connected by vectors (lower left, arrows). These two strings can then be compared as to identity of their loci (see previous Section) to calculate the Sp and Ss similarity indices between EMs

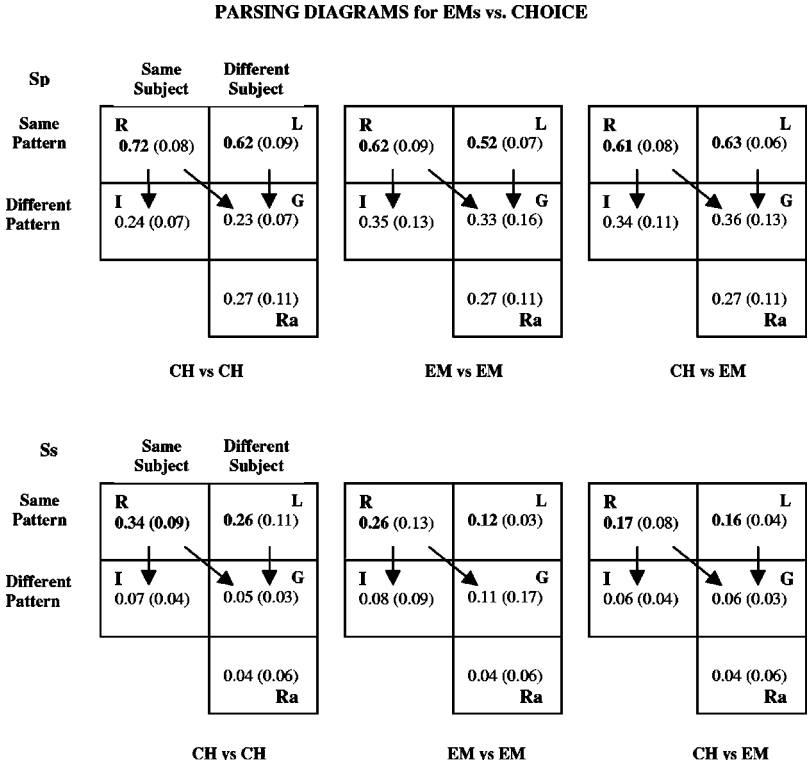


Fig. 8 Parsing diagrams; EMs compared with choice, CH. Sp (upper) and Ss (lower) parsing diagrams for CH compared with EM study. Intramodal read-out comparisons (left and middle panels) as well as intermodal read-out comparisons (right panel).

and CH procedures.[44] As an extra bonus for the reader, consider that this cave painting, an artistic work created 31 000 years ago has perhaps been equaled but not surpassed in the ensuing millennia of human social prehistory and history. The scanpath theory has awakened new interest in the neurology of artistic communication.[53–55]

As explained in the previous section, many pairwise comparison indices are collected and sorted using the Y-matrices. Averaged results are then organized in the parsing diagram (Fig. 8: Sp, upper row; Ss, lower row). EM comparisons (middle column) document that the R repetitive values, 0.62 and 0.26, are significantly different from random, Ra, and from global, G, the two bottom anchors (bolding or heavy arrows indicate an ANOVA $p<0.01$). Note that while Sp-L has a relatively high value, indicating that different subjects selected similar ROIs, the Ss-L value is lower, suggesting that different subjects utilized different sequences for the same picture and similar loci across subjects. The Ss-Ra values throughout are much lower than the Sp-Ra values, since there are many ways to establish sequences among similar loci. Almost identical results are found for CH comparisons (left column), with somewhat higher coherence, perhaps due to the more deliberate TD selection mental process for cursor clicks vs natural EMs.

Now, when we compare EMs and CH (right column) we find related distributions of similarity indices; R values are large and significantly different from G and Ra. The R-Sp index is large, indicating similarity of objects across modes; that is, it coheres for both read-out modes. Thus it appears that almost none of the structural binding is related to read-out mode differences. However, because the cross-mode R-Ss value equal to 0.17 is less than the R-Ss values for either CH or EMs, we must, in this case, partition the sequential binding between inherent and readout components.

4.1.2 *Choice, CH, compared with walking, WK; grid viewing*

In another set of experiments, we presented grids (see Figs. 4 and 9) to be memorized and to be recalled. Again two modes of response were compared—CH, using a cursor moving over a blank grid presented on a computer screen, or walking, WK, over a large grid outlined over the laboratory floor (Fig. 9).[32]

Squares with alphabetical symbols represent grid patterns that the subject could look at for a period of three seconds for each of two presentations (Fig. 9, two left-most columns). Subjects were then asked to move the cursor sequentially to each of the visually imaged locations of the symbols and to click the mouse buttons (this took about ten seconds) to indicate the remembered alphabetically-labeled grid squares (Fig. 9, four right-most columns), thus providing an output string of remembered alphabetically labeled grid squares. There was a fixed initial position from which the subjects started each time. Experiments were also carried out with subjects instructed to walk freely over a large grid placed on the floor; again, there was a fixed initial position from which they started each time. They were also instructed to stand with two feet in the appropriate grid-spaces for a brief moment, to indicate each labeled locus; in

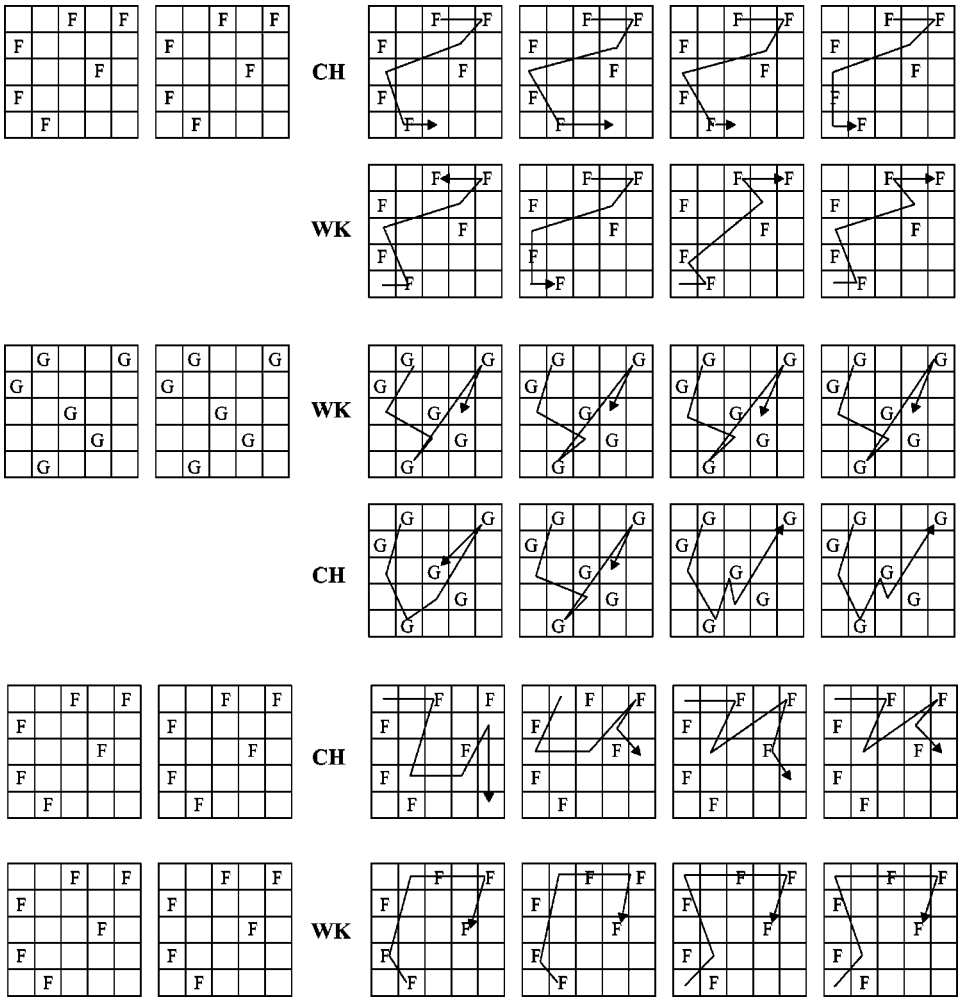


Fig. 9 CH compared with walking protocol, WK. Experimental protocol for cross-modal comparison between choice, CH, and walk, WK, read-out modes. Note similarity of patterning when a second display of stimuli patterns was not given (absent grids in both sets of upper panels); note difference in patterning when refreshment of stimulus pattern allowed a new memory scheme to be formed (lower panel).

this way, the experimenter could record the sequences of stops. Again, this provided an output string of remembered alphabetically labeled grid squares. CH and walking, WK were alternated without additional refreshment (Fig. 9, blank regions, second and fourth rows).

To buttress the qualitative results as shown in Fig. 9, we provide quantitative assessments from the similarity indices. The results came from many pair-wise comparisons for four subjects, naive with respect to the purpose of the experiment, but performing quite well in the task; their similarity indices were sorted using the Y-matrices and averaged in the parsing diagrams (Fig. 10: Sp, upper row, and Ss, lower row). CH comparisons (Fig. 10, left column) document that the R repetitive values, 0.90 and 0.77, are significantly different from Ra and from G, the two bottom anchors. Bold values or heavy arrows indicate an ANOVA $p<0.01$, meaning that the values statistically differed from Ra (bold) and from G (arrows). Again, the Ss-Ra values throughout are much lower than the Sp-Ra values, since there are many ways to establish sequences among similar loci. Almost identical results, 0.91 and 0.81, are found for WK comparisons (Fig. 10, middle column) with respect to the R values, both Sp and Ss.

When we alternate CH and WK without refreshment (Fig. 10, right column), we find somewhat modified distributions of similarity indices. R values, 0.87 and 0.60 are large and significantly different from G and Ra. That Sp-R is identical to the values for CH×CH and for WK×WK indicates that structural binding relied upon an inherent component. That Ss-R is less than the values for CH×CH and for WK×WK indicates that both the inherent and the readout components were important for sequential binding.

CH and WK could be presented with refreshment (Fig. 9, fifth and sixth rows). This refreshment (two leftmost grids, sixth row) allowed a modified, reinitialized, sequential pattern to be developed in the subject's representation (compare fifth and sixth rows).

4.1.3 *Modular cortical organization: the new phrenology*

A sketch of the lateral view of the human cortex (Fig. 11, upper panel) is presented to help understand the logic of

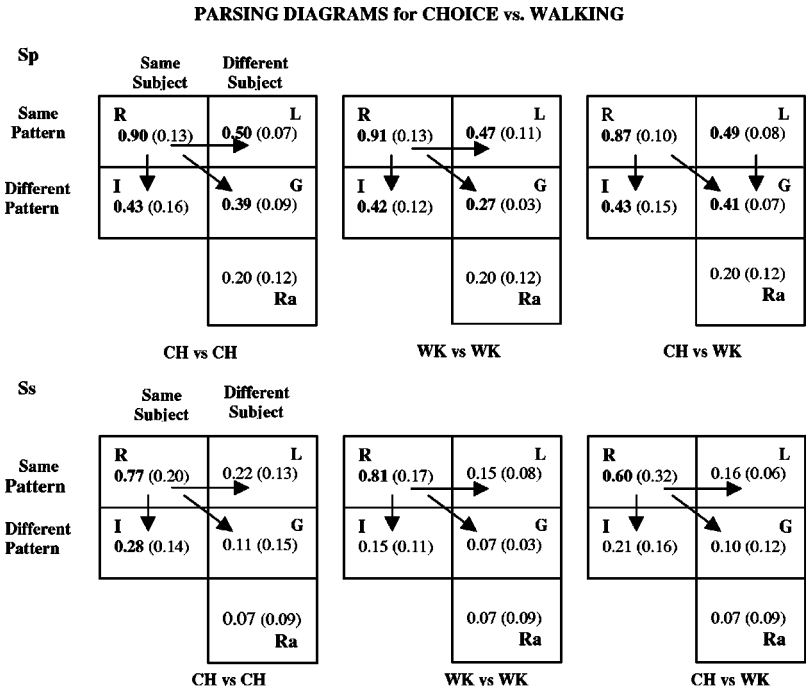


Fig. 10 Visual imagery: Parsing diagram for walking, WK vs choice, CH. Sp (upper) and Ss (lower) parsing diagrams for the choice, CH, compared with walking, WK, study. Intramodal read-out comparisons (left and middle panels) as well as intermodal read-out comparisons (right panel).

these different readout experiments. We are trying here to distinguish between inherent sequential binding, likely located in the prefrontal cortex, from variable sequential binding, dependent upon readout mode. The modes we are exploring are EM fixations, as in the classical scanpath experiments, choice, CH, using mouse-cursor positioning and clicking, and WK, locomotion over a grid on the laboratory floor.

Information about localization in the cortex comes from a variety of sources. Classical neuroanatomy and analyses of neurological syndromes have existed for several centuries, and had achieved a considerable degree of sophistication. Experimental ablation and electrical stimulation physiological studies next came into play. Modern methods, ranging from intrusive single-unit neurophysiology, to current PET and fMRI are daily supplementing earlier studies.[56] We have collected in Sec. 5.5 a few significant references to the neurophysiology in higher-level functions, that are pertinent to new concepts of modular cortical organization. Note (Fig. 11, upper panel) the *what* ventral pathway from visual cortex, VC, to the temporal cortex, TC, (especially left side) to which we attribute Semantic Binding. Similarly, note the *where* dorsal pathway from VC to the parietal cortex, PC (especially on the right side), to which we attribute structural binding. Known connections from PC to the pre-frontal cortex, PFC, have been shown to be related to temporal sequencing, and to which we attribute inherent sequential binding. Connections continue to the frontal eye fields, FEF, to which we attribute one form of read-out sequential binding (although a complete review of this fascinating area is beyond the scope of the present paper, a number of articles are referred to in the Discussion and Sec. 5.5). Thus the scanpath theory and the grid experimental results force one to take the functional connections

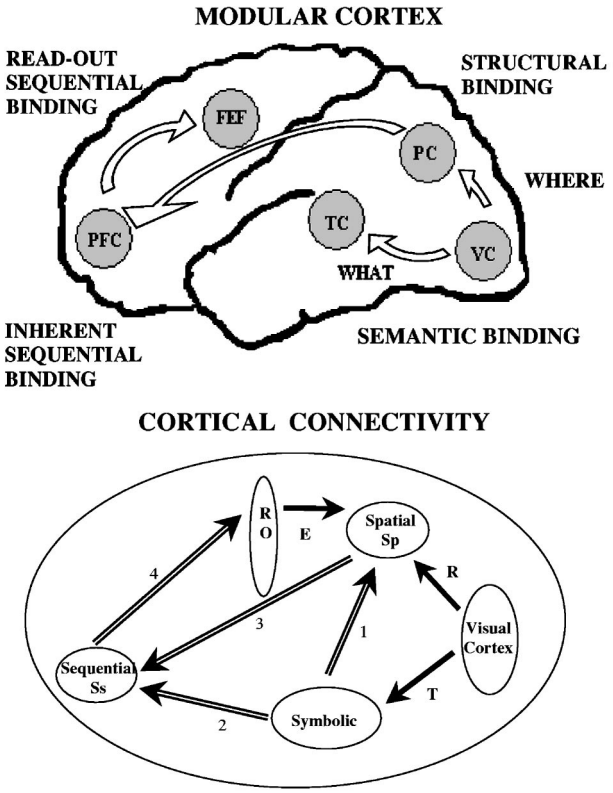


Fig. 11 Modular cortex and connectivity. Recent studies in neurophysiology and fMRI have established a *new phrenology*, the modular cortex (upper), with different functions assigned to specific regions of the cortex. Connectivity, explored in our experiment on inherent and read-out sequential binding, and as well, on the influence of symbolic binding, is indicated as numbered arrows joining labeled regions (lower).

Experimental type	Sp	Diff & N %	Binding	Ss	Diff & N %	Binding
CH X CH (pictures)	0.72			0.34		
EM X EM (pictures)	0.62			0.26		
average	0.67			0.30		
		0.06 = 15 %	Read-Out		0.13 = 50 %	Read-Out
CH X EM (pictures)	0.61			0.17		
		0.34 = 85 %	Inherent		0.13 = 50 %	Inherent
Random	0.27			0.04		
CH X CH (imagery)	0.90			0.77		
W X W (imagery)	0.91			0.81		
average	0.91			0.79		
		0.04 = 6 %	Read-Out		0.19 = 26 %	Read-Out
CH X W (imagery)	0.87			0.60		
		0.67 = 94 %	Inherent		0.53 = 74 %	Inherent
Random	0.20			0.07		

Inherent vs. Readout Sequential Binding

	TOP ANCHOR	CONTROL EXPERIMENT	MAIN EXPERIMENT	BOTTOM ANCHOR
SECOND STIMULUS	None	Same	Same	Different
SECOND LABEL	None	Same	Different	Same
OTHER INFORMATION	Interrupt	Interrupt	New Label	Diff 2nd stimulus
EFFECT		Allows re-initialization	Encourages re-initialization	Spatial dominates over Semantic
SECOND RESPONSE	Same	Same Some differences	Different Some Similarities	Different
SS (COHERENT)	0.86	0.71	0.46	0.07
SP (COHERENT)	0.94	0.80	0.76	0.36

Summary of Symbolic Experiment

Fig. 12 Inherent vs readout sequential binding and summary. Sequential read-out experimental findings can be summarized as almost 100% inherent binding, for spatial or structural similarity of patterns (upper table, middle column). However, sequential bindings are markedly influenced by read-out mode; only two-thirds of the binding is inherent (upper table, right column). Symbolic memory has an important influence on sequential binding, producing an average 50% loss of coherence (lower table, compare 0.46 with 0.71, next to bottom row) when the labeling is changed. Since the same loci were represented with a different label, the structural binding, of course, remained the same (compare 0.76 with 0.80, lower table, bottom row).

between different parts of the modular cortex seriously in order to proceed from the TD noniconic visual memory schema to a matching of the BU iconic signal pathways in the visual cortex.

4.1.4 *Inherent vs read-out sequential binding*

Now, we may consider the connectivity of the modules of the modular cortex (Fig. 11), as an aid in defining the logic of these experimental results. The summary tables (Fig. 12) and the parsing diagrams (see above, Figs. 8 and 10) have been explicitly designed to show more intuitively the logic and quantitative statistical significance of our results. A summary of coefficients comparing within-mode and across-mode experiments (Fig. 12, upper table, Sp, left columns, and Ss, right columns) provided the basic similarity coefficients from the parsing diagrams (Figs. 8 and 10). These were then normalized (Diff & N % columns) as percentages, by setting the bottom Ra anchor to 0% (e.g., 0.27, Fig. 12, upper table, upper left panel), and the within-mode R values as 100% (e.g., 0.67, Fig. 12, upper table, upper left panel). The two experimental protocols, CH×EMs (upper panels) and CH×walking, WK (lower panels), have yielded reasonably consistent results.

For Sp, we subtract the R values for across-mode from the within-mode values; the resulting normalized percentage numbers are 85% and 94% for the inherent component of structural binding. The fact that for structural binding the inherent component dominates, indicates that the parietal lobe structural memory is put too early in the process to be disturbed or altered by read-out mode differences.

For Ss, we again subtract the R values for across-mode from within-mode values; the resulting normalized percentage numbers are 50% and 74% for the inherent component of structural binding. We see that for sequential binding, although the inherent component is larger (two-thirds), the readout component is significant (one-thirds) and thus both components are important in sequential binding. This may be interpreted as allowing the prefrontal lobe inherent sequential memory to be somewhat altered by readout mode located further back in the frontal lobe; these regions are, of course, different for EM, for hand movement and for locomotion.

To summarize, structural binding is inherent, i.e., independent of readout modes. Sequential binding has strong components for both inherent and for readout binding; that

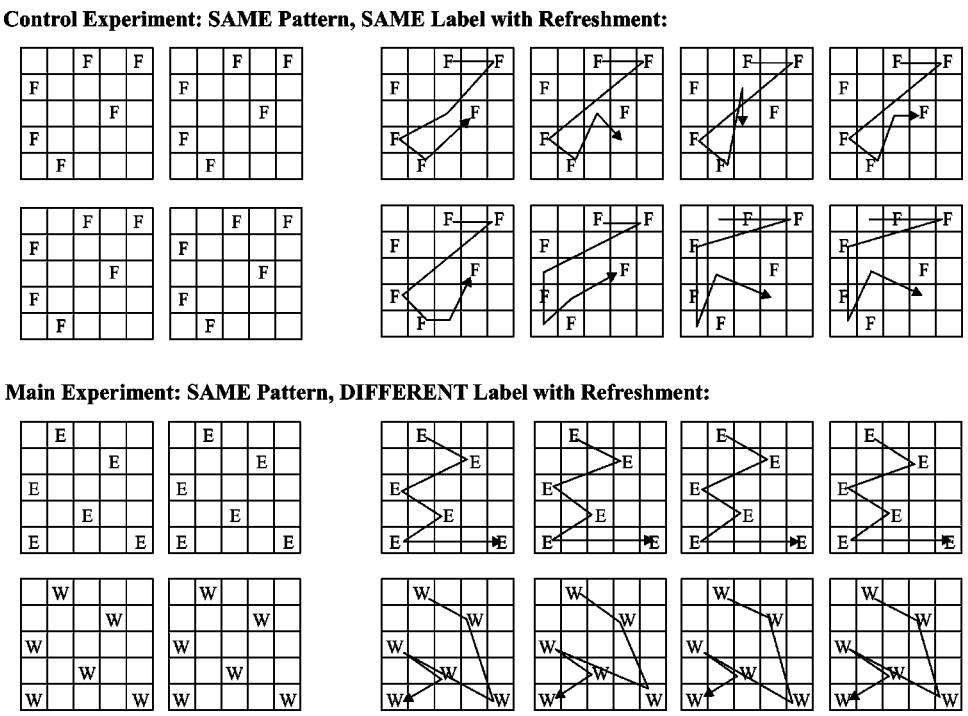


Fig. 13 Control experiment and main experiment. Experimental protocol for the control (upper panel) and main experiment (lower panel) to analyze the influence of symbolic binding. A major result is the influence of dissimilarity of labeling on the dissimilarity of the sequential pattern. Clearly, the spatial loci are the same, and thus, the structural similarity remains high. Note that refreshment in the form of two additional looking stimuli is presented in both experiments (lower pair of grids in each of the panels).

is, the readout mode contributes strongly (about one-third) to the memory of the sequence.

4.2 *Symbolic Binding Results*

How can we experiment on symbolic binding? The naming of a pattern, or its symbolic binding, plays an important role in this scanpath memory process. Quantitative experiments were carried out by Yang and Stark[57] to explore this phenomenon. Subjects were asked to remember lettered grids under a variety of conditions. Often, they were presented only with the letter, or symbol, of the pattern, and asked to remember the grid pattern that they were previously able to reconstruct. Interruptions, such as becoming familiar with and reconstructing other grid patterns, were most often interjected between the learning phase and the test-of-memory phase.

4.2.1 *Control experiment: same pattern, same label with refreshment*

An important control experiment was to test the ability of subjects to carry out pattern reconstruction from memory (Fig. 13, upper panel, upper row). Here, subjects were presented for a few seconds with two identical lettered grids, immediately followed by four successive blank grids wherein the subject attempted to reconstruct the previously seen patterns. Subjects were then interrupted with other tasks. Next, the identical lettered grid was presented (Fig. 13, upper panel, lower row), and the subjects attempted again to reconstruct the labeled pattern, onto four successive blank grids. Subjects, as we would have expected from this control experiment, were able to carry out this task very well, with Ss values of 0.71, and Sp values of 0.80. Which is the effect of the label (or the symbolic binding) on the scanpath memory process?

4.2.2 *Main experiment: same pattern, different label with refreshment*

The main experiment again tested whether subjects attempted to reremember a newly presented pattern identical to an old remembered pattern, with an important, significant difference (Fig. 13, lower panel). The new, identical pattern was labeled with a different symbol or letter!

In both of these sequences, the subject was able to perform consistently over the reconstruction in the four blank grids (Fig. 13, lower panel, upper row and lower row, four right grids). However, the new symbol encouraged the subject to reinitialize the memory pattern. Thus, the Ss value fell to 0.46, when the first and second presentations were compared. (Note differences in sequential patterning, Fig. 13, lower panel, upper row, compared to lower row.) Of course, the localization of the clicks, Sp=0.76, was of equal accuracy as the control experiment described above. Only the sequence was newly established because of the new label.

Top and bottom anchor experiments enabled us to establish a range of values for the Sp and Ss similarity indices. An experiment quite similar to the control experiment was performed, with its main difference that it allowed for no representation or refreshment of the lettered pattern for the second set of the memory test blank grids (Fig. 14, upper

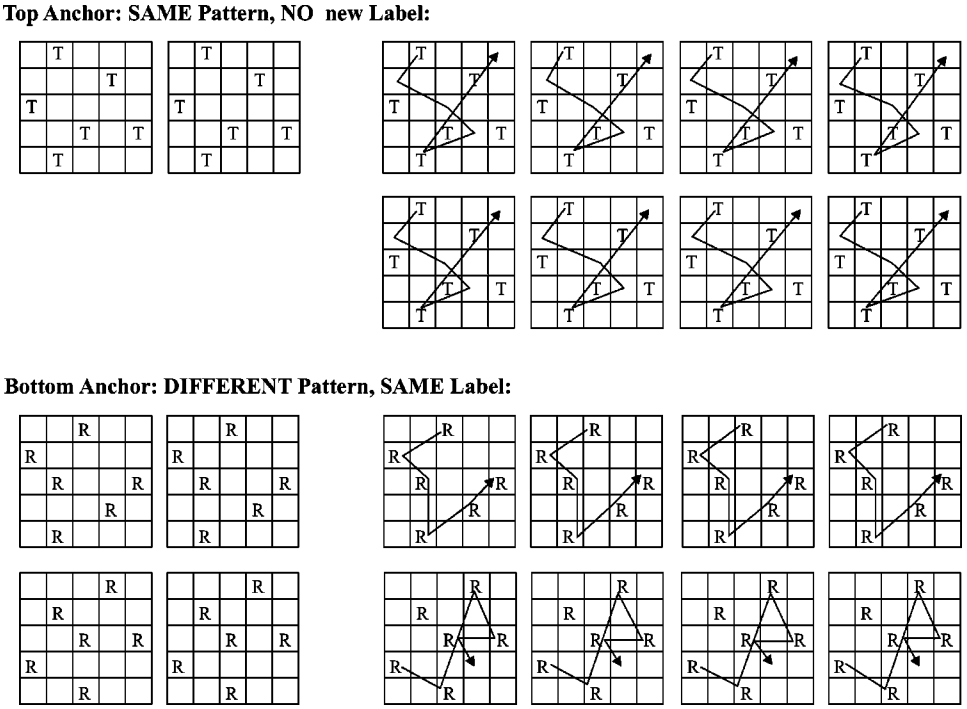


Fig. 14 Top anchor and bottom anchor experiments. Experimental protocol to establish the range of values for Sp and Ss similarity. Top anchor (upper panel) shows high correlation when no refreshment is permitted (two absent grids, lower row, upper panel). Bottom anchor (lower panel) shows absence of structural and sequential similarity when a different pattern is presented with the same label.

panel). Thus, the memory trace remained more or less the same without additional information relating to the pattern or the letter being presented. This gave us the highest values, Ss=0.86 and Sp=0.94, meaning that both the structural and sequential components of memory were very well preserved between the two sections. We thus consider this to be the top anchor of the similarity scales.

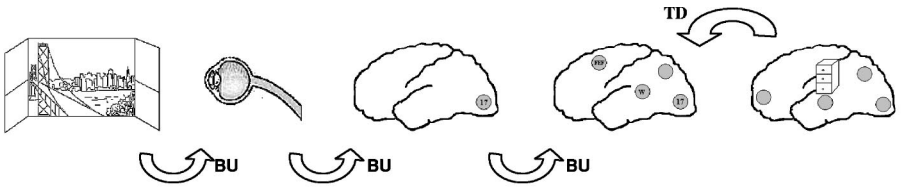
For the *bottom anchor*, we used the same letter symbol, but in a completely different grid pattern (Fig. 14, lower panel). As might be expected, the two sets of memory tests, with four blank grids each, showed little intertrial coherence or similarity of their patterns, with Ss=0.07, and Sp=0.36. The structural intertrial coherence and the sequential coherence are very low, and close to random values; thus, we can use this experiment as a bottom anchor. Although the subjects were *tricked* by having the same label for different patterns, still the structural pattern dominated over the symbolic label.

4.2.3 *Summary of symbolic experiment*

Again, the summary table (Fig. 12, lower) has been explicitly designed to show more intuitively the logic and quantitative statistical significance of our results. We have been able to establish that in the control experiment, the representation of the same stimulus with the same label a second time allows some reinitialization of the sequential memory trace (Ss=0.71). However, the values remain quite close to the top anchor values (Ss=0.86), where the absence of a second presentation did not allow for even mild changes of the memory pattern. Of course, with time and interruption, as was carried out in our experimental protocol, some decay of the memory pattern occurs. Note that the bottom anchor experiment with the same label, but with a different stimulus (allowing spatial to dominate over semantic-symbolic; see Fig. 12, lower table, row labeled *effect*) clearly documents the great loss of both locational and sequential coherence in the memory pattern (Sp=0.36, and Ss=0.07).

The main experimental result is that when the same stimulus grid is presented a second time but with a new label, this new label encourages reintialization of the memory trace (Ss=0.46) (see Fig. 12, lower table, row labeled *effect*). Thus, the second set of responses is quite different (about 50% loss of coherence for the sequence, Ss, but essentially no loss of coherence for the pattern, Sp); recall that the same loci were represented with the new label, and thus, Sp should remain quite consistent. The quantitative result, 50% loss of Ss coherence due to a changed label, comes from averages of many experiments done with a variety of subjects. Subjects varied, and even the same subject would produce much higher or lower coherence in different trials. More experiments are necessary to determine whether a quasiswitching occurs between coherent and noncoherent results.

This experiment documents the crucial role symbolic labeling plays in memories of spatial patterns.[58–60] It also raises questions and points out suggestive interpretations for connectivity between operations in different parts of the modular cortex [Fig. 11, lower panel: 1—symbolic to spatial; 2—symbolic to sequential; 3—spatial to sequential; 4—sequential to read-out (RO); R—visual cortex to spatial; T—visual cortex to symbolic; EC—efferent copy to spatial].



APPEARANCE	SENSATION	SENSORY ORGANIZATION	PERCEPTION	REPRESENTATION
Chaos	Impression	CNS processing for velocity: v = dx/dt	A planned forceful, determined activity	Noumenon Symbolic
Class of appearances	Doctrine of specific nerve 'endings'	Pre-attentive psychophysics	Perception Intuition	Ideal Notion
Bottom up stuff (not 'things'!)	Bottom up physiology without space and time	Bottom up neurophysiology with space/time computation	Top down active looking scanpath as operational phase of perception per se	Top down cognitive model

Where does TD meet BU?

Levels I, II, and III meet level IV in the retinotopic visual cortex.

Fig. 15 Philosophical approach to perception. Five stages of the perceptual process (five columns) are illustrated with icons (upper), also showing BU and TD processes (curved arrows).

5 Philosophy and Neurology of Perception

The form or representation of visual memory in the brain is an important goal in understanding how the scanpath theory of perception works. Philosophers have developed a metaphorical approach to the various stages of perception (Kant). These beginnings of our understanding are acute in their insights, and we have further developed them with the scanpath mechanism in mind. In another trend of human study, recent neurological and neurophysiological studies have emphasized the modular cortex. MRI, magnetic resonance imaging, studies of the human brain while cooperating subjects are carrying out perceptual tasks, have brought new objective evidence to this frontier. We will review the various parts of the cortex, visual, parietal, and motor, to indicate the locations for different components of the mechanisms postulated by the scanpath theory. In addition, insights from the connectivity of cortical neurons (Braitenberg) and from many neurophysiological experiments have contributed. Finally, cognitive psychologists have become interested in attention shifts (most usually associated with eye movements) and have developed ideas along the lines of the scanpath theory for top-down vision. Thus, the scanpath theory is no longer an isolated voice.

5.1 *Perception and Sensory Organization*

Philosophers have long speculated about the nature of visual perception and several have suggested that we see in our *mind's eye* and have clarified the several stages that must be accomplished by verbally defined brain mechanisms. The scanpath brings scientific evidence to these issues but of course we have greatly benefited from the thoughts of Plato, Decartes, Kant, McCulloch, and we have tried to summarize their views. Using four terms defined by the philosopher Kant and adding a fifth component, a visual perceptual schema has been developed to incorporate the relevant concepts of experimental metaphysics. The senior author appreciates early discussions with Professor W.H. Zangemeister,[61] that led to a preliminary version of Fig. 15.[31,62,63] We start (Fig. 15, column one) with the world of *APPEARANCE*, the *chaos* of early Greek philosophers; in our terminology it is called *BU stuff*. At one time, we used *things* for the so-called *real* outside world, but an anonymous discussant pointed out that by the time the brain had done figure-ground separation to identify an object as distinguished from background, and applied knowledge about physical coherence of the object, much of the perception of the object had already been accomplished. The next stage (Fig. 15, column two), *SENSATION*, represents the inflows of energy onto body sense endings. It now appears that the filtering for *specific nerve energies* is actually accomplished by *specific nerve endings*, and specific nuclei on which they project.

We call the next stage (Fig. 15, third column), *SENSORY ORGANIZATION*, BU physiology, wherein the Kantian internal constructs of space and time are added. The frog's eye, using *bug* detectors, can calculate the velocity of a small moving spot accurately enough to keep frogs very well in bugs.[64,65] While it took 350 million years of vertebrate evolution to arrive at the frog's eye, yet another 350 million years was necessary to arrive at the brains of McCulloch and Lettvin, so that they would be capable of demonstrating the elegance of this aspect of sensory organization of the frog's eye. Since velocity requires both space and time computation, it is clear that these Kantian internal processes have been captured by evolution.[66,67]

If we jump ahead (Fig. 15, rightmost column five) to *REPRESENTATION*, the *ideals* of Plato and the *notions* of Berkeley, we see that our term, *TD cognitive models*, may perhaps be symbolized with a file drawer icon. We will return to the question of representation in a subsection below. Such models, acting TD onto the critical stage (*PERCEPTION per se*, Fig. 15, fourth column) can be seen to be planned, forceful, determined sets of activities.[68,69] In our model for perception, the TD active looking scanpath plays its role as the operational phase of perception per se. The set of five columns (Fig. 15), dissecting the overall perceptual process, leads to an important question we can pose for the neurophysiologist, *Where does TD meet BU?* Our conjecture is, where TD iconic inputs to levels I, II, and III of the visual cortex meet BU iconic visual signal information going to levels IV and V in the retinotopic visual cortex (Fig. 16, lower, right panel). This is the site of the *iconic matching* process.

5.2 *Visual Cortex: Where Does Top-Down, TD, Meet Bottom-Up, BU?*

In their famous paper, Pitts and McCulloch conjectured that the inflow information from eye and lateral geniculate would reach the striate cortex (Fig. 16, lower, right panel).[70,71] This was a BU theory as was the later frog's eye paper.[65]

We have now modified this BU approach to add TD perception. The visual cortex has a retino-topical organization that is apt for matching a TD iconic subfeature representation with incoming BU sensory signal flows. Likely some interactive feedback process could match these two maps, one TD, the other BU, to some criterion of fit (see Secs. 5.3 and 5.5 below). This, then, permits the scanpath, if confirmed to this point, to continue to the next ROI or subfeature of the representation. In this way, the TD model moves, fixates, and foveates the eye, to bring forward successive subfeatures for checking. An absence of fitting forces a new model and a revised scanpath. In this way, the scanpath as an operational mechanism plays an active role in the overall perceptual process.[72]

5.3 *Representation and the Braitenberg Cortex*

Our results are interpretable in terms of a set of models or schemata (Fig. 16). These models suggest visual patterns of thinking about (i) procedures for visual perception and recognition, (ii) the macroscopic, and (iii) microscopic neuroanatomical underpinnings of these memory processes as they are interpreted according to current neurological knowledge, and (iv) quantitative and normalized values for relative strengths to the several components of memory binding (Fig. 16). The connectivity of the cortex is vast; studies by Valentino Braitenberg and other neuroanatomists, from Golgi and Cajal on, have illuminated many aspects of this constrained mesh and link the microscopic and macroscopic views.

Some principles of cortical connectivity are listed, largely abstracted and perhaps with erroneous simplifications, from Professor Valentino Braitenberg.[73–76]

1. The idea of columns of cortical cells may have been initiated by Warren McCulloch (1945); then Vernon Mountcastle with the somatosensory cortex, and David Hubel and Torsten Wiesel with the visual cortex, strongly supported this concept in the 1950s.
2. Six layers or levels of cortical neurons.
3. Almost all are pyramidal cells; exceptions seem to be truncated or inverted pyramidal cells, with eponymous names.
4. Ponder the constraint of a pair of cortical cells not to connect more than twice with one another—one connection in each direction. How then is any pair more connected than any other pair? By virtue of their connectiveness to a common group of cells. Each cell has about 20 000 outputs (axons) and about 20 000 inputs (dendritic knobs or synapses). If two cells share none of their other connections then they are *unrelated*; if almost all of the connections are to the same cells then they are closely *related*.
5. For each pyramidal cell the 20 000 inputs and outputs connect almost exclusively to other cortical cells. On average, only about one output proceeds toward an output motor relay and only about one input arrives from a sensory waystation. Let not the skeptic deny that most of our brain computation is within the cortex and not with multiple stages of processing input and output. Of course, the design of present-day experiments forces this input–output view with paradigms to test responsivity to stimuli and to observe regular output responses to stimulation.
6. Lateral axons and their web of connectivity show a decreasing density of connections with distance. Apical axons to apical dendrites appear relatively independent of distance, leading to widespread connectivity. Indeed the size and number of a column in a human brain is sufficient for each column to be connected with every other column of the cortex. For the mouse with its much smaller brain, this is still true. The lesser number of cells in a column still suffice for this order of connectivity to the lesser number of columns.
7. Braitenberg and Almut Shutz showed in hamsters that the tremendous growth of dendritic knobs and axonal connections occurred developmentally before these creatures experienced the outside world. Similarly, development in Coghill's salamander embryos went on apace and functional connectivity reached an appropriate stage even when the salamanders were raised in anesthetic solution until tested.[77]

5.4 *Function Magnetic Resonance Imaging, fMRI*

Since the classical studies of Hubel and Wiesel, a number of approaches have developed to further our understanding of the neuroanatomical and neurophysiological substrates of cortical connectivity.

We are excited about the new experimental paradigms using fMRI and PET as non-invasive techniques for working with alert, cooperating humans engaged in perceptual and other high-level tasks. For the benefit of readers, we include a few references to current, ongoing work in fMRI. In general, the neurological studies and neurophysiological findings regarding the modular cortex have been strength-

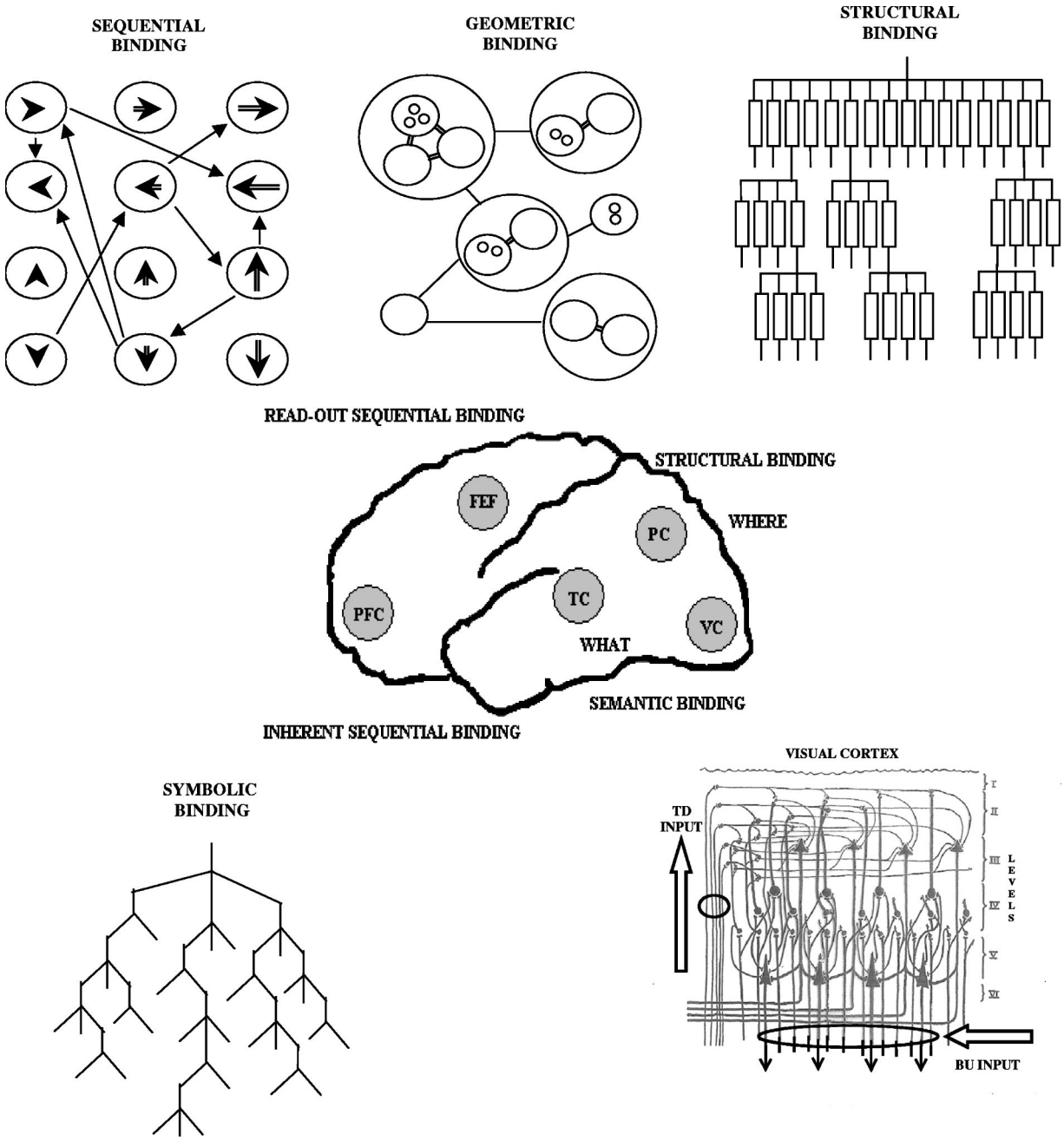


Fig. 16 Cortical representation of perceptual processes. Although only the microanatomy of the visual cortex is known well enough to support a graph theoretical model, yet we have suggested a variety of such graphs for structural, sequential, and symbolic binding, with loci as per labels in the modular cortex. Geometrical binding is used in our modeling schema, for syntactical interaction between foveal ROIs and peripheral segments. Different forms of the graphs do not represent any knowledge about feasible or understood properties of the brain, but rather stress our ignorance. Six levels or layers of the visual cortex (lower, right panel), known from neuroanatomy, are suggested as the iconic matching region, where TD input to the visual cortex, at layers I, II, and III, interact with BU input going to layers IV and V, from retina via geniculate and optic tracts (modified from Pitts and McCulloch, 1947).

ened and deepened with these ongoing human researches.[9,23,24,28,36,48,49,51,52,56,59,60,78–86]

5.5 *Macroscopic Cortical Processes: Where does Memory Dwell?*

Recently, especially with the advent of functional magnetic resonance, fMRI, and its associated imaging technology, there has been an increase in localization studies on awake cooperating humans that has led to a new *phrenology*—this time hopefully based upon more scientific evidence.[87,88]

There is an extensive literature on perception and EMs to which the important fields of visual cognition and memory have contributed in the recent past. We would like to point out that the publication of the scanpath theory in 1971 preceded cognitive science as a named field of research and, of course, we have tried to keep up with im-

portant contributions during the intervening 30 years as the long list of references in this review attest. We have also included in our references several texts such as Palmer's that should make it easy for the reader to access the wider literature.[89]

Classical neurophysiological techniques have enriched our knowledge of the animal brain, and by analogy, of the human brain; these are reinforced by longstanding and recent neurological studies. Of course, animal experiments are difficult, especially considering that it is impossible to have verbal interaction with the subjects of the experiment, so essential for studying higher-level functions. As mentioned in the paper, the design of neurophysiological experiments has been forced into paradigms that are exclusively input–output studies. That is, a visual stimulus has consequences which can be measured in various locations in the animal brain. Contrariwise, it is all but impossible to measure what the TD functions of the animal brain are signalling to these regions. However, with great ingenuity, a number of inroads have been made.[87–96]

Most neurophysiology is neuroanatomy, that is, locations of regions that show activity during particular sensory processing. Pioneering and future studies in actual microcortical neurophysiology are expanding with multiple electrode, optical and molecular biological approaches.[56,90,96–106]

Of particular interest to our own studies are pathways connecting the visual cortex to other cortices. The ventral pathway from visual cortex, VC, to the temporal cortex, TC (especially left-side), is the *what* pathway to which we attribute semantic binding; in similar fashion, the dorsal pathway from VC to the parietal cortex, PC (especially on the right side), is the *where* pathway to which we attribute structural binding. Spatial vision and memory and their uses in animal and human behavior are crucial functions that have been widely studied.[107,108] There are strong known connections (Pribram's law) from PC to the prefrontal cortex, PFC, that are related to temporal sequencing, and to which we attribute inherent sequential binding. Then connections continue to the frontal eye fields, FEF, to which we attribute read-out sequential binding. Of course, there are different motor areas for different behaviors used to indicate imaged loci and sequences in our experiments and in normal behavior, more generally.[37] Indeed, our experiments were designed to test some of these physiological–anatomical conjectures and to determine whether there were differences between inherent and read-out sequential bindings that could be captured in an experiment. In 1970, after a lecture by Professor Bela Julesz on his famous random-dot stereograms, L.W. Stark asked if he thought psycho-anatomical procedures obeyed a transitivity rule. The question illustrated the possible complexity of cortical connectivity; our diagrams are only a simplistic beginning.

We also use as metaphor, a robotic computer vision study (see Sec. 6) that has a complete TD model of the robot working environment, the robot kinematics and dynamics, the pose of the robots and the monitoring cameras (see Sec. 6). Here the model directs and limits the scanning of the video images to known positions of the ROIs in the 2D camera projections of the 3D-operating world. The model may be displayed on a computer screen for the supervisory controller to observe. Now, we ask a hypothetical question, *Where in the computer is the model located?* The answer makes us realize that the model is a collection of noniconic programs and parameters, widely distributed in active memory, in rotating memories, in registers, caches and pointers of the running programs and most often cannot be definitively located. This metaphor tempers our attempts to fix memory loci in the brain.

We know little about microscopic cortical processes.[98,105] As Hubel and Wiesel have pointed out, their classical work served to locate processes, rather than to establish how these processes occur. Similarly, the new phrenology substantiating the modular cortex, and as well, our studies, serve to fix anatomical locations. We therefore have used a variety of graphs to express our ignorance of function (Fig. 16, multiple graphs for different functions).[109–114] We do not at all suggest that the differences in these graphic displays represent known functional differences for the macroscopic modules. Pioneering and future studies of microscopic cortical function are and will be an exciting subject. What we emphasize is that the different memory functions, or bindings, in different parts of the modular cortex, must be carried out by some cellular networks, as postulated by McCulloch and Pitts. The cellular anatomical diagram for the visual cortex alone serves to give body to the above discussion.[60,105,114]

A number of residual questions naturally arise. One is how cognitive models are originally formed, and an answer to this has been sketched out in an earlier paper.[31] Another asks what is the difference between BU collecting of visual information and TD cognitive processes leading to signature scanpath? Earlier we suggested that BU signal information is brought to the retina, geniculate, and optic tract to layers 4 and 5 of the visual cortex, area 17. At the same time, the internal cognitive model brings up symbolic sequential and spatial information for many part of the brain to form an icon image of the points of interests to layers, 1, 2, and 3 of the visual cortex, to match the BU foveation signal. Thus we have tried to make explicit the distinction between the BU and the TD aspects of the model.

6 Implementation of a TD Computer Vision Model

For some years, we have been working with a top-down robotic vision system. Here, the internal top-down schema consists of three-dimensional knowledge of the robotic working environment, the pose kinematics and dynamics of the mobile robots, and various paths, obstacles, and tools required for the functioning of the robots. Simulation of this TD computer-vision model enabled us to have an explicit set of algorithmic mechanisms to give body to the scanpath theory and to explicitly illustrate the feasibility of the scanpath approach. It further lends itself to specific engineering considerations, such as signal-noise ratios, redundancy, and robustness in performance.

Because the scanpath theory rests upon continuing studies of the human brain, we necessarily lack a complete operational model. There are some neural models with BU approaches[115,116] and as well, a general appreciation by the computer vision world of an important future role for *image understanding*.[117–122] For some years now, we have developed a vigorous, explicit and functioning model of the TD scanpath scheme to aid our researches into robotic vi-

sion. In fact, the scanpath model for robotic computer vision has proven very successful in a number of laboratory engineering studies. In addition, it has been used in a giant civil engineering project where unmanned robotic vehicles such as back-hoes and bulldozers have built earthen dams to divert pyroclastic volcanic flow in Japan.[123–137] We now explain this model in some detail.

6.1 *TD Robotic Vision*

The complete model of the robots (Fig. 17, upper and middle panels) consists of compacting vehicles carrying out civil engineering dam building.[138] A set of visual enhancements, VEs, made up of prominent lights is easily detected by distant cameras. For the image processing aspect of the scheme, it is important to underline that the model of the robot includes the knowledge of the placement of these luminaires. In addition, the known camera loci, directions, and optical parameters enable prediction of the 2D projection of the scene onto any particular camera image plane.[125–128,139] The display mode (Fig. 17, upper and middle panels) indicates the model expectation of each luminaire location by showing white boxes outlining ROIs (upper right); expected locations may not be the actual locations and thus may require feedback correction.

As with TD scanpath control, the robot model predicts where the vehicle will move and this provides anticipatory information for locating the ROIs. According to the scanpath model, the image processing algorithms will move in sequence from ROI to ROI in the camera plane to carry out the various image-processing procedures (heavy white arrows, Fig. 17, middle panel, left). Thus, the alternation that occurs in the case of computer vision is similar to the alternation between TD motor-control of EMs for successive foveations, and the matching of the BU visual signals to the TD iconic model in the visual cortex in the case of human vision.

After image-processing steps, the display mode indicates the locus of the actual measured VE with a cross (middle right). The vector of actual locations is then passed to the feedback control mode.[125,133,134,136,138]

6.2 *Stages in TD Model-Based Image Processing*

The cornerstone of the algorithms is the model that represents the TD information the system has about the robotic working environment, RWE.[128] This model consists first of the robot kinematics; in addition, the dynamic component describes how the model configuration changes over time. The remote camera component describes the pose and geometry of each camera; finally, the locations and sizes of objects that may interact with the robots are also stored in the RWE. As discussed above, the VEs aid in image processing, and their geometrical representation is also reflected in the model. Now, given this environment, the visual algorithms perform four main steps in sequence.

6.2.1 *TD pre-filtering*

The TD model predicts the expected incoming signals: the 3D locations of the VEs using the last known pose of the robots, the kinematic model of the robots, and the control signal history. Using known camera pose and geometry, the 3D locus prediction is then projected onto a camera frame of reference. An ROI with the resulting location is then assigned. The estimated apparent size of the feature is calculated in a similar manner, and alters the ROI size for that feature. Implicitly, the image processing algorithm output of the ROI is logically bound to the feature at this stage.[140]

The viewability or detectability of the predicted features is thus aided by the known estimated locations of all objects in the RWE. For example, occlusion or possible overlapping of the VEs by the robot itself or by other known objects in the environment can be predicted, and a sampling dependability factor then generated. Sensitivity of robot pose to the 2D loci of features viewed from a given camera may be calculated; the jacobian matrix provides static and dynamic weightings. Both sensitivity and dependability factors are used to determine the significance of the sampling in a particular ROI. The dependability factor is of particular importance if the sampling cost of each ROI is high, for example, if the sampling speed is slow with respect to a limited time window for image processing and thus a sampling priority has to be assigned to each ROI. The sampling sequence of the ROI, similar to the scanpath sequence, is then generated based upon these factors.

6.2.2 *BU image processing*

The TD model applies an appropriate BU image processing algorithm, IPA, suitable for the ROI and its feature of interest. In the case illustrated, features are spherical light sources and the IPAs utilized are adaptive thresholding followed by a centroid calculation. Video camera images, even under the best conditions, are often very noisy (note the 3D gray-level pixel distribution, Fig. 17, lower panels, left). Indeed our design of the luminaires was an engineering attempt to provide adequate signal/noise ratios. By restricting the image processing only to the small ROI area, the amount of noise impinging upon our signal processing is greatly reduced; thus, the adaptive thresholding techniques yield robust results.[141] The ROIs are indicated as rectangular vertical boxes (Fig. 17, lower panels, right), where the top border of each is the actual adaptive threshold, utilized in each local area. The VEs can be clearly distinguished as narrow hilltops (Fig. 17, lower panels) above these adaptive thresholds, and contribute to the robustness of the BU image processing procedures controlled in this TD fashion. Indeed, the fovea of the retina and its magnified cortical representation (Fig. 1) must also possess local adaptive advantages of a similar sort.

6.2.3 *TD plus BU post-filtering*

The next procedure verifies the integrity of the individual centroid measurements in the contest of the overall TD model. For a feature type, in this instance a spherical VE, the system generates a criterion, such as moment invariance, to test the possibility of error due to unexpected effects. Thus, significant ellipticity of features in an ROI would be marked as unreliable, and thus, weighted less in the data integration part of the program.

6.2.4 *TD data integration*

The RWE model is next synchronized with the feedback measurements so as to produce a consistent updated model. Each ROI locus creates two constraints in the estimation of robotic pose. These, the dependability factor, TD model-pose and occlusion information, and the reliability factor,

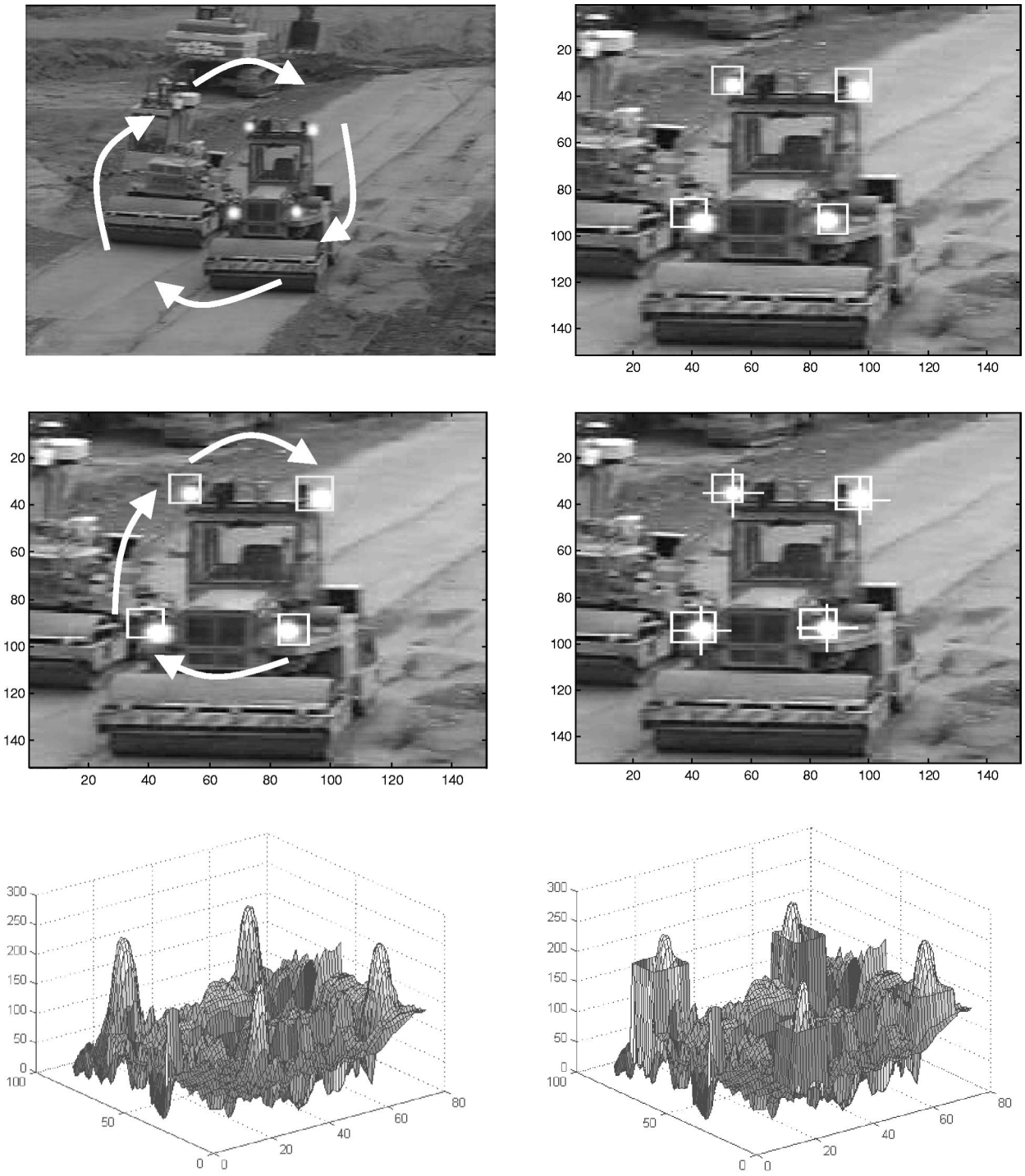


Fig. 17 TD scanpath scheme for robotic vision. Four image processing steps showing robot vehicles with VEs (upper left) and model ROI-predicted locations (white squares, upper right). Note scanpath sequence for computer image processing of ROIs (white arrows, middle left) yielding centroid-calculated loci (white crosses, middle right). Pixel intensity diagram forming a 3D representation of the video image (lower panels). Note hilltops representing VEs. By predicting ROI loci using TD model (rectangular boxes, right), it is possible to do adaptive thresholding only in a small localized region, and thus, achieve important signal-to-noise ratio improvements. Clearly, foveal fixation in normal human vision achieves the same functionality.

judged by the fit of the image-prediction to the BU processed image signal, are fed into an optimization routine that finds the optimal robot pose such that constraint violations are minimized. If fitting error is high, indicating a failure in the IPA procedures, re-initialization of the IPA subsystem is performed.[136] Otherwise, the resultant estimated robot pose is accepted, updates the model, and the next IPA iteration is performed. These stages of the com-

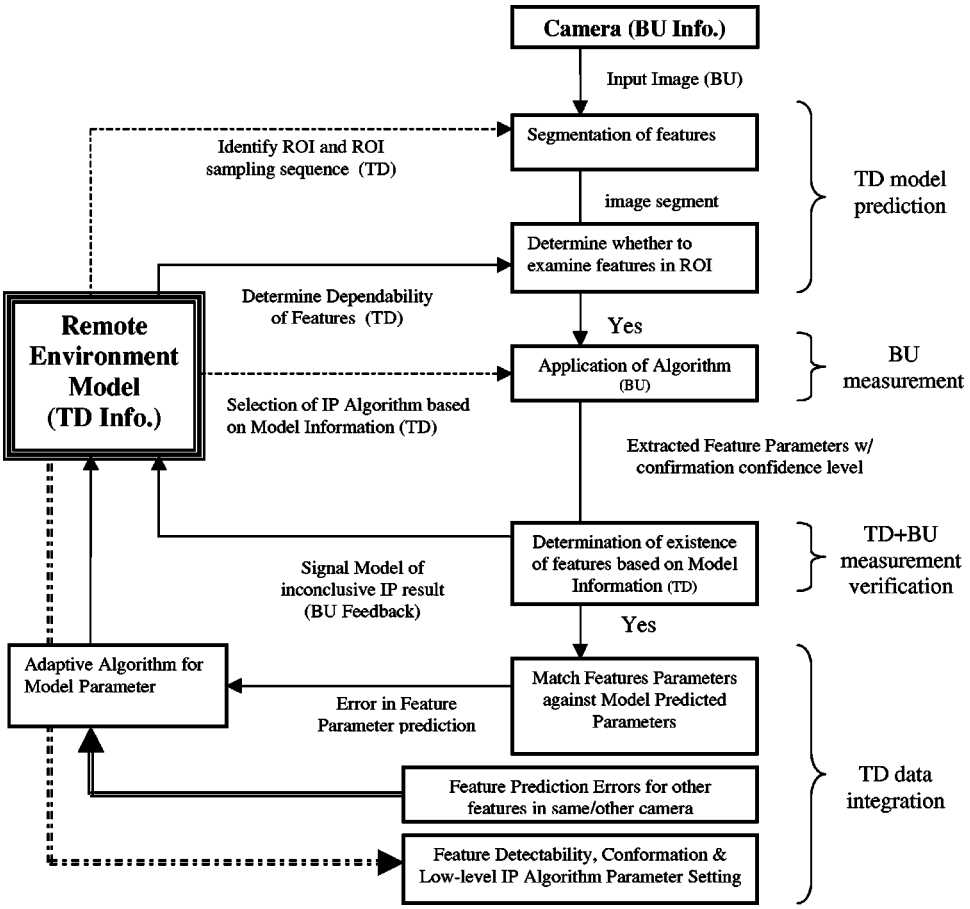


Fig. 18 Block diagram explicating telerobotic vision scheme. Flow diagram scheme to aid in understanding steps of our telerobotic TD scanpath approach to image processing and to supervisory control.

puter vision scheme have not only a sequential structure but also multiple interactions (Fig. 18).

An important question to ask is, *just how is visual knowledge kept in the brain?* Recall that re-cognition is re-knowing. Earlier we showed how different parts of the modular cortex house different components of the entire spatial cognitive schema, the internal vision in the mind's eye. While the visual memory in the telerobotic scheme is primitive compared to the visual memory of real objects in real scenes in the human brain, yet it shares many critical aspects. For example the spatial relationship of the robot working environment, the location of tools and potential interaction with other robots must all be stored (in mathematical and not iconic form) in the mind's eye of the computer. In addition, the kinematics and pose of the robots must be similarly stored in order for control to be accomplished. Finally, in close analogy to the scanpath, a series of visual enhancements, VEs, must be located in the 2D projection of the camera plane so that visual attention can be directed from one expected location of VE to the next and so on in sequence. If the VEs are in fact found at the expected loci then all is well, and the control computation proceeds. If not, then perhaps redundancies can compensate for a missing VE or two. If even this is not sufficient, then the system downgrades to a reinitialization process to find and make the most satisfactory new match.

7 Summary

In the following abbreviated summary, we group our findings as TD Vision, Representation, Philosophy, and Computer Vision.

7.1 *TD Vision*

This paper has considered the TD aspects of human vision to be equally (or more) important to vision as a whole than are the usual textbook presentations of BU vision with constellations of psychophysical and neurophysiological experimental paradigms. Recall that classical experimental designs themselves prejudice the vision scientist to think in unidirectional input-output terms. We began with the essential role that EMs play, and their function in the TD control of the flow of visual information. The scanpath theory proposes that an internal spatial-cognitive model controls perception and the active looking EMs, of the scanpath sequence. Evidence supporting the scanpath theory[2–4] includes experiments with ambiguous figure and visual imagery.[20,22] Application to dynamic scenes, although only beginning, yet has many lessons for further visual studies.[25] We have also provided an introduction to the experimental procedures including careful calibration of EMs, definition of ROIs, and the analysis and comparison programs for studying scanpaths.

There is a big difference between descriptive accounts of EMs and a theory formulated so as to be quantitatively tested. We have inserted material explaining how the scanpath theory can and has been tested not only qualitatively, but in a quantitative manner that explicitly lends itself to statistical analysis. In particular, we have been able to document that the predictions of the theory for ambiguous figures and for eye movements, EMs, during visual imagery have been confirmed. When we applied our tests to dynamic visual scenarios, dynamic smooth pursuit EMs were generated by the observer tracking moving targets. In this case, the straightforward application of the theory, predicting only fixational EMs and saccades, could not be confirmed and the theory had to be modified. These smooth pursuit EMs play a large role in scanpaths while subjects are observing dynamic scenarios and have a very interesting characteristic—they maintain the fovea over the moving object as long as this is possible and as long as the moving object is one the top-down spatial cognitive model continues to pay attention to. Given that the sensory system *sees* the moving object fixed on the fovea, we were able to modify the scanpath theory to include smooth pursuit as a sort of *superfixaton*. We hope that the above examples show that as experimental scientists we have put the scanpath theory to experimental test and we have reacted to complex situations by enlarging and modifying the theory.

The scanpath research and the recent studies on memory binding[32] described in this paper aid an understanding of the dual role played by TD and BU visual processes (Fig. 19). The TD representation in the mind's eye, and as elaborated in this paper, throughout the brain, controls not only EMs, but the placement of spatially defined iconic models in the visual cortex. Here they are matched with BU signal information, arriving from the so-called *real* world. These BU signals are known to have a distorted log-polar iconic form (Figs. 1 and 19), up to and including the primary visual cortex. First note that the picture (Fig. 19, top) in the *mind's eye* is close to the picture (bottom) in the *real* world, or our species, heavily dependent upon vision, would have disappeared. Successive EM fixations produce retinal images (three are shown, just above the *real* world picture); from this ensues cortical magnification of foveal regions and cortical minification of peripheral low resolution segments. At the same time, the TD representation sends similar iconic representations (three are also shown here, just below the mind's eye picture) to the visual cortex for matching; TD scanpath EMs predict the spatial loci for these matches. At the visual cortex (Fig. 19, middle), iconic matching of TD and BU occurs.[142–145]

7.2 *Representation*

More recently, we have used the repetitive scanpath EM sequence to approach problems of the representation of the visual image in the brain. We suppose that there are several levels of *binding*—semantic or symbolic binding, structural binding for the spatial locations of the ROIs and sequential binding for the dynamic execution program that yields the sequence of EMs. The scanpath sequence has enabled experimental dissection of these various bindings that appear to play independent roles and are likely located in different parts of the modular brain. Cortical localization has advanced recently with fMRI studies on cooperating humans and supporting animal experiments, but this is not to ignore important subcortical loci with likely major functions.

Experiments carefully described in this paper show that symbolic binding strongly influences sequential binding, but cannot overwhelm spatial or structural memory. Sequential bindings themselves appear to be partitioned between inherent and *read-out* memory. The inherent sequential memory component is closely linked to structural binding, whereas the read-out components are apparently modified by each of the different motor systems we have explored—EMs, hand control of a cursor on a computer screen, and locomotion over a grid on the laboratory floor.

7.3 *Philosophy*

The background of visual perception has ancient roots in philosophy. Although philosophers have long speculated that *we see in our mind's eye* and that we can have no certain knowledge of the external chaos or classes of appearances in which we find ourselves, yet until the scanpath theory no strong scientific evidence was available to support their conjectures. (The senior author was influenced by the strong TD structures he himself built into so-called *artificial intelligence* programs so that the remaining *self-organizing* was largely a matter of optimization, itself influenced by successive modifications.)[112,113,146,147]

7.4 *Computer Vision*

Use of the TD scanpath for robotic computer vision has proved itself in a series of applications. Clearly TD information can change an ineffective vision system into a robust feedback mechanism for control of telerobots. Another positive aspect has been a complete and explicit demonstration of how a scanpath mechanism works in the artificial system. This model may also suggest further directions for extending the sparse experimental data about the human brain mechanisms controlling our own vision.[46]

Appendix: String-Editing Algorithm

The string-editing algorithm is a discrete dynamic programming method.[1] Using the operations—insertion, In, deletion, De, and replacement, Re, the algorithm of Wagner[45] finds the minimum distance or cost to convert from string2(i) to string1(j); this defines the matrix (Fig. 20, upper matrix). The two strings label the rows, string2(i), and columns, string1(j), of the matrix. Insertions result in horizontal shifts, deletions in vertical ones, and replacements produce shifts along the diagonal. Each operation may add to the cost; the coefficients of the matrix are the hypothetical costs to reach that cell. The middle matrix describes the stage just after the insertion of a *B* in the preceding step with an added cost of 1 (circled coefficient); the next step enables the *Cs* to match without added cost. At the end (lower matrix) deletion of *A* (note vertical shift) finally matched string 2(i) to string 1(j), at a minimum cost of 2 (circled coefficient). Thus, the normalized distance is equal to 0.4 (2 divided by string length, 5); Ss, the sequential similarity index, is equal to 0.6 (1–0.4).

A short program (Fig. 20, equations listed to right of matrices) enabled these calculations.

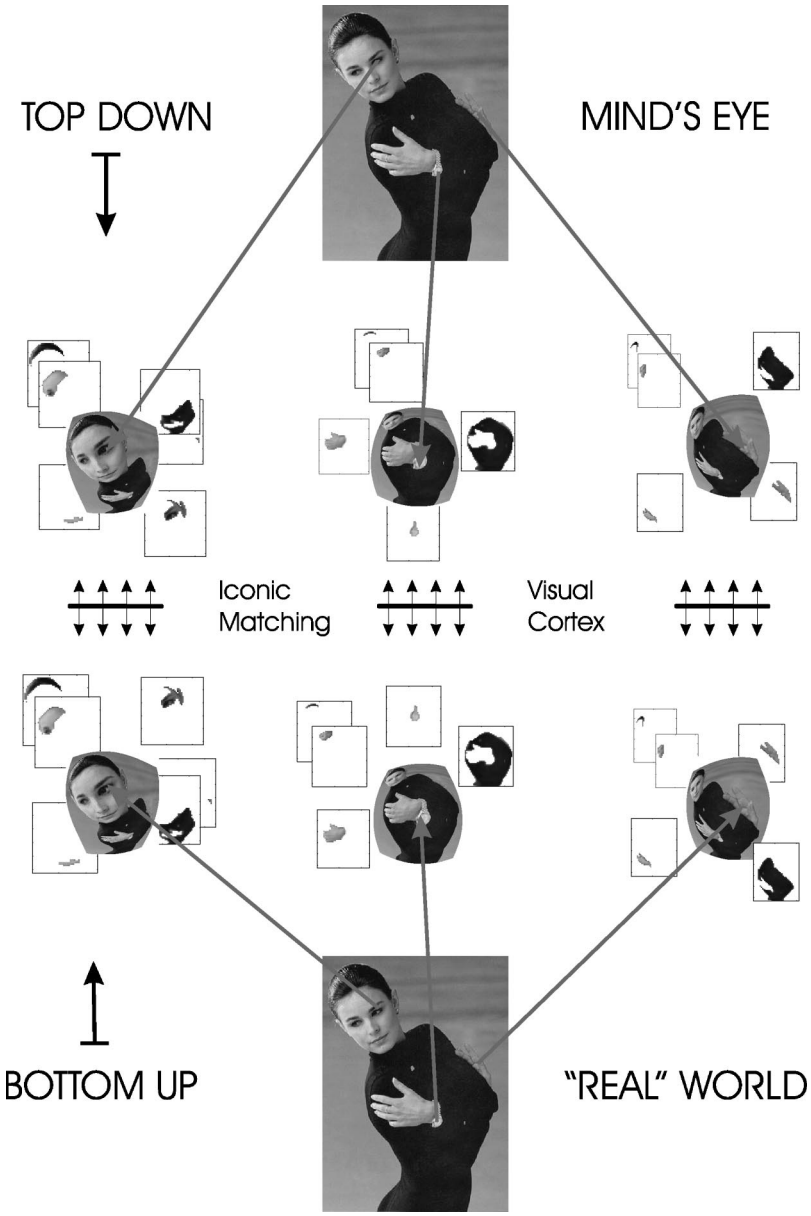


Fig. 19 Iconic matching in visual cortex TD representation and BU signals. BU retinal image is shifted with each EM fixation to provide a centered and magnified foveal projection in the visual cortex. These may be matched by predicted TD iconic representations from the mind's eye image. Continuous periphery is shown broken into segments, also suitable for TD symbolic coding.

1. Initialize D-matrix to zero. An additional possible step (not herein employed) is to truncate longer string.
2. Distance of first to null—do this by deleting each character in string j one by one; at most this will equal string length. Distance of second to null and first to second also calculated and will also equal at most string length.
3. Using dynamic programming, proceed from row to row and from top to bottom to calculate minimum distances; this fills out the D-matrix.
4. This triple computation, in addition to replacements, takes into account the effect of deletions and insertions in sidewise shifting of string elements, and thus traveling along a minimum-cost discrete path in the D-matrix. Wagner proved that these operations will find the optimum solution;[46] this extended the discrete dynamic programming algorithm.
5. Lowermost right corner of D-matrix will be the minimum total cost of making string2(i) identical to string1(j).
6. An additional possible step (not herein employed) is to assign non-unity costs to each operation.

Our use of string editing in matching loci and sequences in images is a bit unusual. However, once we have established a finite state automaton and equivalently, a Markov model (see Fig. 4, lower row), the sequences are inherently

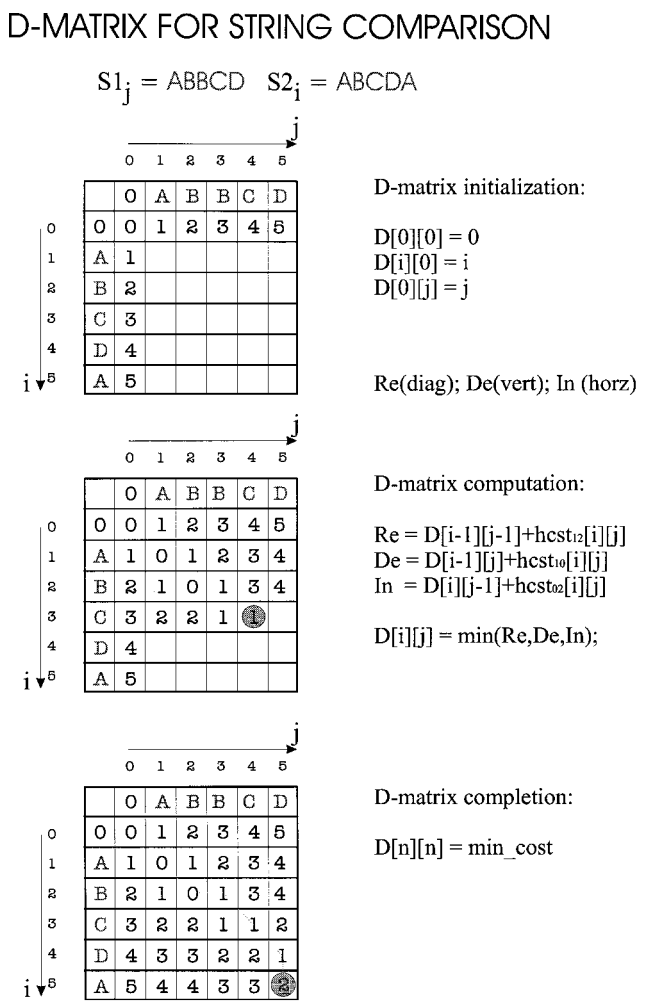


Fig. 20 String editing algorithm. Successive stages of the discrete dynamic programming algorithm (matrices at left) document minimum cost optimization of string editing distance, and thus, accurate measure of string sequential similarity; computational equations at each stage (to right of matrices).

in a form appropriate for application of the string-editing algorithm. The widest use of string-editing algorithms is perhaps in spellcheck programs. The use in matching of double-stranded chromosomes and sequences of nucleic acids within them, is an important current application. By using perhaps as yet undiscovered principles of biomechanical mechanisms, it may be possible to assign weightings, or nonunity costs, to such strand-distorting actions such as those caused by insertions or deletions. For example, the redundant looping often seen in chromosomes may not be permissible for loops that are too short.

Acknowledgments

The senior author thanks two of his oldest friends: Professor Valentino Braitenberg, Emeritus Professor of the Max Planck Institute of Biological Cybernetics in Tuebingen, whose studies of cortical connectivity and cerebral and cerebellar cortical models have been an important stimulus; and Professor Elwin Marg, of the School of Optometry at the University of California, Berkeley, whose passionate interest in visual neurophysiology and pioneering single-unit studies of human brain have likewise been important. We wish to acknowledge support from Dr. Stephen Ellis and Dr. Michael Sims, scientific monitors for research grants with NASA-Ames Research Center; also Dr. Kamran Siminou of Neuro-Optics and Dr. Ken Kawamura of Fujita Research. Our laboratory colleagues served as subjects during preliminary control experiments, and often contributed discussion comments with bite bars in their mouths. Professor V. V. Krishnan and Professor G. M. Gauthier provided helpful detailed critiques. We also want to thank Dr. Bernice Rogowitz and Dr. Thrasyvoulos Pappas for their important suggestions and discussions, and Tina Choi at the University of California, Berkeley, for editorial assistance.

References

1. R. Bellman and E. S. Lee, ''History and development of dynamic programming,'' *IEEE Control Syst. Mag.* **4**(4), 24–28 (1984).
2. D. Noton and L.W. Stark, ''Scanpaths in saccadic eye movements while viewing and recognizing patterns,'' *Vision Res.* **11**(9), 929–942 (1971).
3. D. Noton and L. W. Stark, ''Eye movements and visual perception,'' *Sci. Am.* **224**(6), 34–43 (1971).
4. D. Noton and L. W. Stark, ''Scanpaths in eye movements during pattern perception,'' *Science* **171**(3968), 308–311 (1971).
5. M. Crosby, ''How do we read algorithms?,'' *Computer* **23**, 25–35 (1990).
6. M. Jeannerod, P. Gerin, and J. Pernier, ''Deplacements et fixation du regard dans l'exploration libre d'une scene visuelle [French],'' *Vision Res.* **8**, 81–97 (1968).
7. P. Locher and C. Nodine, ''The role of scanpaths in the recognition of random shapes,'' *Percept. Psychophys.* **15**, 308–314 (1974).
8. M. Mandler and J. Whiteside, ''The role of scanpaths in recognition of random dot patterns,'' *J. Undergraduate Psychology Res.* **3**, 84–90 (1976).
9. A. Mackworth, ''How to see a simple world: An exegesis of some computer programs for scene analysis,'' *Machine Intelligence* **8**, 510–537 (1978).
10. N. Mackworth and J. Bruner, ''How adults and children search and recognize pictures,'' *Human Development* **13**, 149–177 (1970).
11. N. Mackworth and A. Morandi, ''The gaze selects informative details within picture,'' *Percept. Psychophys.* **2**, 547–552 (1967).
12. R. Parker, ''Picture processing during recognition,'' *J. Exp. Psychol.* **4**, 284–293 (1978).
13. P. Schifferli, ''Etude par enregistrement photographique de le motricite oculaire dans l'exploration, dans la reconnaissance et dans la representation visuelles [French],'' *Mon. Psychiat. Neurol.* **126**, 65–118 (1953).
14. J. Senders, D. Fisher, and R. Monty, *Eye Movements and the Higher Psychological Processes*, Lawrence Erlbaum Associates, Hillsdale, NJ (1978).
15. A. Yarbus, *Eye Movements and Vision*, Plenum, New York (1967).
16. M. Mackeben and K. Nakayama, ''Express attentional shifts,'' *Vision Res.* **33**(1), 85–90 (1993).
17. R. McPeek, V. Maljkovic, and K. Nakayama, ''Saccades require focal attention and are facilitated by a short-term memory system,'' *Vision Res.* **39**(8), 1555–1566 (1999).
18. K. Nakayama and J. Joseph, ''Attention, pattern recognition, and pop-out visual search,'' in *The Attentive Brain*, R. Parasuraman *et al.*, Eds. (MIT Press, Cambridge, MA, 1998), pp. 279–298.
19. S. R. Ellis and L. W. Stark, ''Reply to Piggins,'' *Perception* **8**(6), 721–722 (1979).
20. L. W. Stark and S. R. Ellis, ''Scanpaths revisited: cognitive models direct active looking,'' in *Eye Movement: Cognition and Visual Perception*, D. Fisher *et al.*, Eds., pp. 193–226, Lawrence Erlbaum Associates, Hillside, NJ (1981).
21. S. Brandt, L. W. Stark, S. Hacisalihzade *et al.*, ''Experimental evidence for scanpath eye movements during visual imagery,'' Proceedings of the 11th IEEE/EMBS, Vol. 1, pp. 278–279, Seattle, WA (1989).
22. S. A. Brandt and L. W. Stark, ''Spontaneous eye movements during visual imagery reflect the content of the visual scene,'' *J. Cognitive Neuroscience* **9**(1), 27–38 (1997).
23. S. Kosslyn, *Image and Mind*, Harvard University Press, Cambridge, MA (1980).
24. S. Kosslyn, *Image and Brain: The Resolution of the Imagery Debate*, MIT Press, Cambridge, MA (1994).
25. T. T. Blackmon, Y. F. Ho, and D. Chernyak *et al.*, ''Dynamic scan-

paths: eye movement analysis methods,'' *IS&T/SPIE's Symposium on Electronic Imaging* (1999).
26. M. Weirda and W. Maring, ''Interpreting eye movements of traffic participants,'' in *Visual Search 2-Proceedings of the Second International Conference on Visual Search*, D. Brogan, A. Gale, and K. Carr, Eds., pp. 287–301, Taylor and Francis, London (1993).
27. M. Lawden, H. Bagelmann, and T. Crawford *et al.*, ''An effect of structured backgrounds on smooth pursuit eye movements in patients with cerebral lesions,'' *Brain* **118**(1), 37–48 (1995).
28. S. Kosslyn and D. Osherson, Visual Cognition, MIT Press, Cambridge, MA (1995).
29. J. Singer, S. Greenberg, and J. Antrobus, ''Looking at the mind's eye: experimental studies of ocular motility during day dreaming,'' *Trans. NY Acad. Sci.* **33**, 694–709 (1971).
30. L. W. Stark, Y. S. Choi, and Y. Yu, ''Visual imagery and virtual reality: New evidence supporting the scanpath theory explains the illusion of completeness and clarity,'' Visual Science: Papers in Honor of J. Enoch (1996).
31. L. W. Stark and Y. S. Choi, ''Experimental metaphysics: The scanpath as an epistemological mechanism,'' in *Visual Attention and Cognition*, edited by W. H. Zangemeister, H. S. Stiehl, and C. Freksa, Eds., pp. 3–69, Elsevier, Amsterdam (1996).
32. L. W. Stark, C. M. Privitera, H. Yang *et al.*, ''Scanpath memory binding: multiple read-out experiments,'' *IS&T/SPIE's Symposium on Electronic Imaging*, San Jose, CA **3644**, 495–510 (1999).
33. J. Wolfe, ''Visual memory: what do you know about what you saw?,'' *Curr. Biol.* **8**(9), R303–R304 (1998).
34. Z. He and K. Nakayama, ''Surfaces versus features in visual search,'' *Nature* **359**(6392), 231–233 (1992).
35. K. Nakayama, Z. He, and S. Shimojo, ''Visual surface representation: A critical link between lower-level and higher-level vision,'' in *Visual Cognition: An Invitation to Cognitive Science*, Stephen M. Kosslyn, Daniel N. Osherson *et al.*, Eds., Vol. 2, pp. 1–70, MIT Press, Cambridge, MA (1995).
36. C. Ploner, B. Gaymard, and N. Ehrle *et al.*, ''Spatial memory deficits in patients with lesions affecting the medial temporal neocortex,'' *Ann. Neurol.* **45**(3), 312–319 (1999).
37. J. Wolfe, G. Alvarez, and T. Horowitz, ''Attention is fast but volition is slow,'' in preparation.
38. E. Llewellyn-Thomas, ''Movements of the eye,'' *Sci. Am.* **219**, 88–95 (1968).
39. T. T. Blackmon, Y. F. Ho, K. Matsunaga *et al.*, ''Eye movements while viewing dynamic and static stimuli,'' *19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (1997).
40. L. W. Stark, S. R. Ellis, H. Inoue *et al.*, ''Cognitive models direct scanpath eye movements: Evidence obtained by means of computer processing of perceptual eye movements,'' *XII International Conference on Medical and Biological Engineering, Jerusalem, Israel* (1979).
41. Y. S. Choi, A. Mosley, and L. W. Stark, ''String editing analysis of human visual search,'' *Optom. Vision Sci.* **72**(7), 439–451 (1995).
42. S. Hacisalihzade, L.W. Stark, and J. Allen, ''Visual perception and sequences of eye movement fixations: A stochastic modeling approach,'' *IEEE Trans. Syst. Man Cybern.* **22**, 474–481 (1992).
43. J. Kruskal, ''An overview of sequence comparison: time warps, string edits, and macromolecules,'' *SIAM Rev.* **25**, 201–237 (1983).
44. C. M. Privitera and L. W. Stark, ''Evaluating image processing algorithms that predict regions of interest,'' *Pattern Recogn. Lett.* **19**, 1037–1043 (1998).
45. C. M. Privitera, N. Krishnan, and L. W. Stark, ''Clustering algorithms to obtain regions-of-interest (ROIs),'' *SPIE: Electronic Imaging, San Jose, CA* **3959**, 634–643 (1999).
46. R. Wagner and M. Fischer, ''The string-to-string correction problem,'' *J. Assoc. Comput. Mach.* **21**(1), 168–173 (1974).
47. L. Stelmach, W. Tam, and P. Hearty, ''Static and dynamic spatial resolution in image coding: An investigation of eye movements,'' *Proc. SPIE* **1453**, 147–152 (1992).
48. R. T. Born and R. B. Tootell, ''Segregation of global and local motion processing in primate middle temporal visual area,'' *Nature* **357**(6378), 497–499 (1992).
49. J. Culham, S. Brandt, and P. Cavanagh *et al.*, ''Cortical fMRI activation produced by attentive tracking of moving targets,'' *J. Neurophysiol.* **80**(5), 2657–2670 (1998).
50. B. N. Flagg, ''Children and television: Effects of stimulus repetition on eye activity,'' in *Eye Movements and the Higher Psychological Functions*, J. W. Senders, D. F. Fisher, and R. A. Monty, Eds., pp. 279–291, Lawrence Erlbaum Associates, Hillsdale, NJ (1978).
51. R. Tootell, J. Reppas, and A. Dale *et al.*, ''Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging,'' *Nature* **375**(6527), 139–141 (1995).
52. R. Tootell, J. Reppas, and K. Kwong *et al.*, ''Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging,'' *J. Neurosci.* **15**(4), 3215–3230 (1995).
53. J. Elderfield, ''Seeing Bonnard,'' in *Bonnard*, pp. 33–52 Museum of Modern Art, New York (1998).
54. W. H. Zangemeister, K. Sherman, and L. W. Stark, ''Evidence for a global scanpath strategy in viewing abstract compared with realistic image,'' *Neuropsychologia* **33**(8), 1009–1025 (1995).
55. S. Zeki and K. Moutoussis, ''Temporal hierarchy of the visual perceptive systems in the Mondrian world,'' *Philos. Trans. R. Soc. London, Ser. B* **264**(1387), 1415–1419 (1997).
56. S. Zeki and A. Bartels, ''The autonomy of the visual systems and the modularity of conscious vision,'' *Philos. Trans. R. Soc. London, Ser. B* **353**(1377), 1911–1914 (1998).
57. H. Yang and L. W. Stark, ''How do we recognize images?,'' *IS&T/SPIE's Symposium on Electronic Imaging, San Jose, CA* (2000).
58. M. Tanenhaus, M. Spivey-Knowlton, and K. Eberhard *et al.*, ''Integration of visual and linguistic information in spoken language comprehension,'' *Science* **268**(5217), 1632–1634 (1995).
59. M. Tempini, C. Price, and O. Josephs *et al.*, ''The neural systems sustaining face and proper-name processing,'' *Brain* **121**(11), 2103–2118 (1998).
60. S. Thompson-Schill, M. D'Esposito, and G. Aguirre *et al.*, ''Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation,'' *Proc. Natl. Acad. Sci. U.S.A.* **94**(26), 14792–14797 (1997).
61. L. W. Stark, W. H. Zangemeister, B. Hannaford *et al.*, ''Use of models in brainstem reflexes for clinical research,'' in *Clinical Problems of Brainstem Disorders*, pp. 172-184, Thieme, New York (1986).
62. I. Kant, *Prolegomena to Any Future Metaphysics*, Bobbs-Merrill Company, Inc., New York (1949).
63. B. Russell, *A History of Western Philosophy*, Simon and Schuster, New York (1945).
64. D. Ingle, ''Prey-catching behavior of Anurans toward moving and stationary objects,'' *Vision Res.* **3**, (Suppl) 447–456 (1971).
65. J. Lettvin, H. Maturana, and W. McCulloch *et al.*, ''What the frog's eye tells the frog's brain,'' *Proc. IRE* **47**, 1940–1959 (1959).
66. L. Itti and C. Koch, ''Comparison of feature combination strategies for saliency-based visual attention systems,'' *SPIE: Electronic Imaging* **3644**, 473–482 (1999).
67. E. Niebur and C. Koch, ''Computational architectures for attention,'' in *The Attentive Brain*, R. Parasuraman, Ed., pp. 163–186, MIT, Cambridge, MA (1998).
68. K. Pribram, *Languages of the Brain*, Prentice-Hall, Englewood Cliffs, NJ (1971).
69. J. Searle, *Intentionality, an Essay in the Philosophy of Mind* Cambridge University Press, Cambridge (1983).
70. W. McCulloch, *Embodiments of Mind* (MIT, Cambridge, MA, 1965).
71. W. Pitts and W. McCulloch, ''How we know universals: The perception of auditory and visual forms,'' *Bull. Math. Phys.* **9**, 127–147 (1947).
72. J. Henderson and A. Hollingsworth, ''Higher level scene perception,'' *Annu. Rev. Psychol.* **50**, 243–271 (1999).
73. V. Braitenberg, *On the Texture of Brains: An Introduction to Neuroanatomy for the Cybernetically Minded* (Springer-Verlag, New York, 1977).
74. V. Braitenberg, ''The cerebral cortex as site of associative memory,'' *Sistemi Intelligenti* **2**(2), 213–227 (1990).
75. V. Braitenberg, *Vehicles, Experiments in Synthetic Psychology* (MIT Press, Cambridge, MA, 1994).
76. V. Braitenberg and A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*, 2nd ed., Springer, Berlin (1998).
77. G. Coghill, *Anatomy and the Study of Behavior*, Cambridge University Press, Cambridge (1929).
78. J. Coull, R. Frackowiak, and C. Frith, ''Monitoring for target objects: activation of right frontal and parietal cortices with increasing time on task,'' *Neuropsychologia* **36**(12), 1325–1334 (1998).
79. B. de Jong, R. Frackowiak, and A. Willemsen *et al.*, ''The distribution of cerebral activity related to visuomotor coordination indicating perceptual and executional specialization,'' *Cognitive Brain Research* **8**(1), 45–59 (1999).
80. P. Fletcher, T. Shallice, and C. Frith *et al.*, ''The functional roles of prefrontal cortex in episodic memory. II. Retrieval,'' *Brain* **121**(7), 1249–1256 (1998).
81. S. Kastner, P. DeWeerd, and R. Desimone *et al.*, ''Mechanisms of directed attention in the human extrastriate cortex as revealed by Functional MRI,'' *Science* **2**, 108–111 (1998).
82. A. Kleinschmidt, C. Buchel, and S. Zeki *et al.*, ''Human brain activity during spontaneously reversing perception of ambiguous figures,'' *Philos. Trans. R. Soc. London, Ser. B* **265**(1413), 2427–2433 (1998).
83. E. O'Sullivan, I. Jenkins, and L. Henderson *et al.*, ''The functional anatomy of remembered saccades: a PET study,'' *NeuroReport* **6**(16), 2141–2144 (1995).
84. M. Posner and M. Raichle, *Images of Mind*, Scientific American Library, New York (1994).
85. M. Rugg, P. Fletcher, and K. Allan *et al.*, ''Neural correlates of memory retrieval during recognition memory and cued recall,'' *Neuroimage* **8**(3), 262–273 (1998).

86. R. Tootell, D. Malonek, and A. Grinvald, ''Optical imaging reveals the functional architecture of neurons' processing, shape, and motion in owl monkey area MT,'' *Philos. Trans. R. Soc. London, Ser. B* **258**, 109–119 (1994).
87. S. Palmer, ''Visual perception and world knowledge: notes on a model of sensory-cognitive interaction,'' in *Explorations in Cognition*, D. A. Norman and D. E. Rumelhart, Eds., pp. 279–307, Freeman, San Francisco (1975).
88. S. Palmer, ''Reference frames in the perception of spatial structure,'' in *Cognition, Information Processing, and Psychophysics: Basic issues*, H. G. Geissler, Stephen W. Link *et al.*, Eds., pp. 141–174, Lawrence Erlbaum Associates, Hillsdale, NJ (1992).
89. S. Palmer, *Vision Science: Photons to Phenomenology*, MIT Press, Cambridge, MA (1999).
90. C. Colby, J. Duhamel, and M. Goldberg, ''Oculocentric spatial representation in parietal cortex,'' *Cerebral Cortex* **5**(5), 470–481 (1995).
91. A. Meystel, I. Rybak, and S. Bhasin *et al.*, ''Multiresolution stroke sketch adaptive representation and neural network processing system for graylevel image recognition,'' *Proc. SPIE : Intelligent Robots and Computer Vision XI: Biological, Neural Net, and 3-D Methods* **1826**, 261–278 (1992).
92. S. Palmer, ''The effects of contextual scenes on the identification of objects,'' *Memory Cognition* **3**(5), 519–526 (1975).
93. S. Palmer and R. Kimchi, ''The information processing approach to cognition,'' in *Approaches to Cognition: Contrasts and Controversies*, Terry J. Knapp, Lynn C. Robertson *et al.*, Eds., pp. 37–43, Lawrence Erlbaum Associates, Hillsdale, NJ (1986).
94. S. Palmer, J. Neff, and D. Beck, ''Amodal completion,'' in *Indirect Perception*, I. Rock, Ed., pp. 63–75, MIT Press, Cambridge, MA (1997).
95. I. Rybak, A. Golovan, and V. Gusakova, ''Behavioral model of visual perception and recognition,'' *Proc. SPIE Human Vision, Visual Processing, and Digital Display IV* **1913**, 548–560 (1993).
96. M. Umeno and M. Goldberg, ''Spatial processing in the monkey frontal eye field. I. Predictive visual responses,'' *J. Neurophysiol.* **78**(3), 1373–1383 (1997).
97. R. Andersen, ''Coordinate transformations and motor planning in posterior parietal cortex,'' in *The Cognitive Neurosciences*, M. S. Gazzaniga, Ed., pp. 519–532, MIT Press, Cambridge, MA (1995).
98. R. Desimone, ''Neural circuits for visual attention in primate brain,'' in *Neural Networks for Vision and Image Processing*, G. A. Carpenter and S. Grossberg, Eds., MIT Press, Cambridge, MA (1992).
99. R. Devalois, ''Color vision mechanisms in the monkey,'' *J. Gen. Physiol.* **43**(Suppl), 115–128 (1960).
100. C. Galletti, P. Fattori, and P. Battaglini *et al.*, ''Functional demarcation of a border between areas V6 and V6A in the superior parietal gyrus of the macaque monkey,'' *Eur. J. Neurosci.* **8**(1), 30–52 (1996).
101. A. McAllister, D. Lo, and L. Katz, ''Neurotrophins regulate dendritic growth in developing visual cortex,'' *Neuron* **15**, 791–803 (1995).
102. D. Nelson and L. Katz, ''Emergence of functional circuits in ferret visual cortex visualized by optical imaging,'' *Neuron* **15**, 23–34 (1995).
103. R. Rogers, B. Sahakian, and J. Hodges, *et al.*, ''Dissociating executive mechanisms of task control following frontal lobe damage and Parkinson's disease,'' *Brain* **121**(5), 815–842 (1998).
104. A. Sillito, T. Salt, and J. Kemp, ''Modulatory and inhibitory processes in the visual cortex,'' *Vision Res.* **25**(3), 375–381 (1985).
105. A. Sillito and K. Grieve, ''A re-appraisal of the role of layer VI of the visual cortex in the generation of cortical end inhibition,'' *Exp. Brain Res.* **87**, 521–529 (1991).
106. A. Sillito and H. Jones, ''Context-dependent interactions and visual processing in V1,'' *J. Physiol. (Paris)* **90**(3–4), 205–209 (1996).
107. R. Klatzky, ''Allocentric and egocentric spatial representations: definitions, distinctions, and interconnections,'' in *Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, C. Freksa, C. Habel, and K. F. Wender, Eds., pp. 1–17, Springer-Verlag, Berlin (1998).
108. R. Klatzky, J. Loomis, and R. Golledge, *et al.*, ''Acquisition of route and survey knowledge in the absence of vision,'' *J. Motor Behavior* **22**(1), 19–43 (1990).
109. C. Freksa, ''Using orientation information for qualitative spatial reasoning,'' in *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space (International Conference GIS-From Space to Territory, Pisa, Italy)*, A. Frank, I. Campari, and U. Formentini, Eds., pp. 162–178, Springer-Verlag, Berlin (1992).
110. C. Freksa, ''Foundations of computer science: Potential theory-cognition,'' in *Spatial and Temporal Structures in Cognitive Processes: Foundations of Computer Science*, C. Freksa, M. Jantzen, and R. Valk, Eds., pp. 379–387, Springer-Verlag, Berlin (1997).
111. K. Schill, E. Umkehrer, S. Beinlich, *et al.*, ''Knowledge-based scene analysis with saccadic eye movements,'' *SPIE: Electronic Imaging* **3644**, 520–531 (1999).
112. L. W. Stark, ''Neural nets, random design and reverse engineering,'' *Proceedings of the IEEE International Conference on Neural Networks*, Vol. 3, pp. 1313–1320, San Francisco, CA (1993).
113. L. W. Stark, ''ANNs and MAMFs: transparency or opacity?,'' *Proceedings of the European Neural Network Society*, pp. 123–129, Sorrento, Italy (1994).
114. W. H. Zangemeister, H. Stiehl, and C. Freksa, *Visual Attention and Cognition* (Elsevier, Amsterdam, 1996).
115. I. Rybak, V. Gusakova, and A. Golovan, *et al.*, ''A model of attention-guided visual perception and recognition,'' *Vision Res.* **38**, 2387–2400 (1998).
116. C. M. Privitera and L. W. Stark, ''Algorithms for defining visual regions-of-interest: comparison with eye fixations,'' *IEEE PAMI* **22**(9), 970–982 (2000).
117. Y. Aloimonos and J. Y. Herve, ''Exploratory active vision: theory,'' *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 10–15, Los Alamitos, CA (1992).
118. R. Bolle, Y. Aloimonos, and C. Fermuller, ''Toward motion picture grammars,'' in *Proceedings of the Third Asian Conference on Computer Vision, Hong Kong*, T.-C. Pong Chin, Ed., Vol. 2, pp. 283–290, Springer-Verlag, Berlin (1998).
119. G. A. Carpenter, S. Grossberg, and G. W. Lesher, ''The what-and-where filter. A spatial mapping neural network for object recognition and image understanding,'' *Comput. Vis. Image Underst.* **69**(1), 1–22 (1998).
120. C. Carson, S. Belongie, H. Greenspan *et al.*, ''Region-based image querying,'' *Proceedings, IEEE Workshop on Content-Based Access of Image and Video Libraries, San Juan, Puerto Rico*, pp. 42–49 (1997).
121. D. Crevier and R. Lepage, ''Knowledge-based image understanding systems: a survey,'' *Comput. Vis. Image Underst.* **67**(2), 161–185 (1997).
122. G. Foresti and G. Pieroni, ''Exploiting neural trees in range image understanding.,'' *Pattern Recogn. Lett.* **19**(9), 869–878 (1998).
123. T. T. Blackmon and L. W. Stark, ''Model-based supervisory control in telerobotics,'' *Presence* **5**, 205–223 (1996).
124. P. Buttolo, D. Kung, and B. Hannaford, ''Manipulation in real, virtual, and remote environments,'' *IEEE Trans. Syst. Man Cybern.* **SMC-5**, 4656–4661 (1995).
125. Y. F. Ho and L. W. Stark, ''Top-down image processing and supervisory control limitations in robotics: A simulation study,'' *8th International Conference on Advanced Robotics (ICAR '97)* pp. 933–938, Monterey, CA (1997).
126. Y. F. Ho and L. W. Stark, ''Visual tracking of tele-operated robots using model-based algorithms,'' *IS&T/SPIE's Symposium on Electronic Imaging, San Jose, CA* (1999).
127. Y. F. Ho and L. W. Stark, ''Model-based visual detection and verification system,'' *SPIE: Electronic Imaging, San Jose, CA* (1999).
128. Y. F. Ho and L. W. Stark, ''Scanpath-Based Model for Visual Tracking of Tele-Robots,'' *IS&T/SPIE's Symposium on Electronic Imaging, San Jose, CA* **3959**, 644–658 (2000).
129. W. S. Kim, M. Takeda, and L. W. Stark, ''On-the-screen visual enhancements for a telerobotics vision system,'' *Proc. IEEE Int. Conf. Syst. Man Cybern.*, pp. 126–130 (1988).
130. W. S. Kim, S. R. Ellis, and M. Tyler, *et al.*, ''Quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking tasks,'' *IEEE Trans. Syst. Man Cybern.* **16**, 61–72 (1987).
131. W. S. Kim, F. Tendick, and L. W. Stark, ''Visual enhancement in pick-and-place tasks: human operators controlling a simulated cylindrical manipulator,'' *IEEE J. Rob. Autom.* **3**, 418–425 (1987).
132. A. Liu, G. Tharp, and S. Lai *et al.*, ''Some of what one needs to know about using head-mounted displays to improve teleoperator performance,'' *IEEE Trans. Rob. Autom.* **9**(5), 638–48 (1993).
133. A. Nguyen and L. W. Stark, ''Model control of image processing: pupillometry,'' *Comput. Med. Imaging Graphics* **17**(1), 21–33 (1993).
134. L. W. Stark, B. Mills, A. Nguyen *et al.*, ''Instrumentation and robotic image processing using top-down model control,'' in *Robotics and Manufacturing*, Jamshidi *et al.*, Eds., pp. 657–682, ASME, New York (1988).
135. L. Sutro and J. Lerman, ''Robot vision,'' in *First National Conference on Remote Manned Systems*, E. Heer, Ed., pp. 251–282, Pasadena (1973).
136. Y. Yu and L. W. Stark, ''An active model-based algorithm for correspondence and estimation of pose parameters of objects,'' *IEEE Int. Conf. Syst. Man Cybern. Intelligent Systems for the 21st Century*, Vancouver, British Columbia **1**, 269–274 (1995).
137. Y. F. Ho, H. Masuda, and H. Oda *et al.*, ''Distributed control for tele-operations,'' *IEEE Trans. Mech.* **5**, 100–109 (2000).
138. Y. F. Ho and L. W. Stark, ''Design simulation of a web-based supervisory control system,'' in *International Conference on Web-Based Modeling and Simulation*, A. Uhrmacher, A. G. Bruzzone, and E. H. Page, Eds., Vol. 31, pp. 211–217, The Society for Computer Simulation (1999).
139. K. Miyata and L. W. Stark, ''Active camera control: seeing around obstacles,'' Power Electronics and Motion Control: *IEEE Proc. Int.*

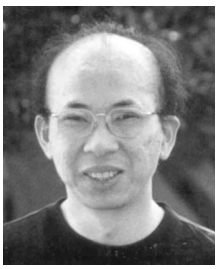
Conf. Industrial Electronics, Control, Instrumentation, Automation **2**, 752–756 (1992).
140. R. Bajcsy and E. Krotkov, ''Active vision for reliable ranging: cooperating focus, stereo, and vergence,'' *Int. J. Comput. Vis.* **11**, 187–203 (1993).
141. W. Uttal, T. Baruch, and L. Allen, ''The effect of combinations of image degradations in a discrimination task,'' *Percept. Psychophys.* **57**(5), 668–681 (1995).
142. D. H. Ballard, M. M. Hayhoe, and J. B. Pelz, ''Visual representations in natural tasks,'' Proceedings of the IEEE Computer Society Workshop on Visual Behaviors, pp. 1–9, Los Alamitos, CA (1994).
143. M. Driels and J. Acosta, ''The duality of haptic and visual search for object recognition,'' *Proc. IEEE Int. Symposium on Intelligent Control*, pp. 255–260, New York (1992).
144. J. Gould, ''Pattern recognition and eye-movement parameters,'' *Percept. Psychophys.* **2**, 399–407 (1967).
145. R. Groner, F. Walder, and M. Groen, ''Looking at faces: Local and global aspects of scanpaths,'' in *Theoretical and Applied Aspects of Eye Movement Research*, A. G. Gale and F. Johnson, Eds., pp. 523–533, North-Holland, Amsterdam (1984).
146. L. W. Stark, M. Okajima, and G. Whipple, ''Computer pattern recognition techniques: electrocardiographic diagnosis,'' *Commun. ACM* **5**, 527–532 (1962).
147. L. W. Stark, ''Top-down and bottom-up image processing,'' *Int. Conf. Neural Networks* **4**, 2294–2299 (1997).

Lawrence W. Stark has been a professor at the University of California at Berkeley since 1968, where he divides his teaching efforts among the EECS and ME departments in engineering and the Physiological Optics and Neurology units in biology and medicine. His research interests are centered in bioengineering, with an emphasis on human brain control of movement and vision, and symbiotic interactions of this knowledge with the rapidly developing fields of robotic vision and control. He pioneered the application of control and information theory to neurological systems, and the use of on-line digital computers for real-time acquisition of data, for self-organizing pattern recognition, and for physiological modeling. Stark has published several books and numerous research papers during his career that also included faculty positions at Yale (1954–1960), MIT (1960–1965), and Illinois (1965–1968). His educational degrees are AB (Columbia, 1945), MD (Albany, 1948), ScDhc (SUNY, 1988), and PhDhc (Tokushima, 1992). His many former graduate students and post-doctoral fellows form a world-wide school of bioengineering and biocybernetics.

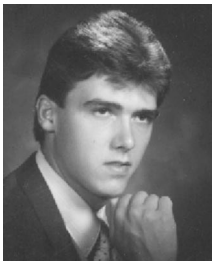
Claudio M. Privitera received the Laurea degree in Computer Science from the University of Pisa, Italy, in 1991. From 1992 to 1995 he held a doctoral fellowship within the National Research Program on Bioelectronic Engineering working at DIST, Department of Informatics, Systems, and Telecommunications of the University of Genoa, Italy. In 1994 he visited Laboratorie Scribens, Ecole Polytechnique, Universite de Montreal, Montreal, P.Q., Canada. From 1995 to 1996 he has been a postdoctoral research fellow at ICSI, the International Computer Science Institute of Berkeley, working in the artificial intelligence group. In 1996 he joined the Neurology and Telerobotics Units of the University of California at Berkeley and he has been teaching since then in the Mechanical Engineering Department. His research interests cover several aspects of biological and computational vision, image processing and pattern recognition. His research interests are also in the area of neural computation in motor control, and robotic and artificial intelligence.

Huiyang Yang: Biography and photograph not available.

Michela Azzariti received the Laurea degree in Statistics and Economics from the University of Rome, Italy, in 1993. From 1993 to 1997 she worked in a private institution as a data analyst and marketing manager collaborating with the Department of Statistics and Economics of the University of Rome (La Sapienza). In 1997 she joined, as a visiting scholar, the Neurology and Telerobotics Units of the University of California at Berkeley where she has been working on different psychophysical experiments on human vision and memory. She is also responsible for the statistical analysis of the Unit's experimental data.



Yeuk Fai Ho received his PhD in Spring 2000 after working in the Neurology and Telerobotics Unit of the Mechanical Engineering Department, University of California, Berkeley; previously he received a BS in Mechanical Engineering from Berkeley and an MS in Mechanical Engineering from UCLA. His research interests include the design of autonomous telerobotic control with visual feedback, and modeling, simulation and analysis of human visual processes.



Theodore T. Blackmon is the Chief Technology Officer and co-founder of Reality Capture Technologies, Inc. Prior to launching Reality Capture, Dr. Blackmon was a research scientist at NASA Ames Research Center, where he served as technology lead for the Intelligent Mechanisms Group within the Computational Sciences Division. During this period, he led development efforts on several significant and highly visible NASA robotic mission applications, including the 'Virtual Reality for Mars Pathfinder' project and the 'Chernobyl 3D Mapping System' project. Dr. Blackmon received his MS (1994) and PhD (1998) in Mechanical Engineering from U.C. Berkeley, with a major in control systems and minors in robotics and human neuromotor control. His thesis work at Berkeley involved the experimental investigation of human neurological control of hand, arm, and eye movements while interacting in virtual environments and telerobotic human-machine interfaces. Dr. Blackmon received his BS in Mechanical Engineering from Penn State University in 1992.

Dimitri Chernyak is a graduate student in the Vision Science Program at UC Berkeley. His research interests are centered on modeling of attention, eye movements and object recognition. Born in Moscow, Russia, he enrolled in the Biology Department of Moscow State University, and later received a BA from Brandeis University in Mathematics and Biology. Later he was awarded an MA in Cognitive and Neural Systems from Boston University. His research has been supported by grants from the NIH, NASA, and the American Optometry Foundation's Ezell Fellowship. He has given several talks at national and international conferences concerning his experimental and theoretical work.