THEORY AND DECISION LIBRARY

SERIES A: PHILOSOPHY AND METHODOLOGY
OF THE SOCIAL SCIENCES

SEEING, THINKING AND KNOWING

Meaning and Self-Organisation in Visual Cognition and Thought

EDITED BY ARTURO CARSETTI

KLUWER ACADEMIC PUBLISHERS

SEEING, THINKING AND KNOWING

THEORY AND DECISION LIBRARY

General Editors: W. Leinfellner (*Vienna*) and G. Eberlein (*Munich*)

Series A: Philosophy and Methodology of the Social Sciences

Series B: Mathematical and Statistical Methods

Series C: Game Theory, Mathematical Programming and Operations Research

SERIES A: PHILOSOPHY AND METHODOLOGY OF THE SOCIAL SCIENCES

VOLUME 38

Scope: This series deals with the foundations, the general methodology and the criteria, goals and purpose of the social sciences. The emphasis in the Series A will be on well-argued, thoroughly analytical rather than advanced mathematical treatments. In this context, particular attention will be paid to game and decision theory and general philosophical topics from mathematics, psychology and economics, such as game theory, voting and welfare theory, with applications to political science, sociology, law and ethics.

The titles published in this series are listed at the end of this volume.

SEEING, THINKING AND KNOWING

Meaning and Self-Organisation in Visual Cognition and Thought

Edited by

Arturo Carsetti

University of Rome "Tor Vergata",
Rome, Italy

KLUWER ACADEMIC PUBLISHERS

NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 1-4020-2081-3
Print ISBN: 1-4020-2080-5

Created in the United States of America

Visit Springer's eBookstore at: http://www.ebooks.kluweronline.com
and the Springer Global Website Online at: http://www.springeronline.com

TABLE OF CONTENTS

PART I

SEEING AND THINKING: A NEW APPROACH

PART II

FORMS AND SCHEMES OF PERCEPTUAL AND COGNITIVE SELF-ORGANISATION

ACKNOWLEDGEMENTS

This book would not have been possible without a lot of help. My immediate debt is to Werner Leinfellner. He encouraged me to edit this book and I have greatly benefited from discussions with him about the general orientation of the volume. I am also grateful to Elizabeth Leinfellner for her interest and her suggestions and to Jean Petitot for his invaluable help in preparing the Colloquium "Seeing and Thinking. Reflections on Kanizsa's Studies in Visual Cognition". The book owes its existence first of all to the successful realisation of this Colloquium.

I have been helped enormously by conversations with Stephen Grossberg, Gail Carpenter, J. Goetschl and F.M. Wuketits.

I am once again indebted to my collaborators Andrea Cataldi and Enrica Vizzinisi for their help at the editorial level. I would like also to thank Jolanda Voogd, Charles Erkelens and Ingrid Krabbenbos of Kluwer for their editorial comments and suggestions which contributed to the quality of the presentation of the book.

My deep thanks to the authors for their invaluable co-operation and to my students at the University of Rome "Tor Vergata" for their stimulus and for their patience.

This book is dedicated to the late Gaetano Kanizsa, a good friend and a great scholar. I learned a lot from him while I was teaching in Trieste. His deep understanding of the self-organisation processes underlying visual cognition represents, for many aspects, the Ariadne's thread of the entire volume.

This page intentionally left blank

ARTURO CARSETTI

INTRODUCTION

According to Putnam to talk of "facts" without specifying the language to be used is to talk of nothing; "object" itself has many uses and as we creatively invent new uses of words "we find that we can speak of 'objects' that were not 'values of any variable' in any language we previously spoke"[1]. The notion of object becomes, then, like the notion of reference, a sort of open land, an unknown territory. The exploration of this land appears to be constrained by use and invention. But, we may wonder, is it possible to guide invention and control use? In what way, in particular, is it possible, at the level of natural language, to link together program expressions and natural evolution?

To give an answer to these onerous questions we should immediately point out that cognition (as well as natural language) has to be considered first of all as a peculiar function of active biosystems and that it results from complex interactions between the organism and its surroundings. "In the moment an organism perceives an object of whatever kind, it immediately begins to 'interpret' this object in order to react properly to it ... It is not necessary for the monkey to perceive the tree in itself... What counts is survival"[2].

In this sense, if it is clearly true that we cannot talk of facts without specifying the language to be used, it is also true that we cannot perceive objects (and their relations) if we are unable continuously to "reconstruct" and invent new uses of the cognitive tools at work at the level of visual cognition. As Ruse remarks, the real world certainly exists, but it is the world as we interpret it. In what way, however, can we interpret it adequately? How can we manage to "feel" that we are interpreting it according to the truth (albeit partially)? In other words, if perceiving and, from a more general point of view, knowing is interpreting, and if the interplay existing between a complex organism and its environment is determined by the *compositio* of interpretative functions, actual emergence of meaning, and evolution, in what way can humans describe the inner articulation of this mysterious interplay, mirroring themselves in this *compositio*? Does this very *compositio* possess a "transcendental" character? How, generally speaking, can our brains give rise to our minds?[3] What types of functions and rules should we identify (and invent) in order to describe and contemporarily construct those evolutionary paths of cognitive processes (and in particular of visual cognition) that progressively constitute the very fibres of our being ? How can we model that peculiar texture of life and knowledge in flux that characterises our mental activity?

In order to do this we need, as is evident, ever-new simulation models considered in turn, from an abstract point of view, as specific mathematical objects. These models should adequately represent specific systems capable of autonomously self-organising in response to a rapidly changing and unpredictable world. In other words, we are really faced with the necessary outlining of models able to explain (and unfold) behavioural and brain data and

A. Carsetti (ed.), Seeing, Thinking and Knowing, 7-26.

contemporarily to interact with the dynamics that they try to describe in order to "prime" new patterns of possible integration. These models, once established, may act as operative channels and may be exactly studied as mathematical objects, i.e., they articulate as self-organising models. When we perceive and "interpret", our mind actually constructs models so that the brain can be forged and guided in its activity of reading the world.

In this sense, informational data must be considered as emergent properties of a dynamical process taking place at the level of the mind. From a general point of view, behaviour must be understood as an ensemble of emergent properties of networks of neurones. At the same time, the neurones and their interactions are necessary in order to correctly define the laws proper to the networks whose emergent properties map onto behaviour. Thus, on the one hand, we need a powerful theoretical language: the language, in particular, of dynamical systems and, on the other, we contemporarily need self-organising models able to draw, fix and unfold the link between real emergence and mental construction as well as the link between the holistic fabric of perception and the step by step developmental processes in action.

The chapter by S. Grossberg "Neural Models of Seeing and Thinking" aims at a very clear exploration of the role played by the neural models at the level of Cognitive Science and in particular at the level of visual cognition. According to Grossberg the brain is organised in parallel processing streams. These streams are not independent modules however; as a matter of fact strong interactions occur between perceptual qualities. "A great deal of theoretical and experimental evidence suggests that the brain's processing streams compute *complementary* properties. Each stream's properties are related to those of a complementary stream, much as a lock fits its key or two pieces of a puzzle fit together"[4]. "How, then, do these complementary properties get synthesised into a consistent behavioural experience? It is proposed that *interactions* between these processing streams overcome their complementary deficiencies and generate behavioural properties that realize the unity of conscious experiences. In this sense, pairs of complementary streams are the functional units, because only through their interactions can key behavioural properties be competently computed. These interactions may be used to explain many of the ways in which perceptual qualities are known to influence each other"[5].

Each stream can possess multiple processing stages, a fact which, according to Grossberg, suggests that these stages realize a process of hierarchical resolution of uncertainty. The computational unity is thus not a single processing stage but a minimal set of processing stages that interact within and between complementary processing streams. "The brain thus appears as a self-organising measuring device in the world and of the world"[6].

The neural theory FACADE illustrated by Grossberg in his article suggests how and why perceptual boundaries and perceptual surfaces compute complementary properties. In particular, Grossberg, using the famous Kanizsa square, shows that a percept is due to an interaction between the processing streams that form perceptual boundaries and surfaces. In this sense: "a boundary formation process in the brain is indeed the mechanism whereby we perceive geometrical objects such as lines, curves, and textured objects. Rather than being defined in terms of such classical units as points and lines, these boundaries arise as a coherent pattern of excitatory and inhibitory signals across a mixed co-operative-competitive feedback network that is defined by a non-linear dynamical system describing the cellular interactions from the retina through LGN and the VI Interblob and V2 Interstripe areas"[7].

These interactions select the best boundary grouping from among many possible interpretations of a scene. The winning grouping is represented either by an equilibrium point or a synchronous oscillation of the system, depending on how system parameters are chosen. FACADE theory suggests how the brain may actually represent these properties using non-linear neural networks that do a type of online statistical inference to select and complete the statistically most-favoured boundary groupings of a scene while suppressing noise and inconsistent groupings.

The boundary completion and the surface filling-in as suggested by Grossberg in his article thus represent a very different and innovative approach with respect to the classical geometrical view as established in terms of surface differential forms. Let us just remark that according to a conservative extension of this perspective, from an epistemological point of view, simulation models no longer appear as "neutral" or purely speculative. On the contrary, true cognition appears to be necessarily connected with successful forms of reading, those forms, in particular, that permit a specific coherent unfolding of the deep information content of the Source. Therefore the simulation models, if valid, materialise as "creative" channels, i.e., as autonomous functional systems (or self-organising models), as the same roots of a new possible development of the entire co-evolutive system represented by mind and its Reality.

The following two chapters are equally centred on an in-depth analysis of Kanizsa's experiments, although according to different theoretical and modelistic perspectives.

The aim of Petitot's chapter "Functional Architecture of the Visual Cortex and Variational Models for Kanizsa'a Modal Subjective Contours" is to present a neuro-geometrical model for generating the shape of Kanizsa's modal subjective contours. This model is based on the functional architecture of the primary areas of the visual cortex. The key instrument utilised by Petitot is the idea of variational model as introduced by S. Ullman in 1976. This idea was improved and enlarged in 1992 by D. Mumford by the outlining of a fundamental model based on the physical concept of *elastica*. Mumford essentially aimed to define curves simultaneously minimising the length and the integral of the square of the curvature κ, i.e. the energy $E = \int (\alpha \kappa + \beta)^2 ds$ where ds is the element of arc length along the curve.

Petitot presents a slightly different variational model based on the concept of "geodesic curves" in V1 that results more realistic at the neural level. As is well known, at the level of visual cortex "due to their structure, the receptive fields of simple cells detect a *preferential orientation*. Simplifying the situation, we can say they detect pairs (a, p) of a spatial (retinal) position a and a local orientation p at a. They are organised in small modules called *hypercolumns* (Hubel and Wiesel) associating retinotopically to each position a of the retina R a full exemplar P_a of the space of orientations p at a"[8].

A simplified schema of this structure (with a 1-dimensional base R) is represented by a *fibration* of base R and fiber P. In geometry pairs (a,p) are called contact elements. Their set V = $\{(a,p)\}$ need to be strongly structured to allow the visual cortex to compute contour integration.

Petitot underlines the importance of the discovery, at the experimental level, of the cortico-cortical horizontal connections. These connections exactly allow the system to compare orientations in two different hypercolumns corresponding to two different retinal positions *a* ad *b*. Besides the horizontal connections we may also individuate vertical connections. According to William Bosking (1997) the system of long-range horizontal connections can be summarised as preferentially linking neurones with co-oriented, coaxially aligned receptive fields. Starting from these experimental findings Petitot finally can show that: "what geometers call the *contact structure* of the fibration π: $R \times P \rightarrow R$ is neurologically implemented"[9].

Thus, he can directly affirm that the integrability condition is a particular version of the Gestalt principle of "good continuation". As emphasised by Field, Hayes, and Hess (1993) ""Elements are associated according to joint constraints of position and orientation"…. "The orientation of the elements is locked to the orientation of the path; a smooth curve passing through the long axis can be drawn between any two successive elements" "[10]. Hence the possibility of the individuation of a discrete version of the integrability condition.

According to these results, Petitot can conclude his analysis saying that "due to the very strong geometrical structure of the functional architecture (hypercolumns, pinwheels, horizontal connections), the neural implementation of Kanizsa's contours is deeply linked with sophisticated structures belonging to what is called contact geometry and with variational models analogue to models already well known in physics"[11].

In other words, a neurally plausible model of Kanizsa-curves at the V1 level reveals itself as linked first of all to the articulation of specific geodesic curves. In this way it is possible to progressively identify some of the principal factors of that "perceptual geometry" that, at the moment, presents itself as the real basis, from a genetic (and genealogical) point of view, of classical Geometry.

In their chapter "Gestalt Theory and Computer Vision" A. Desolneux, L.Moisan and J.M. Morel also start from an in-depth analysis of Kanizsa's contribute to Gestalt theory. Their approach, however, essentially concerns a mathematical model characterised in discrete and not in continuous terms. In their opinion both Gestalt Theory and classical Information Theory have attempted to answer the following question: how is it possible to individuate global percepts starting from the local, atomic information contained in an image?

The authors distinguish two kinds of laws at the gestaltic level: 1) the grouping laws (like vicinity and similarity) whose aim is to build up partial gestalt; 2) the gestalt principles whose aim it is to operate a synthesis between the partial groups obtained by the elementary grouping laws.

The obtained results show that " …-there is a simple computational principle (the so-called Helmholtz principle), inspired from Kanizsa's masking by texture, which allows one to compute any partial gestalt obtainable by a grouping law …- this computational principle can be applied to a fairly wide series of examples of partial gestalts, namely alignments, clusters, boundaries, grouping by orientation, size or grey level"[12].

The authors also show that "…the partial gestalt recursive building up can be led up to the third level (gestalts built by three successive grouping principles)"[13]. In particular, they show that all partial gestalts are likely to lead to wrong scene interpretations. In this way it is possible to affirm that wrong detections are explainable by a conflict between gestalts. Hence the research for principles capable of resolving some of these conflicts in an adequate way such as, for instance, the Minimal Description Length principles.

The central core of the analysis is represented by the definition of several quantitative aspects, implicit in Kanizsa's definition of masking and by the attempt to show that one particular kind of masking, Kanizsa's masking by texture, suggests precise computational procedures. Actually, "The pixels are the computational atoms from which gestalt grouping procedures can start. Now, if the image is finite, and therefore blurry, how can we infer sure events as lines, circles, squares and whatsoever gestalts from discrete data? If the image is blurry all of these structures cannot be inferred as completely sure; their exact location must remain uncertain. This is crucial: all basic geometric information in the image has a precision"[14].

Moreover, the number N_{conf} of possible configurations for partial gestalts is finite because the image resolution is bounded. Starting from these simple considerations the authors apply a general perception principle called Helmholtz principle: "This principle yields computational grouping thresholds associated with each gestalt quality. It can be stated in the following generic way. Assume that atomic objects. $O_1, O_2, ..., O_n$ are present in an image. Assume that k of them, say $O_1,..., O_k$, have a common feature, say, same colour, same orientation, position etc. We are then facing the dilemma: is this common feature happening by chance or is it significant and enough to group $O_1,..., O_k$? In order to answer this question, we make the following mental experiment: we assume *a priori* that the considered quality has been randomly and uniformly distributed on all objects $O_1,.., O_n$. Then we (mentally) assume that the observed position of objects in the image is a random realisation of this uniform process. We finally ask the question: is the observed repartition probable or not? If not, this proves *a contrario* that a grouping process (a gestalt) is at stake. Helmholtz principle states roughly that in such mental experiments, the numerical qualities of the objects are assumed to be equally distributed and independent"[15]. The number of "false alarms" (NFA) of an event measures the "meaningfulness" of this event: the smaller it is, the more meaningful the event is.

This kind of measure is perfectly coherent with an ancient measure of semantic information content as introduced by R. Carnap e Y. Bar Hillel in 1952. Actually, in order to model holistic perception and the *Gestalten* in action we have to take into account not only the syntactic measures of information but also the semantic ones; only in this way shall we be able to give an explanation in computational terms of that peculiar (intensional) meaningfulness that characterises real perception.

If we aim to explain in modelistic terms the reality of visual perception we have, however, not only to take into account the different intensional aspects of meaningfulness (as well as the intentional ones) but also the phenomenal consciousness (actually, at the neural level, visual consciousness appears as distributed in space and time, as S. Zeki has recently pointed out).

It is precisely to the problem of a possible simulation of phenomenal consciousness that the chapters by J. K. O'Regan, E. Myin and A. Noë and by A. Di Ferdinando and D. Parisi are devoted.

In what way can we explain how physical processes (neural and computational) can produce experience: i.e. phenomenal consciousness? Gestalt is not only a mathematical or computational construction: it is something that *lives*, of which we have direct and holistic experience.

As is well known, several scholars have argued that phenomenal consciousness cannot be explained in functional or neural terms. According to O'Regan's, Myin's and Noë's opinion as expressed in the article "Towards an Analytic Phenomenology: the Concepts of Bodiliness and Grabbiness" the problem is misplaced: feel is not generated by a neural mechanism at all, rather it is exercising what the neural mechanism allows the organism to do. "An analogy can be made with "life": life is not something which is generated by some special organ in biological systems. Life is a *capacity* that living systems possess. An organism is alive when it *has the potential* to do certain things, like replicate, move, metabolise, etc. But it need not be doing any of them right now, and still it is alive... When we look out upon the world, we have the impression of seeing a rich, continuously present visual panorama spread out before us. Under the idea that seeing involves exercising a skill however, the richness and continuity of this sensation are not due to the activation in our brains of a neural representation of the outside world. On the contrary, the ongoingness and richness of the sensation derive from the knowledge we have of the many different things we can do (but need not do) with our eyes, and the sensory effects that result from doing them (O'Regan 1992). Having the impression of a whole scene before us comes, not from every bit of the scene being present in our minds, but from every bit of the scene being immediately available for "handling" by the slightest flick of the eye"[16].

According to this point of view, we no longer need to postulate a neural process that generates phenomenal consciousness, this kind of consciousness must, on the contrary, be considered as a skill people exercise.

Thinking, however, is different from seeing: thinking has no perceptual quality. The fundamental difference between mental phenomena that have no feel (like for instance knowledge) and mental phenomena that have feel (like sensations) can be explained through the introduction of the concepts of bodiliness and grabbiness.

"Bodiliness is the fact that when you move your body, incoming sensory information immediately changes. The slightest twitch of an eye muscle displaces the retinal image and produces a large change in the signal coming along the optic nerve. Blinking, moving your head or body will also immediately affect the incoming signal"[17].

On the other hand, grabbiness is the fact that sensory stimulation can grab our attention away from what we were previously doing. Bodiliness and grabbiness are objectively measurable quantities that determine the extent to which there is something it is like to have a sensation.

It is the order of what I can do potentially to organise which determines the horizon of my seeing, and the horizon inserts itself within this type of self-organisation. When we look at a scene, we see, we forecast and we self-organise at the same time. If we want to build a robot that feels we have to provide the robot with mastery of the laws that govern the way its actions affect its sensory input. We have to wire up its sensory receptors and we have to give it "access to the fact that it has mastery of the skills associated with its sensory exploration"[18].

The point of view according to which internal representations are action-based is also assumed by A. Di Ferdinando and D. Parisi, in their chapter "Internal Representations of Sensory Input Reflect the Motor Output with which Organisms Respond to the Input".

"What determines how sensory input is internally represented? The traditional answer is that internal representations of sensory input reflect the properties of the input. This answer is based on a passive or contemplative view of our knowledge of the world which is rooted in the philosophical tradition and, in psychology, appears to be almost mandatory given the fact that, in laboratory experiments, it is much easier for the researcher to control and manipulate the sensory input which is presented to the experimental subjects than the motor output with which the subjects respond to the input. However, a minority view which is gaining increasing support (Gibson, 1986; O'Regan and Noë, in press) is that internal representations are instead action-based, that is, that the manner in which organisms internally represent the sensory input reflects the properties of the actions with which the organisms respond to the sensory input rather than the properties of the sensory input"[19].

The authors present, in particular, a series of computer simulations using neural networks that tend to support the action-based view of internal representations. In their opinion, internal representations are not symbolic or semantic entities, they are patterns of activation states in the network's internal units which are caused by input activation patterns and which in turn cause activation pattern in the network's output units. The authors distinguish between micro-actions and macro-actions, the latter are sequences of microactions that allow the organism to reach some goal. Internal representations exactly encode the properties of macro-actions. The properties of the visual input are retained on the internal representations only insofar as they are relevant for the action to be executed in response of the visual input. Hence the necessity to resort to concepts like adaptation and assimilation. Hence, on the other hand, the importance, with respect to this frame of reference, of the Adaptive Resonance Theory as outlined by Grossberg, a theory that, in particular, predicts that all conscious states are resonant states of the brain.

The theme concerning the interaction between man and machine and the possible construction of a robot able to observe human motion also constitutes the central core of the chapter by L. Goncalves, E. Di Bernardo and P. Perona "Movemes for Modeling Biological Motion Perception". As the authors write "Perceiving human motion, actions and activities is as important to machines as it is to humans. People are the most important component of a machine's environment. Endowing machines with biologically-inspired senses, such as vision, audition, touch and olfaction appears to be the best way to build user-friendly and effective interfaces. Vision systems which can observe human motion and, more importantly, understand human actions and activities, with minimal user cooperation are an area of particular importance"[20].

However: "While it is easy to agree that machines should "look" at people in order to better interact with them, it is not immediately obvious which measurements should a machine perform on a given image sequence, and what information should be extracted from the human body. There are two classes of applications: "metric" applications where the position of the body has to be reconstructed in detail in space-time (e.g. used as input for positioning an object in a virtual space), and "semantic" applications where the meaning of an action (e.g. "she is slashing through Rembrandt's painting") is required. The task of the vision scientist/engineer is to define and measure "visual primitives" that are potentially useful for a large number of applications. These primitives would be the basis for the design of perceptual user interfaces …substituting mouse motions and clicks, keystrokes etc. in existing applications, and perhaps enabling entirely new applications. Which measurements should we take?[21]

"In looking for a model of human motion one must understand the constraints to such motion. First of all: our motions are constrained both by the kinematics and by the dynamics of our body. Our elbows are revolute joints with one degree of freedom (DOF), our shoulders are ball joints with three DOF etc. Moreover, our muscles have limited force, and our limbs have limited acceleration. Knowledge of the mechanical properties of our bodies is helpful in constraining the space of solutions of biological motion perception. However, we postulate that there is a much more important constraint: *the motion of our body is governed by our brain*. Apart from rare moments, when we are either competing in sport or escaping an impending danger, our movements are determined by the stereotypical trajectories generated by our brain… the dynamics of our body at most acts as a low-pass filter"[22].

However, generating trajectories is a complex computational task. Neurophysiological evidence suggest that our nervous system encodes complex motions and discrete sequences of elementary trajectories. "This suggests a new computational approach to biological motion perception and to animation. One could define a set of elementary motions or *movemes* which would roughly correspond to the 'elementary units of motion' used by the brain. One could represent complex motions by concatenating and combining appropriate movemes. These movemes would be parameterized by 'goal' parameters in Cartesian space. This finds analogies in other human behaviours: the "phonemes" are the elementary units both in speech perception and production; in handwriting one thinks of 'strokes' as the elementary units"[23].

From a general point of view, in order to encode microactions we need a language, we need tools to compress information. Movemes greatly compress motion information. They appear to be a natural and rich representation which the brain might employ in perceiving biological motion. Many questions, however, arise. In what way can we semantically model and handle the processes of information compression as they articulate at the cognitive level? How many movemes might there be? How is it possible to build a complete catalogue? What types of syntactical laws govern their generation? What about the link between this generation and the unfolding of form constraints?

Lorenceau's chapter "Form Constraints in Motion Integration, Segmentation and Selection" also concerns the realm of form and motion selection. In this article, however, particular attention is devoted to the problems concerning the action expressed by the constraints at the level of form construction. According to Lorenceau, "Perception is a process by which living organisms extract regularities from the physical fluxes of varying physical characteristics in the external world in order to construct the stable representations that are needed for recognition, memory formation and the organisation of action. The exact nature of the process is still not well understood as the type of regularities that are indeed used by sensory systems can be manifold. However, perception is not a process by which living organisms would reproduce the physical fluxes such as to build an internal representation identical to its physical counterpart. One issue then is to understand the kind of physical regularities that are relevant for perceiving and recognising events in the outside world"[24].

We may, following the gestaltist approach, develop experimental paradigms to define and isolate the general rules underlying perceptual organisation.

"In vision, figure/ground segregation and perceptual grouping of individual tokens into a "whole" appeared to strongly rely on several rules such as good continuity, proximity, closure, symmetry, common fate, synchronicity etc. Most importantly, these principle define spatial and temporal relationships between "tokens", whatever the exact nature of these tokens: dots, segments, colour, contours, motion, etc. Implicitly, the general model underlying the Gestaltist approach is a geometrical one, stressing the spatial relations between parts rather than concerned with the intrinsic processing of the parts themselves. However, the attempt of the gestalt school to offer a plausible neuronal perspective that could explain their observations on perceptual organisation failed, as the Gestaltists were thinking in term of an isomorphism between external and internal geometrical rules whereby spatial relationships between neurones would mimic the geometry of the stimulus. Electrophysiological and anatomical studies did not revealed such an isomorphism"[25]. In his paper Lorenceau takes into account recent neurophysiological findings that suggest how geometrical principles may be implemented in the brain, also discussing hypotheses about the physiological mechanisms that may underlie perceptual grouping. Lorenceau, in particular, points out that "In support of a functional link between neurones through horizontal connections in primary visual cortex, a number of recent psychophysical studies uncovered strong contrast dependent centre-surround interactions, either facilitatory or suppressive, that occur when one or several oriented test stimuli are analysed in the presence of surrounding oriented stimuli. For instance, contrast sensitivity is improved by similar flankers, collinear and aligned with the test stimulus. Changing the relative distance, orientation, spatial frequency or contrast of the flankers modulates the change in sensitivity, allowing the analysis of the architecture of these spatial interactions"[26].

From a general point of view, he underlines the fact that feedback and long range connections within a single area provide a possible physiological substratum to compute some of the geometrical properties of the incoming image. With respect to the interplay existing between form and motion the results presented by the author demonstrate the critical role played by geometrical information in global motion computation.

"Local singularities such as vertices, junctions or line-ends appears to exert a strong control on the balance between integration and segmentation as salient contour terminators appear to be used to parse the image into parts. Global geometrical image properties also appear to provide strong constraints on the integration process, as integrating moving contours into a global motion is a simple task for some configurations (diamonds) while it is difficult for others (crosses and chevrons)"[27].

Actually: "Closure of the diamond by amodal completion (Kanisza, 1979), together with the filling-in of is its interior this may engenders, would serve effectively the segregation of the diamond from its background. Consequently, judging the diamond's direction of rotation would be much easier than for open shapes which generate poorer responses at the level of object representation…. The effects of boundary completion, filling-in and figure/ground segregation, can all be considered broadly under the rubric of form processing. Our data suggest that the role of form information is to regulate whether motion integration should go ahead or not"[28].

Lastly, Lorenceau points out that numerous studies now support the idea that geometrical relationships between visual elements or "tokens" play a fundamental role in the perceptual organisation of form and motion. The available anatomical and physiological evidence suggests that the neural circuitry described in the primary visual cortex possesses some of the properties needed to process the geometric features of the retinal input. With respect to this framework it is in any case necessary to underline that the interactions between form and motion are bi-directional. Neural circuitry appears able to individuate invariants and to connect these invariants according to different stages, phases and rules.

When we speak in terms of tokens, syntactic and informational rules, invariants (and attractors) and so on, we actually try to describe (and creatively unfold) an informational code utilised by our mind in order to realize an adequate recovery-reading of depth information. Actually, we continuously invent and construct new models and visual languages in order to perceive, i.e. to interpret according to the truth in a co-evolutive context.

J. Ninio concludes his intriguing and "passionate" chapter "Scintillations, Extinctions and Other New Visual Effects" by means of these simple words: "I find more and more satisfaction, as Kanizsa did, in elaborating striking images. Whereas, in his case, the images must have been the outcome of a completely rational line of thinking, in my case they came by surprise. They were – at least for Fig. 7 and 8, the unexpected reward of a very systematic work of variations in the geometry of the stimuli"[29].

In these words lies the soul of the chapter. Its "intelligence" is first of all in the images presented by the author: new effects, new possible "openings" of our mind, new ways of seeing. Like Kanizsa, Ninio too possesses the art of "constructing with maestria" and in the article this is predicated on the account of a lifetime. By means of his images-effects Ninio offers new "cues" for the "invention" of new languages, of new models of simulation and, at the same time, for the individuation of new "ways to be given", of new intensional settings.

M. Olivetti, R. Di Matteo, C. del Gratta, A. de Nicola, A. Ferretti and G.L. Romani in their chapter "Commonalities Between Visual Imagery and Imagery in Other Modalities: an Investigation by Means of fMRI "remark, first of all, that: "The attempt to shadow the differences between seeing and thinking by stressing their similarities is not an epistemologically correct operation, because by using principles related to another domain (like thinking) to explain vision may induce a pre-packed explication. Instead, stressing differences may lead to the discovery of new rules governing only one of the two processes under investigation (Kanizsa, 1991). We report this provoking statement of Kanizsa, while approaching our research on mental imagery for two main reasons: 1) the main part of the psychological research on imagery is devoted to visual imagery, implicitly assuming that imagery derived from other sensory modalities will present characteristics that are similar to those of visual imagery; 2) a lot of studies on visual imagery are devoted to assess whether primary perceptual circuits are implied also in imagery and, therefore, to assess how much seeing is similar to imaging. In this study we accepted Kanizsa's suggestion by trying to assess differences between visual and other-senses imagery in order to detect their peculiarities and the grade of their overlap"[30].

As the authors underline, the studies examining the relationship between imagery and processes related to modalities other than vision are very rare. The chapter is devoted to study how much imagery according to various sensory modalities is tied to the processing of visual features. It tries to identify first of all the common substrate of visual images and images generated according to other sensory modalities. "It consists of a fMRI block design while participants were requested to generate mental images cued by short sentences describing different perceptual object (shapes, sounds, odours, flavours, self-perceived movements and internal sensations). Imagery cues were presented visually and were contrasted with sentences describing abstract concepts, since differences in activation during visual imagery and abstract thoughts were often assessed in literature"[31].

From this study, it is possible to derive three key findings: "First, common brain areas were found to be active in both visual imagery and imagery based on other sensory modalities. These common areas are supposed to reflect either the verbal retrieval of long-term representations or the segregation of long-term representations into highly interactive modality specific regions. Second, each imagery modality activates also distinct brain areas, suggesting that high-level cognitive processes imply modality-specific operations. This result is coherent with the domain-specific hypothesis proposed for the functioning of the fronto-parietal associative stream (Rushworth & Owen, 1998; Miller, 2000). Third, primary areas were never found to be active, suggesting that different, though interactive, neural circuits underlie low-level and high-level processes. Although this claim is only indicative, as in this study no direct comparisons were made between imagery and perceptual/motor processes, it outlines the lack of primary cortex activation for imagery in those modalities that were not accompanied by any corresponding sensory stimulation due to either the visual presentation of the stimuli or to the noisy apparatus"[32].

The second part of the volume is devoted to a thorough analysis of a number of conceptual tools that revealed themselves particularly useful in interpreting cognitive and mental phenomena. Microgenesis, Synergetics, Self-Organisation Theory, Semantics, Evolutionary Theory etc., are extensively utilised in the different chapters in order to clarify the mysterious relationships existing between emergence of meaning, self-organisation processes, emotion, differentiation and unfolding processes, symbolic dynamics etc. Actually, in order to outline more sophisticated models of cognitive activities (and in particular of that inextricable plot constituted by "seeing *and* thinking") we have to examine and individuate specific theoretical methods capable, for instance, of taking into account also the intentional and semantic aspects of that specific, mental and biological process linking together growth with symbolic differentiation which characterises human cognition.

In his chapter "Microgenesis, Immediate Experience and Visual Processes in Reading", V. Rosenthal clearly illustrates and discusses the concept of microgenesis as correlated to specific processes of unfolding and differentiation.

"Microgenetic development concerns the psychogenetic dynamics of a process that can take from a few seconds (as in the case of perception and speech) up to several hours or even weeks (as in the case of reading, problem solving or skill acquisition). It is a living process that dynamically creates a structured coupling between a living being and its environment and sustains a knowledge relationship between that being and its world of life *(Umwelt)*. This knowledge relationship is protensively embodied in a readiness for further action, and thereby has practical meaning and value. Microgenetic development is thus an essential form of cognitive process: it is a dynamic process that brings about readiness for action. Microgenesis takes place in relation to a thematic field which, however unstable and poorly differentiated it might be, is always given from the outset. To this field, it brings stabilised, differentiated structure and thematic focalization, thereby conferring value and meaning to it. Figure/ground organisations are an illustration of a typical microgenetic development. Yet, one should bear in mind that however irresistible an organisation might appear, it is never predetermined but admits of alternative solutions, that a 'figure' embodies a focal theme, and that a 'ground' is never phenomenologically or semantically empty"[33].

At the level of microgenesis form, meaning and value cannot be considered as separate entities, on the contrary, perception is directly meaning-and value-laden with actual meaning developing along the global-to-local dynamics of microgenesis. Meaning, in this sense, is not the end product of perception but rather part and parcel of the perceptual process. Actually: "…theories which separate sensory, semantic, motivational and emotional processes, and view perception as a construction of abstract forms out of meaningless features (only to discover later their identity and meaning), face in this respect insurmountable paradoxes. If semantics post-dates morphology, then it cannot affect form reconstruction, and if semantics is concomitant with form reconstruction, how can it influence morphological processing prior to 'knowing' what the latter is about? Finally, since morphological and semantic processes are viewed as incommensurable, how can they be brought to cooperate together without recourse to yet another, higher-order process? Invoking such a process would either amount to conjuring up a sentient device of the homunculus variety or would stand in contradiction to the very postulate of the distinctness and independence of meaning and form"[34].

Perception, according to Rosenthal, necessarily refers to the assumption of the consistency and meaningfulness of the world in which we live. It anticipates meaningful structures and categorises them on a global dynamic basis.

"The segmentation of the perceptual field into individual objects is thus the result of perceptual differentiation, and not the objective state of affairs that perception would merely seek to detect and acknowledge. In this sense, microgenesis is the process that breaks up the holistic fabric of reality into variably differentiated yet *meaningful* objects, beings and relations"[35].

Thus, the proposition that perception is based on reconstruction from elementary components raises more problems than it may be expected to solve. According to Rosenthal which quotes Husserl, the "now" has retentions and protentions i.e. there is a continuous and dynamic structure to experience.

Various phenomena of perceptual completion, whether figures, surfaces or regions, provide an interesting illustration of microgenetic dynamics at work in perception. "Consider the famous example of the Kanizsa square where a collinear arrangement of edges of four white 'pacmen' (inducers) on a black background gives rise to the perception of a black square whose area appears slightly darker than the background. In addition, the surface of the square appears to the observer to be in front of four disks that it partly occludes. Since the square is perceived in spite of the absence of corresponding luminance changes (i.e. forming complete boundaries), and thus does not reflect any real distal object, it can only be created by the visual system which purportedly completes, closes, and fills in the surfaces between 'fragments', so as to make the resulting 'subjective' region emerge as figura standing in the ground. Yet, as Kanizsa (1976; 1979) aptly showed, this and other examples of so-called subjective contours demonstrate the basic validity of Gestalt principles of field organisation, in particular of its figure/ground structure and of *Prägnanz*, whereby incomplete fragments are, upon completion, transformed into simpler, stable and regular figures. Although this phenomenon is often described in terms of contour completion, it clearly demonstrates a *figural* effect, whereby the visual system imposes a figural organisation of the field (and hence figure completion), and where *the contour results from perceiving a surface*, not the other way around, again as Kanizsa suggested. Moreover, these *subjective figures* illustrate the categorial and anticipatory character of microgenetic development, such that the perceptual system anticipates and actively seeks meaningful structures and immediately categorises them on a global dynamic basis"[36]. The crucial role of meaningfulness is demonstrated by the fact that no subjective figures arise in perception when the spatial arrangement of inducers fails to approximate a 'sensible form' or when the inducers are themselves meaningful forms. Actually, in Husserlian terms, meaning "shapes" the forms creatively. In order, however, to understand how this shaping takes place we need more information about the genealogical aspects of this mysterious process.

Lastly, Rosenthal presents a specific illustration of certain principles of microgenetic theory in the field of reading. A further step in the outlining of a genetic phenomenological science of embodied cognition.

According to Y.M. Visetti, ordinary perception does not constitute a foundation for linguistic but rather an essential correlate and a particular illustration of the construction of meaning. As he writes in his chapter "Language, Space and the Theory of Semantic Forms" perception "… has to be considered as instantiating a *general* structure of cognition, and not only as resorting to a purely sensorial and peripheral organisation. As a slogan, we could say that 'to perceive is from a single move to act and to express'. Perception already gives access to, and sketches, a meaning. It implies not only the presence of things, but a perspective of the subject, and a suggestion of acting. Perception in space is not grasping pure configurations or shapes, nor only a basis for other, subsequent 'associative' or 'metaphorical' interpretations: it is from the outset a dynamic encounter of 'figures' with no necessary dissociation between forms and values, apprehended in the course of actions, and deeply qualified by a specific mode of access or attitude. It is this notion of a *qualified relation* (which is a way of 'accessing', of 'giving' of 'apprehending'....) that we want to transpose into semantics, in order to view it as a kind of perception and/or construction of forms. At this level, any distinction between abstract or concrete, or between interior or exterior perception, is irrelevant. In the same way as there is more than topology or geometry in our multiple relations to ambiant space, we can say that 'figures' are objective counterparts, phenomenological manifestations of the *relations* we have with them"[37].

In such a framework "schemes" are not formal types, but "potentials" to be actualised. In the same way forms are to be considered as the result of dynamical stabilisation processes, i.e. as units in an ongoing continuous flow. As Visetti remarks recent advances in the theory of complex systems allow us to conceive a unified setting for language activity considered as a construction of forms in a semantic field. Hence the revisitation of an Humboltdian conception of language which considers it not as a finished product but as a self-organized activity.

One of the major aims of Wimmer's chapter "Emotion-Cognition Interaction and Language" is to show that language and the required underlying levels of emotion and cognition appears as an interacting phenomenon. None of these three functions can be isolated. There is neither a pure emotion nor a pure cognition nor any kind of ideal language without any relationship to the underlying levels of being. The roots of Wimmer's considerations may be found in the Evolutionary Epistemology and in the Genetic Epistemology. According to a well known K. Lorenz's statement, life itself can be considered as a "knowledge gaining process". In Piaget's Genetic Epistemology, on the other hand, we can find a similar type of naturalistic account: life itself is considered as a self-regulatory process.

In accordance with this theoretical setting, Wimmer, paraphrasing I. Kant, formulates a basic hypothesis: "Affects without cognitions are blind and cognitions without affects are empty" "What does this mean and what contributes this hypothesis to language related issues? The core of the argument is the assumption that from an evolutionary-phylogenetical viewpoint, the distinction between affect and cognition seems to be artificially drawn leading to wrong conclusions. The sharp distinction between affect and cognition has deep roots in our cultural heritage, leading back to ancient Greek philosophy. (comp. Averill 1996; Gardiner/Metcalf/Beebe-Center 1937) Beside these historical roots also recent neuroanatomical and neurophysiological research indicates a distinction between brain areas and mechanisms responsible for affective and cognitive processes. (Panksepp 1998; MacLean 1990) In contrast to these considerations an evolutionary approach leads to the assumption that there exists one common root of emotion and cognition. A root which in its early and very basic form is very close to homoeostatic – regulatory mechanisms. The root metaphor is very helpful in proposing a picture of one common root, which branches off in different branches always remaining closely related to the basic root. Even (in phylogenetical dimensions) the very young ability of language usage can be traced back to this basis root"[38].

If we consider cognition (and consequently perception) as the result of a coupled process, intrinsically related each time, from a phenomenological point of view, to the constitution of an inner horizon (or of a multiplicity of horizons) emotion clearly appears radically hinged on specific cognitive schemes.

Also for M. Stadler and P. Kruse in the chapter "Appearance of Structure and Emergence of Meaning in the Visual System" the brain is a self-organising system. Cognitive processes are actually based on the elementary neural dynamics of the brain. In this sense the synergetic approach can be concretised in three empirical hypotheses: " -It is possible to demonstrate non-linear phase transitions in cognition. For example continuous changes in stimulus conditions are able to trigger sudden reorganisations in perception. Perceptual organisation cannot be reduced to the properties of the stimulus.

-Stable order in cognition is the result of underlying neuronal dynamics and therefore critically bound to instability. For example any percept is the result of a process of dynamic order formation. Because of the underlying dynamics perception is in principle multistable. Each stable percept can be destabilised and each instable percept can be stabilised.

-Meaning is an order parameter of the elementary neuronal dynamics. For example in the instability of ambiguous displays the basic order formation of perception can be influenced by subtle suggestive cues"[39].

The fact that meaning may influence the structure of brain processes is predicted by the synergetic model of mind-brain interaction. In order to establish a good model for macrodynamic brain-mind processes we need to define specific order parameters which emerge out of the elementary dynamics and which transform the basic instability into coherent stable patterns.

When we consider forms as the result of dynamical stabilisation processes also utilising the more recent advances in the theory of complex systems, we have the concrete possibility to conceive some new methodological tools in order to investigate the problem of form construction, i.e. to outline a theory both phenomenological and physical relative to the actual emergence of meaning: and we have just seen that meaning "shapes" the forms creatively. However, in order to understand in what way this "shaping" takes place we need, as we have just said, more information about the genealogical aspects of the process. Moreover, we also need a semantic and dynamic handling of the processes of information compression as they express themselves at the biological level. In particular we need more and more adequate measures of meaningful complexity, capable, for instance, of taking into account also the dynamic and interactive aspects of depth information. In short, we need new models of cognition. Functional and co-evolutive models not static or interpretative ones. At the level of this kind of models, emergence (in a co-evolutive landscape) and truth (in an intensional setting) for many aspects will necessarily coincide.

The chapter by A, Carsetti "The Embodied Meaning: Self-Organisation and Symbolic Dynamics in Visual Cognition" seeks to present some aspects of contemporary attempts to "reconstruct" a genealogy of vision through a precise revisitation of some of Husserl's original intuitions. Also in this case, this revisitation is operated in the general framework of the contemporary theoretical development of Self-organisation Theory.

In Carsetti's opinion "…vision is the end result of a construction realised in the conditions of experience. It is "direct" and organic in nature because the product of neither simple mental associations nor reversible reasoning, but, primarily, the "harmonic" and targeted articulation of specific attractors at different embedded levels.

The resulting texture is experienced at the conscious level by means of self-reflection, we really sense that it cannot be reduced to anything else, but is primary and self-constituting. We see visual objects; they have no independent existence in themselves but cannot be broken down into elementary data. Grasping the information at the visual level means managing to hear, as it were, inner speech. It means reconstructing in the negative, in an inner generative language, through progressive assimilation, selection and real metamorphosis (albeit partially and roughly) the articulation of the complex "genealogical" apparatus which works at the deep semantic level and moulds and subtends the presentation of the functional patterns at the level of the optical sieve. Vision as emergence aims first of all to grasp the paths and the modalities that determine the selective action, the modalities specifically relative to the revelation of the afore-mentioned apparatus at the surface level according to different and successive phases of generality…. The afore-mentioned paths and modalities thus manage to "speak" through my own fibres. It is exactly through a similar self-organising process, characterised by the presence of a double-selection mechanism, that the brain can partially manage to perceive depth information in an objective way. The extent to which the simulation model succeeds, albeit partially, in encapsulating the secret cipher of this articulation through a specific chain of programs determine the irruption of new creativity as well as the model's ability to see with the eyes of the mind.

To assimilate and see the system must first "think" internally the secret structures of the possible, and then posit itself as a channel (through the precise indication of forms of potential coagulum) for the process of opening and revelation of depth information. This process then works itself gradually into the system's fibres, *via* possible selection, according to the coagulum possibilities offered successively by the system itself.

The revelation and channelling procedures thus emerge as an essential and integrant part of a larger and coupled process of self-organisation. In connection with this process we can ascertain the successive edification of an I-subject conceived as a progressively wrought work of abstraction, unification, and emergence. The fixed points which manage to articulate themselves within this channel, at the level of the trajectories of neural dynamics, represent the real bases on which the "I" can reflect and progressively constitute itself. The I-subject can thus perceive to the extent in which the single visual perceptions are the end result of a coupled process which, through selection, finally leads the original Source to articulate and present itself, by means of cancellations and "irruptions", within (and through) the architectures of reflection, imagination and vision. These perceptions are (partially) veridical, direct, and irreducible. They exist not in themselves, but, on the contrary, for the "I", but simultaneously constitute the primary departure-point for every successive form of reasoning perpetrated by the observer. As an observer I shall thus witness *Natura Naturata* since I have connected functional forms in accordance with a successful and coherent "score".

It is precisely through a coupled process of self-organisation of the kind that it will finally be possible to manage to define specific procedures of reconstruction and representation within the system, whereby the system will be able to identify a given object within its context, together with its *Sinn*. The system will thus be able to perceive the visual object as immersed within its surroundings, as a self-sustaining reality, and, at the same time, feel it living and acting objectively within its own fibres. In this way it will be possible for the brain to perceive depth information according to the truth (albeit partially)"[40].

At the end of this short and incomplete presentation of the main guidelines of the book, let us now make just a few final remarks.

According to the suggestions presented by the authors in the different chapters the world perceived at the visual level appears as constituted not by objects or static forms, but by processes appearing imbued with meaning. As Kanizsa stated, at the visual level the line *per se* does not exist: only the line which enters, goes behind, divides, etc.: a line evolving according to a precise holistic context, in comparison with which function and meaning are indissolubly interlinked. The static line is in actual fact the result of a dynamic compensation of forces. Just as the meaning of words is connected with a universe of highly-dynamic functions and functional processes which operate syntheses, cancellations, integrations, etc. (a universe which can only be described in terms of symbolic dynamics), in the same way, at the level of vision, we must continuously unravel and construct *schemata*; must assimilate and make ourselves available for selection by the co-ordinated information penetrating from external reality. We have, at the same time, to inventively explore the secret architecture of non-standard grammars governing visual perception. Lastly, we must interrelate all this with the internal selection mechanisms through a precise "journey" into the regions of intensionality.

In accordance with these intuitions, we may tentatively consider, from the more general point of view of contemporary Self-organisation theory, the network of meaningful programs living at the level of neural systems as a complex one which forms, articulates, and develops, functionally, within a "coupled universe" characterised by the existence of a double selection. This network gradually posits itself as the basis for the emergence of meaning and the simultaneous, if indirect, surfacing of an "acting I": as the basic instrument, in other words, for the perception of real and meaningful processes, of "objects" possessing meaning, aims, intentions, etc.: above all, of objects possessing an inner plan and linked to the progressive expression of a specific cognitive action.

The brain considered as an "intelligent" network which develops with its meaning articulates as a growing neuronal network through which continuous restructuring processes are effected at a holistic level, thus constituting the indispensable basis of visual cognitive activity. The process is first of all, as stated above, one of canalisation and of revelation (according *in primis* to specific reflection procedures) of precise informational (and generative) fluxes-principles. It will necessarily articulate through *schemata* which will stabilise within circuits and flux determinations. In this sense the brain progressively constitutes itself as a self-organising measuring device in the world and of the world. When, therefore, the model-network posits itself as a 'I-representation' (when the arch of simulation "reaches completion"), and views the world-Nature before it, it sees the world in consonance with the functional forms on which its realisation was based, i.e. according to the architecture proper to the circuits and the patterns of meaning which managed to become established. The result is Nature written in mathematical formulae: Nature read and seen *iuxta propria principia* as a great book (library) of natural forms by means of symbolic characters, grammatical patterns and specific mathematical modules.

From a general point of view, at the level of the articulation of visual cognition, we are actually faced with the existence of precise forms of co-evolution. With respect to this dynamic context, we can recognise, at the level of the aforementioned process of inventive exploration, not only the presence of patterns of self-reflection but also the progressive unfolding of specific fusion and integration functions. We can also find that the *Sinn* that embodies in specific and articulated rational intuitions guides and shapes the paths of the exploration selectively. It appears to determine, in particular, by means of the definition of precise constraints, the choice of a number of privileged patterns of functional dependencies with respect to the entire relational growth. As a result, we can inspect a precise spreading of the development dimensions, a selective cancellation of relations and the rising of specific differentiation processes.

We are faced with a new theoretical landscape characterised by the successive unfolding (in a co-evolutive context) of specific mental processes submitted to the action of well-defined selective pressures and to a continuous "irruption" of depth information. This irruption, however, reveals itself as canalised by means of the action of precise constraints that represent the end product of the successive transformation of the original *gestalten.* Actually, the *gestalten* can "shape" the perceptual space according to a visual order only insofar as they manage to act (on the basis of the transformation undergone) as constraints concerning the generative (and selective) processes at work. Selection is creative because it determines ever-new linguistic functions, ever-new processing units which support the effective articulation of new coherence patterns. The development of this creativity, however, would not be possible without the above mentioned transformation and the inner guide offered by the successful *compositio* of the different constraints in action. On the other hand, the very irruption could not take place if we were not able to explore the non-standard realm in the right way, i.e. if we were not capable of outlining adequate non-standard models and continuously comparing, in an "intelligent" way, our actual native competence with the simulation recipes at work.

We can perceive the objective existence of specific (self-organising) forms only insofar as we transform ourselves into a sort of arch or gridiron for the articulation, at the second-order or higher-order level and in accordance with specific selective procedures, of a series of conceptual plots and fusions, a series that determines a radical *metamorphosis* of our intellectual capacities. It is exactly by means of the actual reflection on the new-generated abstract constructions as well as of the mirror represented by the successful invention of adequate simulation models that I shall finally be able to inspect the realisation of my autonomy, the progressive embodiment of my mental activities in a "new" coherent and self-sustained system.

Meaning can selectively express itself only through, a) the nested realisation, at the abstract level, of specific "fusion" processes, b) the determination of specific schemes of coherence able to support this kind of realisation, c) a more and more co-operative and unified articulation at the deep level of the primary informational fluxes. It shapes the forms in accordance with precise stability factors, symmetry choices, coherent contractions and ramified completions. We can inspect (and "feel") this kind of embodiment, at the level of "categorial intuition", insofar as we successfully manage to reconstruct, identify and connect, at the generative level, the attractors of this particular dynamic process. It is exactly by means of the successive identification and guided *compositio* of these varying attractors that we can manage to imprison the thread of meaning and identify the coherent texture of the constraints concerning the architecture of visual thoughts. In this way we shall finally be able to obtain a first self-representation of our mental activities, thus realising a form of effective autonomy. A representation that exactly concerns the "narration" relative to the progressive opening of the eyes of our mind and the correlated constitution of the *Cogito* and its rules.

REFERENCES

1) Cf.: PUTNAM, H., *Representations and Reality*, Cambridge, 1983, p. 137.

2) Cf.: WUKETITS, F.M., "Functional Realism" in A. Carsetti (Ed.), *Functional Models of Cognition*, Dordrecht, 1999, p. 30.

3) Cf.: GROSSBERG, S., "Linking Mind to Brain: the Mathematics of Biological Intelligence", *Notices of AMS,* 2000, p. 1361.

4) Cf.: GROSSBERG, S., "Linking Mind to Brain: the Mathematics of Biological Intelligence“, op. cit., p. 1364.

5) Cf.: GROSSBERG, S., "Linking Mind to Brain: the Mathematics of Biological Intelligence“, op. cit., p. 1364.

6) Cf.: GROSSBERG, S., "Neural Models of Seeing and Thinking" this volume, p. 52.

7) Cf.: GROSSBERG, S., "Linking Mind to Brain: the Mathematics of Biological Intelligence", op. cit., p. 1366.

8) Cf.: PETITOT, J., "Functional Architecture of the Visual Cortex and Variational Models for Kanizsa's Modal Subjective Contours", this volume, p. 59.

9) Cf.: PETITOT, J., "Functional Architecture of the Visual Cortex and Variational Models for Kanizsa's Modal Subjective Contours", this volume, p. 63.

10) Cf.: PETITOT, J., "Functional Architecture of the Visual Cortex and Variational Models for Kanizsa's Modal Subjective Contours", this volume, p. 66.

11) Cf.: PETITOT, J., "Functional Architecture of the Visual Cortex and Variational Models for Kanizsa's Modal Subjective Contours", this volume, p. 69.

12) Cf.: DESOLNEUX, A., MOISAN, L. and MOREL, J.M., "Gestalt Theory and Computer Vision “, this volume, p. 71.

13) Cf.: DESOLNEUX, A., MOISAN, L. and MOREL, J.M., "Gestalt Theory and Computer Vision ", this volume, p. 72.

14) Cf.: DESOLNEUX, A., MOISAN, L. and MOREL, J.M., "Gestalt Theory and Computer Vision “, this volume, p. 84.

15) Cf.: DESOLNEUX, A., MOISAN, L. and MOREL, J.M., "Gestalt Theory and Computer Vision ", this volume, p. 86.
16) Cf.: O'REGAN, J.K., MYIN, E. and NOË, A., "Towards an Analytic Phenomenology: the Concepts of "Bodiliness" and "Grabbiness", this volume, p. 104.
17) Cf.: O'REGAN, J.K., MYIN, E. and NOË, A., "Towards an Analytic Phenomenology: the Concepts of "Bodiliness" and "Grabbiness", this volume, p. 106.
18) Cf.: O'REGAN, J.K., MYIN, E. and NOË, A., "Towards an Analytic Phenomenology: the Concepts of "Bodiliness" and "Grabbiness", this volume, p. 112.
19) Cf.: DI FERDINANDO, A. and PARISI, D., "Internal Representations of Sensory Input Reflect the Motor Output with which Organism Respond to the Input", this volume, p. 115.
20) Cf.: GONCALVES, L., DI BERNARDO, E. and PERONA, P., "Movemes for Modeling Biological Motion Perception", this volume, p. 143.
21) Cf.: GONCALVES, L., DI BERNARDO, E. and PERONA, P., "Movemes for Modeling Biological Motion Perception", this volume, p. 144.
22) Cf.: GONCALVES, L., DI BERNARDO, E. and PERONA, P., "Movemes for Modeling Biological Motion Perception", this volume, p. 147.
23) Cf.: GONCALVES, L., DI BERNARDO, E. and PERONA, P., "Movemes for Modeling Biological Motion Perception", this volume, p. 148.
24) Cf.: LORENCEAU, J., "Form Constraints in Motion Integration, Segmentation and Selection ", this volume, p. 171.
25) Cf.: LORENCEAU, J., "Form Constraints in Motion Integration, Segmentation and Selection ", this volume, pp. 171-172.
26) Cf.: LORENCEAU, J., "Form Constraints in Motion Integration, Segmentation and Selection ", this volume, p. 176.
27) Cf.: LORENCEAU, J., "Form Constraints in Motion Integration, Segmentation and Selection ", this volume, p. 186.
28) Cf.: LORENCEAU, J., "Form Constraints in Motion Integration, Segmentation and Selection ", this volume, p. 186.
29) Cf.: NINIO, J., "Scintillations, Extinctions, and Other New Visual Effects", this volume, p. 194.
30) Cf.: OLIVETTI, M., DI MATTEO, R., DEL GRATTA, C., DE NICOLA, A., FERRETTI, A., and ROMANI, G.L., "Commonalities Between Visual Imagery and Imagery in Other Modalities; an Investigation by Means of fMRI", this volume, p. 203.
31) Cf.: OLIVETTI, M., DI MATTEO, R., DEL GRATTA, C., DE NICOLA, A., FERRETTI, A., and ROMANI, G.L., "Commonalities Between Visual Imagery and Imagery in Other Modalities; an Investigation by Means of fMRI", this volume, p. 205.
32) Cf.: OLIVETTI, M., DI MATTEO, R., DEL GRATTA, C., DE NICOLA, A., FERRETTI, A., and ROMANI, G.L., "Commonalities Between Visual Imagery and Imagery in Other Modalities; an Investigation by Means of fMRI", this volume, p. 214.
33) Cf.: ROSENTHAL, V., "Microgenesis, Immediate Experience and Visual Processes in Reading", this volume, p. 222.
34) Cf.: ROSENTHAL, V., "Microgenesis, Immediate Experience and Visual Processes in Reading", this volume, p. 223.
35) Cf.: ROSENTHAL, V., "Microgenesis, Immediate Experience and Visual Processes in Reading", this volume, p. 224.
36) Cf.: ROSENTHAL, V., "Microgenesis, Immediate Experience and Visual Processes in Reading", this volume, p. 233.
37) Cf.: VISETTI, Y.M., "Language, Space and the Theory o Semantic Forms", this volume, p. 258.
38) Cf.; WIMMER, M., "Emotion-Cognition Interaction and Language", this volume, p. 280.
39) Cf.; STADLER, M. and KRUSE, P., "Appearance of Structure and Emergence of Meaning in the Visual System", this volume, p. 296.
40) Cf.: CARSETTI, A., "The Embodied Meaning: Self-Organisation and Symbolic Dynamics in Visual Cognition", this volume, pp. 311-2.

PART I

SEEING AND THINKING:

A NEW APPROACH

This page intentionally left blank

STEPHEN GROSSBERG

NEURAL MODELS OF SEEING AND THINKING

INTRODUCTION: SEEING AND THINKING BASED ON BOUNDARIES AND SURFACES

Helmholtz proposed that seeing and thinking are intimately related. He articulated this claim as part of his doctrine of *unconscious inference.* Kanizsa, in contrast, proposed that seeing and thinking often function according to different rules. These alternative intellectual positions remain as an enduring controversy in visual science. Why is the relationship between seeing and thinking so hard to disentangle?

Recent neural models of visual perception have clarified how seeing and thinking operate at different levels of the brain and use distinct specialized circuits. But these processes also interact imtimately via feedback and use similar laminar cortical designs that are specialized for their distinct functions. In addition, this feedback has been predicted to be an essential component in giving rise to conscious visual percepts, and recent data have provided support for this prediction. Thus, although seeing and thinking are carried out by different parts of the brain, they also often interact intimately via feedforward and feedback interactions to give rise to conscious visual percepts.

The distinction between seeing and thinking is sometimes caste as the distinction between seeing and knowing, or between seeing and recognizing. The fact that these processes are not the same can be understood by considering a suitable juxtaposition of boundary and surface percepts. The FACADE (Form-And-Color-And-DEpth) theory of how the brain gives rise to visual percepts has clarified the sense in which these boundary and surface percepts compute *complementary* properties, and along the way how and why properties of seeing and recognizing are different (Grossberg, 1984, 1994, 1997; Grossberg and Kelly, 1999; Grossberg and McLoughlin, 1997; Grossberg and Mingolla, 1985a, 1985b; Grossberg and Pessoa, 1998; Grossberg and Todorovic, 1988; Kelly and Grossberg, 2000; McLoughlin and Grossberg, 1998). FACADE theory proposes that perceptual boundaries are formed in the LGN—interblob— interstripe—V4 stream, whereas perceptual surfaces are formed in the LGN—blob—thin stripe—V4 stream (Grossberg, 1994); see Figure 1. Many experiments have supported this prediction (Elder and Zucker, 1998; Lamme et al., 1999; Rogers-Ramachandran and Ramachandran, 1998).

A. Carsetti (ed.), Seeing, Thinking and Knowing, 29-54.

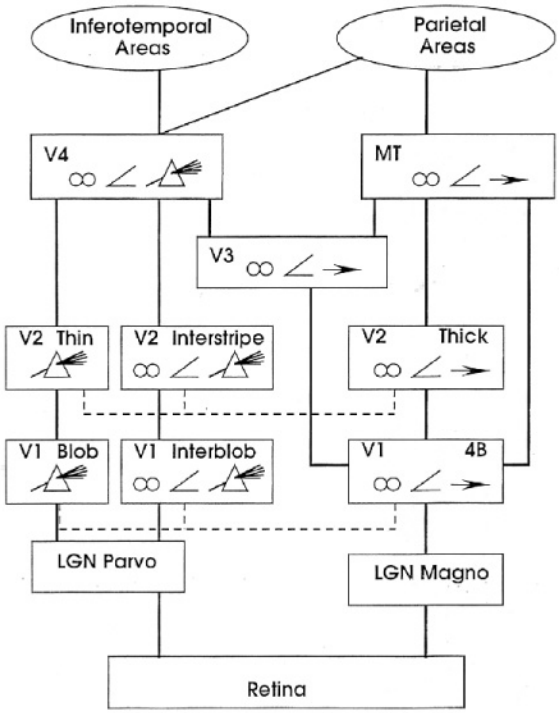

Figure 1. Schematic diagram of processing streams in visual cortex in the macaque monkey brain. Icons indicate the response selectivities of cells at each processing stage: rainbow = wavelength selectivity, angle symbol = orientation selectivity, spectacles binocular selectivity, and right-pointing arrow = selectivity to motion in a prescribed direction. [Adapted with permission from DeYoe and van Essen (1988).]

Figure 2a illustrates three pairs of complementary properties using the illusory contour percept of a Kanizsa square (Kanizsa, 1974). Such a percept immediately raises the question of why our brains construct a square where there is none in the image. There are several functional reasons why our brains have developed strategies to construct complete representations of boundaries and surfaces on the basis of incomplete information. One reason is that there is a *blind spot* in our retinas; namely, a region where no light-sensitive photoreceptors exist. This region is blind because of the way in which the pathways from retinal photoreceptors are collected together to form the optic nerve that carries them from the retina to the LGN in Figure 1. We are not usually aware of this blind spot because our brains complete boundary and surface information across it. The actively completed parts of these percepts are visual illusions, because they are not derived directly from visual signals on our retinas. Thus many of the percepts that we believe to be "real" are visual illusions whose boundary and surface representations just happen to look real. I suggest that what we call a visual illusion is just an unfamiliar combination of boundary and surface information. This hypothesis is illustrated by the percepts generated in our brains from the images in Figure 2.

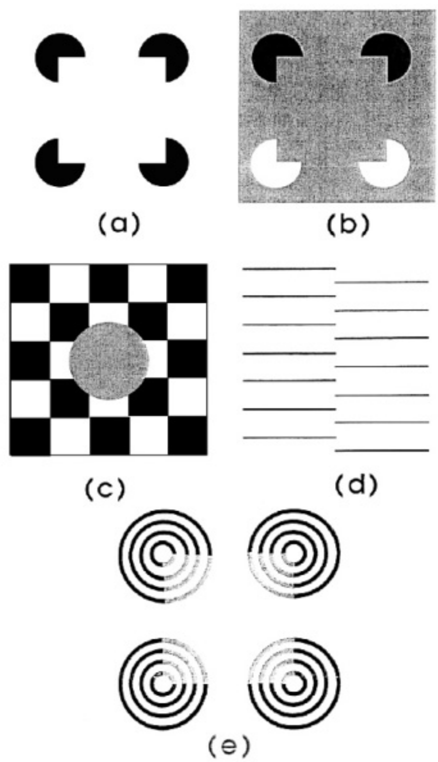

Figure 2. Visual boundary and surface interactions: (a) A Kanizsa square. (b) A reverse-contrast Kanizsa. (c) An object boundary can form around the gray disk even though its contrast reverses relative to the background along its perimeter. (d) An invisible, or amodel, vertical boundary. (e) An example of neon color spreading.

In response to the image in Figure 2a, boundaries form *inwardly* between cooperating pairs of colinear edges of the four pac man, or pie shaped, inducers. Four such contours form the boundary of the perceived Kanizsa square. (If boundaries formed outwardly from a single inducer, then any speck of dirt in an image could crowd all our percepts with an outwardly growing web of boundaries.) These boundaries are *oriented* to form in a collinear fashion between (almost) colinear and (almost) like-oriented inducers. The square boundary in Figure 2a can be both seen and recognized because of the enhanced illusory brightness of the Kanizsa square. By contrast, the square boundary in Figure 2b can be recognized even though it is not visible; that is, there is no brightness or color difference on either side of the boundary. Figure 2b shows that some boundaries can be recognized even though they are invisible, and thus that seeing and recognizing cannot be the same process. FACADE theory predicts that "all boundaries are invisible" within the boundary stream, which is proposed to occur in the interblob cortical processing stream (Figure 1). This prediction has not yet been directly tested neurophysiologically, although several studies have shown that the distinctness of a perceptual grouping, such as an illusory contour, can be dissociated from the visible stimulus contrast with which it is associated (Hess et al., 1998; Petry and Meyer, 1987).

Why is the square boundary in Figure 2b invisible? This property can be traced to the fact that its vertical boundaries form between black and white inducers that possess opposite contrast polarity with respect to the gray background. The same is true of the boundary around the gray disk in Figure 2c, which is another figure that was originally proposed by Kanizsa, but to make a different point. In this figure, the gray disk lies in front of a textured background whose contrasts with respect to the disk reverse across space. In order to build a boundary around the entire disk, despite these contrast reversals, the boundary system pools signals from opposite contrast polarities at each position. This pooling process renders the boundary-system output *insensitive to contrast polarity.* The boundary system therefore loses its ability to represent visible colors or brightnesses, as its output cannot signal the difference between dark and light. It is in this sense that "all boundaries are invisible". Figure 2d illustrates another invisible boundary that can be consciously recognized. Figure 2 hereby illustrates that seeing and recognizing must use different processes, since they can be combined or dissociated, in response to relatively small changes in the contrasts of an image, holding its geometrical relationships constant, or indeed by changing the geometrical relationships of an image, while holding its contrasts constant.

If boundaries are invisible, then how do we see anything? FACADE theory predicts that visible properties of a scene are represented by the surface processing stream, which is predicted to occur within the blob cortical stream (Figure 1). A key step in representing a visible surface is "filling-in". Why does this step occur? An early stage of surface processing compensates for variable illumination, or "discounts the illuminant" (Grossberg and Todorovic, 1988; Helmholtz, 1910/1925; Land 1977), in order to prevent illuminant variations, which can change from moment to moment, from distorting all percepts. Discounting the illuminant attenuates color and brightness signals, except near regions of sufficiently rapid surface change, such as edges or texture gradients, which are relatively uncontaminated by illuminant variations. Later stages of surface formation fill in the attenuated regions with these relatively uncontaminated color and brightness signals, and do so at the correct relative depths from the observer through a process called surface capture. This multi-stage process is an example of *hierarchical resolution of uncertainty,* because the later filling-in stage overcomes uncertainties about brightness and color that were caused by discounting the illuminant at an earlier processing stage.

How do the illuminant-discounted signals fill-in an entire region? Filling-in behaves like a diffusion of brightness across space (Arrington, 1994; Grossberg and Todorovic, 1988; Paradiso and Nakayama, 1991). Figure 2e leads to a percept of "neon color spreading" in which, filling-in spreads *outwardly* from the individual gray inducers in all directions. Its spread is thus *unoriented.* How is this spread of activation contained? FACADE theory predicts that signals from the boundary stream to the surface stream define the regions within which filling-in is restricted. In response to Figure 2e, the brain forms boundaries surround the annuli, except for small breaks in the boundaries where the gray and black contours intersect, and also forms the square illusory boundary. Some of the gray color can escape from their annuli through these breaks into the square region in the surface stream. This prediction has not yet been tested neurophysiologically. Without these boundary signals, filling-in would dissipate across space, and no surface percept could form. Invisible boundaries therefore indirectly assure their own visibility through their interactions with the surface stream.

In Figure 2a, the square boundary is induced by four black pac man disks that are all less luminant than the white background. In the surface stream, discounting the illuminant causes these pac men to induce local brightness contrasts within the boundary of the square. At a subsequent processing stage, these brightness contrasts trigger surface filling-in within the square boundary. The filled-in square is visible as a brightness difference because the filled-in activity level within the square differs from the filled-in activity of the surrounding region. Filling-in can lead to visible percepts because it is *sensitive to contrast polarity.* These three properties (outward, unoriented, sensitive to contrast-polarity) are complementary to the corresponding properties (inward, oriented, insensitive to contrast-polarity) of boundary completion.

In Figure 2b, the opposite polarities of the two pairs of pac men with respect to the gray background lead to approximately equal filled-in activities inside and outside the square, so the boundary can be recognized but not seen. In Figure 2d, the white background can fill-in uniformly on both sides of the vertical boundary, so no visible contrast difference is seen.

These remarks just begin the analysis of filling-in. Even in the seemingly simple case of the Kanizsa square, one often perceives a square hovering in front of four partially occluded circular disks, which seem to be completed behind the square. FACADE theory predicts how surface filling-in is organized to help such figure—ground percepts to occur, in response to both 2-D pictures and 3-D scenes (Grossberg, 1994, 1997).

In summary, boundary and surface formation illustrate two key principles of brain organization: hierarchical resolution of uncertainty and complementary interstream interactions. Hierarchical resolution of uncertainty is illustrated by surface filling-in: discounting the illuminant creates uncertainty by suppressing surface color and brightness signals, except near surface discontinuities. Higher stages of filling-in complete the surface representation using properties that are complementary to those by which boundaries are formed, guided by signals from these boundaries (Arrington, 1994; Grossberg, 1994; Grossberg and Todorovic, 1988; Paradiso and Nakayama, 1991).

Before going further, it should be emphasized that the proposal that boundaries and surfaces are computed by the interblob and blob streams is not the same as the proposal of Hubel and Livingstone (1985) that orientation and color are computed by these streams.

Some differences between these proposals are: (1) Illusory contours are boundaries that form over regions that receive no oriented bottom-up inputs. Boundaries that can form over regions that receive no oriented inputs cannot be viewed as part of an "orientation" system. (2) Boundaries are predicted to be invisible, or amodal, throughout the boundary stream, whether real or illusory. This is a concept that goes far beyond any classical notion of an "orientation" system. (3) Surfaces are predicted to fill-in amodally, or invisibly, in cortical area V2, but modally, or visibly, in cortical area V4. A surface system that can fill-in amodal percepts cannot be viewed as a "color " system. (4) Boundaries are predicted to organize the separation of occluded objects from their occluding objects. The analysis of figure-ground separation also goes far beyond any direct concept of an "orientation" system. Because of such differences, FACADE theory has been able to propose explanations of many experiments that cannot be explained just using classical concepts of orientation and color. Within FACADE theory, the circuits that form perceptual boundaries are called the Boundary Contour System, or BCS, and the circuits that form perceptual surfaces are called the Feature Contour System, or FCS. This nomenclature arose from the realization that, in both systems, an early stage of processing extracts contour-sensitive information as a basis for further processing. In the BCS, these contours are completed as invisible perceptual boundaries, In the FCS, these contours are extracted by the process of "discounting the illuminant" and form the basis for percepts of visible filled-in surface "features."

The percepts derived from Figure 2 clarify that seeing and thinking are different processes, but do not indicate where the thinking processes take place. Much evidence suggests that objects are recognized in the inferotemporal cortex, which may be primed by top-down inputs from prefrontal cortex; see Figure 2. Thus, if an amodal boundary in the interblob boundary stream of visual cortex is projected to a familiar recognition category in the inferotemporal cortex, it can be recognized even if it cannot be seen in area V4 of the blob surface stream. Recognition is not, however, merely a matter of feedforward activation of the inferotemporal cortex. Nor is seeing just a matter of feedforward activation of area V4. Much experimental and theoretical evidence, notably as explained within Adaptive Resonance Theory, or ART, suggests that top-down matching and attentional processes are normally part of the events that lead to conscious recognition, even of an amodal percept (Grossberg, 1980, 1999c).

BOUNDARY COMPLETION AND ATTENTION BY THE LAMINAR CIRCUITS OF VISUAL CORTEX

How does visual cortex complete boundaries across gaps due to internal brain imperfections, such as the retinal blind spot, or due to incomplete contours in external inputs, such as occluded surfaces, spatially discrete texture elements, illusory contour stimuli, or even missing pixels in impressionist paintings? This information is shown below to lead to new insights about processes like figure-ground perception that clarify many of the percepts that Kanizsa earlier observed in this field. In particular, the BCS model proposes how long-range horizontal cooperation interacts with shorter-range competition to carry out percep-

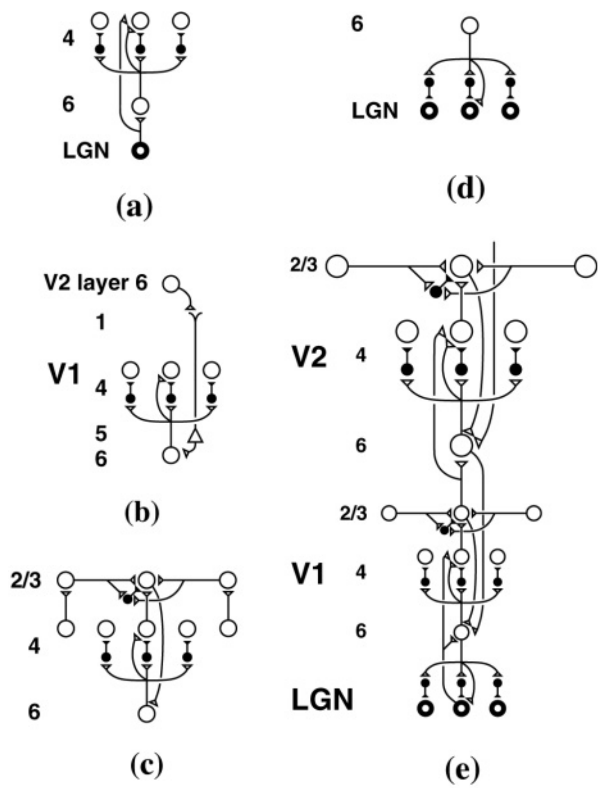

Figure 3. Boundary and attentional laminar circuits in interblob cortical areas Vi and V2. Inhibitory interneurons are shown filled-in black. (a) The LGN provides bottom-up activation to layer 4 via two routes: It makes strong connections directly into layers 4 and 6. Layer 6 then activates layer 4 via a 6 → 4 on-center off-surround network. In all, LGN pathways to layer 4 form an on-center off-surround network, which contrast-normalizes layer 4 cell responses. (b) *Folded feedback* carries attentional signals from higher cortex into layer 4 of Vi, via the 6 → 4 path. Corticocortical feedback axons tend to originate in layer 6 of the higher area and to terminate in the lower cortex's layer 1, where they excite apical dendrites of layer 5 pyramidal cells whose axons send collaterals into layer 6. Several other routes through which feedback can pass into V1 layer 6 exist. Having arrived in layer 6, the feedback is then "folded" back up into the feedforward stream by passing through the 6 → 4 on-center off-surround path. (c) Connecting the 6 → 4 on-center off-surround to the layer 2/3 grouping circuit: like-oriented layer 4 simple cells with opposite contrast polarities compete (not shown) before generating half-wave rectified outputs that converge onto layer 2/3 complex cells in the column above them. Groupings that form in layer 2/3 enhance their own positions in layer 4 via the 6 → 4 on-center, and suppress other groupings via the 6 → 4 off-surround. There exist direct layer 2 / 3 → 6 connections in macaque Vi, as well as indirect routes via layer 5. (d) Top-down corticogeniculate feedback from V1 layer 6 to LGN also has an on-center off-surround anatomy, similar to the 6 → 4 path. The on-center feedback selectively enhances LGN cells that are consistent with the activation that they cause, and the off-surround contributes to length-sensitive (endstopped) responses that facilitate grouping perpendicular to line ends. (e) The model V1/V2 circuit: V2 repeats the laminar pattern of V1 circuitry at a larger spatial scale; notably, the horizontal layer 2/3 connections have a longer range in V2. V1 layer 2/3 projects to V2 layers 6 and 4, just as LGN projects to layers 6 and 4 of V1. Higher cortical areas send feedback into V2 which ultimately reaches layer 6, just as V2 feedback acts on layer 6 of V1. Feedback paths from higher cortical areas straight into V1 (not shown) can complement and enhance feedback from V2 into V1. [Reprinted with permission from Grossberg and Raizada (2000).]

tual grouping. The cooperating cells were predicted to satisfy a *bipole* property (Cohen and Grossberg, 1984; Grossberg, 1984; Grossberg and Mingolla, 1985a, 1985b). Such "bipole cells" realize inward and oriented boundary completion by firing when they receive inputs from pairs of (almost) like oriented and (almost) colinear scenic inducers. Von der Heydt et al. (1984) reported the first neurophysiological evidence in support of this prediction, by observing bipole cell properties in cortical area V2. Subsequent psychophysical studies have also provided additional evidence in support of a bipole property during perceptual grouping. Field et al. (1993) called this property an "association field," and Shipley and Kellman (1992) called it a "relatability condition".

More recently, the BCS was extended to clarify how and why visual cortex, indeed all sensory and cognitive neocortex, is organized into layered circuits (Grossberg, 1999b; Grossberg and Raizada, 2000; Raizada and Grossberg, 2001). This LAMINART model predicts how bottom-up, top-down, and horizontal interactions within the cortical layers realize: (1) perceptual grouping; (2) attention; and (3) stable development and learning. The model proposes how mechanisms that achieve property (3) imply (1) and (2). That is, constraints on stable development of cortical circuits in the infant determine properties of learning, perception, and attention in the adult.

Figure 3 summarizes how known laminar cortical circuits may carry out perceptual grouping and attention. This summary omits binocular interactions for simplicity; but see below. The lateral geniculate nucleus, or LGN, directly activates V1 layers 4 and 6 (Figure 3a). Layer 6, in turn, sends on-center off-surround inputs to the simple cells of layer 4. Layer 6 can strongly inhibit layer 4 through the off-surround, but the excitatory and inhibitory inputs in the on-center are proposed to be approximately balanced, with perhaps a slight excitatory bias. Layer 6 can thus modulate the excitability of layer 4 cells, but not drive them to fire vigorously. This balance has been shown through modeling simulations to help the cortex develop its connections in a stable way (Grossberg and Williamson, 2001). The direct LGN-to-4 connections are proposed to drive layer 4 cells to reach suprathreshold activation levels. The direct and indirect pathways from LGN-to-4 together form an on-center off-surround network. Under the assumption that layer 4 cells obey membrane, or shunting, equations, such an on-center off-surround network can contrast-normalize the responses of layer 4 cells (Douglas et al., 1995; Grossberg, 1973, 1980; Heeger, 1992), and thus preserve their sensitivity to input differences over a wide dynamic range.

Figure 3b illustrates how the modulatory layer 6-to-4 circuit can also be used by top-down signals from V2 layer 6 to attentionally modulate the excitability of V1 layer 4 cells, while inhibiting layer 4 cells that are not in the attentional focus.

Boundary completion that obeys a bipole property occurs within layer 2/3, as illustrated in Figure 3c. Layer 4 cells activate layer 2/3 complex cells, which communicate with their layer 2/3 neighbors via long-range horizontal excitatory connections and shorter-range inhibitory interneurons. The strengths of these excitatory and inhibitory interactions are predicted to be approximately balanced. This balance has also been proposed to help ensure that the cortex develops its connections in a stable way (Grossberg and Williamson, 2001). Because of this balance, activation of a single layer 2/3 cell causes its horizontal excitatory and inhibitory connections to be approximately equally activated. The inhibition cancels the excitation, so a boundary cannot shoot out from individual scenic inducers. When two or more (approximately) like-oriented and (approximately) colinear layer 2/3 cells are activated, the excitation that converges on cells in between can summate, but the inhibitory interneurons form a recurrent inhibitory net that normalizes its total activity. As a result, total excitation exceeds total inhibition, and the cells can fire. A boundary can hereby be completed inwardly, but not outwardly.

If a scene has unambigous groupings, then this horizontal interaction can rapidly complete boundaries along a feedforward pathway from layer 4 to layer 2/3 and then horizontally across layer 2/3, from which outputs to higher cortical areas are emitted. This property is consistent with recent data showing that very fast recognition of visual scenes is possible (Thorpe et al., 1996). On the other hand, it is also well-known that some scenes take longer to recognize. Within the model, in response to scenes with multiple possible groupings, competitive interactions within layers 2/3 and 4 can keep the layer 2/3 activities small, and thus prevent large output signals from being rapidly emitted to higher cortical areas. These smaller layer 2/3 activities are large enough, however, to generate positive feedback signals between layers 2/3-6-4-2/3 of their own cortical area (see Figure 3c). The positive feedback signals can quickly amplify the activities of the strongest grouping, which can then generate large outputs from layer 2/3, while its strong layer 6-to-4 off-surround signals inhibit weaker groupings. These intracortical feedback signals convert the cells across the layers into functional columns (Mountcastle, 1957) and show that the classical Hubel and Wiesel (1977) proposal that there are feedforward interactions from layer 4 simple cells to layer 2/3 complex cells is part of a more complex circuit which also ties these cells together using nonlinear feedback signals.

The above discussion shows that the layer 6-to-4 circuit has at least three functions: It contrast-normalizes bottom-up inputs, selects groupings via intracortical feedback from layer 2/3 without causing a loss of analog sensitivity, and primes attention via intercortical feedback from higher cortical areas. This intimate connection between grouping and attention enables attention to flow along a grouping, and thereby selectively enhance an entire object, as Roelfsema et al. (1998) have shown in macaque area Vi. Because attention acts through an on-center off-surround circuit, it can "protect" feature detectors from inhibition by distractors by using its off-surround, as Reynolds et al. (1999) have shown in areas V2 and V4. Because both cooperation and competition influence groupings, the effects of colinear inducers can be either facilitatory or suppressive at different contrasts, as Polat et al. (1998) have shown in area V1. Grossberg and Raizada (2000) and Raizada and Grossberg (2001) have quantitatively simulated these and related neurophysiological data using the LAMINART model.

The model proposes that a top-down on-center off-surround network from V1 layer 6 to the LGN (Figure 3d) can act in much the same way as the top-down signals from V2 layer 6 to V1. The existence of this type of top-down modulatory corticogeniculate feedback was originally predicted in Grossberg (1976), and has recently been supported by neurophysiological data of Sillito et al. (1994). Grossberg (1976) also predicted that such top-down modulatory feedback helps to stabilize the development of both bottom-up and top-down connections between the LGN and V1. This prediction has not yet been neurophysiologically tested, although it is consistent with evidence of Murphy et al. (1999) showing that the top-down signals from an oriented cortical cell tend to distribute themselves across the LGN in an oriented manner that is consistent with such a learning process. Figure 3e synthesizes all of these LGN and cortical circuits into system architecture, which shows that the horizontal interactions within V2 layer 2/3 can have a broader spatial extent than those in V1 layer 2/3. The V2 interactions are proposed to carry out perceptual groupings like illusory contours, texture grouping, completion of occluded objects, and bridging the blind spot. The V1 interactions are proposed to improve signal-to-noise of feature detectors within the V1 cortical map.

The LAMINART model has been extended to clarify how depthful boundaries are formed (Grossberg and Howe, 2002; Howe and Grossberg, 2001) and how slanted surfaces are perceived in 3-D (Swaminathan and Grossberg, 2001). This generalization is consistent with laminar anatomical and neurophysiological data. For example, suppose that a scenic feature activates monocular simple cells in layer 4 via the left eye and right eye. These simple cells are sensitive to the same contrast polarity and orientation. Because they are activated by different eyes, they are positionally displaced with respect to one another on their respective retinas. These monocular simple cells activate disparity-selective binocular simple cells in layer 3B. Binocular simple cells that are sensitive to the same disparity but to opposite contrast polarities then activate complex cells in layer 2/3A. This two-stage process enables the cortex to binocularly match features with the same contrast polarity, yet to also form boundaries around objects in front of textured backgrounds, as in Figure 2c.

3-D VISION AND FIGURE-GROUND SEPARATION

How are depthful boundary and surface representations formed? How are percepts of occluding and occluded objects represented in depth? FACADE theory proposes how such percepts arise when boundary and surface representations interact together during 3-D vision and figure-ground perception. The rest of this section summarizes some of key design problems that need to be solved before outlining model mechanisms that embody a proposed solution of these problems. A key insight is that the bipole property that controls perceptual grouping also initiates figure-ground separation. How FACADE theory explains figure-ground separation will then be illustrated with a simulation example of Bregman-Kanizsa figure-ground separation.

1. 3-D Surface Capture and Filling-in. How are the luminance and color signals that are received by the two eyes transformed into 3-D surface percepts? FACADE theory posits that multiple depth-selective boundary representations exist and interact with multiple surface filling-in domains to determine which surfaces in depth can be seen. The same filling-in processes which enable us to see perceptual qualities like brightness and color are hereby predicted to also determine the relative depth of these surfaces. In particular, depth-selective boundaries selectively *capture* brightness and color signals at the subset of filling-in domains with which they interact. Filling-in of these captured signals leads to surface percepts at the corresponding relative depths from the observer. The hypothesis that the same filling-in process controls brightness, color, and depth predicts that perceived depth and brightness can influence one another. In fact, the property of "proximity-luminance covariation" means that brighter surfaces can look closer (Egusa, 1983). In particular, brighter Kanizsa squares can look closer than their pac man inducers (Bradley and Dumais, 1984).

2. Binocular Fusion, Grouping, and da Vinci Stereopsis. Granted that surface capture can achieve depth-selective filling-in, how are the depth-selective boundaries formed that control surface capture? Our two eyes view the world through slightly different perspectives. Their different views lead to relative displacements, or disparities, on their retinas of the images that they register. These disparate retinal images are binocularly matched at disparity-sensitive cells, as noted above in the discussion of binocular matching within cortical layer 3B (Grossberg and Howe, 2002). The disparity-sensitive cells in the interblobs of area Vl are used to form depth-selective boundaries in the interstripes of area V2. These boundaries capture surface filling-in signals at the corresponding filling-in domains in the thin stripes of area V2, among other places.

When two eyes view the world, part of a scene may be seen by only one eye. No disparity signals are available here to determine the depth of the monocularly viewed features, yet they are seen at the correct depth, as during Da Vinci stereopsis (Nakayama and Shimojo, 1990). FACADE theory proposes how depth-selective filling-in of a nearby binocularly viewed region spreads into the monocularly viewed region to impart the correct depth. This proposal also explains related phenomena like the "equidistance tendency," whereby a monocularly viewed object in a binocular scene seems to lie at the same depth as the retinally most contiguous binocularly viewed object (Gogel, 1965). How this works leads to a number of new hypotheses, including how horizontal and monocularly-viewed boundaries are added to all boundary representations, and how an "asymmetry between near and far" adds boundaries from nearer to farther surface representations (Grossberg, 1994; Grossberg and Howe, 2002; Grossberg and McLoughlin, 1997). Without these mechanisms, all occluding objects would look transparent.

3. Multiple Scales into Multiple Boundary Depths. When a single eye views the image of an object in depth, the same size of the retinal image may be due to either a large object far away or to a small object nearby. How is this ambiguity overcome to activate the correct disparity-sensitive cells? The brain uses multiple receptive field sizes, or scales, that achieve a "size-disparity correlation" between retinal size and binocular disparity. It has often been thought that larger scales code nearer objects and smaller scales more distant objects. For example, a nearer object can lead to a larger disparity that can be binocularly fused by a larger scale. In fact, each scale can fuse multiple disparities, although larger scales can fuse a wider range of disparities (Julesz and Schumer, 1981). This ambiguity helps to explain how higher spatial frequencies in an image can sometimes look closer, rather than more distant, than lower spatial frequencies in an image, and how this percept can reverse during prolonged viewing (Brown and Weisstein, 1988). FACADE theory explains these reversals by analyzing how multiple spatial scales interact to form depth-selective boundary groupings (Grossberg, 1994).

Multiple spatial scales also help to explain how shaded surfaces are seen. In fact, if boundaries were sensitive only to the bounding edge of a shaded surface, then shaded surfaces would look uniformly bright and flat after filling-in occurs. This does not occur because boundaries respond to shading gradients as well as to edges. Within each scale, a "boundary web" of small boundary compartments can be elicited in response to a shading gradient. Although the boundaries themselves are invisible, their existence can indirectly be detected because the boundaries in a boundary web trap contrasts locally. The different contrasts in each of the small compartments leads to a shaded surface percept. Different scales may react differently to such a shading gradient, thereby leading to a different boundary web of small boundary compartments at each depth. Each boundary web can capture and selectively fill-in its contrasts on a distinct surface representation in depth. The ensemble of these filled-in surface representations can give rise to a percept of a shaded surface in depth.

4. Recognizing Objects vs. Seeing their Unoccluded Parts. In many scenes, some objects lie partially in front of other objects and thereby occlude them. How do we know which features belong to the different objects, both in 3-D scenes and 2-D pictures? If we could not make this distinction, then object recognition would be severely impaired. FACADE theory predicts how the mechanisms which solve this problem when we view 3-D scenes also solve the problem when we view 2-D pictures (Grossberg, 1994, 1997).

Figure 4. (a) Uppercase gray B letters that are partially occluded by a black snakelike occluder. (b) Same B shapes as in (a) except the occluder is white and therefore merges with the remainder of the white background. [Adapted with permission from Nakayama, Shimojo, and Silverman (1989).]

In the Bregman-Kanizsa image of Figure 4a, the gray B shapes can be readily recognized even though they are partially occluded by the black snakelike occluder. In Figure 4b, the occluder is removed. Although the same amount of gray is shown in both images, the B shapes are harder to recognize in Figure 4b. This happens because the boundaries that are shared by the black occluder and the gray B shapes in Figure 4a are assigned by the brain to the black occluder. The bipole property plays an important role in initiating this process. The occluder boundaries form within a boundary representation that codes a nearer distance to the viewer than the boundary representation of the gray shapes. With the shared boundaries removed from the gray B shapes, the B boundaries can be completed behind the positions of the black occluder as part of a farther boundary representation. The completion of these boundaries also uses the bipole property. These completed boundaries help to recognize the B's at the farther depth. In Figure 4b, the shared boundaries are not removed from the gray shapes, and they prevent the completion of the gray boundaries.

To actually do this, the brain needs to solve several problems. First, it needs to figure out how geometrical and contrast factors work together. In Figure 4a, for example, the T-junctions where the gray shapes intersect the black occluders are a cue for signaling that the black occluder looks closer than the gray shapes. However, if you imagine the black occluder gradually getting lighter until it matches the white background in Figure 4b, it is clear that, when the occluder is light enough, the gray shapes will no longer appear behind the occluder. Thus, geometrical factors like T-junctions are not sufficient to cause figure-ground separation. They interact with contrast relationships within the scene too.

The brain also needs to figure out how to complete the B boundaries "behind" the occluder in response to a 2-D picture. In particular, how do different spatial scales get differentially activated by a 2-D picture as well as a 3-D scene, so that the occluding and occluded objects can be seen in depth? Moreover, if the B boundaries can be completed and thereby recognized, then why do we not *see* completely filled-in B shapes too, including in the regions behind the black occluder? This state of affairs clarifies that there is a design tension between properties needed to recognize opaque objects, including where they are occluded, and our ability to see only their unoccluded surfaces. Here again, "the asymmetry between near and far" plays a key role, as noted below.

5. From Boundary-Surface Complementarity to Consistency. Such subtle data make one wonder about how the brain evolved to behave in this way. FACADE theory predicts how simple mechanisms that realize a few new perceptual principles can explain figure-ground data when they interact together. One such principle is that boundary and surface computations are complementary, as noted above. How, then, do we see a single percept wherein boundaries and surfaces are consistently joined? How does *complementarity* become *consistency?* FACADE theory proposes how consistency is realized by a simple kind of feedback that occurs between the boundary and surface streams. Remarkably, this feedback also explains many properties of figure-ground perception. Figure-ground explanations can hereby be reduced to questions about complementarity and consistency, rather than about issues concerning the ecological validity, or probability, of these percepts in our experience. The remainder of the chapter sketches some of these principles and mechanisms, followed by explanations and simulations of the BregmanKanizsa and Kanizsa Stratification percepts.

3-D BOUNDARY AND SURFACE FORMATION

Figure 5 is a macrocircuit of FACADE theory in its present form. This macrocircuit will be reviewed to clarify how it embodies solutions to the five design problems that have just been summarized. Monocular processing of left-eye and right-eye inputs by the retina and LGN discounts the illuminant and generates parallel signals to the BCS and FCS. These signals activate model cortical simple cells via pathways 1 in Figure 5, and monocular filling-in domains (FIDOs) via pathways 2. Model simple cells have oriented receptive fields and come in multiple sizes. Simple cell outputs are binocularly combined at disparity-sensitive complex and complex end-stopped (or hypercomplex) cells via pathways 3. Complex cells with larger receptive fields can binocularly fuse a broader range of disparities than can cells with smaller receptive fields, thereby realizing a "size-disparity correlation." Competition across disparity at each position and among cells of a given size scale sharpens complex cell disparity tuning (Fahle and Westheimer, 1995). Spatial competition (end-stopping) and orientational competition convert complex cell responses into spatially and orientationally sharper responses at hypercomplex cells.

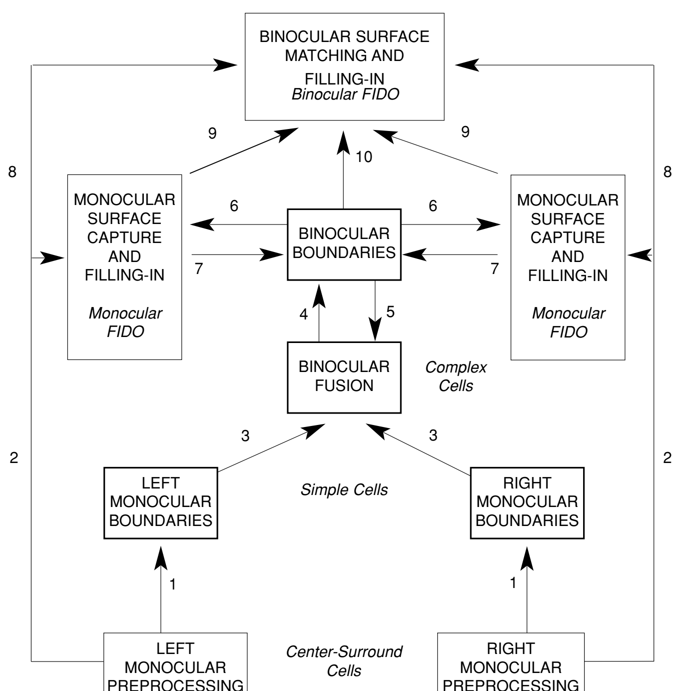

Figure 5. FACADE macrocircuit showing interactions of the Boundary Contour System (BCS) and Feature Contour System (FCS). See text for details. [Reprinted with permission from Kelly and Grossberg (2000).]

How are responses from multiple receptive field *sizes* combined to generate boundary representations of relative *depths* from the observer? Hypercomplex cells in area Vi activate bipole cells in area V2 via pathway 4. The bipole cells carry out long-range grouping and boundary completion via horizontal connections that occur in layer 2/3 of area V2 interstripes. Bipole grouping collects together outputs from hypercomplex cells of all sizes that are sensitive to a given depth range. The bipole cells then send excitatory feedback signals via pathways 5 back to all hypercomplex cells that represent the same position and orientation, and inhibitory feedback signals to hypercomplex cells at nearby positions and orientations; cf., layer 2/3-6-4 inhibition in Figure 3e. The feedback groups cells of multiple *sizes* into a BCS representation, or copy, that is sensitive to a range of *depths.* Multiple BCS copies are formed, each corresponding to different (but possibly overlapping) depth ranges.

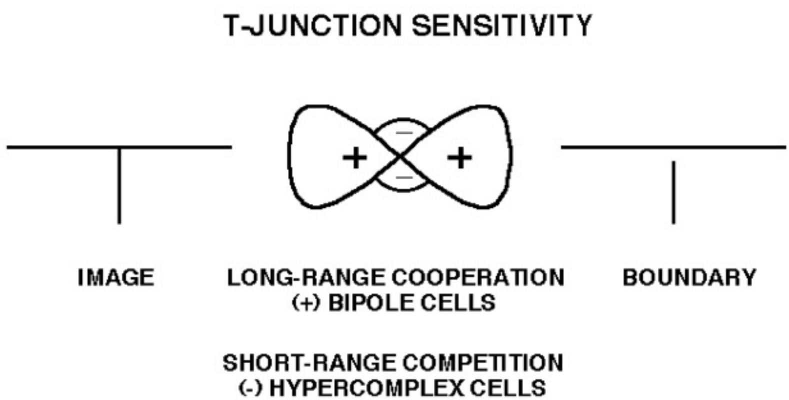

Figure 6. T-junction sensitivity in the BCS: (a) T-junction in an image. (b) Bipole cells provide long-range cooperation (+), whereas hypercomplex cells provide shorter-ranger competition (-). (c) An end-gap in the vertical boundary arises due to this combination of cooperation and competition. [Reprinted with permission from Grossberg (1997).]

Bipole cells play a key role in figure-ground separation. Each bipole cell has an oriented receptive field with two branches (Figure 6). Long-range excitatory bipole signals in layer 2/3 combines with shorter-range inhibitory signals in layers 4 and 2/3 to make the system sensitive to T-junctions (Figure 6). In particular, horizontally-oriented bipole cells that are located where the top of the T joins its stem receive excitatory inputs to both of their receptive field branches. Vertically-oriented bipole cells that process the stem of the T where it joins the top receive excitatory support only in the one branch that is activated by the stem. Because of this excitatory imbalance, inhibition of the stem by the top can cause a gap in the stem boundary, termed an *end-gap* (Figure 6). During filling-in, boundaries contain the filling-in process. Where end-gaps occur, brightness or color can flow out of a figural region, much as it flows out of the annuli in Figure 2e during neon color spreading. FACADE theory predicts that this escape of color or brightness via filling-in is a key step that initiates figure-ground separation (Grossberg. 1994, 1997; Grossberg and McLoughlin. 1997; Kelly and Grossberg. 2000). Figure 7 shows a simulation from Kelly and Grossberg (2000) which illustrates end gaps in response to the Bregman-Kanizsa image. End-gaps occur where the horizontal occluder touches the partially occluded B shape, at both near (Figure 7a) and far (Figure 7b) depths.

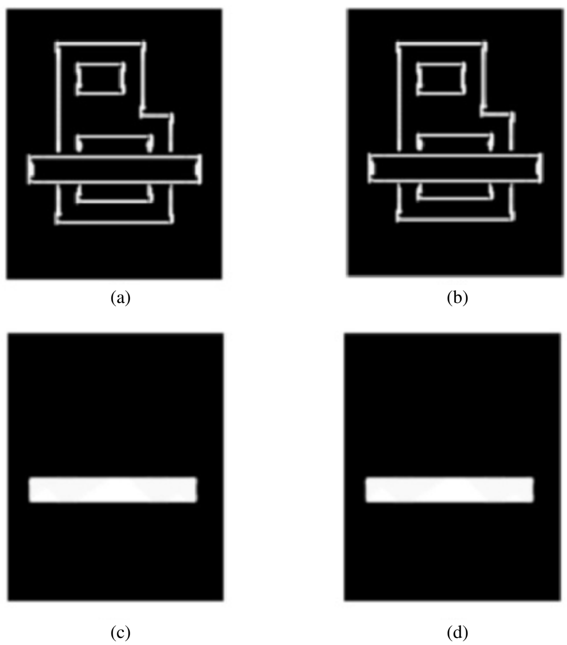

Figure 7. Binocular boundaries for monocular filling-in in response to a Bregman-Kanizsa image: (a) near depth and (b) far depth. Filled-in Monocular FIDOs before boundary pruning occurs: (c) near depth and (d) far depth. [Reprinted with permission from Kelly and Grossberg (2000).]

How do multiple depth-selective BCS copies capture brightness and color signals within depth-selective FCS surface representations? This happens in at least two stages. The first stage of *monocular filling-in domains,* or FIDOs, may exist in V2 thin stripes. Each monocular FIDO is broken into three pairs of opponent filling-in domains (black/white, red/green, blue/yellow) that receive achromatic and chromatic signals from a single eye. A pair of monocular FIDOs, one for each eye, corresponds to each depth-selective BCS copy, and receives its strongest boundarygating signals from this BCS copy. Each monocular FIDO may also receive weaker boundary signals from BCS copies that represent depths near to that of its primary BCS copy. In this way, a finite set of FIDOs can represent a continuous change in perceived depth, much as three classes of retinal cones can be used to represent a continuum of perceived colors.

Surface capture is triggered when boundary-gating BCS signals interact with illuminantdiscounted FCS signals. Pathways 2 in Figure 5 input discounted monocular FCS signals to *all* monocular FIDOs. Only some FIDOs will selectively fill-in these signals, and thereby lift monocular FIDO signals into depth-selective surface representations for filling-in. The boundary signals along pathways 6 in Figure 5 determine which FIDOs will fill-in. These boundary signals selectively capture FCS inputs that are spatially coincident and orientationally aligned with them. Other FCS inputs are suppressed. These properties arise when double-opponent and filling-in processes within the FIDOs interact with oriented boundary-gating signals from the boundary representations. How this happens, and how it can explain data about binocular fusion and rivalry, among other percepts, are discussed in Grossberg (1987).

Because these filled-in surfaces are activated by depth-selective BCS boundaries, they inherit the depths of their boundaries. 3-D surfaces may hereby represent depth as well as brightness and color. This link between depth, brightness, and color helps to explain "proximityluminance covariation," or why brighter surfaces tend to look closer; e.g., Egusa (1983).

Not every filling-in event can generate a visible surface. Because activity spreads until it hits a boundary, only surfaces that are surrounded by a *connected* BCS boundary are effectively filled-in. Otherwise, the spreading activity can dissipate across the FIDO. This property helps to explain data ranging from neon color spreading to how T-junctions influence 3-D figure-ground perception (Grossberg, 1994). Figures 7c and 7d illustrate how filling-in occurs in response to the Bregman-Kanizsa boundaries of Figures 7a and 7b. The connected boundary surrounding the occluder can contain its filled-in activity, but activity spreads through the end-gaps of the B boundaries, thereby dissipating across space, at both near (Figure 7c) and far (Figure 7d) depths.

An analysis of how the BCS and FCS react to 3-D images shows that too many boundary and surface fragments are formed as a result of the size-disparity correlation. This redundancy is clear in Figure 7. As noted above, larger scales can fuse a larger range of disparities than can smaller scales. How are the surface depths that we perceive selected from this range of possibilities across all scales? The FACADE theory answer to this question follows from its answer to the more fundamental question: How is perceptual *consistency* derived from boundary-surface *complementarity?* FACADE theory predicts how this may be achieved by feedback between the boundary and surface streams, that is predicted to occur no later than the interstripes and thin stripes of area V2. This mutual feedback also helps to explain why blob and interblob cells share so many receptive field properties even though they carry out such different tasks. In particular, boundary cells, which summate inputs from both contrast polarities, can also be modulated by surface cells, which are sensitive to just one contrast polarity.

Boundary-surface consistency is realized by a contrast-sensitive process that detects the contours of successfully filled-in regions within the monocular FIDOs. Only successfully filled-in regions *can* activate such a contour-sensitive process, because other regions either do not fill-in at all, or their filling-in dissipates across space. These filled-in contours activate FCS-to-BCS feedback signals (pathways 7 in Figure 5) that strengthen boundaries at their own positions and depths, while inhibiting redundant boundaries at farther depths. Thus the feedback pathway forms an on-center off-surround network whose inhibition is biased towards farther depths. This inhibition from near-to-far is called "boundary pruning." It illustrates a perceptual principle called the "asymmetry between near and far." This principle shows itself in many data, including 3-D neon color spreading (Nakayama et al., 1990). Grossberg (1994, 1999a) discusses how to explain such data.

How does boundary pruning influence figure-ground separation? Boundary pruning spares the closest surface representation that successfully fills-in a region, and inhibits redundant copies of occluding object boundaries that would otherwise form at farther depths. When these redundant occluding boundaries are removed, the boundaries of partially occluded objects can be completed behind them within BCS copies that represent farther depths, as we perceive when viewing Figure 4a but not 4b. Moreover, when the redundant occluding boundaries collapse, the redundant surfaces that they momentarily supported at the monocular FIDOs also collapse. Occluding surfaces hereby form in front of occluded surfaces.

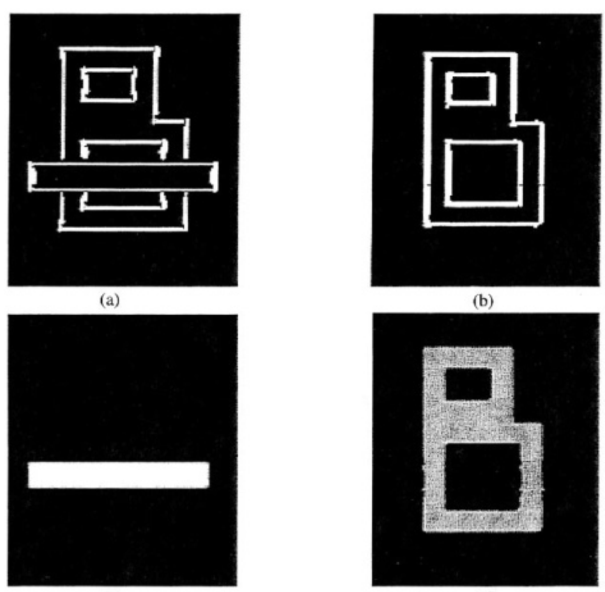

Figure 8. Amodal boundary and surface representations in response to a Bregman-Kanizsa image. Binocular boundaries after boundary pruning occurs: (a) near depth and (b) far depth. Filled-in amodal surface representations at the Monocular FIDOs: (c) near depth and (d) far depth. [Reprinted with permission from Kelly and Grossberg (2000).]

Figures 8a and 8b illustrate boundary pruning and its asymmetric action from near-to-far. The near boundaries in Figure 7a are retained in Figure 8a. But the far boundary of the occluder in Figure 7b is inhibited by boundary pruning signals from the contour of the near filled-in surface representation in Figure 7c. When these occluder boundaries are eliminated, the B boundary can be colinearly completed, as in Figure 8b. Because the boundaries of both the horizontal occluder and the B are now connected, they can contain their filled-in activities within the Monocular FIDOs, as shown in Figures 8c and 8d.

Boundary pruning also helps to explain data about depth/brightness interactions, such as: Why do brighter Kanizsa squares look closer (Bradley and Dumais, 1984)? Why is boundary pruning relevant here? A Kanizsa square's brightness is an emergent property that is determined after *all* brightness and darkness inducers fill-in within the square. This emergent brightness within the FIDOs then influences the square's perceived depth. Within FACADE, this means that the FIDO's brightness influences the BCS copies that control relative depth. This occurs via the BCS-to-FCS feedback signals, including pruning, that ensure boundary-surface consistency (Grossberg, 1997, Section 22).

Visible brightness percepts are not represented within the monocular FIDOs. Model V2 representations of binocular boundaries and monocular filled-in surfaces are predicted to be *amodal,* or perceptually invisible. These representations are predicted to directly activate object recognition (i.e., Thinking!) mechanisms in inferotemporal cortex and beyond, since they accurately represent occluding and occluded objects. In particular, boundary pruning enables boundaries of occluded objects to be completed within the BCS, which makes them easier to recognize, as is illustrated for the Bregman-Kanizsa display in Figure 8. The monocular FIDO surface representations fill-in an occluded object within these completed object boundaries, even behind an opaque occluding object. We can hereby *know* the color of occluded regions without *seeing* them. How, then, do we *see* opaque occluding surfaces? How does the visual cortex generate representations of occluding and occluded objects that can be easily recognized, yet also allow us to consciously see, and reach for, only the unoccluded parts of objects?

FACADE theory proposes that the latter goal is realized at the *binocular FIDOs,* which process a different combination of boundary and surface representations than is found at the monocular FIDOs. The surface representations at the monocular FIDOs are depth-selective, but they do not combine brightness and color signals from both eyes. Binocular combination of brightness and color signals takes place at the binocular FIDOs, which are predicted to exist in cortical area V4. It is here that *modal,* or visible, surface representations occur, and we see only unoccluded parts of occluded objects, except when transparent percepts are generated by special circumstances.

To accomplish binocular surface matching, monocular FCS signals from both eyes (pathways 8 in Figure 5) are binocularly matched at the binocular FIDOs. These matched signals are redundantly represented on multiple FIDOs. The redundant binocular signals are pruned by inhibitory contrast-sensitive signals from the monocular FIDOs (pathways 9 in Figure 5). As in the case of boundary pruning, these *surface pruning* signals arise from surface regions that successfully fill-in within the monocular FIDOs. These signals inhibit the FCS signals at their own positions and farther depths. As a result, occluding objects cannot redundantly fill-in surface representations at multiple depths. Surface pruning is another example of the asymmetry between near and far. Figure 9 illustrates how surface pruning works for the Bregman-Kanizsa image. Figure 9a shows the signals that initiate filling-in at the near Binocular FIDO, and Figure 9b shows them at the far Binocular FIDO. Surface pruning eliminates signals from the occluder in Figure 9b.

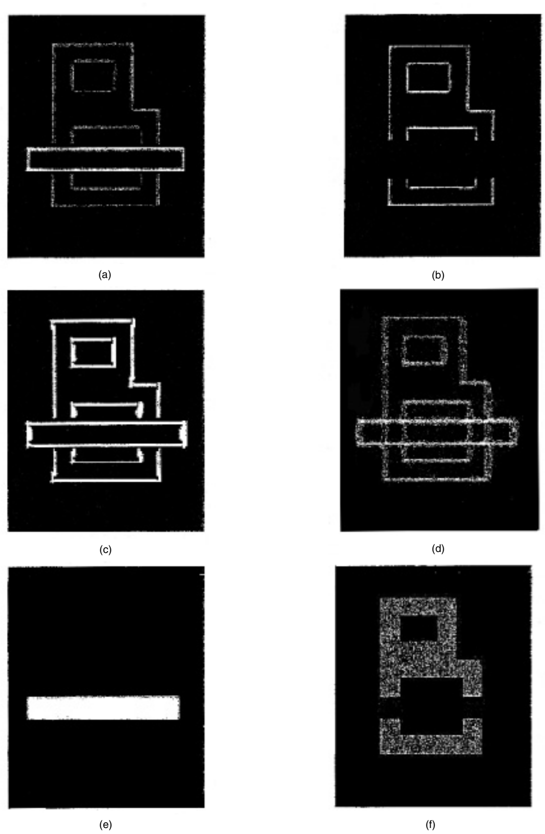

Figure 9. Enriched boundary and modal surface representations. Binocular FIDO filling-in signals at (a) near depth and (b) far depth. Enriched boundaries at the (c) near depth and (d) far depth. Filled-in Binocular FIDO activity consisting of two modal surfaces at two different depths: (e) near depth and (f) far depth. [Reprinted with permission from Kelly and Grossberg (2000).]

As in the monocular FIDOs, FCS signals to the binocular FIDOs can initiate filling-in only where they are spatially coincident and orientationally aligned with BCS boundaries. BCS-to-FCS pathways 10 in Figure 5 carry out depth-selective surface capture of the binocularly matched FCS signals that survive surface pruning. In all, binocular FIDOs fill in FCS signals that: (a) survive within-depth binocular FCS matching and across-depth FCS inhibition; (b) are spatially coincident and orientationally aligned with BCS boundaries; and (c) are surrounded by a connected boundary (web).

One further property completes this summary: At the binocular FIDOs, nearer boundaries are added to FIDOs that represent their own and farther depths. This asymmetry between near and far is called *boundary enrichment.* Enriched boundaries prevent occluding objects from looking transparent by blocking filling-in of occluded objects behind them. The total filled-in surface representation across all binocular FIDOs represents the visible percept. It is called a FACADE representation because it multiplexes the properties of Form-And-Color-And-DEpth that give FACADE theory its name. Figures 9c and 9d show the enriched near and far Binocular FIDO boundaries, respectively, for the Bregman-Kanizsa image. Note the superposition of occluder and occluding boundaries in Figure 9d. Figures 9e and 9f show the filled-in near and far modal surface representations that the surface signals in Figure 9a and 9b cause within the boundaries of Figures 9c and 9d. Note that only the unoccluded surface of the B is “visible” in the Binocular FIDO representation, even though the entire B surface is completed within the amodal Monocular FIDO representation in Figure 8d.

KANIZSA STRATIFICATION

Kanizsa Stratification images (Kanizsa, 1985) can also lead to depthful figure-ground percepts (e.g., Figure 10). Here the percept is one of a square weaving over and under the cross. This image is interesting because a single globally unambiguous figure-ground percept of one object being in front (cross or thin outline square) does not occur. On the left and right arms of the cross in Figure 10, the contrastive vertical black lines are cues that the outline square is in front of the cross arms. The top and bottom regions consist of a homogeneously white figural area, but most observers perceive two figures, the cross arms in front of the thinner outline square. This is usually attributed to the fact that a thinner structure tends to be perceived behind a thicker one most of the time (Petter, 1956; Tommasi et al., 1995). The figure-ground stratification percept is bistable through time, flipping intermittently between alternative cross-in-front and square-in-front percepts. Kanizsa used this sort of percept to argue against the Helmholtz “unconscious inference” account which would not expect interleaving to occur, due to its low probability during normal perceptual experience. FACADE theory uses the same mechanisms as above to explain how perceptual stratification of a homogeneously-colored region occurs, and how the visual system knows which depth to assign the surface color in different parts of the display. Many other percepts have also been explained by using the same small set of concepts and mechanisms.

Figure 10. An example of perceptual stratification. [Reprinted with permission from Kanizsa, (1985).]

An outline of the FACADE explanation is as follows. The thin vertical black lines create T-junctions with the cross. The stems of the T boundaries are broken by bipole feedback, thus separating the thin outline square from the cross (see Figure 11a). At the top and bottom arms of the cross, vertical bipole cells link the sections of the cross arms together, thereby creating a T-junction with the sections of the square. The vertical bipole cells of the cross win out over the horizontal bipole cells of the squares. This happens because the cross is wider than the square. Thus vertical bipole cells have more support from their receptive fields than do the horizontal bipole cells at the cross-square intersection. The boundaries of the square are hereby inhibited, thereby creating end gaps. As a result, the cross arms pop in front and the square is seen behind the cross (Figure 11b and 11c).

Figure 11. (a) Near-depth boundaries in response to the Kanizsa stratification image. Binocular filling-in domain activity at the (b) near depth and (c) far depth.

The bistability of the stratification percept may be explained in the same way that the bistability of the Weisstein effect (Brown and Weisstein, 1988) was explained in Grossberg (1994). This explanation used the habituative transmitters that occur in the pathways 3 between complex cells and hypercomplex cells (Figure 5). Transmitter habituation helps to adapt active pathways and thereby to reset boundary groupings when their inputs shut off. This transmitter mechanism has been used to simulate psychophysical data about visual persistence, aftereffects, residual traces, and metacontrast masking (Francis, 1997; Francis and Grossberg, 1996a, 1996b; Francis, Grossberg, and Mingolla, 1994), developmental data about the self-organization of opponent simple cells, complex cells, and orientation and ocular dominance columns within cortical area V1 (Grunewald and Grossberg, 1998; Olson and Grossberg, 1998), and neurophysiological data about area Vi cells (Abbott, Varela, Sen, and Nelson, 1997). The bistability of the stratification percept can hereby be traced to more basic functional requirements of visual cortex.

CONCLUSION

The present chapter describes how the FACADE theory of 3-D vision and figure-ground perception helps to explain some of the most widely known differences between seeing and thinking. Along the way, the theory provides explanations of the percepts that are generated by some of Kanizsa' s most famous displays. These explanations gain interest from the fact that they reflect fundamental organizational principles of how the brain sees. In particular, they illustrate some of the complementary properties of boundary and surface computations in the interblob and blob cortical processing streams of visual cortex.

These insights lead to a revision of classical views about how visual cortex works. In particular, visual cortex does not consist of independent processing modules. Rather, hierarchical and parallel interactions between the boundary and surface streams synthesize consistent visual percepts from their complementary strengths and weaknesses. Boundaries help to trigger depth-selective surface filling-in, and successfully filled-in surfaces reorganize the global patterning of boundary and surface signals via feedforward and feedback signals. Boundary-gated filling-in plays a key role in surface perception, ranging from lower-level uses, such as recovering surface brightness and color after discounting the illuminant and filling-in the blind spot, to higher-level uses, such as completing depthful modal and amodal surface representations during 3-D vision and figure-ground separation.

Boundary and surface representations activate leamed object representations which, in turn, prime them via top-down modulatory attentional signals. This priming property emphasizes that the visual cortex is not merely a feedforward filter that passively detects visual features, as was proposed by many scientists who thought of the visual brain as a Fourier filter or as a feedforward hierarchy of bottom-up connections that form increasingly complex and large-scale receptive fields. Rather, the visual brain is an integrated system of bottom-up, top-down, and horizontal interactions which actively completes boundary groupings and fills-in surface representations as its emergent perceptual units. This interactive perspective has enabled recent neural models to quantitatively simulate the dynamics of individual cortical cells in laminar cortical circuits *and* the visual percepts that emerge from their circuit interactions. Such results represent a concrete proposal for beginning to solve the classical Mind/Body Problem, and begin to do justice to the exquisite sensitivity of our visual percepts to the scenes and images through which we know the visual world.

Prof. Stephen Grossberg
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215, U.S.A.

REFERENCES

ABBOTT, L.F., VARELA, J.A., SEN, K., and NELSON, S.B. (1997). Synaptic depression and cortical gain control. *Science, 275,* 220-224.

ARRINGTON, K.F. (1994). The temporal dynamics of brightness filling-in. *Vision Research,* 34, 337 1-3387.

BRADLEY, D.R. and DUMAIS, S.T. (1984). The effects of illumination level and retinal size on the depth stratification of subjective contour figures. *Perception,* 13,155-164.

BROWN, J.M. AND WEISSTEIN, N. (1988). A spatial frequency effect on perceived depth. *Perception and Psychophysics,* 44, 157-166.

BROWN, J.M. and WEISSTEIN, N. (1988). A spatial frequency effect on perceived depth. *Perception and Psychophysics,* 43, 53-56.

COHEN, M.A. and GROSSBERG, 5. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics,* 36, 428-456.

DEYOE, E.A. and VAN ESSEN, D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience,* 11, 2 19-226.

DOUGLAS, R.J., KOCH, C., MAHOWALD, M., MARTIN, K.A.C., and SUAREZ, H.H. (1995). Recurrent excitation in neocortical circuits. *Science,* 269, 981-985.

EGUSA, H. (1983). Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual fields. *Perception,* 12, 167-175.

ELDER, J.H. and ZUCKER, S.W. (1998). Evidence for boundary-specific grouping. *Vision Research,* 38, 143-152.

FAHLE, M. and WESTHEIMER, G. (1995). On the time-course of inhibition in the stereoscopic perception of rows of dots. *Vision Research, 35,* 1393-1399.

FIELD, D.J., HAYES, A., and HESS, R.F., (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research,* 33, 173-193.

FRANCIS, G. (1997). Cortical dynamics of lateral inhibition: Metacontrast masking. *Psychological Review,* 104, 572-594.

FRANCIS, G. & GROSSBERG, S. (1996a). Cortical dynamics of boundary segmentation and reset: persistence, afterimages, and residual traces. *Perception, 25,* 543-567.

FRANCIS, G. & GROSSBERG, S. (1996b). Cortical dynamics of form and motion integration: Persistence, apparent motion, and illusory contours. *Vision Research,* 36, 149-173.

FRANCIS, G., GROSSBERG, S., and MINGOLLA, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research,* 34, 1089-1104.

GOGEL, W.C. (1965). Equidistance tendency and its consequences. *Psychological Bulletin,* 64, 153- 163.

GROSSBERG, 5. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52,* 2 17-257.

GROSSBERG, 5. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics,* 23, 187-202.

GROSSBERG, 5. (1980). How does a brain build a cognitive code? *Psychological Review,* 87, 1-5 1.

GROSSBERG, 5. (1984). Outline of a theory of brightness, color, and form perception. In E. Degreef and J. van Buggenhaut (Eds.), Trends in mathematical psychology. Amsterdam: North-Holland.

GROSSBERG, 5. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. *Perception and Psychophysics,* 41, 117-158.

GROSSBERG, 5. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics, 55,* 48-120.

GROSSBERG, 5. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review,* 104, 6 18-658.

GROSSBERG, S., (1999a). A comment on “Assimilation of achromatic color cannot explain the brightness effects of the achromatic neon effect” by Marc K Albert. *Perception,* 28, 129 1-1302.

Grossberg, S. (1999b). How does the, cerebral cortex work? Leaming, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision,* 12, 163-186.

GROSSBERG, S. (1999c). The link between brain learning, attention, and consciousness. *Consciousness and Cognition,* 8, 1-44.

GROSSBERG, S. and HOWE, P.D.L. (2002). A laminar cortical model of stereopsis and 3-D surface perception. Submitted for publication.

GROSSBERG, S. and KELLY, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research,* 39, 3796-38 16.

GROSSBERG, S. and MCLOUGHLIN, N. (1997). Cortical dynamics of 3-D surface perception: Binocular and half-occluded scenic images. *Neural Networks,* 10, 1583-1605.

GROSSBERG, S. and MINGOLLA, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review,* 92, 173-211.

GROSSBERG, S. and MINGOLLA, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics,* 38, 141-17 1.

GROSSBERG, S. and PESSOA, L. (1998). Texture segregation, surface representation, and figure-ground separation. *Vision Research,* 38, 2657-2684.

GROSSBERG, S. and RAIZADA, R.D.S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research,* 40, 14 13-1432.

GROSSBERG, S. and TODOROVIC, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics,* 43, 241-277.

GROSSBERG, S. and WILLIAMSON, J.R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral Cortex,* 11, 37-58.

GRUNEWALD, A. and GROSSBERG, 5. (1998). Self-organization of binocular disparity tuning by reciprocal corticogeniculate interactions. *Journal of Cognitive Neuroscience,* 10, 199-2 15.

HEEGER, D.J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience,* 9, 18 1-197.

HELMHOLTZ, H.L.F. VON (19 10/1925). Treatise on Physiological Optics. Dover Press.

HESS, R.F., DAKIN, S.C., and FIELD, D.J. (1998). The role of “contrast enhancement” in the detection and appearance of visual contours. *Vision Research,* 38, 783-787.

HOWE, P.D.L. and GROSSBERG 5. (2001). Laminar cortical circuits for stereopsis and surface depth perception. *Society for Neuroscience Abstracts,* 27, 164.17.

HUBEL, D.H. and LIVINGSTONE, M.S. (1985). Complex-unoriented cells in a subregion of primate area 18. *Nature,* 315, 325-327.

HUBEL, D.H. and WIESEL, T.N. (1977). Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London (B),* 198, 1-59.

JULESZ, B. and SCHUMER, R.A. (1981). Early visual perception. *Annual Review of Psycholgy,* 32, 572-627.

KANIZSA, G. (1974). Contours without gradients or cognitive contours. *Italian Journal of Psychology,* 1,93-113.

KANIZSA, G. (1985). Seeing and thinking. *Revista di Psicologia,* 49, 7-30. Praeger/Greenwood.

KELLY, F. and GROSSBERG, 5. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception and Psychophysics,* 62, 1596-16 18.

LAMME, V.A.F., RODRIGUEZ-RODRIGUEZ, V., and SPEKREIJSE, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex,* 9, 406-413.

LAND, E.H. (1977). The retinex theory of color vision. *Scientific American,* 237, 108-128.

McLoughlin, N.P and Grossberg, 5. (1998). Cortical computation of stereo disparity. *Vision Research,* 38, 91-99.

Mountcastle, V.B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology,* 20, 408-434.

Murphy, P.C., Duckett, S.G., and Sillito, A.M. (1999). Feedback Connections to the Lateral Geniculate Nucleus and Cortical Response Properties. *Science,* 286, 1552-1554.

Nakayama, K. and Shimojo, 5. (1990). DaVinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research,* 30, 1811-1825.

Nakayama, K., Shimojo, S., and Ramachandran, V.S. (1990). Transparency: Relation to depth, subjective contours, luminance, and neon color spreading. *Perception,* 19, 497-5 13.

Nakayama, K., Shimojo, S., & Silverman, G.H. (1989) Stereoscopic depth: Its relation to image segmentation, grouping and the recognition of occluded objects. *Perception,* 18, 55-68.

Olson, S.J. and Grossberg, 5. (1998). A neural network model for the development of simple and complex cell receptive fields within cortical maps of orientation and ocular dominance. *Neural Networks,* 11, 189-208.

Paradiso, M.A. and Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research,* 31, 1221-1236.

Petry, S. and Meyer, G. (Eds.) (1987). The Perception of Illusory Contours. Springer-Verlag.

Petter, G. (1956). Nuove ricerche sperimentali sulla totalizzazione oercettiva. *Rivista du Psicologia, 50,* 213-227.

Polat, U., Mizobe, K., Pettet, M.W., Kasamatsu, T., and Norcia, A.M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature,* 391, 580-584.

Raizada, R.D.S. and Grossberg, 5. (2001). Context-sensitive binding by the laminar circuits of Vi and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition,* 8, 43 1-466.

Reynolds, J., Chelazzi, L., and Desimone, R., (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience,* 19, 1736-1753.

Roelfsema, P.R., Lamme, V.A.F., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395,* 376-38 1.

Rogers-Ramachandran, D.C. and Ramachandran, V.S. (1998). Psychophysical evidence for boundary and surface systems in human vision. *Vision Research.* 38, 7 1-77.

Shipley, T.F. and Kellman, P.J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception and Psychophysics, 52,* 97-106.

Shipley, T.F. and Kellman, P.J. (1992). Perception of partly occluded objects and illusory figures: Evidence for an identify hypothesis. *Journal of Experimental Psychology: Human Perception and Performance,* 18, 106-120.

Sillito, A.M., Jones, H.E., Gerstein, G.L., and West, D.C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature,* 369, 479-482.

Swaminathan, G. and Grossberg, 5. (2001). Laminar cortical circuits for the perception of slanted and curved 3-D surfaces. *Society for Neuroscience Abstracts,* 27, 619.49.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature,* 381, 520-522.

Tommasi, L., Bressan, P., & Vallortigara, G. (1995) Solving occlusion indeterminacy in chromatically homogeneous patterns. *Perception,* 24, 391-403.

von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). illusory contours and cortical neuron responses. *Science,* 224, 1260-1262.

FUNCTIONAL ARCHITECTURE OF THE VISUAL CORTEX AND VARIATIONAL MODELS FOR KANIZSA'S MODAL SUBJECTIVE CONTOURS

INTRODUCTION

We will present a neuro-geometrical model for generating the *shape* of Kanizsa's modal subjective contours – that we will call in the following *K*-contours. It will be based on the functional architecture of the primary areas of the visual cortex.

As we are interested by a mathematical clarification of some very basic phenomena, we will restrict ourselves to a very small part of the problem, involving only the functional architecture of the first cortical area V1. We will see that the model is already quite sophisticated. Many other aspects (e.g., the role of V2) would have of course to be taken into account in a more complete model.

I. TOWARDS VARIATIONAL MODELS OF KANIZSA'S ILLUSORY CONTOURS

The object under study will not be classical straight *K*-contours but *curved* ones (*K-curves*) where the sides of the internal angles of the pacmen are *not aligned* (see figure 1).

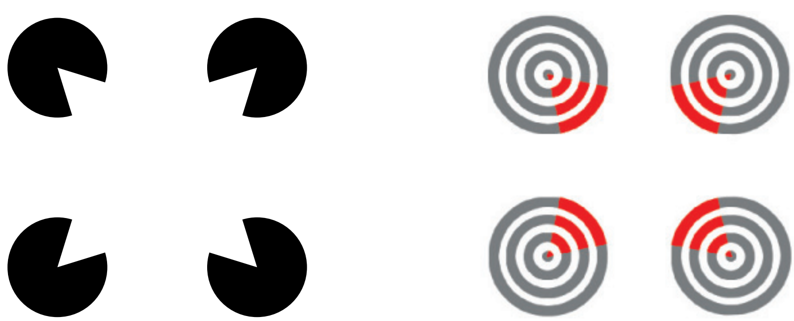

Figure 1. An example of a Kanizsa curved illusory modal contour. The second figure shows the well known "neon effect" (diffusion of color inside the area bounded by the virtual contours).

In an important paper, Shimon Ullman (1976) of the MIT AI Lab introduced the key idea of *variational models.*

"A network with the local property of trying to keep the contours 'as straight as possible' can produce curves possessing the global property of minimizing total curvature."

A. Carsetti (ed.), Seeing, Thinking and Knowing, 55-69.

He was followed by Horn (1983) who introduced a particular type of curves, the curves of least energy. Then in 1992, David Mumford introduced in computer vision, but only for *amodal* contours, a fundamental model based on the physical concept of *elastica*. Elastica are well known in classical Mechanics. They are curves minimizing at the same time the length and the integral of the square of the curvature κ, i.e. the energy

$$E = \int(\alpha\kappa+\beta)^2 ds$$

where ds is the element of arc length along the curve.

We will present here a slightly different variational model, based on the concept of "geodesic curves" in V1 and more realistic at the neural level. Let us begin with some experimental results.

II. AN EXPERIMENT ON *K*-CURVES (WITH JACQUES NINIO)

With our colleague Jacques Ninio of the Ecole Normale Supérieure (Paris) we worked out an experiment aiming at measuring the exact position of the extremum of a *K*-curve. For that purpose, we looked at families of *K*-curves generated by

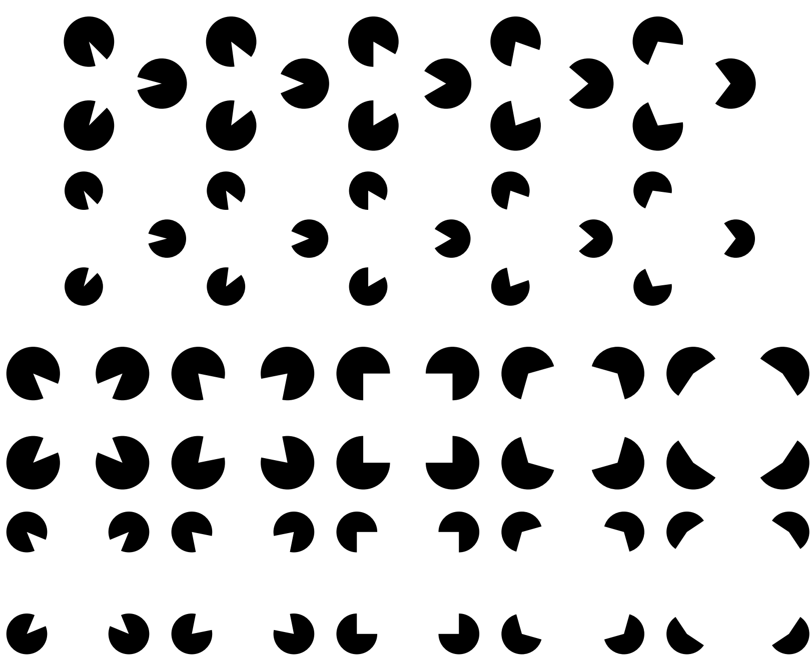

Figure 2. Curved Kanizsa triangles and squares used for the experiment with J. Ninio.

2 configurations: triangle or square;
2 sizes of configuration;
2 sizes of pacmen;
4 orientations;
5 angles (see figure 2).

There are different methods for measuring the extremum of a *K*-contour. For instance, one can use the "subthreshold summation" method: the threshold for the detection of a small segment parallel to the *K*-contour decreases when the segment is exactly located on the *K*-contour (see figure 3).

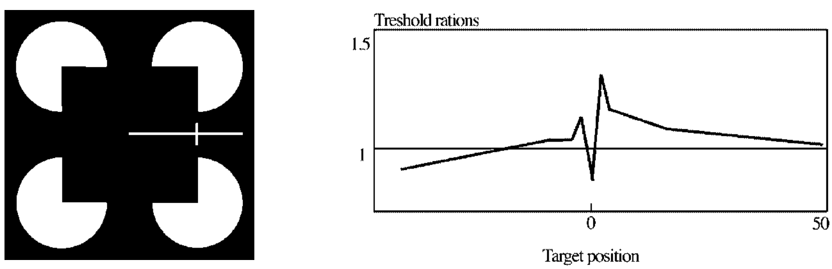

Figure 3. The subthreshold summation method: the threshold for the detection of a small segment parallel to the *K*-contour decreases when the segment is exactly located on the *K*-contour (from Dresp, Bonnet, 1955).

As for us, we used another method for detecting the extremal point of a *K*-curve: the subject was asked to place a marker (the extremity of an orthogonal line, a small segment, the symmetry axis of a small stripe) as exactly as possible at the extremum (see figure 4).

For different cases (triangle / square and small / large pacmen size) we compare three positions:

1. the piecewise linear position (intersection of the corresponding sides of the two pacmen);
2. the position chosen by the subjects;
3. the circle position (extremum of the arc of circle tangent to the sides of the two pacmen).

Let us take for instance the case (see figure 5) of the square with small pacmen (parameter *ps* = pacmen size = 1). The graphics plots the distance *d* of the extremum of the *K*-contour to the center of the configuration as a function of the aperture angle (figure 5b). *d* is measured by its ratio to the piecewise rectilinear case (which corresponds to $d = 1$) (figure 5a). 5 aperture angles are considered:

#2 corresponds to the classical case of a straight *K*-contour ($d_2 = 1$);
#1 to a slightly convex one ($d_1 > d_2 = 1$);
#0 to a more convex one ($d_0 > d_1 > d_2 = 1$);

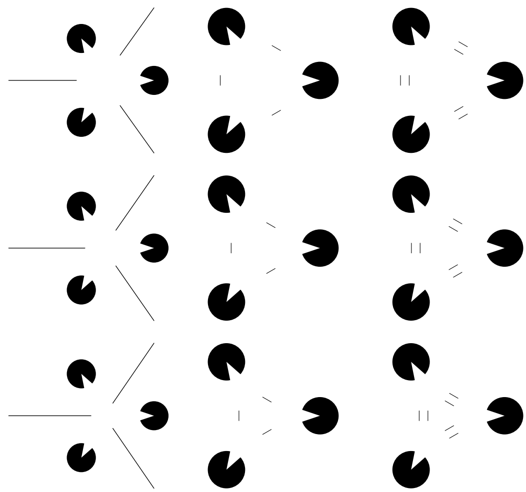

Figure 4. The method of detection of the extremal point of a curved *K*-contour. The subject is asked to place a marker (the extremity of an orthogonal line, a small segment, the symmetry axis of a small stripe) as exactly as possible at the extremum.

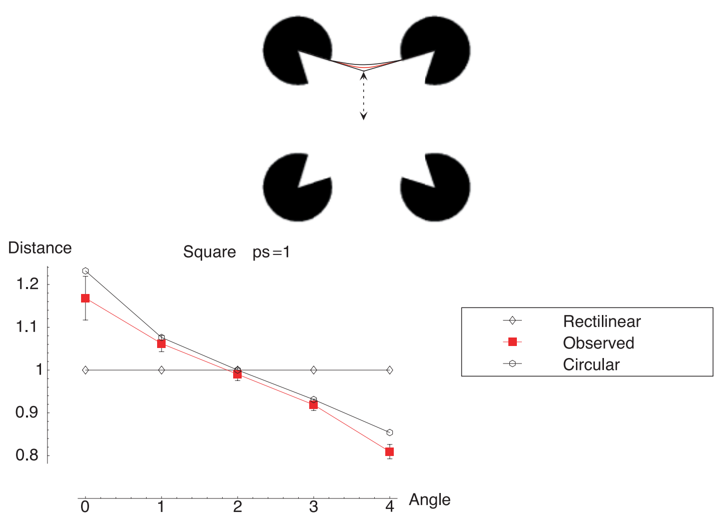

Figure 5. (a) The case of the square with small (parameter *ps* = pacmen size = 1) pacmen. The distance *d* is the distance of the extremum of the *K*-contour to the center of the configuration. *d* is measured by its ratio to the piecewise rectilinear case (which corresponds therefore to *d* = 1). (b) Comparison of three *K*-contours: the piecewise rectilinear one, the one chosen by the subjects, the circle one. The graphics plots the distance *d* as a function of the aperture angle.

#3 to a slightly concave one ($d_3 < d_2 = 1$);
#4 to a more concave one ($d_4 < d_3 < d_2 = 1$).
We see that the observed empirical *K*-contour is located between the piecewise rectilinear one and the circular one, and that the latter is therefore false.

Another important result is that the deflections for the triangle and for the square are *not* the same. This is a typical *global* effect (see figure 6) which is very interesting but won't be taken into account here.

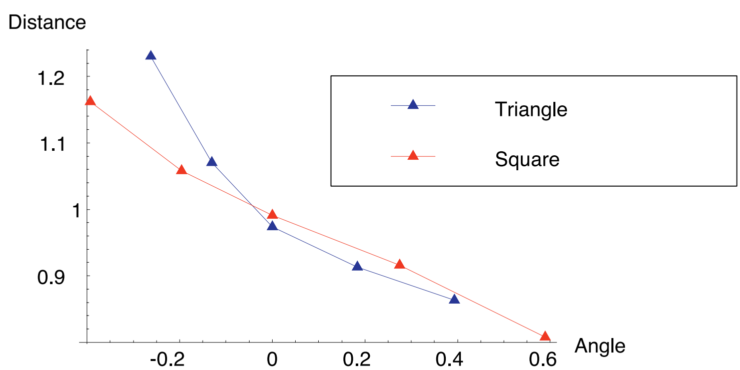

Figure 6. The deflections for the triangle and for the square are *not* the same.

III. NEURAL FUNCTIONAL ARCHITECTURE

We want now to work out a neurally plausible model of *K*-curves at the V1 level. We need first some results concerning the functional architecture of V1.

It is well known (see for instance De Angelis *et al.* 1995) that the receptive profile (the transfert function) of simple cells of V1 are like third order derivatives of Gaussians (with a well described underlying neural circuitry, from lateral geniculate body to layers of V1) (see figure 7).

Such receptive profiles act on the signal as filters by convolution and process a *wavelet analysis*.

Due to their structure, the receptive fields of simple cells detect a *preferential orientation*. Simplifying the situation, we can say they detect pairs (a, p) of a spatial (retinal) position a and a local orientation p at a. They are organized in small modules called *hypercolumns* (Hubel and Wiesel) associating retinotopically to each position a of the retina R a full exemplar P_a of the space of orientations p at a. A very simplified schema of

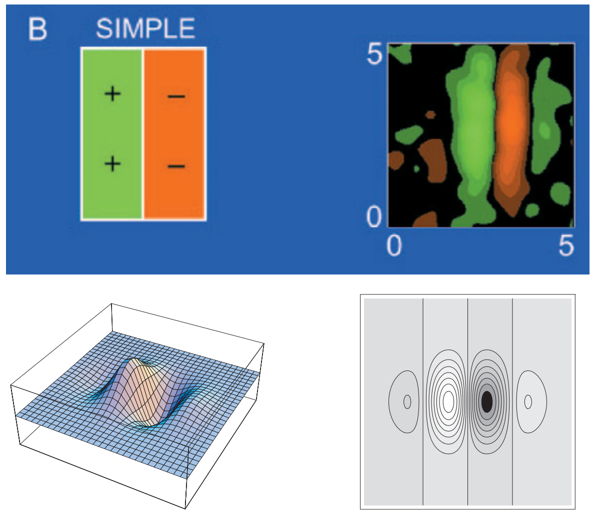

Figure 7. (a) Level curves of the receptive profile of a simple cell of V1 (right) and a simple schema of the receptive field (left) (from De Angelis *et al*. 1995). (b) A third derivative of a Gaussian along a direction. (c) The level curves. They fit very well with the empirical data (a).

this structure (with a 1-dimensional base R) is shown at the figure 8. It is called a *fibration* of base R and fiber P.

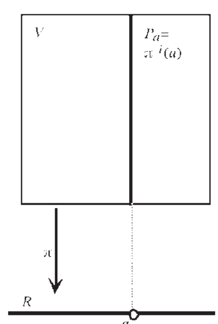

Figure 8. The general schema of a fibration with base space R, fiber P and total space V. The projection π projects V onto R and above every point a of R the fiber P_a is isomorphic to P.

Pairs (a, p) are called in geometry *contact elements*. But, beyond retinotopy formalized by the projection π, their set $V = \{(a, p)\}$ need to be *strongly structured* to allow the visual cortex to compute contour integration. We meet here the problem of the *functional architecture* of V1.

Recent experiments have shown that hypercolumns are geometrically organized in *pinwheels*. The cortical layer is reticulated by a network of singular points which are the centers of the pinwheels, around these singular points all the orientations are distributed along the rays of a "wheel", and the wheels are glued together in a global structure. The experimental method is that of in vivo optical imaging based on activity-dependent intrinsic signals (Bonhöffer & Grinvald, 1991) which allows to acquire images of the activity of the superficial cortical layers. Gratings with high contrast and different (e.g., 8) orientations are presented many times (20-80) with, e.g., a width of 6.25° for the dark strips and of 1.25° for the light ones, and velocity of 22.5°/s. A window is then opened above V1 and the cortex is illuminated with an orange light. One sums the images of V1's activity for the different gratings, constructs differential maps, and eliminates the low frequency noise. The maps are then normalized (by dividing the deviation relative to the mean value at each pixel by the global mean deviation) and the orientations are coded by colors (iso-orientation lines become therefore iso-chromatic lines).

In the celebrated figure 9 due to William Bosking, one can identify three classes of points:

1. regular points, where the orientation field is locally trivial;
2. singular points at the center of the pinwheels, where a full set of iso-orientation lines converge, two adjacent singular points being of opposed chiralities;
3. saddle-points at the center of the cells defined by the singular points.

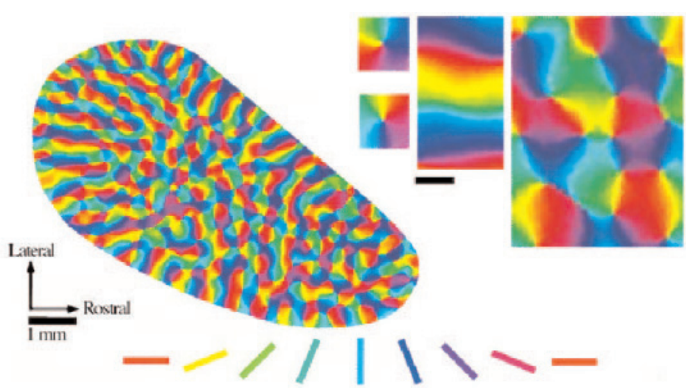

Figure 9. The pinwheel structure of V1 for a tree shrew. The different orientations are coded by colors. Examples of regular points and singularities of opposed chiralities are zoomed in. (From Bosking et al. 1997).

A cristal-like model of such a network of pinwheels is shown at the figure 10.

As we have already noticed, the fonctional architecture associating retinotopically to each position a of the retina R an exemplar P_a of the space of the orientations implements a well known geometrical structure, namely the *fibration* $\pi : R \times P \rightarrow R$ with base R and fiber P. But such a "vertical" structure idealizing the retinotopic mapping between R and P is definitely insufficient. To implement a *global* coherence, the system must be able to

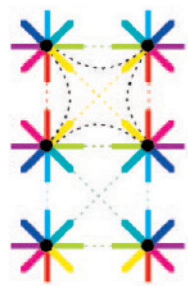

Figure 10. An idealized "cristal-like" model of pinwheels centered on a regular lattice of singular points. Some iso-orientations lines are represented. The saddle points in the centers of the domains are well visible.

compare between them two retinotopically neighboring fibers P_a et P_b over two neighboring retinal points a and b. This is a problem of *parallel transport* whose simplified schema is shown at the figure 11 (to be compared with figure 8). It has been solved experimentally by the discovery of *"horizontal" cortico-cortical connections*.

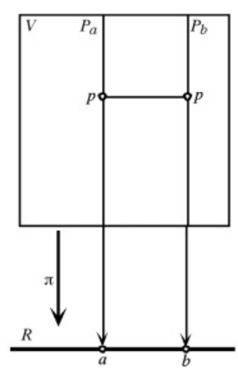

Figure 11. Cortico-cortical horizontal connections allow the system to compare orientations in two different hypercolumns corresponding to two different retinal positions a and b.

Experiments show that cortico-cortical connections connect neurons of the *same* orientation in neighboring hypercolumns. This means that the system is able to know, for b near a, if the orientation p at a is the same as the orientation q at b. The retino-geniculo-cortical "vertical" connections give an internal meaning to relations between contact elements (a, p) and (a, q) (*different* orientations p and q at the *same* point a) while the "horizontal" cortico-cortical connections give an internal meaning to relations between contact elements (a, p) and (b, p) (*same* orientation p at *different* points a and b) (see figure 12).

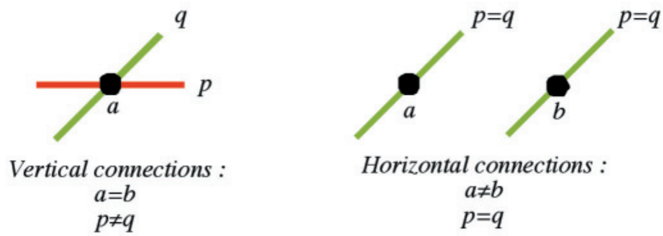

Figure 12. While the retino-geniculo-cortical "vertical'" connections give a meaning to the relations between pairs (a, p) and (a, q) (different orientations p and q at the same point a), the "horizontal'" cortico-cortical connections give a meaning to the relations between pairs (a, p) and (b, p) (same orientation p at different points a and b).

Moreover, as is schematized in figure 13, cortico-cortical connections connect not oly parallel but also *coaxial* orientation cells, that is neurons coding contact elements (a, p) and (b, p) *such that p is the orientation of the axis ab*.

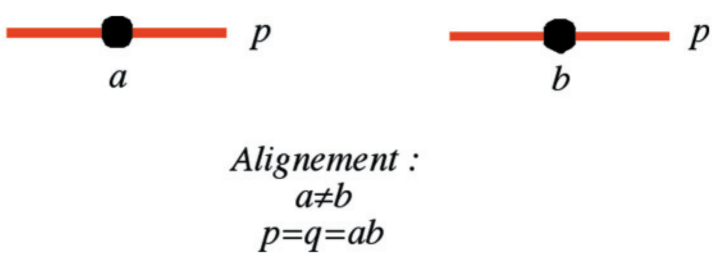

Figure 13. Coaxial orientation cells.

As emphasizes William Bosking (1997):

"The system of long-range horizontal connections can be summarized as preferentially linking neurons with co-oriented, co-axially aligned receptive fields".

We will now show that these results mean that what geometers call the *contact structure* of the fibration $\pi : R\times P\to R$ is neurologically implemented.

IV. THE CONTACT STRUCTURE OF V1 AND THE ASSOCIATION FIELD

We work in the fibration $\pi : V = R\times P\to R$ with base space R and fiber P = set of orientations p. V is an idealized model of the functional architecture of V1. Mathematically, π can be interpreted as the fibration $R\times\mathbf{P}^1$ ($P = \mathbf{P}^1$ = projective line of orientations), or as the fibration $R \times \mathbf{S}^1$ ($P = \mathbf{S}^1$ = unit circle of the orientation angles θ), or as the space $R \times \mathbf{R}$ of *1-jets* of curves *C* in *R* ($P = \mathbf{R}$ = the real line of the tangent $p = \tan(\theta)$ of the orientation angles). In the following, we will use the later model. A coordinate system for *V* is therefore given by triplets (x, y, p) where $p = \tan(\theta)$.

If *C* is a regular curve in *R* (a contour), it can be *lifted* to *V* through the map $\Gamma: C \rightarrow V = R \times P$ wich associates to every point *a* of *C* the contact element (a, p_a) where p_a is the *tangent* of *C* at *a*. *Γ* represents *C* as the *enveloppe* of its tangents (see figure 14).

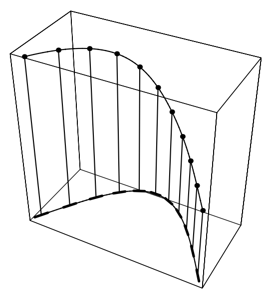

Figure 14. The lifting of a curve *Γ*, $y = f(x)$, in the base space *R* to the space *V*. Above every point $(x, y = f(x))$ of *Γ* we take the tangent direction $p = f'(x)$.

If $a(s)$ is a parametrization of *C*, we have $p_a = a(\ s)$ (where a' symbolizes the derivative $y'(s)/x'(s)$) and therefore $\Gamma = (a(s), p(s)) = (a(s), a(\ s))$. If $y = f(x)$ is a (local) equation of *C*, then a (local) equation of the lifting *Γ* in *V* is $(x, y, p) = (x, y, y')$.

To every curve *C* in *R* is associated a curve *Γ* in *V*. *But the converse is definitely false*. Indeed, let $\Gamma = (a(s), p(s))$ be a (parametrized) curve in *V*. The projection $a(s)$ of *Γ* is of course a curve *C* in *R*. But in general *Γ will not be* the lifting of its projection *C*. *Γ* will be the lifting of $C = \pi(\Gamma)$ *iff* $p(s) = a(\ s)$. In differential geometry, this condition is called a Frobenius *integrability condition*. Technically, it says that to be a *coherent* curve in *V*, *Γ* must be an *integral curve* of the contact structure of the fibration π. We show in figure 15, besides the integrable example of figure 14, three exemples of *non* integrable curves *Γ* which are not the lifting of their projection *C*.

Geometrically, the integrability condition means the following. Let $t = (x, y, p; 1, y', p')$ be a tangent vector to *V* at the point $(a, p) = (x, y, p)$. If $p = y'$ we get $t = (x, y, p; 1, p, p')$. It is easy to show that this condition means exactly that *t* is in the *kernel* of the 1-form $\omega = dy - pdx$. This kernel is a plane, called the *contact plane* of *V* at (a, p), and the integrability condition for a curve *Γ* in *V* says exactly that *Γ* is tangent at each of its point (a, p) to the contact plane at that point. It is in that sense that *Γ* is an *integral curve* of the contact structure of *V*.

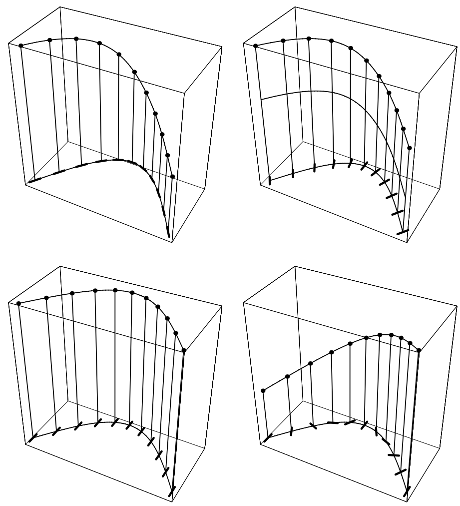

Figure 15. The association field as a condition of integrability. (a) The integrability condition is satisfied. (b), (c), (d) the condition is not satisfied. In (b) we add a constant angle to the tangent (i.e. $p = f'(x) + p_0$). In (c) p is constant while f' is not. In (d) p rotates faster than f'.

The integrability condition is a version of the Gestalt principle of "good continuation". Its psychophysical counterpart has been experimentally analyzed by David Field, Anthony Hayes and Robert Hess (1993) and explained using the concept of *association field*. Let (a_i, p_i) be a sequence of contact elements embedded in a background of distractors. The authors show that they generate a perceptively salient curve (pop-out) iff the p_i are *tangent* to the curve interpolating the a_i. This is due to the fact that the activation of a simple cell detecting a contact element (a, p) *pre*-activates via the horizontal cortico-cortical connections, cells detecting contact elements (c, q) with c roughly aligned with a in the direction p and q close to p. The preactivation is strongly enhanced if the cell (c, q) is sandwiched between a cell (a, p) and a cell (b, p) (see figure 16).

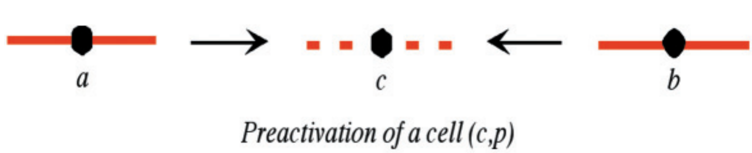

Figure 16

The pop-out of the curve generated by the (a_i, p_i) is a typical Gestalt phenomenon which results from a *binding* induced by the co-activation. It manifests a very particular type of grouping. As was emphasized by Field, Hayes, and Hess (1993):

"Elements are associated according to joint constraints of position and orientation." (p. 187)
"The orientation of the elements is locked to the orientation of the path; a smooth curve passing through the long axis can be drawn between any two successive elements." (p. 181)

This is clearly a discrete version of the integrability condition.

V. A VARIATIONAL MODEL OF MODAL KANIZSA ILLUSORY CONTOURS

In such a framework, we can solve some aspects of the Kanizsa problem in a *principled* way. Two pacmen of respective centers a and b with a specific aperture angle define two contact elements $A = (a, p)$ and $B = (b, q)$ of V. A K-curve interpolating between A and B is

1. a curve C from a to b in R with tangent p at a and tangent q at b;
2. a curve minimizing a sort of "energy" (variational problem).

But as far as the integration of C is processed in V, we must lift the problem to V. We must therefore find *in* V a curve Γ interpolating between (a, p) and (b, q) *in* V, and wich is at the same time:

1. "as straight as possible", that is "geodesic" in V;
2. an integral curve of the contact structure.

In general Γ will not be a straight line because it will have to satisfy the integrability condition. It will be "geodesic" *only in the class of integral curves*.

Mathematically, the problem is not trivial. We have to solve constrained Euler-Lagrange equations in the jet space V. We must first define appropriate Lagrangians on V based on Riemannian metrics which reflect the weakening of the horizontal cortico-cortical connections when the discrepancy between the boundary values θ_A and θ_B increases. If the angle θ is measured relatively to the axis AB (θ has therefore an *intrinsic* geometric meaning), the weakening must vanish for $\theta = 0$ and $\theta = \pi$ and diverge for $\theta = \pi/2$. The function $p = f' = \tan\theta$ being the simplest function sharing this properties, it seems justified to test first the Euclidean metric of V. We will use a frame Oxy of R where the x-axis is identified with AB. The invariance under a change of frame is expressed by the action of the Euclidean group $E(2)$ on V.

We look therefore for curves of minimal length in V *among those which lift curves in* R, that is which satisfy the Frobenius integrability condition and are integrals of the contact structure. We will call them "Legendrian geodesics". Let $(x, y, p; \xi, \eta, \pi)$ be coordinates in the tangent space TV of V. We have to minimize the length of Γ expressed by the fonctional $\int_{x_A}^{x_B} ds$ where the element of arc length ds is given by $ds^2 = dx^2 + dy^2 + dp^2$.

The energy to minimize is therefore $E = \int_{x_A}^{x_B} L(x)dx$ where the Lagrangien L is given, for a curve Γ of the form $(x, y = f(x), p = f'(x))$, by the formula $L(x)dx = ds$, that is by:

$$L(x) = \sqrt{\xi^2 + \eta^2 + \pi^2} = \sqrt{1 + f'(x)^2 + f''(x)^2}$$

We have to solve the Euler-Lagrange (E-L) equations constrained by the integrability constraint $p = f'(x)$, i.e. $\Sigma = 0$ with $\Sigma = p - \eta$. These constrained E-L equations are :

$$\begin{cases} \left(\dfrac{\partial}{\partial y} - \dfrac{d}{dx}\dfrac{\partial}{\partial \eta}\right)(L + \lambda\Sigma) = 0 \\ \left(\dfrac{\partial}{\partial p} - \dfrac{d}{dx}\dfrac{\partial}{\partial \pi}\right)(L + \lambda\Sigma) = 0 \end{cases}$$

where $\lambda(x)$ is a function, called a *Lagrange multiplier*. The idea is that the E-L equations with the constraint $\Sigma = 0$ are the same as the non constrained E-L equations for the modified Lagrangian $L + \lambda\Sigma$.

After some tedious computations, we get the following differential equations for the function $p = f'$, where c and d are two integration constants:

$$1 + p(x)^2 = (cp(x) + d)\sqrt{1+p(x)^2 + p'(x)^2}$$

As the solution is given by an *elliptic integral,* Legendrian geodesics are integrals of elliptic functions.

We can greatly simplify the solution of the equation when the function f is *even*, and the curve Γ *symmetric* under the symmetry $x \to -x$. Indeed, this condition implies immediately $c = 0$, whence, putting $k = 1/d$, the simpler differential equation for $p = f'$:

$$(p')^2 = (1+p^2)[k^2(1+p^2) - 1]$$

The parameter k is correlated to curvature: $k^2 - 1 = \kappa(0)^2$.

We get therefore:

$$x = \text{cst} + \int_0^{p(x)} \frac{1}{\sqrt{(1+t^2)\left(1 + \dfrac{k^2}{k^2-1}t^2\right)}}dt$$

which is a well known elliptic integral of the first kind.

We show in figure 17 how the solution f evolves when k varies from 1 to 1.65 by steps of 0.5.

Figure 18 shows in the base space R and in the jet space V how the Legendrian geodesic corresponding to $k = 1.5$ is situated relatively to the arc of circle, the arc of parabola and the piecewise linear solution defined by the same boundary conditions.

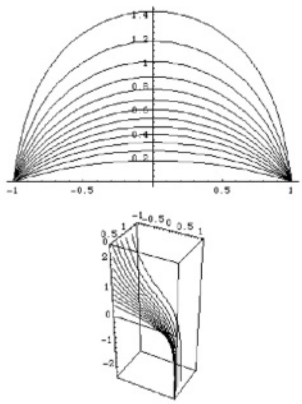

Figure 17. Evolution of Legendrian geodesics when the boundary tangents become more and more vertical. (a) Curves C in the base space R. (b) Curves Γ in the jet space V.

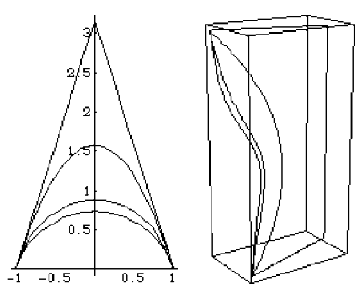

Figure 18. Position of the Legendrian geodesic ($k = 1.5$) relatively to the arc of circle, the arc of parabola and the piecewise linear solution defined by the same boundary conditions. (a) In the base space R. (b) In the jet space V.

The following table shows that the geodesic minimizes the length:

Curves	Geodesic	Arc of circle	Arc of parabola	Piecewise linear
Length	7.02277	7.04481	7.50298	12.9054

CONCLUSION

Due to the very strong geometrical structure of the functional architecture (hypercolumns, pinwheels, horizontal connections), the neural implementation of Kanizsa's contours is deeply linked with sophisticated structures belonging to what is called contact geometry and with variational models analogue to models already well known in physics.

Prof. Jean Petitot
CREA
1 rue Descartes
75005, Paris, France

REFERENCES

BONHÖFFER, T., GRINVALD, A., 1991: "Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns", *Nature*, 353, 429-431.

BOSKING, W., ZHANG, Y., SCHOFIELD, B., FITZPATRICK, D., 1997: "Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex", *J. of Neuroscience*, 17, 6, 2112-2127.

DE ANGELIS, G., OHZAWA, I., FREEMAN, R., 1995: "Receptive Fields Dynamics in the Central Visual Pathways", *Trends in Neuroscience*, 18, 451-458.

DRESP, B., BONNET, C., 1995: "Subthreshold summation with illusory contours", *Vision Research*, 35, 1071-1078.

FIELD, D., HAYES, A., HESS, R., 1993: "Contour Integration by the Human Visual System: Evidence for a Local ' Association Field '", *Vision Research*, 33, 2, 173-193.

HORN, B. K. P., 1983: "The curve of least energy", *ACM Transactions in Mathematical Software*, 9, 4, 441-460.

MUMFORD, D., 1992: "Elastica and Computer Vision", *Algebraic Geometry and Applications*, Springer.

PETITOT, J., TONDUT, Y., 1999: "Vers une neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux", *Mathématiques, Informatique et Sciences Humaines*, 145, 5-101.

SPILLMAN, L., DRESP, B., 1995: "Phenomena of illusory form: can we bridge the gap between levels of explanation?", *Perception*, 24, 1333-1364.

ULLMAN, S., 1976: "Filling-in the gaps: the shape of subjective contours and a model for their generation", *Biological Cybernetics*, 25, 1-6.

This page intentionally left blank

AGNÈS DESOLNEUX, LIONEL MOISAN AND JEAN-MICHEL MOREL

GESTALT THEORY AND COMPUTER VISION

1. INTRODUCTION

The geometric Gestalt theory started in 1921 with Max Wertheimer's founding paper [31]. In its 1975 last edition, the Gestalt Bible *Gesetze des Sehens* by Wolfgang Metzger [22] gave a broad overview of the results of a fifty years research. At about the same date, Computer Vision was an emerging new discipline, at the meeting point of Artificial Intelligence and Robotics. The foundation of signal sampling theory by Claude Shannon [28] was already twenty years old, but computers were able to deal with images with some efficiency only at the beginning of the seventies. Two things are noticeable:

• Computer Vision did not use at first the Gestalt theory results: the founding book of David Marr [21] involves much more neurophysiology than phenomenology. Also, its programme and the robotics programme [12] founded their hopes on binocular stereo vision. This was in total contradiction to, or ignorance of the results explained at length in Metzger's chapters on *Tiefensehen*. Indeed, these chapters demonstrate that binocular stereo vision is a *parent pauvre* in human depth perception.

• Conversely, Shannon's information theory does not seem to have influenced Gestalt research, as far as we can judge from Kanizsa's and Metzger's books. The only bright exception is Attneave's attempt [2] to give a shape sampling theory adapted to shape perception.

This lack of initial interaction is surprising. Indeed, both disciplines have attempted to answer the following question: how to arrive at global percepts (let them be visual objects or gestalts[1]) from the local, atomic information contained in an image?

In this paper, we shall propose an analysis of the Wertheimer programme adapted to computational issues. We shall distinguish two kinds of laws:

- the practical grouping laws (like vicinity or similarity), whose aim it is to build up *partial gestalts*,

- the gestalt principles like masking or *articulazione senza resti*, whose aim it is to operate a synthesis between the partial groups obtained by the elementary grouping laws.

We shall review some recent methods proposed by the authors of the present paper in the computation of partial gestalts (groups obtained by a single grouping law). These results show that

• there is a simple computational principle (the so-called Helmholtz principle), inspired from Kanizsa's masking by texture, which allows one to compute any partial gestalt ob-

A. Carsetti (ed.), Seeing, Thinking and Knowing, 71-101.

tainable by a grouping law (section 4). Also, a particular use of the *articulazione senza resti*, which we call maximality, yields optimal partial gestalts;

• this computational principle can be applied to a fairly wide series of examples of partial gestalts, namely alignments, clusters, boundaries, grouping by orientation, size or grey level;

• the experiments yield evidence that in natural world images, partial gestalts often collaborate.

We push one of the experiments to prove that the partial gestalt recursive building up can be led up to the third level (gestalts built by three successive grouping principles). In contrast, we also show by numerical counterexamples that all partial gestalts are likely to lead to wrong scene interpretations. As we shall see, the wrong detections are always explainable by a conflict between gestalts. We eventually show some experiment suggesting that Minimal Description Length principles [26] may be adequate to resolve some of the conflicts between gestalt laws.

Our plan is as follows. We start in section 2 with an account of Gestalt theory, centered on the initial 1923 Wertheimer programme. In section 3, we focus on the problems raised by the synthesis of groups obtained by partial grouping laws: we address the conflicts between these laws and the masking phenomenon, which we discuss in the line of Kanizsa. In section 4, we point out several quantitative aspects implicit in Kanizsa's definition of masking and show that one particular kind of masking, Kanizsa's *masking by texture*, suggests computational procedures. Such computational procedures are explained in section 5. We end this paper in section 6 by the discussion of a list of numerical experiments on digital images.

2. GROUPING LAWS AND GESTALT PRINCIPLES

2.1 GROUPING LAWS

Gestalt theory starts with the assumption of active "grouping" laws in visual perception (see [14], [31]). These groups are identifiable with subsets of the retina. We shall talk in the following of "points" or groups of points which we identify with spatial parts of the planar rough percept. In image analysis, we shall identify them as well with the points of the digital image. Whenever points (or previously formed groups) have one or several characteristics in common, they get grouped and form a new, larger visual object, a gestalt. The list of elementary grouping laws given by Gaetano Kanizsa in *Grammatica del Vedere* page 45 and following *[14]* is *vicinanza, somiglianza, continuita di direzione, completamento amodale, chiusura, larghezza constante, tendenza alla convessita, simmetria, movimento solidale, esperienza passata*, that is: vicinity, similarity, continuity of direction, amodal completion, closure, constant width, tendency to convexity, symmetry, common motion, past experience. This list is actually very close to the list of grouping laws considered in the founding paper of Wertheimer [31]. These laws are supposed to be at work for every new percept. The amodal completion, one of the main subjects of Kanizsa's books, is, from the geometric viewpoint, a variant of the good continuation law[2]. Figure 1 illustrates many of the grouping laws stated above. The subjects asked to describe briefly such a figure give an account of it as “three letters X” built in different ways.

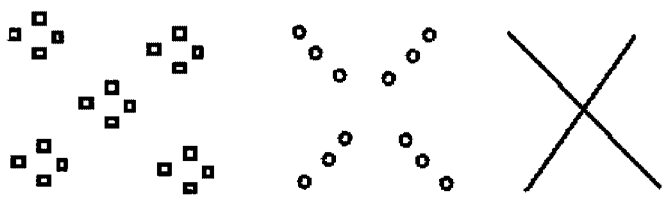

Figure 1. The building up of gestalt: *X-shapes. Each one is built up with branches which are themselves groups of similar objects; the objects, rectangles or circles are complex gestalts, since they combine color constancy, constant width, convexity, parallelism, past experience etc.*

Most grouping laws stated above work from local to global. They are of mathematical nature, but must actually be split into more specific grouping laws to receive a mathematical and computational treatment:

• *Vicinity* for instance can mean: connectedness (*i.e.* spots glued together) or clusters (spots or objects which are close enough to each other and apart enough from the rest to build a group). This vicinity gestalt is at work in all sub-figures of figure 2.

• *similarity* can mean: similarity of color, shape, texture, orientation,... Each one of these gestalt laws is very important by itself (see again figure 2).

• *continuity of direction* can be applied to an array of objects (figure 2 again). Let us add to it alignments as a grouping law by itself (constancy of direction instead of continuity of direction).

• *constant width* is also illustrated in the same figure 2 and is very relevant for drawings and all kinds of natural and artificial forms.

• Notice, in the same spirit, that *convexity*, also illustrated, is a particularization of both closure and continuity of direction.

• *past experience:* In the list of partial gestalts which are looked for in any image, we can have generic shapes like circles, ellipses, rectangles, and also silhouettes of familiar objects like faces, cats, chairs, etc.

Figure 2. Illustration of gestalt principles. *From left to right and top to bottom: color constancy + proximity, similarity of shape and similarity of texture; good continuation; closure (of a curve); convexity; parallelism; amodal completion (a disk seen behind the square); color constancy; good continuation again (dots building a curve); closure (of a curve made of dots); modal completion: we tend to see a square in the last figure and its sides are seen in a modal way (subjective contour). Notice also the texture similarity of the first and last figures. Most of the figures involve constant width. In this complex figure, the sub-figures are identified by their alignment in two rows and their size similarity.*

All of the above grouping laws belong, according to Kanizsa, to the so called *processo primario* (primary process), opposed to a more cognitive secondary process. Also, it may of course be asked *why and how* this list of geometric qualities has emerged in the course of biological evolution. Brunswick and Kamiya [4] were among the first to suggest that the gestalt grouping laws were directly related to the geometric statistics of the natural world. Since then, several works have addressed from different points of views these statistics and the building elements which should be conceptually considered in perception theory, and/or numerically used in Computer Vision [3], [25], [10].

The grouping laws usually collaborate to the building up of larger and larger objects. A simple object like a square, whose boundary has been drawn in black with a pencil on a white sheet, will be perceived by connectedness (the boundary is a black line), by constant width (of the stroke), convexity and closure (of the black pencil stroke), parallelism (between opposite sides), orthogonality (between adjacent sides), again constant width (of both pairs of opposite sides).

We must therefore distinguish between *global* gestalt and the *partial* gestalts. The square is a global gestalt, but it is the synthesis of a long list of concurring local groupings, leading to parts of the square endowed with some gestalt quality. Such parts we shall call *partial gestalts*.

Notice also that all grouping gestalt laws are *recursive*: they can be applied first to atomic inputs and then in the same way to partial gestalts already constituted. Let us illustrate this by an example. In figure 3, the same partial gestalt laws, namely alignment, parallelism, constant width and proximity, are recursively applied not less than six times: the single elongated dots first aligned in rows, these rows in groups of two parallel rows, these groups again in groups of five parallel horizontal bars, these groups again in groups of six parallel vertical bars. The final groups appear to be again made of two macroscopic horizontal bars. The whole organization of such figures is seeable at once.

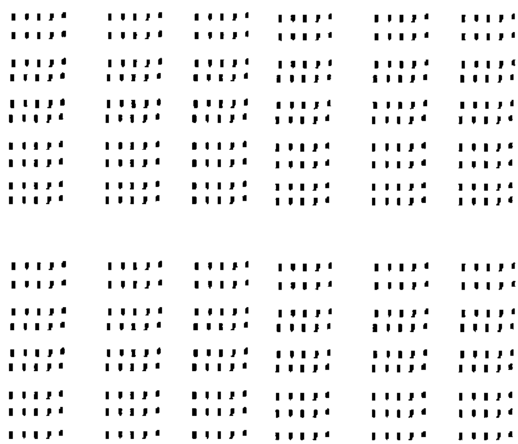

Figure 3. Recursiveness of gestalt laws: *here, constant width and parallelism are applied at different levels in the building up of the final group not less than six times, from the smallest bricks which are actually complex gestalts, being roughly rectangles, up to the final rectangle. Many objects can present deeper and more complex constructions.*

2.2 Global Gestalt principles

While the partial, recursive, grouping gestalt laws do not bring so much doubt about their definition as a computational task from atomic data, the global gestalt principles are by far more challenging. For many of them, we do not even know whether they are properly constitutive laws or rather an elegant way of summarizing various perception processes. They constitute, however, the only cues we have about the way the partial gestalt laws could be derived from a more general principle. On the other hand, these principles are absolutely necessary in the description of the perception process, since they should fix the way grouping laws interact or compete to create the final global percepts, the final gestalts. Let us go on with the gestalt principles list which can be extracted from [14]. We have:

raggruppamento secondo la direzionalità della struttura (Kanizsa, *Grammatica del Vedere*, op. cit., page 54): inheritance by the parts of the overall group direction. This is a statement which might find its place in Platon's Parmenides: “the parts inherit the whole's qualities”.

pregnancy, structural coherence, unity *(pregnanza, coerenza strutturale, carattere unitario, ibidem, page 59),* tendency to maximal regularity (ibidem, p. 60), articulation whole-parts, (in German, *Gliederung),* articulation without remainder (ibidem p. 65): These seven Gestalt laws are not partial gestalts; in order to deal with them from the computer vision viewpoint, one has to assume that all partial grouping laws have been applied and that a synthesis of the groups into the final global gestalts must be thereafter performed. Each principle describes some aspect of the synthesis made from partial grouping laws into the most wholesome, coherent, complete and well-articulated percept.

3. CONFLICTS OF PARTIAL GESTALTS AND THE MASKING PHENOMENON

With the computational discussion to come in mind, we wish to examine the relationship between two important technical terms of Gestalt theory, namely *conflicts* and *masking*.

3.1 CONFLICTS

The gestalt laws are stated as independent grouping laws. Now, they start from the same building elements. Thus, conflicts between grouping laws can occur and therefore conflicts between different interpretations, that is, different possible groups in a given figure. Three cases:

a) two grouping laws act simultaneously on the same elements and give raise to two overlapping groups. It is not difficult to build figures where this occurs, as in figure 4. In that example, we can group the black dots and the white dots by similarity of color. All the same, we see a rectangular grid made of all the black dots and part of the white ones. We also see a good continuing curve, with a loop, made of white dots. These groups do not compete.

b) two grouping laws compete and one of them wins, the other one being inhibited. This case is called masking and will be discussed thoroughly in the next section.

c) conflict: in that case, both grouping laws are potentially active, but the groups cannot exist simultaneously. In addition, none of the grouping laws wins clearly. Thus, the figure is ambiguous and presents two or more possible interpretations.

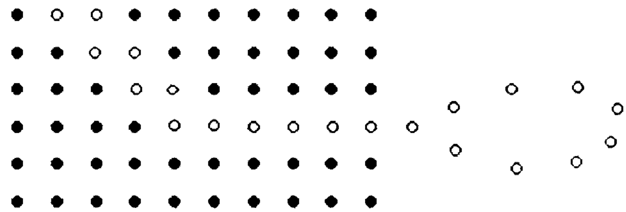

Figure 4. Gestalt laws in simultaneous action whithout conflict: *the white dots are elements of the grid (alignment, constant width) and simultaneously participate of a good continuing curve.*

A big section of Kanizsa's second chapter [14], is dedicated to conflicts of gestalts. Their study leads to the invention of clever figures where an equilibrium is maintained between two conflicting gestalt laws struggling to give the final figure organization. The viewer can direct his attention in both ways, see both organizations and perceive their conflict. A seminal experiment due to Wertheimer[3] gives an easy way to construct such conflicts. In figure 5, we see on the left a figure made of rectangles and ellipses. The prominent grouping laws are: a) shape similarity (L_1), which leads us to group the ellipses together and the rectangles as two conspicuous groups; b) the vicinity law L_2, which makes all of these elements build anyway a unified cluster. Thus, on the left figure, both laws coexist without real conflict. On the right figure instead, two clusters are present. Each one is made of heterogeneous shapes, but they fall apart enough to enforce the splitting of the ellipses group and of the rectangles group. Thus, on the right, the vicinity law L_2 tends to win. Such figures can be varied, by changing (e.g.) progressively the distance between clusters until the final figure presents a good equilibrium between conflicting laws.

Some laws, like good continuation, are so strong that they almost systematically win, as is illlustrated in figure 6. In this figure, two figures with a striking axial symmetry are concatenated in such a way that their boundaries are put in "good continuation". The result is a different

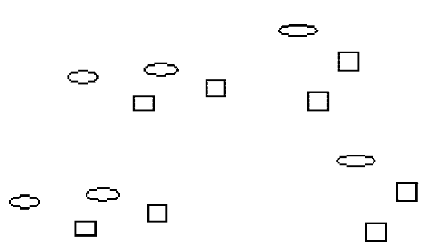

Figure 5. Conflict of similarity of shapes with vicinity. We can easily view the left hand figure as two groups by shape similarity, one made of rectangles and the other one of ellipses. On the right, two different groups emerge by vicinity. Vicinity "wins" against similarity of shapes.

interpretation, where the symmetric figures literally disappear. This is a conflict, but with a total winner. It therefore enters into the category of masking.

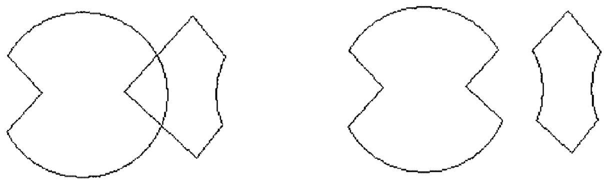

Figure 6. A "conflict of gestalts": two overlapping closed curves or, as suggested on the right, two symmetric curves which touch at two points? We can interpret this experiment as a masking of the symmetry gestalt law by the good continuation law. (From Kanizsa, Grammatica del Vedere p 195, op. cit.)

3.2 MASKING

Masking is illustrated by a lot of puzzling figures, where partial gestalts are literally hidden by other partial gestalts giving a better global explanation of the final figure. The masking phenomenon is the outcome of a conflict between two partial gestalts L_1 and L_2 struggling to organize a figure. When one of them, L_1, wins, a striking phenomenon occurs: the other possible organization, which would result from L_2, is hidden. Only an explicit comment to the viewer can remind her of the existence of the possible organization under L_2: the parts of the figure which might be perceived by L_2 have become invisible, masked in the final figure, which is perceived under L_1 only.

Kanizsa considers four kinds of masking: *masking by embedment in a texture; masking by addition (the Gottschaldt tecnique); masking by substraction (the Street technique)*; masking by manipulation of the figure-background articulation (Rubin, many famous examples by Escher). The first technique we shall consider is *masking in texture*. Its principle is: a geometrically organized figure is embedded into a texture, that is, a whole region made of similar building elements. This masking may well be called *embeddedness* as suggested by Kanizsa[4]. Figure 7 gives a good instance of the power of this masking, which has been thoroughly studied by the schools of Beck and Juslesz [13]. In this clever figure, the basis of a triangle is literally hidden in a set of parallel lines. We can interpret the texture masking as a conflict between an arbitrary organizing law L_2 and the similarity law, L_1. The masking technique works by multiple additions embedding a figure F organized under some law L_2 into many elements which have a shape similar to the building blocks of F.

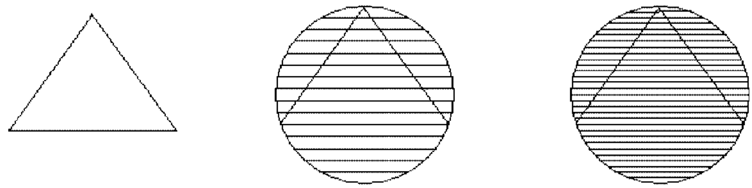

Figure 7. Masking by embedding in a texture. The basis of the triangle becomes invisible as it is embedded in a group of parallel lines. (Galli and Zama, quoted in Vedere e pensare, op. cit.).

The same proceeding is at work in figure 8. In that figure, one sees that a curve made of roughly aligned pencil strokes is embedded in a set of many more parallel strokes.

Figure 8. Masking by embedding in a texture again. On the right, a curve created from strokes by "good continuation". This curve is present, but masked on the left. We can consider it as a conflict between L_2, "good continuation" and L_1: similarity of direction. The similarity of direction is more powerful, as it organizes the full figure (articulazione senza resti).

In the *masking by addition* technique, due to Gottschaldt, a figure is concealed by addition of new elements which create another and more powerful organization. Here, L_1 and L_2 can be any organizing law. In figure 9, an hexagon is thoroughly concealed by the addition to the figure of two parallelograms which include in their sides the initial sides of the hexagon. Noticeably, the "winning laws" are the same which made the hexagon so conspicuous before masking, namely closure, symmetry, convexity and good continuation.

As figure 10 shows, L_1 and L_2 can revert their roles. On the right, the curve obtained by good continuation is made of perfect half circles concatenated. This circular shape is masked in the good continuation. Surprisingly enough, the curve on the left is present in the figure on the right, but masked by the circles. Thus, on the left, good continuation wins against the past experience of circles. On the right, the converse occurs; convexity, closure and circularity win against good continuation and mask it. The third masking technique considered by Kanizsa is substraction (Street technique), that is, removal of parts of the figure. As is apparent in figure 11, where a square is amputated in three different ways, the technique results effective only

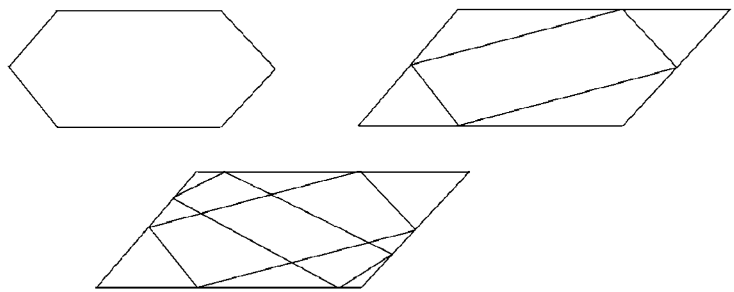

Figure 9. Masking by concealment (Gottschaldt 1926). The hexagon on the left is concealed in the figure on the right, and more concealed below. The hexagon was built by the closure, symmetry, convexity gestalt laws. The same laws are active to form the winner figures, the parallelograms.

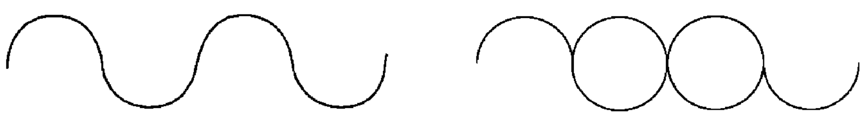

Figure 10. Masking of circles in good continuation, or, conversely, masking of good continuation by closure and convexity. We do not really see arcs of circles on the left, although significant and accurate parts of circles are present: we see a smooth curve. Conversely, we do not see the left "good" curve as a part of the right figure. It is nonetheless present in it.

when the removal creates a new gestalt. The square remains in view in the third figure from the left, where the removal has been made at random and is assimilable to a random perturbation. In the second and fourth figure, instead, the square disappears, although some parts of its sides have been preserved.

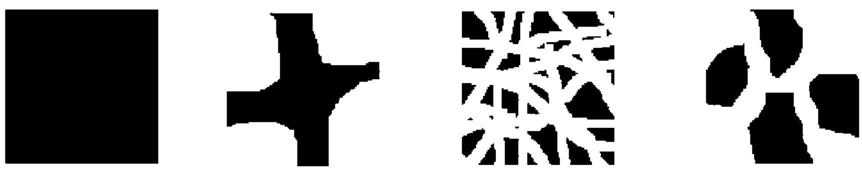

Figure 11. Masking by the Street substraction technique (1931), inspired from Kanizsa (Vedere e pensare p 176, op. cit.). Parts are removed from the black square. When this is done in a coherent way, it lets appear a new shape (a rough cross in the second subfigure, four black spots in the last one) and the square is masked. It is not masked at all in the third, though, where the removal has been done in a random way and does not yields a competing interpretation.

We should not end this section without considering briefly the last category of masking mentioned by Kanizsa, the masking by inversion of the figure-background relationship. This kind of masking is well known thanks to the famous Escher drawings. Its principle is “the background is not a shape” (*il fondo non è forma*). Whenever strong gestalts are present in an image, the space between those conspicuous shapes is not considered as a shape, even when it has itself a familiar shape like a bird, a fish, a human profile. Again here, we can interpret masking as the result of a conflict of two partial gestalt laws, one building the form and the other one, the loser, not allowed to build the background as a gestalt.

4. QUANTITATIVE ASPECTS OF GESTALT THEORY

In this section, we open the discussion on quantitative laws for computing partial gestalts. We shall first consider some numerical aspects of Kanizsa's *masking by texture*. In continuation, we shall make some comments on Kanizsa's paradox and its answer pointing out the involvement of a quantitative image resolution. These comments lead to Shannon's sampling theory.

4.1 QUANTITATIVE ASPECTS OF THE MASKING PHENOMENON

In his fifth chapter of *Vedere e pensare*, Kanizsa points out that “it is licit to sustain that a black homogeneous region contains all theoretically possible plane figures, in the same way as, for Michelangelo, a marble block virtually contains all possible statues”. Thus, these virtual statues could be considered as masked. This is the so called Kanizsa paradox. Figure 12 shows that one can obtain any simple enough shape by pruning a regular grid of black dots. In order

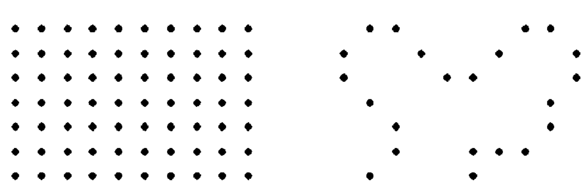

Figure 12. According to Kanizsa's paradox, the figure on the right is potentially present in the figure on the right, and would indeed appear if we colored the corresponding dots. This illustrates the fact that the figure on the left contains a huge number of possible different shapes.

to go further, it seems advisable to the mathematician to make a count: how many squares could we see, for example, in this figure? Characterizing the square by its upper left corner and its side length, it is easily computed that the number of squares whose corners lie on the grid exceeds 400. The number of curves with "good continuation" made of about 20 points like the one drawn on the right of figure 12 is equally huge. We can estimate it in the following way: we have 80 choices for the first point, and about five points among the neighbors for the second point, etc. Thus, the number of possible good curves in our figure is grossly $80 * 5^{20}$ if we accept the curve to turn strongly, and about $80 * 3^{20}$ if we ask the curve to turn at a slow rate. In both cases, the number of possible "good" curves in the grid is very large.

This multiplicity argument suggests that a grouping law can be active in an image, only if its application would not create a huge number of partial gestalts. Or, to say it in another way, we can sustain that the multiplicity implies a masking by texture. Masking of all possible good curves in the grid of figure 12 occurs, just because too many such curves are possible.

In the above figure 8 (subsection 3.2), we can repeat the preceding quantitative argument. In this figure, the left hand set of strokes actually contains, as an invisible part, the array of strokes on the right. This array of strokes is obviously organized as a curve (good contination gestalt). This curve becomes invisible on the left hand figure, just because it gets endowed in a more powerful gestalt, namely parallelism (similarity of direction). As we shall see in the computational discussion, the fact that the curve has been masked is related to another fact which is easy to check on the left hand part of the figure: one could select on the left many curves of the same kind as the one given on the right.

In short, we do not consider Kanizsa's paradox as a difficulty to solve, but rather as an arrow pointing towards the computational formulation of gestalt: In section 5, we shall define a partial gestalt as a *structure which is not masked in texture*.

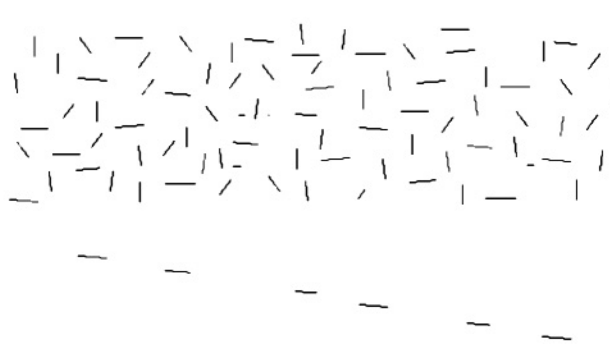

Figure 13. Below, an array of roughly aligned segments. Above, the same figure is embedded into a texture is such a way that it still is visible as an alignment. We are in the limit situation associated with Vicario's proposition: "is masked only what can be unmasked".

We shall therefore not rule out the extreme masking cases, in contradiction with Vicario's principle *"è mascherato solo ciò che può essere smascherato"* (is masked only what can be unmasked). Clearly, all psychophysical masking experiments must be close enough to the "conflict of gestalts" situation, where the masked gestalt still is attainable when the subject's attention is directed. Thus, psychological masking experiments must remain close to the non masking situation and therefore satisfy Vicario's principle. From the computational viewpoint instead, figures 12 and 8 are nothing but very good masking examples.

In this masking issue, one feels the necessity to pass from qualitative to quantitative arguments: a gestalt can be more or less masked. How to compute the right information to quantize this "more or less"? It is actually related to a *precision parameter*. In figure 13, we constructed a texture by addition from the alignment drawn below. Clearly, some masking is at work and we would not notice immediately the alignment in the texture if our attention was not directed. All the same, the alignment remains somewhat conspicuous and a quick scan may convince us that *there is no other alignment of such an accuracy in the texture*. Thus, in that case, alignment is not masked by parallelism. Now, one can suspect that this situation can be explained in yet quantitative terms: the precision of the alignment matters here and should be evaluated. Precision will be one of the three parameters we shall use in computational gestalt.

4.2 SHANNON THEORY AND THE DISCRETE NATURE OF IMAGES

The preceding subsection introduced two of the parameters we shall have to deal with in the computations, namely the *number of possible partial gestalts* and a *precision* parameter. Before proceeding to computations, we must discuss the rough datum itself, namely the computational nature of images, let them be digital or biological. Kanizsa addresses briefly this problem in the fifth chapter of *Vedere e pensare*, in his discussion of the masking phenomenon: "We should not consider as masked elements which are too small to attain the visibility threshold". Kanizsa was aware that the amount of visible points in a figure is finite[5]. He explains in the same chapter why this leads to work with figures made of dots; we can consider this decision as a way to quantize the geometric information.

In order to define mathematically an image, be it digital or biological, in the simplest possible way, we just need to fix a point of focus. Assume all photons converging towards this focus are intercepted by a surface which has been divided into regular cells, usually squares or hexagons. Each cell counts its number of photons hits during a fixed exposure time. This count gives a grey level image, that is, a rectangular, (roughly circular in biological vision) array of grey level values on a grid. In the case of digital images, C.C.D. matrices give regular grids made of squares. In the biological case, the retina is divided into hexagonal cells with growing sizes from the fovea. Thus, in all cases, a digital or biological image contains a finite number of values on a grid. Shannon [28] made explicit the mathematical conditions under which, from this matrix of values, a continuous image can be reconstructed. By Shannon's theory, we can compute the grey level at all points, and not only the points of the grid. Of course, when we zoom in the interpolated image it looks more and more blurry: the amount of information in a digital image is indeed finite and the resolution of the image is bounded. The points of the grid together with their grey level values are called pixels, an abbreviation for picture elements.

The pixels are the computational atoms from which gestalt grouping procedures can start. Now, if the image is finite, and therefore blurry, **how can we infer sure events as lines, circles, squares and whatsoever gestalts from discrete data?** If the image is blurry all of these structures cannot be inferred as completely sure; their exact location must remain uncertain. This is crucial: all basic geometric information in the image has a precision[6]. Figure 13 shows it plainly. It is easy to imagine that if the aligned segments, still visible in the figure, are slightly less aligned, then the alignment will tend to disappear. This is easy cheeked with figure 14, where we moved slightly up and down the aligned segments.

Let us now say briefly which local, atomic, information can be the starting point of computations. Since every local information about a function u at a point (x, y) boils down to its Taylor expansion, we can assume that these atomic informations are:

- the value $u(x, y)$ of the grey level at each point (x, y) of the image plane. Since the function u is blurry, this value is valid at points close to (x, y).

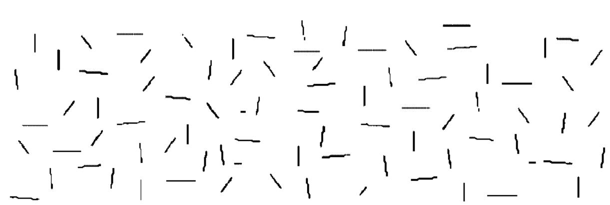

Figure 14. When the alignment present in figure 13 is made less accurate, the masking by texture becomes more efficient. The precision plays a crucial role in the computational gestalt theory outlined in the next section.

- the gradient of u at (x, y), the vector

$$Du(x, y) = \left(\frac{\delta u}{\delta x}, \frac{\delta u}{\delta y}\right)(x, y)$$

• the *orientation* at *(x, y),*

$$Orient(x) = \frac{1}{\|Du(x,y)\|}\left(-\frac{\delta u}{\delta y}, \frac{\delta u}{\delta x}\right)(x,y)$$

This vector is visually intuitive, since it is tangent to the boundaries one can see in an image.

These local informations are known at each point of the grid and can be computed at any point of the image by Shannon interpolation. They are quantized, having a finite number of digits, and therefore noisy. Thus, each one of the preceding measurements has an intrinsic precision. The orientation is invariant when the image contrast changes (which means robustness to illumination conditions). Attneave and Julesz [13] refer to it for shape recognition and texture discrimination theory. Grey level, gradient and orientation are the only local information we shall retain for the numerical experiments of the next section, together with their precisions.

5. COMPUTING PARTIAL GESTALTS IN DIGITAL IMAGES

In this section, we shall summarize a computational theory which permits to find automatically partial gestalts in digital images. This theory essentially predicts *perception thresholds* which can be computed on every image and give a usually clear cut decision between what is seeable as a geometric structure (gestalt) in the image and what is not. Those thresholds are computable thanks to the discrete nature of images. Many more details can be found in [9]. All computations below will involve three fundamental numbers, whose implicit presence in Gestalt theory has just been pointed out, namely

- a relative precision parameter p which we will treat as a probability;
- a number N_{conf} of possible configurations for the looked for partial gestalt. This number is <u>finite</u> because the image resolution is bounded;

- The number N of pixels of the image.

Of course, p and N_{conf} will depend upon the kind of gestalt grouping law under consideration. We can relate p and N to two fundamental qualities of any image: its noise and its blur.

5.1 A GENERAL DETECTION PRINCIPLE

In [6], [7], [8], we outlined a computational method to decide whether a given partial gestalt (computed by any segmentation or grouping method) is reliable or not. As we shall recall, our method gives *absolute thresholds*, depending only on the image size, permitting to decide when a given gestalt is perceptually relevant or not.

We applied a general perception principle which we called Helmholtz principle (figure 15). This principle yields computational grouping thresholds associated with each gestalt quality. It can be stated in the following generic way. Assume that atomic objects O_1, O_2, …, O_n are present in an image. Assume that k of them, say O_1, …, O_k, have a common feature, say, same color, same orientation, position etc.. We are then facing the dilemma: is this common feature happening by chance or is it significant and enough to group O_1,... O_k? In order to answer this question we make the following mental experiment: we assume *a priori* that the considered quality has been randomly and uniformly distributed on all objects O_1,... O_n Then we (mentally) assume that the observed position of objects in the image is a random realization of this uniform process. We finally ask the question: is the observed repartition probable or not ? If not, this proves a *contrario* that a grouping process (a gestalt) is at stake. Helmholtz principle states roughly that in such mental experiments, the numerical qualities of the objects are assumed to be equally distributed and independent. Mathematically, this can be formalized by

Definition 1 (ε-meaningful event [6]) *We say that an event of type "such configuration of geometric objects has such property" is ε-meaningful if the expectation of the number of occurrences of this event is less than ε under the uniform random assumption.*

As an example of generic computation we can do with this definition, let us assume that the probability that a given object O_i has the considered quality is equal to p. Then, under the independence assumption, the probability that at least k objects out of the observed n have this quality is

$$B(p, n, k) = \sum_{i=k}^{n} \binom{n}{k} p^i (1 - p)^{n-i},$$

i.e. the tail of the binomial distribution. In order to get an upper bound for the number of false alarms, *i.e.* the expectation of the number of geometric events happening by pure chance, we can simply multiply the above probability by the number of tests we perform on the image. This number of tests N_{conf} corresponds to the number of different possible positions we could have for the searched gestalt. Then, in most cases we shall consider in the next subsections, a considered event will be defined as ε-meaningful if

$$NFA = N_{conf} B(p, n, k) \leq \varepsilon .$$

We call in the following the left hand member of this inequality the "number of false alarms" (NFA). The number of false alarms of an event measures the "meaningfulness" of this event: the smaller it is, the more meaningful the event is. We refer to [6] for a complete discussion of this definition. To the best of our knowledge, the use of the binomial tail, for alignment detection, was introduced by Stewart [30].

5.2 COMPUTATION OF SIX PARTIAL GESTALTS

Alignments of points.

Points will be called aligned if they all fall into a strip thin enough, and in sufficient number. This qualitative definition is easily made quantitative. The precision of the alignment is measured by the width of the strip. Let S be a strip of width a. Let $p(S)$ denote the prior probability for a point to fall in S, and let $k(S)$ denote the number of points (among the M) which are in S. The following definition permits to compute all strips where a meaningful alignment is observed (see figures 15 and 18).

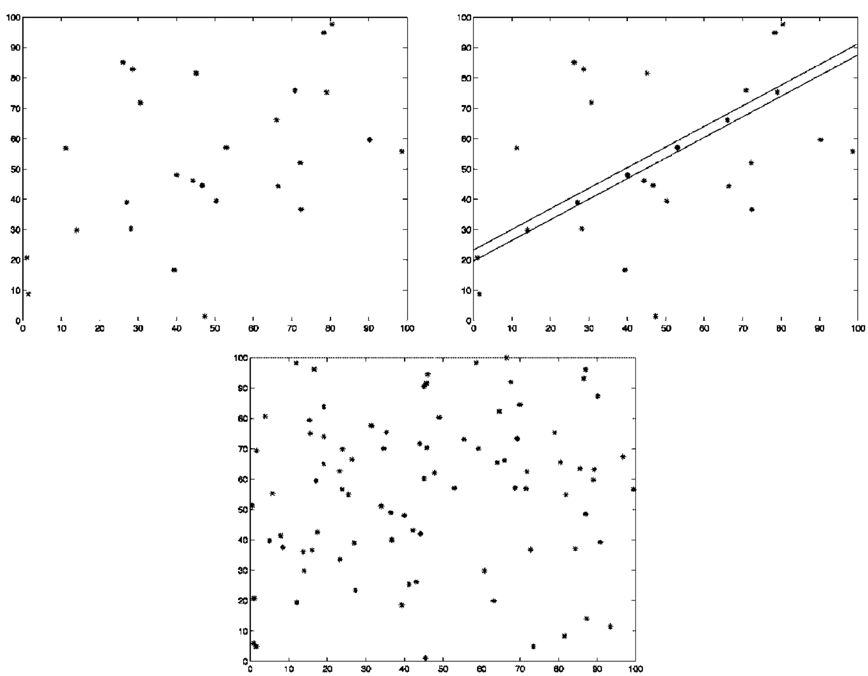

Figure 15. An illustration of Helmholtz principle: non casual alignments are automatically detected by Helmholtz principle as a large deviation from randomness. *Left, 20 uniformly randomly distributed dots, and 7 aligned added. Middle: this meaningful (and seeable) alignment is detected as a large deviation from randomness. Right: same alignment added to 80 random dots. The alignment is no more meaningful (and no more seeable). In order to be meaningful, it would need to contain at least 11 points.*

Definition 2 ([9]) *A strip S is ε-meaningful if*

$$NF(S) = Ns \cdot B(p(S), M, k(S)) \leq \varepsilon,$$

where Ns is the number of considered strips (one has $Ns \simeq 2\pi(R/a)^2$, where R is the half-diameter of Ω and a the minimal width of a strip).

Let us give immediately a summary of several algorithms based on the same principles, most of which use, as above, the tail of the binomial law. This is done in table 1, where we summarize the (very similar) formulae permitting to compute the following partial gestalts: *alignments (of orientations in a digital image), contrasted boundaries, all kinds of similarities for some quality measured by a real number (grey level, orientation,...)*, and of course the most basic one, treated in the last row, namely the *vicinity* gestalt.

Alignments in a digital image

The first row of table 1 treats the alignments in a digital image. As we explained in subsection 4.2, an orientation can be computed at each point of the image. Whenever a long enough segment occurs in the image, whose orientations are aligned with the segment, this segment is perceived as an alignement. We consider the following event: "on a discrete segment of the image, joining two pixel centers, and with length l, at least k independent points have the same direction as the segment with precision p." The definition of the number of false alarms is given in the first row of the table and an example of the alignments, in a digital aerial image, whose number of false alarms is less than 1 is given in figure 16.

Maximal meaningful gestalts and *articulazione senza resti*

On this example of alignments, we can address a problem encountered by the mathematical formalization of gestalt.

GROUP LOOKED FOR	MEASUREMENTS	NUMBER OF FALSE ALARMS
Alignment of directions on a segment [6]		
a discrete segment with points at Nyquist distance (i.e. 2)	k: number of aligned points l: number of points on the segment	$N_{segments} \cdot B(p, l, k)$ $N_{segments} = N^3$ N: number of pixels in the image $p = 1/16$ (angular precision)
Contrasted edges and boundaries [7]		
a level line (or a piece of) with points at Nyquist distance (i.e. 2)	μ : minimum contrast (gradient norm) along the curve l: length of the curve	$N_{level\ lines} \cdot H(\mu)^l$ H is the empirical cumulative distribution of the gradient norm on the image
Similarity of a uniform scalar quality (grey level, orientation, etc.) [9]		
a group of objects having a scalar quality q such that $a \leqslant q \leqslant b$	k: number of points in the group M: total number of objects	$\frac{L(L+1)}{2} \cdot B\left(\frac{b-a+1}{L}, M, k\right)$ L: number of values ($q \in \{1..L\}$)
Similarity of a scalar quality with decreasing distribution (area, length, etc.) [9]		
a group of objects having a scalar quality q such that $a \leqslant q \leqslant b$	k: number of points in the group M: total number of objects	$\frac{L(L+1)}{2} \cdot \max_{p \in \mathcal{D}} B\left(\sum_{i=a}^{b} p(i), M, k\right)$ L: number of values ($q \in \{1..L\}$) $\mathcal{D}$: set of decreasing distributions on $\{1..L\}$
Alignment of points (or objects) [9]		
a goup of points falling in a strip (region enclosed by two parallel lines)	p: relative area of the strip k: number of points falling in the strip	$N_{strips} \cdot B(p, M, k)$ M: total number of points The strips are quantized in position, width and orientation
Vicinity : clusters of points (or objects) [9]		
a group of points falling in a region enclosed by a low-resolution curve	σ: relative area of the region σ': relative area of the thick low-resolution curve k: number of points falling in the region	$N_{regions} \cdot \sum_{i=k}^{M} \binom{M}{i} \sigma^i (1 - \sigma - \sigma')^{M-i}$ M: total number of points $N_{regions} = N^2 qr2^L$: the low resolution curves are quantized in resolution (q), thickness (r), location (N), and bounded in length (L).

Table 1. List of gestalts computed so far.

Assume that on a straight line we have found a very meaningful segment *S*. Then, by enlarging slightly or reducing slightly *S*, we still find a meaningful segment. This means that meaningfulness cannot be a univoque criterion for detection, unless we can point out the "best meaningful" explanation of what is observed as meaningful. This is done by the following definition, which can be adapted as well to meaningful boundaries [7], meaningful edges [7], meaningful modes in a histogram and clusters [9].

Definition 3 ([8]) *We say that an ε-meaningful geometric structure A is maximal meaningful if*

- *it does not contain a strictly more meaningful structure:* $\forall B \subset A,\ NF(B) \geq NF(A)$.
- *it is not contained in a more meaningful structure:* $\forall B \supset A,\ B \neq A,\ NF(B) > NF(A)$.

It is proved in [8] that maximal structures cannot overlap, which is one of the main theoretical outcomes validating the above definitions. This definition formalizes the *articulazione senza resti* principle in the case of a single grouping law.

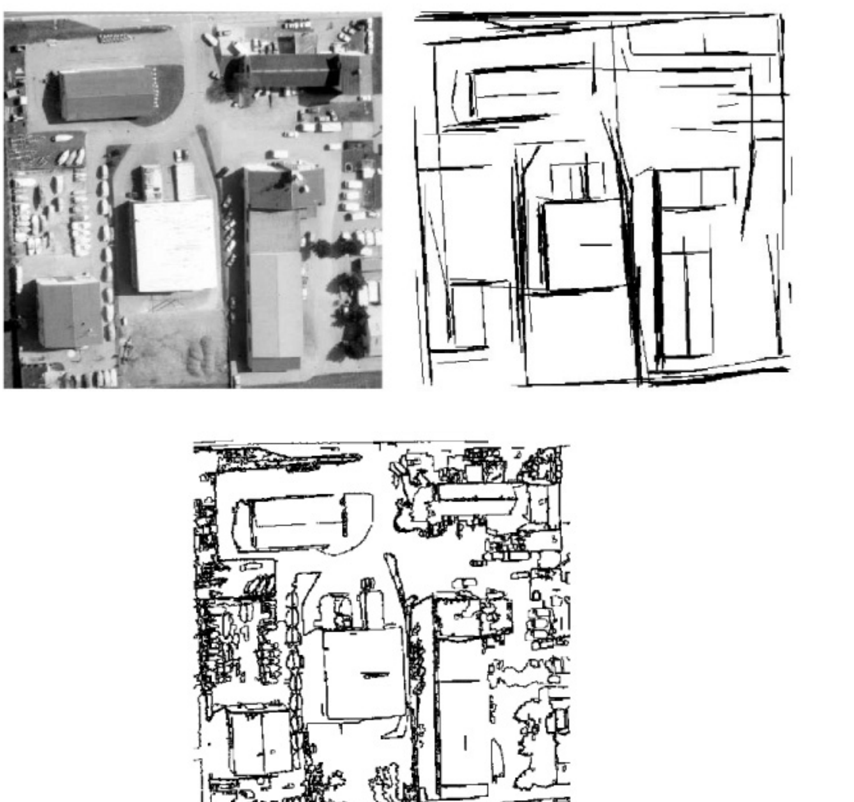

Figure 16. Two partial gestalts, alignments and boundaries. *Left: original aerial view (source: INRIA), middle: maximal meaningful alignments, right: maximal meaningful boundaries.*

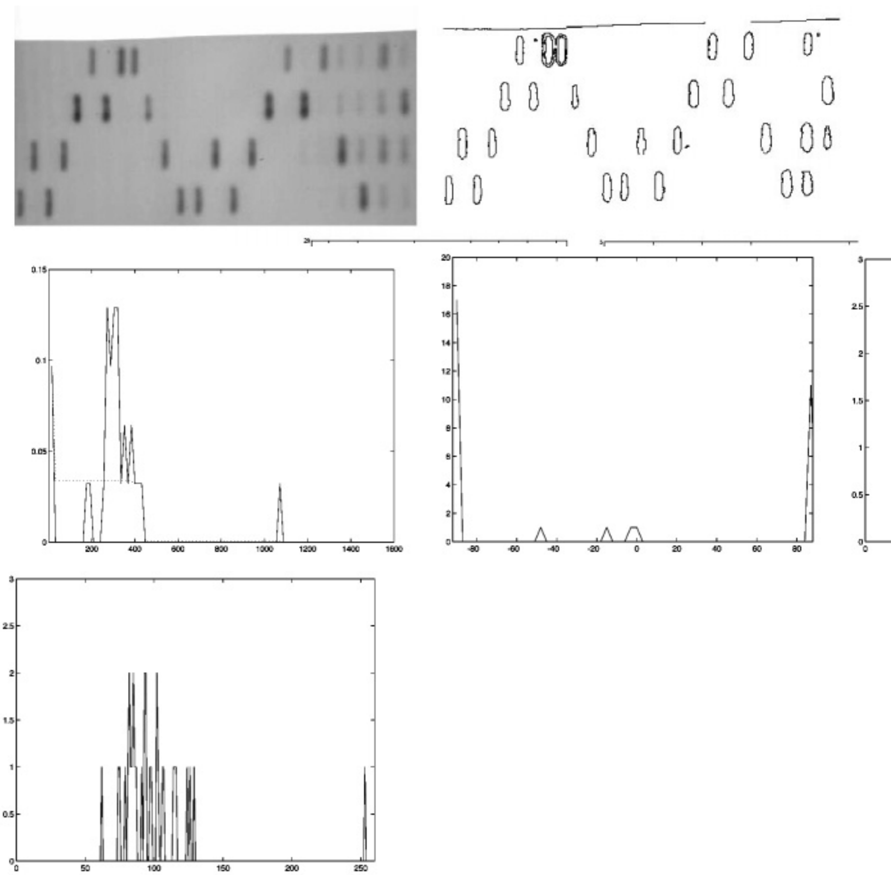

Figure 17. Collaboration of gestalts. The objects tend to be grouped similarly by several different partial gestalts. *First row: original DNA image (left) and its maximal meaningful boundaries (right). Second row, left: histogram of areas of the meaningful blobs. There is a unique maximal mode (256-416). The outliers are the double blob, the white background region and the three tiny blobs. Second row, middle: histogram of orientations of the meaningful blobs (computed as the principal axis of each blob). There is a single maximal meaningful mode (interval). This mode is the interval 85-95. It contains 28 objects out of 32. The outliers are the white background region and three tiny spots. Second row, right: histogram of the mean grey levels inside each block. There is a single maximal mode containing 30 objects out of 32, in the grey level interval 74-130. The outliers are the background hite region and the darkest spot.*

Boundaries

One can define in a very similar way the “boundary” grouping law. This grouping law is never stated explicitly in gestaltism, because it is probably too obvious for phenomenologists. From the computation viewpoint it is not, at all.

The definition of the number of false alarms for boundaries involves again two variables: the length l of the level line, and its minimal contrast μ, which is interpreted as a precision. An example of boundary detection is given on figure 16.

Similarity

The third row of table 1 addresses the main gestaltic grouping principle: points or objects having a feature in common are being grouped, just because they have this feature in common. Assume k objects $O_1, \ldots O_k$, among a longer list $O_1, \ldots, O_M$, have some quality q in common. Assume that this quality is actually measured as a real number. Then our decision of whether the grouping of $O_1, \ldots, O_k$ is relevant must be based on the fact that the values $q(O_1), \ldots, q(O_k)$ make a *meaningfully dense interval* of the histogram of $q(O_1), \ldots, q(O_M)$. Thus, the automatic quality grouping is led back to the question of an automatic, parameterless, histogram mode detector. Of course, this mode detector depends upon the kind of feature under consideration.

We shall consider two paradigmatic cases, namely the case of orientations, where the histogram can be assumed by Helmholtz principle to be flat, and the case of the objects sizes (areas) where the null assumption is that the size histogram is decreasing (see figure 17).

Thus, the third and fourth row of our table permit to detect all kinds of similarity gestalt: objects grouped by orientation, or grey level, or any perceptually relevant scalar quality.

5.3 STUDY OF AN EXAMPLE

We are going to perform a complete study of a digital image, figure 17, involving all computational gestalts defined in table 1. The analyzed image is a common digital image. It is a scan of photograph and has blur and noise. The seeable objects are electrophoresis spots which have all similar but varying shape and color and present some striking alignments. Actually, all of these perceptual remarks can be recovered in a fully automatic way by combining several partial gestalt grouping laws.

First, the *contrasted boundaries* of this electrophoresis image are computed (above, right). Notice

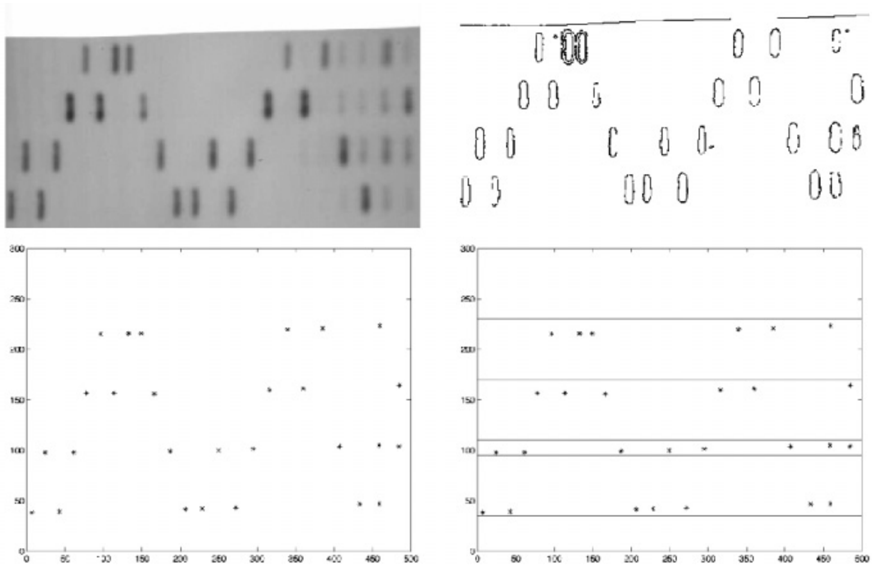

Figure 18. Gestalt grouping principles at work for building an "order 3" gestalt (alignment of blobs of the same size). *First row: original DNA image (left) and its maximal meaningful boundaries (right). Second row: left, barycenters of all meaningful regions whose area is inside the only maximal meaningful mode of the histogram of areas; right, meaningful alignments of these points.*

that all closed curves found are indeed perceptually relevant as they surround the conspicuous spots. Other many possible boundaries in the noisy background have been ruled out and remain "masked in texture". Let us apply a second layer of grouping laws. This second layer will use as atomic objects the blobs found at the first step. For each of the detected boundaries, we compute three qualities, namely

a) *the area* enclosed by the boundary, whose histogram is displayed on the top left of figure 17. There is a unique maximal mode in this figure, which actually groups all and exactly the blobs with similar areas and rules out two tiny blobs and a larger one enclosing two different blobs. Thus, almost all blobs get grouped by this quality, with the exception of two tiny spots and a double spot.

b) *the orientation* of each blob, an angle between $-\pi/2$ and $\pi/2$. This histogram (figure 17, bottom, middle) again shows a single maximal mode, again computed by the formula of the third row of table 1. This mode appears at both end points of the interval, since the dominant direction is $\pm\pi/2$ and these values are identified modulo π. Thus, about the same blobs as in b) get grouped by their common orientation.

c) *the average grey level* inside each blob: its histogram is shown on the bottom right of figure 17. Again, most blobs, but not all get grouped with respect to this quality.

A further structural grouping law can be applied to build subgroups of blobs formed by alignment. This is illustrated in figure 18 (bottom, left), where the meaningful alignments are found. This experiment illustrates the usual strong collaboration of partial gestalts: most salient objects or groups come to sight by several grouping laws.

6. THE LIMITS OF EVERY PARTIAL GESTALT DETECTOR

The preceding section argued in favor of a very simple principle, Helmholtz principle, applicable to the automatic and parameterless detection of any partial gestalt, in full agreement with our perception. In this section, we shall show by commenting briefly several experiments that *tout n'est pas rose*: there is a good deal of visual illusion in any apparently satisfactory result provided by a partial gestalt detector on a digital image. We explained in the former section that partial gestalts often collaborate. Thus, in many cases, what has been detected by a partial gestalt will be corroborated by another one. For instance, boundaries and alignments in the experiment 16 are in good agreement. But what can be said about the experiment of figure 19? In this cheetah image, we have applied the alignment detector explained above. It works wonderfully on the grass leaves when they are straight. Now, we also see some unexpected alignments in the fur. These alignments do exist: these detected lines are tangent to several of the convex dark spots on the fur. This generates a meaningful excess of aligned points on these lines, the convex sets being smooth enough and having therefore on their boundary a long enough segment tangent to the detected line.

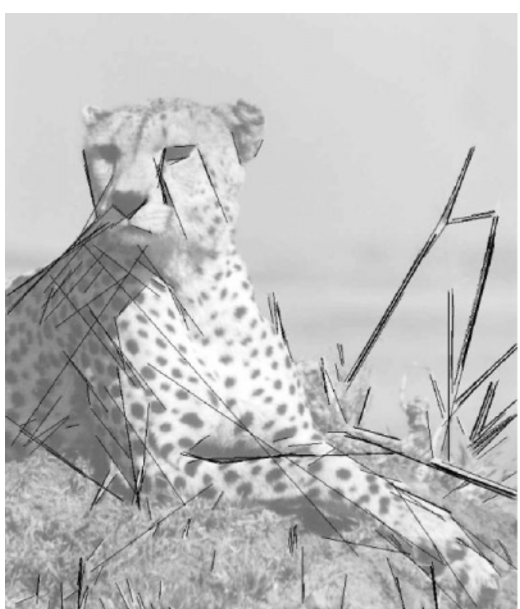

Figure 19. Smooth convex sets or alignments?

We give an illustration of this fact in figure 20. The presence of only two circles can actually create a (to be masked) alignment since the circles have bitangent straight lines. In order to discard such spurious alignments, the convexity (or good continuation) gestalt should be systematically searched when we look for alignments. Then, the alignments which only are tangent lines to several smooth curves, could be inhibited.

We should detect as alignment what indeed is aligned, but only under the condition that the alignment does *not* derive from the presence of several smooth curves... This statement can be generalized: no gestalt is just a positive quality. The outcome of a partial gestalt detector is valid only when *all* other partial gestalts have been tested and the eventual conflicts dealt with.

The same argument applies to our next experimental example, in figure 21. In that case, a dense cluster of points is present. Thus, it creates a meaningful amount of dots in many strips and the result is the detection of obviously wrong alignments. Again, the detection of a cluster should inhibit such alignment detections. We defined an alignment as “many points in a thin strip”, but must add to this definition: “provided these points do not build one or two dense clusters”.

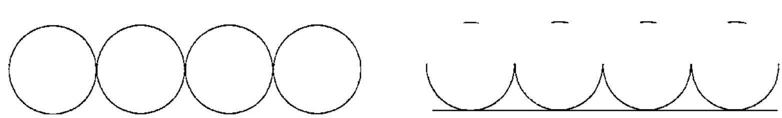

Figure 20. Alignment is masked by good continuation and convexity: the small segments on the right are perfectly aligned. Any alignment detector should find them. All the same, this alignment disappears on the left figure, as we include the segments into circles. In the same way, the casual alignments in the Cheetah fur (figure 19) are caused by the presence of many oval shapes. Such alignments are perceptually masked and should be computationally masked!

One can reiterate the same problematic with another gestalt conflict (figure 22). In this figure, a detector of arcs of circles has been applied. The arc of circle detection grouping law is easily adapted from the definition of alignments in table 1. The main outcome of the experiment is this: since the MegaWave figure contains many smooth boundaries and several straight lines, lots of meaningful circular arcs are found. It may be discussed whether those circular arcs are present or not in the figure: clearly, any smooth curve is locally tangent to some circle. In the same way, two segments with an obtuse angle are tangent to several circular arcs (see figure 23). Thus, here again, a partial gestalt should mask another one. Hence the following statement, which is of wide consequence in Computer Vision: *We cannot hope any reliable explanation of any figure by summing up the results of one or several partial gestalts. Only a global synthesis, treating all conflicts of partial gestalts, can give the correct result.*

In view of these experimental counterexamples, it may well be asked why partial gestalt detectors often work "so well". This is due to the redundancy of gestalt qualities in most natural images, as we explained in the first section with the example of a square. Indeed, most natural or synthetic objects are simultaneously conspicuous, smooth and have straight or convex parts, etc. Thus, in many cases, each partial gestalt detector will lead to the same group definition. Our experiments on the electrophoresis image (figure 17) have illustrated the *collaboration of gestalt* phenomenon[7]. In that experiment, partial gestalts collaborate and seem to be redundant.

This is an illlusion which can be broken when partial gestalts do not collaborate.

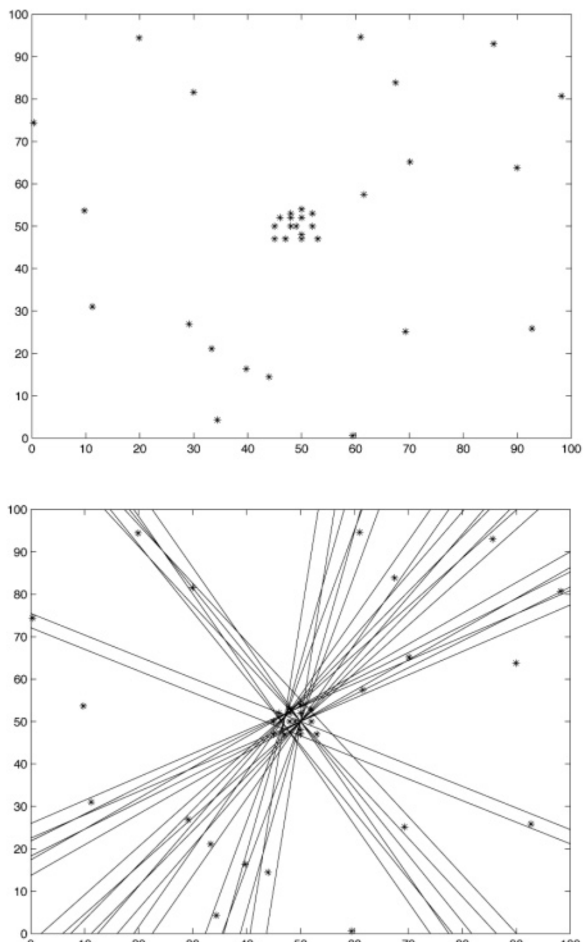

Figure 21. One cluster, or several alignments?

Figure 22. Left: original “MegaWave” image. Right: a circular arc detector is applied to the image. Now, this image contains many smooth curves and obtuse angles to which meaningful circular arcs are tangent. This illustrates the necessity of the interaction of partial gestalts: the best explanation for the observed structures is “good continuation” in the gestaltic sense, i.e. the presence of a smooth curve, or of straight lines (alignments) forming obtuse angles. Their presence entails the detection of arcs of circles which are not the final explanation.

Figure 23. Left: Every obtuse angle can be made to have many points in common with some long arc of circle. Thus, an arc of circle detector will make wrong detections when obtuse angles are present (see Figure 22). In the same way, a circle detector will detect the circle inscribed in any square and conversely, a square detector will detect squares circumscribing any circle.

6.1. CONCLUSION

We shall close the discussion by expressing some hope, and giving some arguments in favor of this hope. First of all, gestaltists pointed out the relatively small number of relevant gestalt qualities for biological vision. We have briefly shown in this paper that many of them (and probably all) can be computed by the Helmholtz principle followed by a maximality argument. Second, the discussions of gestaltists about "conflicts of gestalts", so vividly explained in the books of Kanizsa, might well be solved by a few information theoretical principles. As a good example of it, let us mention how the dilemma alignment-versus-parallelism can be solved by an easy minimal description length principle (MDL) [26], [8]. Figure 24 shows the problem and its simple solution. On the middle, we see all detected alignments in the Brown image on the left. Clearly, those alignments make sense but many of them are slanted. The main reason is this: all straight edges are in fact blurry and therefore constitute a rectangular region where all points have roughly the same direction. Thus, since alignment detection is made up to some precision, the straight alignments are mixed up with slanted alignments which still respect the precision bound. We can interpret the situation as a conflict between alignment and parallelism, as already illustrated in figure 8.

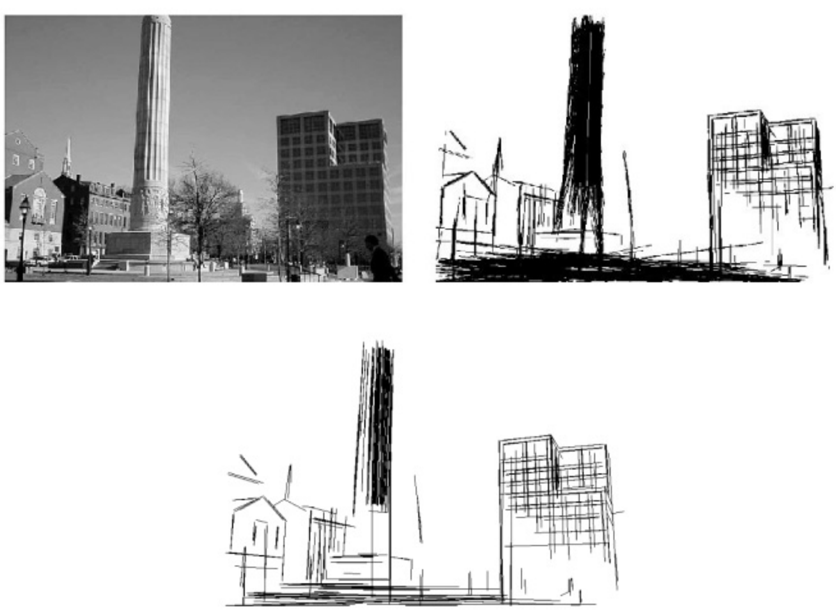

Figure 24. Parallelism against alignment. Left, original Brown image. Middle: maximal meaningful alignments. Here, since many parallel alignments are present, secondary, parasite slanted alignments are also found. Right: Minimal description length of alignments, which eliminates the spurious alignments. This last method outlines a solution to conflicts between partial gestalts.

The spurious, slanted alignments are easily removed by the application of a MDL principle: it is enough to retain for each point only the most meaningful alignment to which it belongs. We then compute again the remaining maximal meaningful alignments and the result (right) shows that the conflict between parallelism and alignment has been solved. Clearly, information theoretical rules of this kind may be applied in a general framework and put order in the proliferation of "partial gestalts". Let us mention an attempt of this kind in [20], where the author proposed a MDL reformulation of segmentation variational methods ([24])

Acknowledgments Work partially supported by Office of Naval Research under grant N00014-97-1-0839, Centre National d'Etudes Spatiales, Centre National de la Recherche Scientifique et Ministère de la Recherche et de la Technologie. We thank the Fondation des Treilles for having hosted the authors during the redaction of this paper.

Prof. Agnès Desolneux
CMLA, ENS,
61 av. du président Wilson
94235 Cachan, France

Prof. Lionel Moisan
CMLA, ENS,
61 av. du président Wilson
94235 Cachan, France

Prof. Jean-Michel Morel
CMLA, ENS,
61 av. du président Wilson
94235 Cachan, France

NOTES

[1] We shall write gestalt and treat it as an English word when we talk about gestalts as groups. We maintain the german uppercase for Gestalt theory.

[2] The good continuation principle has been extensively adressed in Computer Vision, first in [23], more recently in [27] and still more recently in [11]. A recent example of computer vision paper implementing "good continuation", understood a "constant curvature", is [32].

[3] Op. cit.

[4] Vedere e pensare op. cit., p 184

[5] "*non sono da considerare mascherati gli elementi troppo piccoli per raggiungere la soglia della visibilita pur potendo essere rivelati con l'ausilio di una lente di ingrandimento, il che dimostra che esistono come stimoli potenziali. E altrettanto vale per il caso inverso, nel quale soltanto con la diminuzione dell'angolo visivo e la conseguente riduzione della grandezza degli elementi e dei loro interspazi (mediante una lente o la visione a grande distanza) è possibile vedere determinate strutture"*.

[6] It is well known by gestaltists that a right angle "looks right" with some ±3 degrees precision, and otherwise does not look right at all.

[7] Kruger and Wörgötter [19] gave strong statistical computational evidence in favor of a collaboration between partial gestalt laws namely collinearity, parallelism, color, contrast and similar motion.

REFERENCES

[1] J.-P. D'ALÈS, J. FROMENT and J.-M. MOREL, "Reconstruction visuelle et généricité", *Intellectica* 1:28, pp. 11-35, 1999.

[2] F. ATTNEAVE, "Some informational aspects of visual perception", *Psych. Rev.* 61, pp. 183-193, 1954.

[3] A.J. BELL A.J. and T.J. SEJNOWSKI, "Edges are the 'independent components' of natural scenes", *Advances in Neural Information Processing Systens*, 9, 1996.

[4] E. BRUNSWIK and J. KAMIYA, "Ecological cue-validity of 'proximity' and other Gestalt factors", *American Journal of Psychology 66*, pp. 20-32, 1953.

[5] V. CASELLES, B. COLL et J.-M. MOREL, "A Kanizsa programme", *Progress in Nonlinear Differential Equations and their Applications*, 25, pp. 35-55, 1996.

[6] A. DESOLNEUX, L. MOISAN and J.-M. MOREL, "Meaningful Alignments", *International Journal of Computer Vision*, vol 40:1, pp.7-23, 2000.

[7] A. DESOLNEUX, L. MOISAN and J.-M. MOREL, "Edge Detection by Helmholtz Principle", *Journal of Mathematical Imaging and Vision*, vol 14:3, pp 271-284, 2001.

[8] A. DESOLNEUX, L. MOISAN and J.-M. MOREL, "Maximal Meaningful Events and Applications to Image Analysis", preprint CMLA no 200-22, available on http://www.cmla.ens-cachan.fr/Cmla/Publications/, submitted.

[9] A. DESOLNEUX, L. MOISAN and J.-M. MOREL, "Partial Gestalts", preprint CMLA no 200-22, submitted to IEEE Trans. on Pattern Analysis and Machine Intelligence, available on http://www.cmla.ens-cachan.fr/Cmla/Publications/, submitted.

[10] W.S. GEISLER, J.S. PERRY, B.J. SUPER and D.P. GALLOGLY, "Edge co-occurrence in natural images predicts contour grouping performance" *Vision Research 41*, pp. 711-724, 2001.

[11] G. GUY and G. MEDIONI, "Inferring global perceptual contours from local features", *Int. Journal of Comp. Vision* 20:1, pp. 113-133, 1996.

[12] B.K. HORN, *Robot Vision*, MIT Press, 1987.

[13] J.R. BERGEN and B. JULESZ. "Textons, the fundamental elements of preattentive vision and perception of textures", *Bell System Techical Journal* 62:6, pp. 1619-1645, 1983.

[14] G. KANIZSA, *Grammatica del Vedere*, Il Mulino, Bologna, 1980.

[15] G. KANISZA, *Vedere e pensare*, Il Mulino, Bologna, 1991.

[16] G. KANIZSA, *Organization in Vision*, Holt, Rinehart & Winston, 1979.

[17] M. KASS, A. WITKIN and D. TERZOPOULOS, "Snakes: active contour models", *Ist Int. Comp. Vision Conf.* IEEE 777, 1987.

[18] N. KIRYATI, Y. ELDAR and A.M. BRUCKSTEIN, "A Probabilistic Hough Transform", *Pattern Recognition* 24:4, pp. 303-316, 1991.

[19] N. KRÜGER and F. WÖRGÖTTER, "Multi-moda1 estimation of collinearity and parallelism in natural image sequences", submitted to *Network Computation in Neural System*, 2001.

[20] LECLERC, Y. Constructing Simple Stable Descriptions for Image Partitioning. *Int. J. of Comp. Vision*, 3, pp. 73-102, 1989.

[21] D. MARR, *Vision*, Freeman and co., 1982.

[22] W. METZGER, *Gesetze des Sehens.* Waldemar Kramer, 1975.

[23] U. MONTANARI, "On the Optimal Detection of Curves in Noisy Pictures", CACM 14:5, pp. 335-345, 1971.

[24] MUMFORD, D. and J. SHAH. Boundary detection by minimizing functionals IEEE *Conf. on Comp. Vision and Pattern Recognition*, San Francisco, 1985.

[25] B.A. OLSHAUSEN and D.J. FIELD, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images", *Nature* 381:6583, pp. 607-609, 1996.

[26] RISSANEN, J. A universal prior for integers and estimation by Minimum Description Length. *Annals of Statistics* 11 (2), 1983.

[27] A. SHA'ASHUA and S. ULLMAN, "Structural saliency: the detection of globally salient structures using a locally connected network", *Proceedings of the 2nd Int. Conf. on Comp. Vision*, pp. 321-327, 1988.

[28] C.E. SHANNON, "A mathematical theory of communication", *Bell System Technical Journal*, 27, pp. 379-423 and 623-656, 1948.

[29] M. SIGMAN M, G.A. CECCHI, C.D. GILBERT, and M.O. MAGNASCO. “On a common circle: Natural scenes and Gestalt rules”, *Proc. Nat. Acad. Sci.* USA 98:4, pp. 1935-1940, 2001.

[30] C.V. STEWART, “MINPRAN: a new robust estimator for Computer Vision”, *I.E.E.E. Trans. on Pattern Analysis and Machine Intelligence* 17, pp. 925-938, 1995.

[31] M. WERTHEIMER, “Untersuchungen zur Lehre der Gestalt, Il”. *Psychologische Forschung* 4, pp. 301-350, 1923. Translation published as Laws of Organization in Perceptual Forms, in Ellis, W. (1938). A source book of Gestalt psychology (pp. 71-88). London: Routledge & Kegan Paul.

[32] D.M. WUESCHER and K.L. BOYER, “Robust contour decomposition using constant curvature criterion”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:1, pp. 41-51, 1991.

This page intentionally left blank

J. KEVIN O'REGAN, ERIK MYIN AND ALVA NOË

TOWARDS AN ANALYTIC PHENOMENOLOGY: THE CONCEPTS OF "BODILINESS" AND "GRABBINESS"[1]

1. PHENOMENAL CONSCIOUSNESS

In this paper, we present an account of phenomenal consciousness. Phenomenal consciousness is experience, and the *problem* of phenomenal consciousness is to explain how physical processes – behavioral, neural, computational – can produce experience. Numerous thinkers have argued that phenomenal consciousness cannot be explained in functional, neural or information-processing terms (e.g. Block 1990, 1994; Chalmers 1996). Different arguments have been put forward. For example, it has been argued that two individuals could be exactly alike in functional/computational/behavioral measures, but differ in the character of their experience. Though such persons would behave in the same way, they would differ in how things felt to them (for example, red things might give rise to the experience in one that green things give rise to in the other). Similarly, it has been held that two individuals could be functionally/computationally/behaviorally alike although one of them, but not the other, is a mere *zombie*, that is, a robot-like creature who acts *as if* it has experience but is in fact phenomenally unconsciousness. For any being, it has been suggested, the question whether it has experience (is phenomenally conscious) cannot be answered by determining that it is an information-processor of this or that sort. The question is properly equivalent to the question whether *there is anything it is like to be* that individual (Nagel 1974). Attempts to explain consciousness in physical or information-processing terms sputter: we cannot get any explanatory purchase on experience when we try to explain it in terms of neural or computational processes. Once a particular process has been proposed as an explanation, we can then always reasonably wonder, it seems, what it is about *that particular* process that makes it give rise to experience. Physical and computational mechanisms, it seems, require some further ingredient if they are to explain experience. This explanatory shortfall has aptly been referred to as "the explanatory gap" (Levine 1983).

We suggest that the explanatory gap is a product of a way of thinking about consciousness which sets up three obstacles to an explanation, that is, three reasons for holding that the explanatory gap is unbridgeable. In this paper we propose ways of surmounting these obstacles, and in this way try to lay the foundations for a science of phenomenal consciousness.

What is it exactly about phenomenal consciousness which makes it seem inaccessible to normal scientific inquiry? What is so special about "feel"?

A. Carsetti (ed.), Seeing, Thinking and Knowing, 103-114.

2. THE FIRST OBSTACLE: THE CONTINUOUSNESS OF EXPERIENCE

A first remarkable aspect about experience is that it seems 'continuous'. Experiences seem to be "present" to us, and to have an "ongoing" or "occurring" quality which we might picturesquely describe as like the buzzing, whirring, or humming of a machine.

Many scientists believe that to explain the ongoingness of experience we must uncover some kind of neural process or activity that *generates* this ongoingness. But this is a mistake (Dennett 1991; Pessoa, Thompson and Noë 1998;). To see why, consider an analogy. Most people would agree that there is something it is like to drive a car, and different cars have different "feels". You have the Porsche driving feel when you know that if you press the accelerator, the car will whoosh forwards, whereas nothing comparable happens in other cars. In a Porsche, if you just lightly touch the steering wheel, the car swerves around, whereas most other cars react more sluggishly. In general: the feel of driving a car, truck, tank, tractor or golf-cart corresponds to the specific way it behaves as you handle it.

Now as you drive the Porsche, you are having the ongoing Porsche driving feel. But notice that as you drive you can momentarily close your eyes, take your hands off the steering wheel and your foot off the accelerator, yet you are still having the Porsche driving feel even though you are getting virtually no Porsche-related sensory input. This is because the Porsche driving feel does not reside in any particular momentary sensory input, but rather in the fact that you are currently engaged in exercising the Porsche driving skill.

If the feel of Porsche driving is constituted by exercising a skill, perhaps the feel of red, the sound of a bell, the smell of a rose also correspond to skills being exercised. Taking this view about what feel is would have a tremendous advantage: we would have crossed the first hurdle over the explanatory gap, because now we no longer need a magical neural mechanism to generate ongoing feel out of nerve activities. Feel is now not "generated" by a neural mechanism at all, rather, it is exercising what the neural mechanism *allows the organism to do*. It is exercising a skill that the organism has mastery of.

An analogy can be made with "life": life is not something which is generated by some special organ in biological systems. Life is a *capacity* that living systems possess. An organism is alive when it is *has the potential* to do certain things, like replicate, move, metabolize, etc. But it need not be doing any of them right now, and still it is alive.

It may seem very peculiar to conceive of say, the feel of red, as a skill being exercised, but we shall see the possibility of this position, as well as its advantages, in the next sections. The idea and its implications has been developed in our previous papers (O'Regan & Noë 2001a; O'Regan & Noë 2001b; O'Regan & Noë 2001c; Myin & O'Regan 2002; cf. also Clark 2000; Grush 1998; Järvilheto 2001; Myin 2001, Pettit 2003a,b for similar recent views).

A CONSEQUENCE OF THE "SKILL" IDEA: CHANGE BLINDNESS

When we look out upon the world, we have the impression of seeing a rich, continuously present visual panorama spread out before us. Under the idea that seeing involves exercising a skill however, the richness and continuity of this sensation are not due to the activation in our brains of a neural representation of the outside world. On the contrary, the ongoingness and richness of the sensation derive from the knowledge we have of the many different things we can do (but need not do) with our eyes, and the sensory effects that result from doing them (O'Regan 1992). Having the impression of a whole scene before us comes, not from every bit of the scene being present in our minds, but from every bit of the scene being immediately available for "handling" by the slightest flick of the eye.

But now a curious prediction can be made. Only part of the scene can be being "handled" at any one moment. The rest of the scene, although perceived as present, is actually not being handled. If such currently un-handled scene areas were to be surreptitiously replaced, the change should go unnoticed.

Under normal circumstances any change made in a scene will provoke an eye movement to the locus of the change. This is because there are hard-wired detectors in the visual system that react to any sudden change in local luminance and cause attention to focus on the change. (We will come back to this important property of the visual system under the heading of "grabbiness" in Section 3.)

But by inserting a blank screen or "flicker" (Rensink, O'Regan & Clark 2000), or else an eye movement, a blink, "mudsplashes" (O'Regan, Rensink & Clark 1999), or a film cut between successive images in a sequence of images or movie sequence (for a review see Simons 2000), the sudden local luminance changes that would normally grab attention and cause perceptual handling of a changing scene aspect are drowned out by the mass of other luminance changes occurring in the scene. There will no longer be a single place that the observers' attention will be attracted to, and so we would expect that the likelihood of "handling" and therefore perceiving the location where the scene change occurs would be low.

And indeed that is what is found: surprisingly large changes, occupying areas as large as a fifth of the total picture area, can be missed. This is the phenomenon of "change blindness" (demonstrations can be found on http://nivea.psycho.univ-paris5.fr and http://viscog.beckman.uiuc.edu/change/).

3. THE SECOND OBSTACLE: THE QUALITATIVE CHARACTER OF EXPERIENCE

In the previous section we showed that by taking the view that experiences depend on the exercise of skills, we can forego the search for neural processes that are, like the experiences themselves, ongoing. We no longer need to postulate a magical neural process that "generates" phenomenal consciousness, because, we claim, phenomenal consciousness is not generated: rather it is a skill people exercise.

We now come to the second difficulty in explaining experience.

Suppose you are thinking about your grandmother. You can cast your attention on the color of her eyes, the sound of her voice, the smell of her perfume. Nevertheless, thinking about your grandmother is nothing like actually seeing her: thinking has no perceptual phenomenal quality. Why is this? Why is there something it is like to have a perceptual experience (Nagel 1974)? This question forms the second obstacle that would seem to bar our path towards understanding phenomenal consciousness.

The key, we propose, has to do with distinct properties of the kinds of skills that we exercise when we undergo conscious experience and that make these skills different from other skills (practical skills such as the ability to drive, cognitive skills, etc). These aspects are bodiliness and grabbiness.

BODILINESS

If you really are looking at your grandmother and you turn your eyes, blink, or move your body, there will be an immediate and drastic change in the incoming sensory information about your grandmother. On the other hand, nothing at all will happen if you are merely thinking about your grandmother.

Bodiliness is the fact that when you move your body, incoming sensory information immediately changes. The slightest twitch of an eye muscle displaces the retinal image and produces a large change in the signal coming along the optic nerve. Blinking, moving your head or body will also immediately affect the incoming signal. As concerns auditory information, turning your head immediately affects the phase and amplitude difference between signals coming from the two ears, etc.

Bodiliness is one aspect of *sensory* stimulation which makes it different from other forms of stimulation, and contributes to giving it its peculiar quality. Because of bodiliness, sensory information has an "intimate" quality: it's almost as though it were part of your own body.

GRABBINESS

Suppose that minor brain damage destroys your knowledge about your grandmother's eyeglasses. Are you immediately aware that this has happened? No, the loss of the memory of your grandmother's glasses causes no whistle to blow in your mind to warn you. Only when you cast your mind upon the memory of your grandmother do you actually realize that you no longer know what her glasses were like.

But consider what happens if instead of thinking about your grandmother, you are actually looking at her. Even if you are not paying attention to her glasses in particular, if they should suddenly disappear, this would inevitably grab your attention: the sudden change would trigger local motion detectors in your low-level visual system, and an eye saccade would immediately be peremptorily programmed towards the location of the change. Your attentional resources would be mobilized and you would orient towards the change. This "grabbiness" of sensory stimulation, that is, its capacity to cause automatic orienting responses, is a second aspect which distinguishes it from other types of neural activity in the brain. Grabbiness is the fact that sensory stimulation can grab your attention away from what you were previously doing.

TOWARDS AN ANALYTIC PHENOMENOLOGY

Our claim is that bodiliness and grabbiness are jointly responsible for giving the particular qualitative character to the exercise of sensorimotor skills which people have in mind when they talk of the "feel" of sensation or experience. Because of bodiliness, you are in a way "connected" to sensory stimulation: it changes with your minutest body motion. Because of grabbiness, you somehow can't get away from sensory stimulation: it has the capacity to monopolize your attention and keep you in contact with it. Bodiliness and grabbiness ensure that, unlike thoughts and memories, sensory stimulation has a "clinging" quality. Unlike thoughts and memories, experiences follow you around like a faithful dog. Furthermore, like the dog, they force themselves upon you by grabbing your attention whenever anything unexpected happens in the world. We suggest that bodiliness and grabbiness may be the reason why there is something it's like to have a sensation.

Note an important point about the concepts of bodiliness and grabbiness: they are physically measurable quantities. A scientist should be able to come in and measure how much bodiliness and how much grabbiness there is in different types of sensory stimulation. The amount of bodiliness is determined by the degree to which sensory input depends on body motions. The amount of grabbiness is determined by the extent to which an organism's orienting responses and processing resources are liable to be grabbed by the input.

If bodiliness and grabbiness are objectively measurable quantities, and if we are right in saying that they determine whether a sensory input has "feel", then we should be able to predict how much "feel" different mental phenomena have.

We have already seen that memory phenomena, like the memory of your grandmother, or thoughts or knowledge, have little or no bodiliness and no grabbiness. They have little feel, therefore. This seems to correspond with what people say about memory, thoughts and knowledge.

We have also seen that experiences, like the experience of seeing the color of your grandmother's eyes, have bodiliness and grabbiness, and should be perceived as possessing "feel".

Now it is interesting to consider whether there exist intermediate cases. If we are right about the relation between bodiliness, grabbiness and feel, then cases of a little bit of bodiliness and grabbiness should correspond to a little bit of feel.

Indeed a case in point is Porsche driving. In Porsche driving, some of your body movements produce immediate changes in sensory input – pressing the accelerator, touching the wheel, etc. But most of your body movements do not change sensory input related to the Porsche driving experience. Turning your head changes visual input, but the change is not specific to the Porsche driving feel – rather it constitutes the feel characteristic of vision. Sniffing your nose gives you the smell of leather, but that's specific to the sense of smell. Those very particular sensorimotor contingencies which determine the feel of Porsche driving are restricted to a very particular set of behaviors which are specific to *driving*, namely those to do with how touching the wheel or pressing the accelerator affects what the car does. You can't get the feel of a car by just waving your hands around in the air. You have to actually be exercising the car-driving skill.

The situation is quite different from the feel of seeing red or hearing a bell, say, where almost any small body twitch or muscle movement in the perceptual system involved causes drastic sensory changes (high bodiliness). Moreover, if anything in your visual field suddenly turns red, or if suddenly a bell starts ringing near you, you will be immediately alerted (high grabbiness).

We thus expect – and this corresponds well with what people say about the feel of driving – that it makes sense to say that Porsche driving has a feel, but the feel is less intimate, less direct, less "present" than the sensation associated with seeing red or hearing a bell, because the latter have bodiliness and grabbiness to a much higher degree.

Another interesting intermediate case is the feeling of being rich. What is being rich? It is knowing that if you go to your bank you can take out lots of money; it is knowing you can go on an expensive trip and that you needn't worry about the price of dinner.

Thus being rich has a certain degree of bodiliness, because there exist things you can do with your body which have predictable sensory consequences (e.g. you can make the appropriate maneuvers at the cash dispenser and the money comes out). But clearly, again, the link with body motions is not nearly as direct as in true sensory stimulation like seeing, when the slightest motion of virtually any body part creates immediate changes in sensory input. So being rich can hardly be said to have very much bodiliness.

Similarly, being rich also has no grabbiness. If your bank makes a mistake and suddenly transfers all your assets to charity, no alarm-bell rings in your mind to tell you. No internal mind-siren attracts your attention when the stock market suddenly goes bust: you only find out when you purposely check the news.

Further interesting cases concern obsessive thoughts and experiences like worry and anxiety, as well as embarrassment, fear, love, happiness, sadness, loneliness and homesickness. These are more grabby than normal thinking, because you cannot but help thinking about them. Some of these phenomena also have a degree of bodiliness, because there are things you can do to change them: for homesickness you can go home, for happiness you can remove the things that make you happy. Clearly there is "something it's like" to experience these mental phenomena, but the quality they have is not of a sensory nature[2].

It is interesting to consider also the case of proprioception: this is the neural input that signals mechanical displacements of the muscles and joints. Motor commands which give rise to movements thus necessarily produce proprioceptive input, so proprioception has a high degree of bodiliness. On the other hand, proprioception has no grabbiness: body position changes do not peremptorily cause you to attend to them. Thus, as expected from the classification we are putting forward, while we generally know where our limbs are, this position sense does not have a sensory nature.

The vestibular system detects the position and motion of the head, and so vestibular inputs have bodiliness. They also have some grabbiness, since sudden extreme changes in body orientation immediately result in re-adjusting reactions and grab your attention, sometimes provoking dizziness or nausea. In this sense then, the vestibular sense has a limited degree of sensory feel.

The examples given here are simply a first attempt to use the notions of bodiliness and grabbiness to make a classification of phenomenal processes (but see also O’Regan and Noë 2001a). Further work is needed in this direction. Additionally it may be useful to consider the possibility that there are other objective dimensions that may be useful in creating what could be called an “analytic phenomenology” based on objectively measurable quantities like bodiliness and grabbiness. In particular, to deal adequately with pain and emotions we may additionally need the concept of “automaticity”, which measures the degree to which a stimulation provokes an automatic behavior on the part of the organism.

SUMMARY

We have seen that, when added to the idea that feels correspond to having mastery of skills, the concepts of bodiliness and grabbiness allow the fundamental difference to be captured between mental phenomena that have no feel, like memory and knowledge, and mental phenomena that have feel, like sensations. Bodiliness and grabbiness furthermore allow us to understand why some intermediate situations, like driving or being rich can also be qualified as possessing a certain, but lesser, degree of “feel”. Bodiliness and grabbiness are objectively measurable quantities that determine the extent to which there is something it’s like to have a sensation. We suggest that bodiliness and grabbiness therefore allow us to pass the second obstacle to overcoming the explanatory gap. They explain why there is something it is like to feel.

4. THIRD OBSTACLE: MODALITY AND SENSORY QUALITY

To explain the nature of experience it is necessary not only to explain why there is something it is like to have an experience, one must also explain why it is like this, rather than like that (Hurley and Noë 2003; Chalmers 1995).

For example hearing involves a different quality as compared to seeing, which has a different quality as compared to tactile sensation. Furthermore, within a given sensory modality there are differences as well: for example, red has a different quality from green. This is the third major obstacle to an account of phenomenal consciousness.

Explaining these differences in neural terms will not work: Neural activation is simply a way of coding information in the brain. As of now, we have no clue how differences in the code could ever give rise to differences in feel.

But if we consider experiences as skills, then we can immediately see where their differences in phenomenal quality come from: they come from the nature of the different skills you exercise. Just as Porsche driving is a different skill from tractor driving, the difference between hearing and seeing amounts to the fact that among other things, you are seeing if, when you blink, there is a large change in sensory input; you are hearing if nothing happens when you blink, but, there is a left/right difference when you turn your head; the amplitude of the incoming auditory signal varies in a certain lawful way when you approach a sound source, etc. We call these relations between possible actions and resultant sensory effects: sensorimotor contingencies (O'Regan & Noë 2001b).

SENSORY SUBSTITUTION

From this follows a curious prediction. We claim that the quality of a sensory modality does not derive from the particular sensory input channel or neural circuitry involved in that modality, but from the laws of sensorimotor contingency that are involved. It should therefore be possible to obtain a visual feel from auditory or tactile input, for example, provided the sensorimotor laws that are being obeyed are the laws of vision (and provided the brain has the computing resources to extract those laws).

Such "sensory substitution" has been experimented with since (Bach-y-Rita 1967) constructed the first device to allow blind people to see via tactile stimulation provided by a matrix of vibrators connected to a video camera. Today there is renewed interest in this field, and a number of new devices are being tested with the purpose of substituting different senses: visual-to-tactile (Sampaio, Maris, & Bach-y-Rita 2001); echolocation-to-auditory (Veraart, Cremieux, & Wanet-Defalque 1992); visual-to-auditory (e.g. Meijer 1992; Arno, Capelle, Wanet-Defalque, Catalan-Ahumada, & Veraart (1999)); auditory-to-tactile (cf. Richardson and Frost 1977 for review). Such devices are still in their infancy. In particular, no systematic effort has been undertaken up to now to analyze the laws of sensorimotor contingency that they provide. In our opinion it will be the similarity in the sensorimotor laws that such devices recreate which determines the degree to which users will really feel they are having sensations in the modality being substituted.

Related phenomena which also support the idea that the feel of a sensory modality is not wired into the neural hardware, but is rather a question of sensorimotor contingencies comes from the amusing experiment of Botvinick & Cohen (1998), where the "feel" of being touched can be transferred from your own body to a rubber replica lying on the table in front of you (see also interesting work on the body image in tool use by Yamamoto & Kitazawa 2001; Iriki, Tanaka, & Iwamura 1996). The finding of (Roe et al. 1990) according to which embryonically "rewired" ferrets can see with their auditory cortex can also be interpreted within the context of our theory.

INTRAMODAL SENSORY DIFFERENCES

We have seen that the feel of different sensory modalities can be accounted for by the different things you do when you use these modalities. But what about the differences within a given sensory modality: can we use the same arguments?

Within the tactile modality, this idea seems quite plausible. Consider the feel of a hard surface and the feel of a soft surface. Does this difference come from different kinds of tactile receptors being activated, or from the receptors being activated in different ways? No, we argue, since receptor activations are only codes that convey information – they are necessary for feel, but cannot by themselves generate the feel of hard and soft. On the contrary, we claim the difference between hard and soft comes from the different skills that you implicitly put to work when you touch hard and soft surfaces: the fact that when you push on a hard surface it resists your pressure; when you push on a soft surface, it gives way. The feel of hard and soft are constituted by the things you implicitly know about how the surface will react to your ongoing exploration.

Now while this makes sense for tactile exploration, it might seem difficult to apply the same approach to other sensory modalities: what has the difference between red and green for example, got to do with sensorimotor contingencies? How can the feel of red consist in doing something, and the feel of green consist in doing something else?

But consider what happens when you look at a red piece of paper. Depending on which way you turn the paper, it can reflect more of bluish sky light or more of yellowish sunlight from your window, or more of reddish lamplight from your desk. We suggest that one aspect of the feel of red is: knowing the laws that govern the changes in the light reflected off the paper as you turn it (cf. Broackes 1992).

Another aspect of the skill involved in the feel of red concerns retinal sampling. Retinal sampling of a centrally fixated red patch is done by a densely packed matrix of short, medium and long-wavelength sensitive cones. There is also a yellowish macular pigment which covers the central retina. When an eye movement brings the patch into peripheral vision, the cone matrix that samples the patch is interspersed with rods, the distribution is slightly different, and there is no macular pigment. The resultant change in quality of the incoming sensory stimulation is another aspect of what it is like to be looking at a red patch.

5. SUMMARY: HOW WE HAVE CROSSED THE GAP

We have presented arguments showing how three obstacles to understanding experience can be circumvented.

The first obstacle was the fact that experiences appear to be ongoing, occurrent processes inside us. This has led scientists to seek for brain mechanisms which are themselves also ongoing, and whose activity gives rise to feel. But we claim that any such quest is doomed, since the question will always ultimately remain of how activity of a physical system, no matter how complex or abstruse, can give rise to "feel".

Our solution is to show that feel is not directly generated by a brain mechanism, but consists in the active exercising of a skill, like driving or bicycle riding. The ongoingness of feel is not "produced" or "secreted" by brain activity, but resides in the active doing, the give-and-take that is involved in exercising a particular skill.

The second barrier to explaining feel is the question of there being something it is like to have the experience, that is, of the experience having a qualitative character. We showed how the concepts of bodiliness and grabbiness allow the fundamental difference to be captured between mental phenomena that have no feel, like memory and knowledge, and mental phenomena that have feel, like experiences or sensations. Bodiliness and grabbiness are objectively measurable quantities that determine the extent to which there is something it's like to have a sensation. Bodiliness and grabbiness allow us to pass the second obstacle to overcoming the explanatory gap. They explain why there is something it's like to feel.

The third obstacle preventing a scientific explanation of experience was that it was difficult to understand how different types of neural activation could give rise to different types of experience, e.g. experiential differences within and between sensory modalities: after all, neural activations are just arbitrary codes for information, and information in itself has no feel.

A natural solution comes from the idea that differences in the feel of different sense modalities correspond to the different skills that are involved in exercising each modality. This idea can also be made to work within a given sense modality, explaining the what-it-is-like of red versus green in terms of the different things you do when you are exploring red and green.

HOW TO MAKE A ROBOT FEEL

With these tools in hand, can we build a robot that feels?

We provide the robot with mastery of the laws that govern the way its actions affect its sensory input. We wire up its sensory receptors so that they provide bodiliness and we ensure grabbiness by arranging things so that sudden sensory changes peremptorily mobilize the robot's processing resources. Will the robot now have "feel"?

No, one more thing is necessary: the robot must have access to the fact that it has mastery of the skills associated with its sensory exploration. That is, it must be able to make use of these sensory skills in its thoughts, planning, judgment and (if it talks) in its language behavior.

Reasoning, thought, judgment and language are aspects of mind where AI and robotics have not yet reached human levels. But there is no a priori, logical argument that prevents this from being possible in the future. This is because there is no barrier in principle that prevents reasoning, thought, judgment, and language from being described in functional terms. They are therefore in principle amenable to the scientific method and can theoretically be implemented by an information-processing device. Of course, because human reasoning is intricately linked with human culture and social interaction, it may not be possible to satisfactorily replicate human reasoning without also replicating the social and developmental process through which each human goes.

But when we manage to do this, then if we make a robot whose sensory systems possess bodiliness and grabbiness, then the robot will feel. Indeed, it will feel for the same reasons that we do, namely because we have access to our mastery of sensory skills, and because of the bodiliness and grabbiness of sensory inputs.

ACKNOWLEDGMENTS

E. M. wishes to thank, for support, the Fund for Scientific Research-Flanders (Belgium) (FWO project G.0175.01). A.N. gratefully acknowledges the support of faculty research funds granted by the University of California, Santa Cruz as well as that of a UC President's Research Fellowship in the Humanities.

Prof. J. Kevin O'Regan
Laboratoire de Psychologie Expérimentale
Univ. René Descartes
71 av. E. Vaillant
92774, Boulogne-Billancourt, France

Prof. Erik Myin
Centre for Logic and Philosophy of Science
Vrije Universiteit
Pleinlaan 2
B1050, Brussel, Belgium

Prof. Alva Noë
Department of Philosophy
University of California
CA 95060, Santa Cruz, U.S.A.

NOTES

[1] This paper offers a theoretical overview of ideas developed in an a series of recent papers – O'Regan and Noë 2001a, b; c; Myin and O'Regan 2002; Noë and O'Regan 2000; 2002; Noë 2002; O'Regan 1992 – and also in work in progress by the authors.

[2] But note that the grabbiness involved in these phenomena is "mental" or "psychological" rather than sensory: it is not automatic orienting of sensory systems, but rather uncontrollable, obsessive mental orienting.

REFERENCES

ARNO, P., CAPELLE, C., WANET-DEFALQUE, M.-C., CATALAN-AHUMADA, M., & VERAART, C. (1999). "Auditory coding of visual patterns for the blind." *Perception,* 28, 8, pp. 1013-1029.

BACH-Y-RITA, P. (1967). "Sensory plasticity. Applications to a vision substitution system", *Acta Neurologica Scandinavica,* 43, 4, pp. 417-426.

BLOCK, N. (1995). "On a confusion about a function of consciousness", *Behavioral and Brain Sciences,* 18, 2, pp. 227-247.

BOTVINICK, M., & COHEN, J. (1998). "Rubber hands 'feel' touch that eyes see [letter]", *Nature,* 391, 6669, pp. 756.

BROACKES, J. (1992). "The autonomy of colour". In K. Lennon & D. Charles (Eds.), *Reduction, Explanation, and Realism*, Oxford University Press, pp. 421-465.

CHALMERS, D. J. (1996). *The conscious mind: In search of a fundamental theory*, Oxford University Press.

CLARK, A. (2000). "Phenomenal immediacy and the doors of sensation", *Journal of Consciousness Studies,* 7, 4, pp. 21-24.

DENNETT, D. C. (1991). *Consciousness Explained*, Little Brown.

GRUSH, R. (1998). "Skill and spatial content", *The Electronic Journal of Analytic Philosophy* (Vol. 6), http://www.phil.indiana.edu/ejap.

HURLEY, S. L. (1998). *Consciousness in Action*, Harvard.

HURLEY, S. L., & NOË, A. (2003). "Neural plasticity and consciousness". *Biology and Philosophy*, 18, I, in press.

IRIKI, A., TANAKA, M., & IWAMURA, Y. (1996). "Coding of modified body schema during tool use by macaque postcentral neurones", *Neuroreport,* 7, 14, pp. 2325-2330.

JÄRVILHETO, T. (2001). "Feeling as knowing - Part 2. Emotion, consciousness, and brain activity.", *Consciousness & Emotion*, 2, 75-102.

LEVINE, J. (1983). "Materialism and qualia: The explanatory gap.", *Pacific Philosophical Quarterly,* 64, pp. 354-361.

MEIJER, P. B. L. (1992). "An experimental system for auditory image representations", *IEEE Transactions on Biomedical Engineering,* 39, 2, pp. 112-121.

MYIN, E. (2001), "Color and the duplication assumption", *Synthese*, 129**:** 1. pp. 61-77.

MYIN, E., & O'REGAN, J. K. (2002). "Perceptual consciousness, access to modality and skill theories: A way to naturalize phenomenology?", *Journal of Consciousness Studies,* 9, 1, pp. 27-45.

NAGEL, T. (1974). "What is it like to be a bat?", *Philosophical Review,* 83, pp. 435-456.

NOË, A. (2002). "On what we see.", *Pacific Philosophical Quarterly*, 83:1

NOË, A. & O'REGAN, J.K. (2000). "Perception, attention and the grand illusion", *Psyche*, 6 (15). URL: http://psyche.cs.monash.edu.au/v6/psyche-6-15-noe.html.

NOË, A. & O'REGAN, J.K. (2002). "On the brain-basis of visual consciousness". In A. Noë & E. Thompson: *Vision and Mind: Selected Readings in the Philosophy of Perception*. MIT Press.

O'REGAN, J. K. (1992), "Solving the 'real' mysteries of visual perception: the world as an outside memory", *Canadian Journal of Psychology,* 46, pp. 461-488.

O'REGAN, J. K., RENSINK, R. A., & CLARK, J. J. (1999). "Change-blindness as a result of 'mudsplashes'.", *Nature*, 398, p. 34.

O'REGAN, J. K., & NOË, A. (2001a). "Acting out our sensory experience", *Behavioral and Brain Sciences,* 2, 5, pp. 955-975.

O'REGAN, J. K., & NOË, A. (2001b). "A sensorimotor account of vision and visual consciousness", *Behavioral and Brain Sciences,* 24, 5, pp. 883-917.

O'REGAN, J. K., & NOË, A. (2001c). "What it is like to see: A sensorimotor theory of perceptual experience", *Synthese,* 129, 1, pp. 79-103.

PESSOA, L.,THOMPSON, E, & NOË, A. (1998). "Finding Out About Filling In: A Guide to Perceptual Completion for Visual Science and the Philosophy of Perception", *Behavioral and Brain Sciences*, 21: 6, pp. 723-748.

PETTIT, P. (2003a). "Looks red", *Philosophical Issues,* 13, in press.

PETTIT, P. (2003b). "Motion blindness and the knowledge argument.". In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *The Knowledge Argument*, MIT Press.

RENSINK, R. A., O'REGAN, J. K., & CLARK, J. J. (2000). 'On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, 7, 1-3, pp. 127-145**.**

RICHARDSON, B. L., & FROST, B. J. (1977). "Sensory substitution and the design of an artificial ear", *J Psychol,* 96, 2d Half, pp. 259-285.

ROE, A. W., PALLAS, S. L., HAHM, J. O., & SUR, M. (1990). "A map of visual space induced in primary auditory cortex", *Science,* 250, 4982, pp. 818-820.

SAMPAIO, E., MARIS, S., & BACH-Y-RITA, P. (2001). "Brain plasticity: 'visual' acuity of blind persons via the tongue", *Brain Research,* 908, 2, pp. 204-207.

SIMONS, D. (2000). Current approaches to change blindness. *Visual Cognition***,** **7**, pp. 1-16.

VERAART, C., CREMIEUX, J., & WANET-DEFALQUE, M. C. (1992). "Use of an ultrasonic echolocation prosthesis by early visually deprived cats", *Behavioral Neuroscience,* 106, 1, pp. 203-216.

YAMAMOTO, S., & KITAZAWA, S. (2001). "Sensation at the tips of invisible tools", *Nature Neuroscience,* 4, 10, pp. 979-980.

ANDREA DI FERDINANDO AND DOMENICO PARISI

INTERNAL REPRESENTATIONS OF SENSORY INPUT REFLECT THE MOTOR OUTPUT WITH WHICH ORGANISMS RESPOND TO THE INPUT

1. INTRODUCTION

What determines how sensory input is internally represented? The traditional answer is that internal representations of sensory input reflect the properties of the input. This answer is based on a passive or contemplative view of our knowledge of the world which is rooted in the philosophical tradition and, in psychology, appears to be almost mandatory given the fact that, in laboratory experiments, it is much easier for the researcher to control and manipulate the sensory input which is presented to the experimental subjects than the motor output with which the subjects respond to the input. However, a minority view which is gaining increasing support (Gibson, 1986; O'Regan and Noe, in press) is that internal representations are instead action-based, that is, that the manner in which organisms internally represent the sensory input reflects the properties of the actions with which the organisms respond to the sensory input rather than the properties of the sensory input.

In this chapter we describe a series of computer simulations using neural networks that tend to support the action-based view of internal representations. Internal representations in neural networks are not symbolic or semantic entities, like cognitivist representations (Fodor, 1981), but they are patterns of activation states in the network's internal units which are caused by input activation patterns and which in turn cause activation patterns in the network's output units. Our networks are sensory-motor neural networks. Their input units encode sensory input and their output units encode changes in the physical location of the organism's body or body parts, i.e., movements. We train networks to execute a number of sensory-motor tasks and by examining their internal representations at the end of training we determine whether these internal representations co-vary with the properties of the sensory input or with the properties of the motor output.

The chapter describes three sets of simulations. In the first set we distinguish between micro-actions and macro-actions and we show that both micro-actions and macro-actions are real for neural networks. Micro-actions are the successive movements that make up an entire goal-directed action, and each micro-action is encoded in the activation pattern observed in the network's motor output units in a single input/output cycle. Macro-actions are sequences of micro-actions that allow the organism to reach some goal. Our simulations show that internal representations encode, i.e., reflect the properties of, macro-actions. In the second set of simulations we show that if there is a succession of layers of internal units from the sensory input to the motor output the layers which are closer to the sensory input will tend to reflect the properties of the input and those closer to the motor output the properties of the output. However, in the third and final set of simulations we also show that the actions with which the organism responds to the input dictate the form of internal representations as low down the succession of in-

A. Carsetti (ed.), Seeing, Thinking and Knowing, 115-141.

ternal layers and as close to the sensory input as is necessary for producing the appropriate actions in response to the input.

2. MACRO-ACTIONS AND MICRO-ACTIONS

The behavior of organisms can be described at various levels of integration. An organism can be said to be "reaching for an object with its arm", or one can describe the sequence of micro-movements of the organism's arm that allow the organism to reach the object. The first type of description is in terms of macro-actions, the second one in terms of micro-actions (or micro-movements). Macro-actions are composed of sequences of micro-actions and typically one and the same macro-action is realized in different occasions by different sequences of micro-actions. The object can be in different spatial locations or the arm's starting position can vary and, as a consequence, the arm's trajectory will be different and will be composed of different sequences of micro-actions although at the macro-action level the organism is in all cases "reaching for an object".

Behavior can be modeled using neural networks, which are computational models inspired by the physical structure and way of functioning of the nervous system (Rumelhart & McClelland, 1986). Neural networks are sets of units (neurons) that influence each other through their connections (synapses between neurons). Activation states propagate from input units to internal units to output units. The network's behavior, i.e., the way in which the neural network responds to the input by generating some particular output, depends on the network's connection weights. Neural networks are trained in such a way that the initial random connection weights are progressively modified and at the end of training the neural network exhibits the desired behavior.

One can train networks to exhibit the behavior of reaching for objects using a two-segment arm. Some external event, e.g., the light reflected by an object, determines a particular activation pattern in the network's (visual) input units. The activation propagates through the network until it reaches the output units and determines a particular activation pattern in the network's output units which is then translated into a micro-movement of the arm. This is repeated for a succession of input/output cycles until the arm's endpoint (the hand) reaches the object. Once the behavior of reaching for objects has been acquired by the network, the behavior can be described at the macro- and at the micro-level. For example, one can either say "the network is reaching the object in the left portion of the visual space", which corresponds to an entire sequence of output activation patterns, or one can describe each single output activation pattern which controls a single micro-action of the arm as the arm "is reaching the object on the left".

The first problem addressed in this chapter is whether the distinction between macro-actions and micro-actions makes sense only for the researcher who is describing the behavior of the network (or of a real organism) or is also appropriate from the point of view of the network itself (or the organism). Micro-actions obviously are real for the network in that in each cycle one can "read" the activation pattern in the network's output units which determines a micro-action. However, one can doubt that macro-actions are also real for the network in that it is not at all clear where one can find macro-actions in the network's structure or organization. The network can be said to know "how to move the arm to reach the object in the left portion of the visual space" because this is what the network does but this knowledge seems to be purely implicit, not explicit. Explicitly, the network only knows how to produce the slight displacements of the arm which are encoded in the activation pattern of its output units in each input/output cycle. One could even say that while the micro-level is sub-symbolic and quantitative in that it is expressed by the vector of activation states observed in each cycle in the network's output units, the macro-level is symbolic and qualitative in that humans use language to generate descriptions of macro-actions. Neural networks are said to be subsymbolic and quantitative systems and in any case the particular networks we are discussing do not have language. Hence, micro-actions seem to be real for them but macro-actions don't.

We will describe some simulations that attempt to show that both micro-and macro-actions are real for neural networks and that neural networks have an explicit, and not only an implicit, knowledge of macro-actions. We train the neural networks to reach for objects. After the desired behavior has been acquired we examine the internal organization of individual networks using two different methods: we measure the activation level of the network's internal units in response to each possible input and we lesion the network's internal units and connections and observe the changes in behavior that result from these lesions. From these analyses we conclude that both micro-actions and macro-actions are real for neural networks in that both micro-actions and macro-actions are explicitly represented in the networks' internal structure.

2.1 SIMULATIONS

An artificial organism lives in a bidimensional world which contains only two objects, object A and object B (Figure 1).

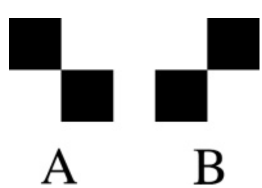

Figure 1. The two objects

At any given time the organism sees either only one of the two objects, A or B, or both objects at the same time. When a single object is seen, the object may appear either in the left or in the right half of the organism's visual field. When the organism sees the two objects together, object A can appear in the left visual field and object B in the right field, or viceversa. The possible contents of the organism's total visual field at different times are shown in Figure 2.

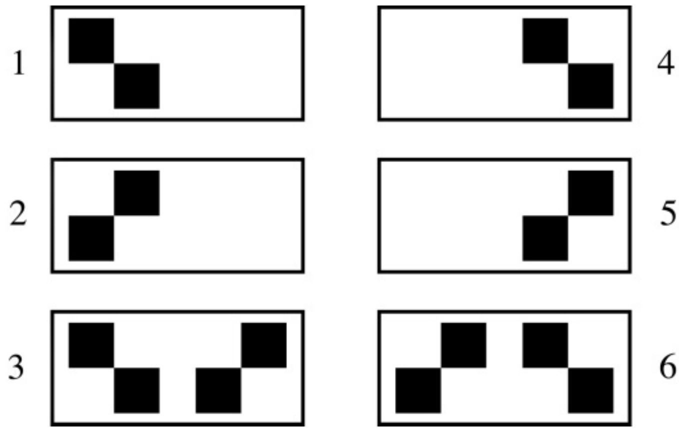

Figure 2. At any given time the organism sees one of these six visual scenes

The organism has a single two-segment arm with which the organism can reach the objects by moving the arm in such a way that the arm's endpoint (the hand) eventually ends up on an object. The objects are always located within reaching distance from the organism (Figure 3). When a single object is presented the organism has to reach for it independently from the location of the object in the left or right field and independently from whether the object is A or B. When both object A and object B are presented the organism has always to reach for object A, ignoring object B, independently from whether object A is in the left or in the right field. In summary, for the first three visual patterns of Figure 2 the organism has to reach for the object on the left side of its visual field, whereas for the last three visual patterns it has to reach for the object on the right side of its visual field.

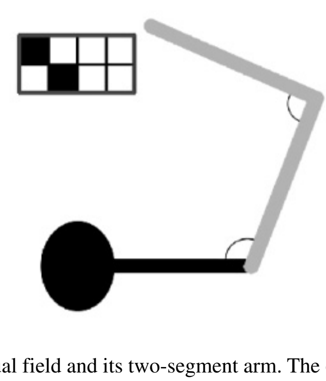

Figure 3. The organism with its total visual field and its two-segment arm. The organism is currently seeing an A object in the left field.

How should the organism's nervous system be internally organized so that the organism is able to exhibit this kind of behavior? We will try to answer this question by simulating the organism's nervous system using a neural network and the acquisition of the behavior we have described using a genetic algorithm for selecting the appropriate connection weights for the neural network.

The neural network that controls the organism's behavior has one layer of input (sensory) units, one layer of output (motor) units, and one layer of internal units. The input layer includes two distinct sets of units, one for the visual input and one for the proprioceptive input which tells the network what is the current position of the arm. The organism has a 'retina' divided up into a left and a right portion. Each portion is constituted by a small grid of 2x2=4 cells. Hence the whole retina is made up of 4+4=8 cells. Each cell of the retina corresponds to one input unit. Hence, there are 8 visual input units. These units can have an activation value of either 1 or 0. An object is represented as a pattern of filled cells appearing in the left or in the right portion of the retina (cf. Figure 1). The cells occupied by the pattern determine an activation value of 1 in the corresponding input unit and the empty cells an activation value of 0. The proprioceptive input is encoded in two additional input units. These units have a continuous activation value that can vary from 0 to 3.14 corresponding to an angle measured in radiants. The organism's arm is made up of two segments, a proximal segment and a distal segment (cf. Figure 3). One proprioceptive input unit encodes the current value of the angle of the proximal segment with respect to the shoulder. The other proprioceptive unit encodes the value of the angle of the distal segment with respect to the proximal segment. In both cases the maximum value of the angle is 180 degrees. The current value of each angle is mapped in the interval between 0 (0° angle) and 3.14 (180° angle) and this number represents the activation value of the corresponding proprioceptive unit. Since the visual scene does not change across a given number of successive input/output cycles whereas the organism moves its arm during this period of time, the visual input for the organism remains identical during this period of time but the proprioceptive input may change if the organism moves its arm.

The network's output layer is made up of two units which encode the arm's movements, one unit for the proximal segment and the other unit for the distal segment. The activation value of each output unit varies continuously from 0 to 1 and is mapped into an angle which can vary from –10° to +10°. This angle is added to the current angle of each of the arm's two segments resulting in a movement of the arm. However, if the unit's activation value happens to be in the interval between 0.45 and 0.55, this value is mapped into a 0° angle, which means that the corresponding arm segment does not move. Hence, after moving the arm in response to the visual input for a while, the network can decide to completely stop the arm by generating activation values between 0.45 and 0.55 in both output units.

The 8 visual input units project to a layer of 4 internal units which in turn are connected with the 2 motor output units. Therefore the visual input is transformed at the level of the internal units before it has a chance to influence the motor output. On the contrary, the proprioceptive input directly influences the motor output. The two input units that encode the current position of the arm are directly connected with the two output units which determine the arm's movements. The entire neural architecture is schematized in Figure 4.

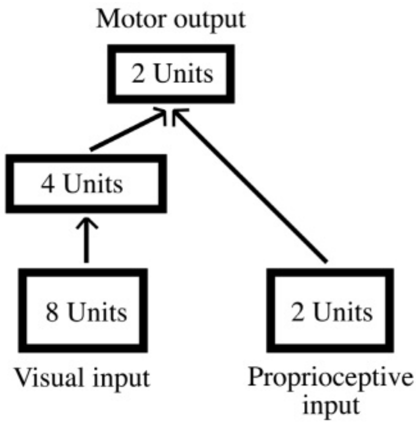

Figure 4. Neural network architecture of the organism

How a neural network responds to the input depends on the network's connection weights. To find the connection weights that result in the behavior of reaching for an A or B object when presented alone and reaching for the A object when presented together with the B object, we have used a genetic algorithm, a computational procedure that mimics evolutionary change (Holland, 1975). An initial population of 100 neural networks is created by assigning randomly selected connection weights to the neural networks. Each individual network lives a life of a maximum of 600 time steps (input/output cycles) divided up into 10 epochs of 60 time steps each. (An epoch is terminated when an object is reached.) During each epoch one of the six possible visual inputs of Figure 2, randomly chosen, is presented to the individual and this visual input remains the same during the entire epoch. However, since the organism can move its arm to reach the object, the proprioceptive input can vary during an epoch. Moreover, since at the beginning of an epoch the arm is positioned in a randomly selected starting position, the initial proprioceptive input varies in each epoch.

At the end of life each individual is assigned a total fitness value which is the average of the fitness values obtained by the individual in each of the 10 epochs. An epoch's fitness value is +1 if the correct object has been reached and is -1 if the incorrect object has been reached, i.e., if the organism reaches the B object with the A object also present. An object is considered as reached if, when the arm stops, the arm's endpoint happens to be located within 10 pixels from the object's center, i.e., from the point in which the two little squares that make up the object touch each other. Furthermore, the fitness value of each epoch is reduced by a small amount which increases with the squared distance between the point in which the arm stops and the object's center. In other words, an individual is rewarded for stopping its arm as close as possible to the object's center.

The fitness formula is the following:

$$\text{Fitness} = \frac{(\text{nCorrect} - \text{nIncorrect}) - k \sum_{i=1}^{10} \text{Distance}_i^{\,2}}{10}$$

where k = 0.001, nCorrect = number of objects correctly reached, nIncorrect = number of objects incorrectly reached, Distance_i = distance between target and final hand position for the epoch *i*, 10 = number of epochs. If in a particular epoch the distance is greater than 100 pixels or the arm does not stop, the distance is considered as equal to 100 pixels.

The 20 networks with the highest fitness are selected for reproduction. The weight values of all the connections of an individual neural network are encoded in the network's genotype. A network which is selected for reproduction generates 5 copies of its genotype and each copy is assigned to one of 5 new networks (offspring). Each copy is slightly modified by adding a quantity randomly selected in the interval between +1 and –1 to the current value of 10% (on average) of the weights (genetic mutations). The 20x5=100 new networks constitute the next generation. The process is repeated for 10,000 generations.

In the early generations the behavior of the organism is not very good but the selective reproduction of the best individuals and the constant addition of new variants by the genetic mutations (reproduction is nonsexual) result in a progressive increase in the average fitness of the population so that after a certain number of generations most individuals in the population exhibit the behavior we have described (Figure 5): when an individual sees a single object, it reaches the object whether it is an A or B object and whether the object appears in its left or right field; when the organism perceives two objects at the same time, it reaches the A object and ignores the B object both when the A object is in the left field and when it is in the right field.

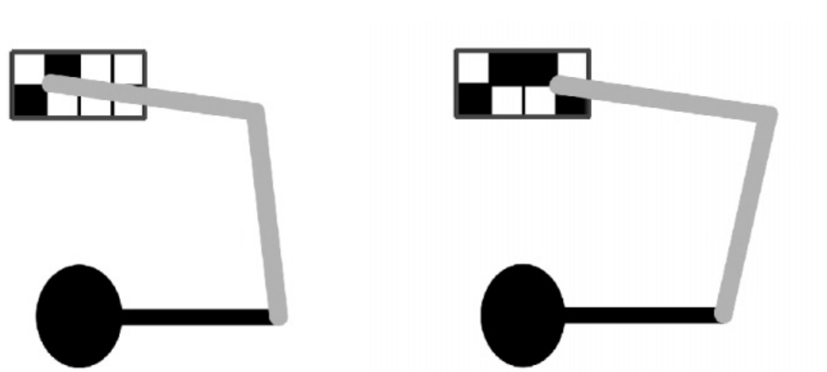

Figure 5. The organism correctly reaches the B object presented alone in the left visual field (left) and the A object presented in the right visual field together with the B object in the left visual field (right)

Figure 6 shows the increase in fitness across 10,000 generations for the single best individual and for all individuals in each generation. The results are the average of 10 replications of the simulations starting with randomly selected initial conditions (different "seeds" for the initial assignment of connection weights, for the initial starting position of the arm in each trial, etc.).

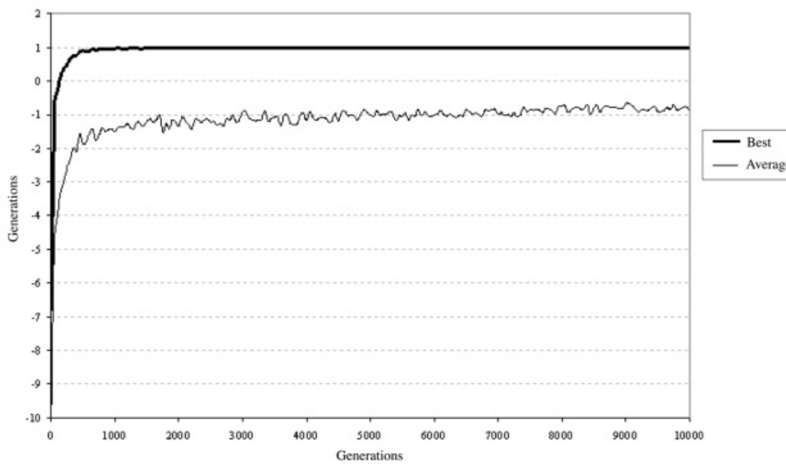

Figure 6. Increase in fitness across 10,000 generations for the single best individual and for all the individuals in each generation (average). The maximum fitness value is 1, which means that in all epochs the correct object has been reached and the arm's endpoint stops exactly on the center of the object.

The robustness of the behavior which has been acquired is demonstrated by a generalization test in which an individual is exposed to all 6 possible visual inputs and to 5 randomly chosen initial positions of the arm for each of the 6 visual inputs, for a total of 30 different inputs. The result of this test, which has been conducted on all the individuals of the last generation for all 10 replications of the simulation, show that the 20 best individuals, that is, those individuals which are selected for reproduction, correctly reach the target object almost always.

2.2 ANALYSIS OF THE INTERNAL ORGANIZATION OF THE NEURAL NETWORKS

To determine how the neural networks of our organisms are internally organized as a result of evolution in order to exhibit the behavior which has been described, we measure the activation level of each of the 4 units of the internal layer of the organisms' neural networks in response to each of the six visual stimuli. This has been done for 10 individuals, i.e., the best individuals of the last generation in each of the 10 replications of the simulation. The results are shown in Figure 7.

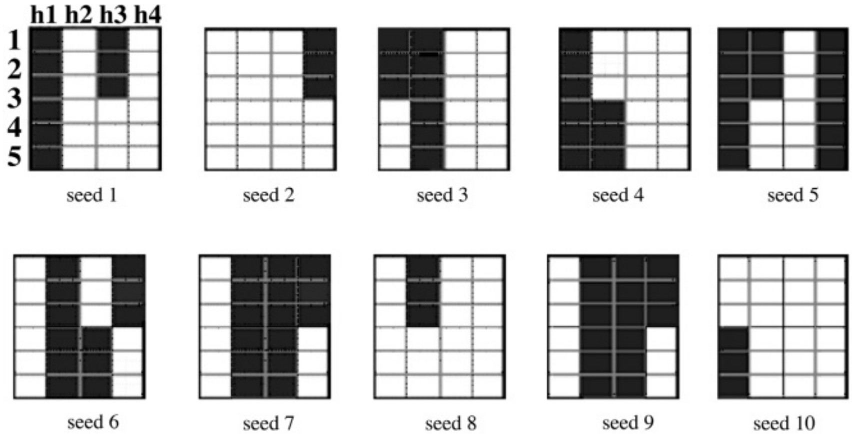

Figure 7. Activation level of the 4 internal units (columns) in response to the six visual stimuli (rows) for the single best individual of the last generation in each of the 10 replications of the simulation. Although the activation level is continuous and is mapped into a grey scale (0 = black and 1 = white), observed activation levels are quite extreme.

The results of this analysis show that in most individuals there are three types of internal units. The first type of internal units (black columns in Figure 7) exhibit an activation level of near zero in response to all possible visual inputs. In other words, these units play no role in determining the neural network's output, i.e., the arm movements. They appear to be useless, at least at the end of evolution. Notice however that these units may play a role during the course of evolution even if they play no role at the end. In fact, if we reduce the number of internal units the same terminal level of performance is eventually reached but it is reached more slowly. In other words, 'starting big', i.e., with more computational resources, may help even if at the end of evolution some of the resources are not used (Miglino and Walker, 2002).

The second type of internal units are units which are invariably highly activated in response to all possible visual inputs (white columns). The activation level of these units also does not co-vary with the input, exactly like the zero activation units we have already described, but these units do have a role in determining the observed behavior. By being constantly activated with an activation level of almost 1 they influence the motor output of the network through the particular weight values of the connections linking them to the output units.

The third and final type of internal units (black and white columns) are those units that have an activation level which is not always the same but varies as a function of the visual input. However, the activation level of these units cannot be said to vary with the visual input in the sense that each of the 6 different visual inputs elicits a different activation level in these units. On the contrary, these units tend to exhibit one of only two different activation levels (and two rather extreme activation levels since one is near zero and the other one near 1) while there are 6 different visual inputs and, furthermore, these two activation levels are correlated with the network's motor output rather than to the network's visual input. If we examine the organism's behavior we see that these internal units have one particular activation level, say 1, when the network is responding to the visual input by moving the arm towards the left visual space, and they have a very different activation level, i.e. 0, when the movement of the arm is toward the right visual space. Notice that our organisms 'move their arm to the left' in response to different visual inputs, i.e., both when there is a single object, A or B, in the left field and when the A object is in the left field and the B object is in the right field. Similarly, they 'move their arm to the right' both when there is a single object, A or B, in the right field and when the A object is in the right field and the B object in the left field. Hence, this third type of internal units tend to reflect the motor output of the network rather than the sensory input. More precisely, they reflect (encode) the macro-actions with which the organism responds to the different sensory inputs.

2.3 LESIONING THE NEURAL NETWORKS

Another type of analysis that may reveal the internal organization of our neural networks consists in lesioning individual internal units and observing the type of disturbed behavior that emerges as a consequence of these lesions. When an internal unit is lesioned all the connections departing from the unit are cut and therefore the lesioned unit ceases to have any role in determining the activation level of the output units and the organism's behavior. We have lesioned one internal unit at a time of the neural networks of the same individuals already examined in the previous section, i.e., the 10 individuals which are the best ones of the last generation in each of the 10 replications of the simulation.

If we lesion the internal units of the first type, i.e., those units which have a constant activation level of 0, there are no consequences for the organism's behavior, both when the object to be reached is in the left portion of the visual field and when it is in the right portion. This is not surprising since these units play no role in determining the organisms' response to the input and therefore lesioning these units has no damaging effects on the organisms' behavior.

If we lesion the second type of internal units, those with a constant activation level of near 1, the negative consequences for the organism's behavior are always very serious and equally distributed across all types of visual inputs and behavioral responses. The organism appears to be completely unable to reach the objects whatever the position, type, and number of objects in its visual field (Figure 8). More specifically, the arm fails to reach the portion of the total space which is visually perceived by the organism (visual space) and in which the objects are found. In other words, when these units are lesioned, the organism appears to be unable to execute the macro-action which consists in "reaching the visual space".

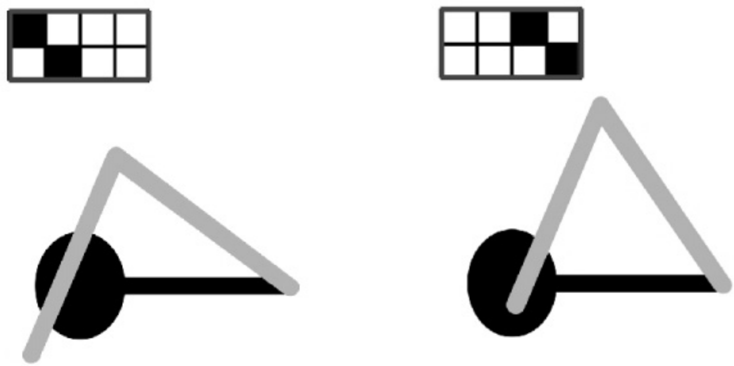

Figure 8. Lesioning an internal unit with a constant activation level of 1 completely disrupts the ability (macro-action) to reach the portion of the space which is visually perceived. The figure shows for a particular individual where the arm stops in response to two sample visual inputs.

A very different type of behavioral damage appears if we lesion the third, selective, type of internal units, i.e., those whose activation level co-varies with the two macro-actions "move the arm toward the left field" and "move the arm toward the right field", respectively. In 9 out of the 10 individuals there is only one unit of this type. Lesioning this unit leads to a form of stereotyped behavior: for different individuals, whatever the visual input, either the organism always moves its arm to the left portion of the visual space or it always moves the arm to the visual space's right portion. Hence, in half the epochs the organism's arm reaches the correct object and in the remaining epochs it either reaches the wrong object (two objects are presented) or the wrong portion of the visual space (only one object is presented) (Figure 9). This appears to be a fortuitous result of the particular position in which the object happens to be located, that is, of whether the correct object happens to lie in the portion of the visual space always reached by the stereotyped behavior of the organism.

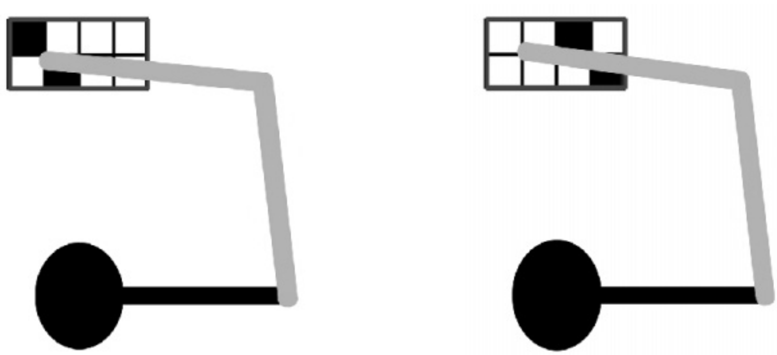

Figure 9. Behavior after lesioning an internal unit whose activation level varies with the visual input in the 9 out of 10 individuals in which there is only one unit of this type. The figure shows where the arm stops in response to two sample visual inputs for an individual in which the internal unit has an activation level of 0 encoding the macroaction "reaching toward the left field" and an activation level of 1 encoding the macro-action "reaching toward the right field".

From the results of these analyses we conclude that the internal units of our networks encode macro-actions. The activation level of the internal units co-varies with the macro-actions of the organism and lesioning these internal units leads to disruption of entire macro-actions.

That the internal layer which receives input from the retina and which therefore constructs internal representations of the visual input, encodes macro-actions can also be shown by contrasting the effects of lesions to the units comprising this internal layer with the effects of lesions to the proprioceptive-to-motor pathway. While the visual-to-motor pathway encodes macro-actions, the proprioceptive-to-motor pathway encodes micro-actions. In other words, the internal layer receiving visual information from the retina tells the network what to do at the macro level, for example “move the arm toward the left portion of the visual space”, while the connection weights from the proprioceptive input units to the motor output units tell the network what to do at the micro level, that is, how to actually implement the macro-action “move the arm toward the left portion of the visual space” given the current and constantly changing position of the arm.

As we have seen, lesions to the visual-to-motor pathway disrupt entire macro-actions. What kind of damage results from lesioning the proprioceptive-to-motor pathway? Since there are no internal units in the proprioceptive-to-motor pathway but the proprioceptive input units directly project to the motor output units, we have lesioned this pathway by introducing some ramdom noise in it. We have added a quantity randomly selected in the interval between –0.2 and +0.2 to the current value of each of the four connection weights linking the 2 proprioceptive input units to the 2 motor output units. The result of this operation is that the behavior of the 10 individuals appears seriously damaged (the percentage of correct responses is 0% for both portions of the visual space) but the behavioral deficit is very different from the deficit observed with lesions to the visual-to-motor pathway. If, for example, the visual input requires reaching for the object in the left visual field, an individual with lesions in the proprioceptive-to-motor pathway is still able to move its arm toward the left visual field (which implies that the macro-action is preserved) but the arm stops when the arm’s endpoint is still somewhat distant from the object. Hence, the object is not actually reached (Figure 10).

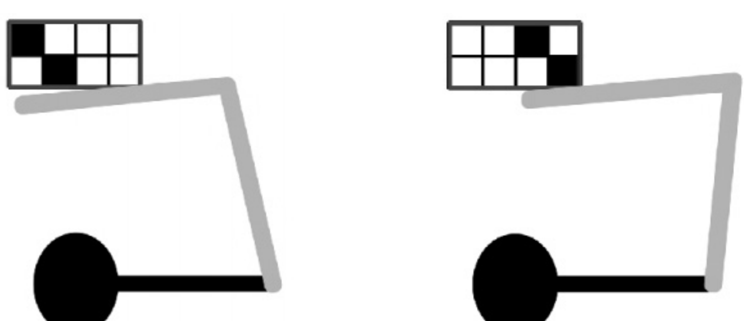

Figure 10. The figure shows where the arm stops after lesioning the proprioceptive-to-motor pathway. The individual’s macro-actions are still intact but their precise realization is disrupted.

This systematic behavioral deficit which is observed after lesioning the proprioceptive-to-motor pathway shows that the organisms still know what is the macro-action that should be produced (and which is encoded in the intact visual-to-motor pathway) but are unable to correctly realize this macro-action at the micro level of the specific changes in successive arm positions because the proprioceptive information that specifies the current arm position is disturbed.

We have measured the average distance from the target in normal, i.e., not lesioned, organisms, in organisms with lesions to the proprioceptive-to-motor pathway, and in organisms with lesions to the internal units that are always activated. Both after lesions to the proprioceptive-to-motor pathway and to the always activated units the performance in terms of correct responses is completely disrupted, i.e., objects are never reached, but whereas in the latter case the arm's endpoint stops very far from the target, in the former case the arm's endpoint stops relatively close to the target (Figure 11).

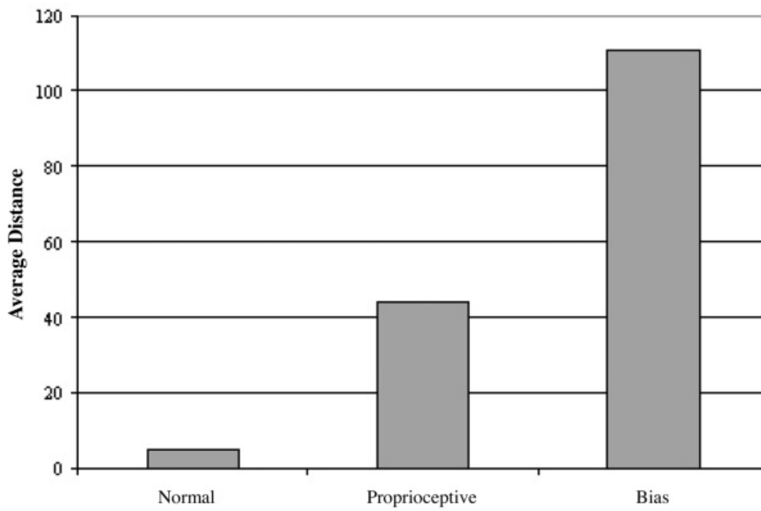

Figure 11. Average distance of the arm's endpoint from the target in normal individuals, in individuals with lesions to the proprioceptive-to-motor pathway, and in individuals with lesions to the always activated units. While after lesions to the latter type of units the individuals are completely unable to reach the portion of the visual space where the objects are found, after lesions to the propioceptive-to-motor pathway they are still able to move their arm toward the correct portion of the visual space, but the arm's endpoint stops when it is more or less removed from the object.

3. FROM VISUAL INPUT TO MOTOR OUTPUT

The activation patterns observed in the internal units of a neural network when some input arrives to the network's input units can be called the network's internal representations. However, although it is the sensory input that causes these internal representations, the simulations we have described demonstrate that a neural network's internal representations tend to reflect the properties of the motor output (at the level of the macro-actions) with which the network responds to the sensory input rather than those of the sensory input itself. The activation patterns observed in the internal units of our organisms co-vary with the macro-action with which the organism responds to a variety of sensory inputs rather than with these different sensory inputs. In other words, networks have action-based internal representations rather than sensory-based representations.

This appears to be a logical consequence of the intrinsic nature of neural networks. In very general terms neural networks can be viewed as systems for transforming activation patterns into other activation patterns. An external cause produces an activation pattern in the network's input units and then the network's connection weights transform this activation pattern into a succession of other activation patterns in the network's successive layers of internal units until the output activation pattern is generated. If we assume that the input pattern encodes the characteristics of some visual stimulus and the output those of some motor response, the successive activation patterns will progressively reflect less and less the characteristics of the visual input and more and more the characteristics of the motor output, and the network's internal representations will become more and more action-based. In this Section we examine how internal representations become progressively less sensory-based and more action-based.

In the simulations described in the first Section there was a single layer of internal units and we have seen that this layer of internal units encodes macro-actions rather than visual information. We might be able to observe a more gradual mapping of visual into motor information if we provide our neural networks with a succession of layers of internal units rather than a single internal layer. For example, with two layers of internal units we should be able to observe that the first layer, i.e., the layer which is closer to the input layer, reflects more closely the characteristics of the visual input, whereas the second layer, i.e., the layer which is closer to the output layer, will reflect the characteristics of the motor output.

3.1 SIMULATIONS

We have run two new simulations using the same task as before but two new network architectures both with two successive layers of internal units, not only one as in the preceding simulation. Both internal layers contain 4 units but in one simulation the entire retina projects to all the units of the lower internal layer (Figure 12a) whereas in the other simulation the lower layer is divided up into two separate sets of 2 units, and the left half of the retina projects to the first 2 units and the right half to the other 2 units (Figure 12b). All the 2+2=4 units of the lower internal layer send their connections to all the 4 units of the higher internal layer, which are connected to the output units. The two new architectures are schematized in Figure 12.

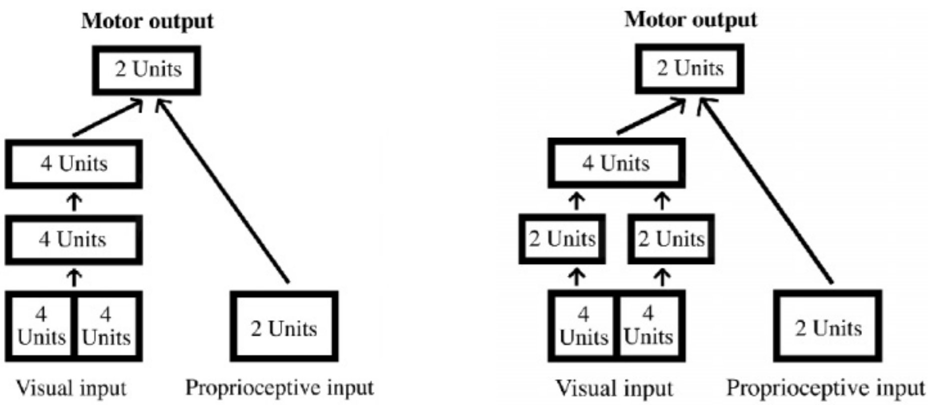

Figure 12. The two new network architectures used for the reaching task

Both these more complex architectures learn the task equally well as the basic architecture with a single layer of internal units. In both cases the evolutionary curve of performance across 10,000 generations is practically indistinguishable from that of Figure 6.

3.2 ANALYSIS OF THE INTERNAL ORGANIZATION OF THE NEURAL NETWORKS

As we have done for the previous simulation, we examined the activation level of the internal units of the best individual of the last generation in each of the 10 replications of the two new simulations. There are 8 internal units in the new architectures, 4 in the lower layer and 4 in the higher layer. The results of this examination are shown in Figures 13 and 14.

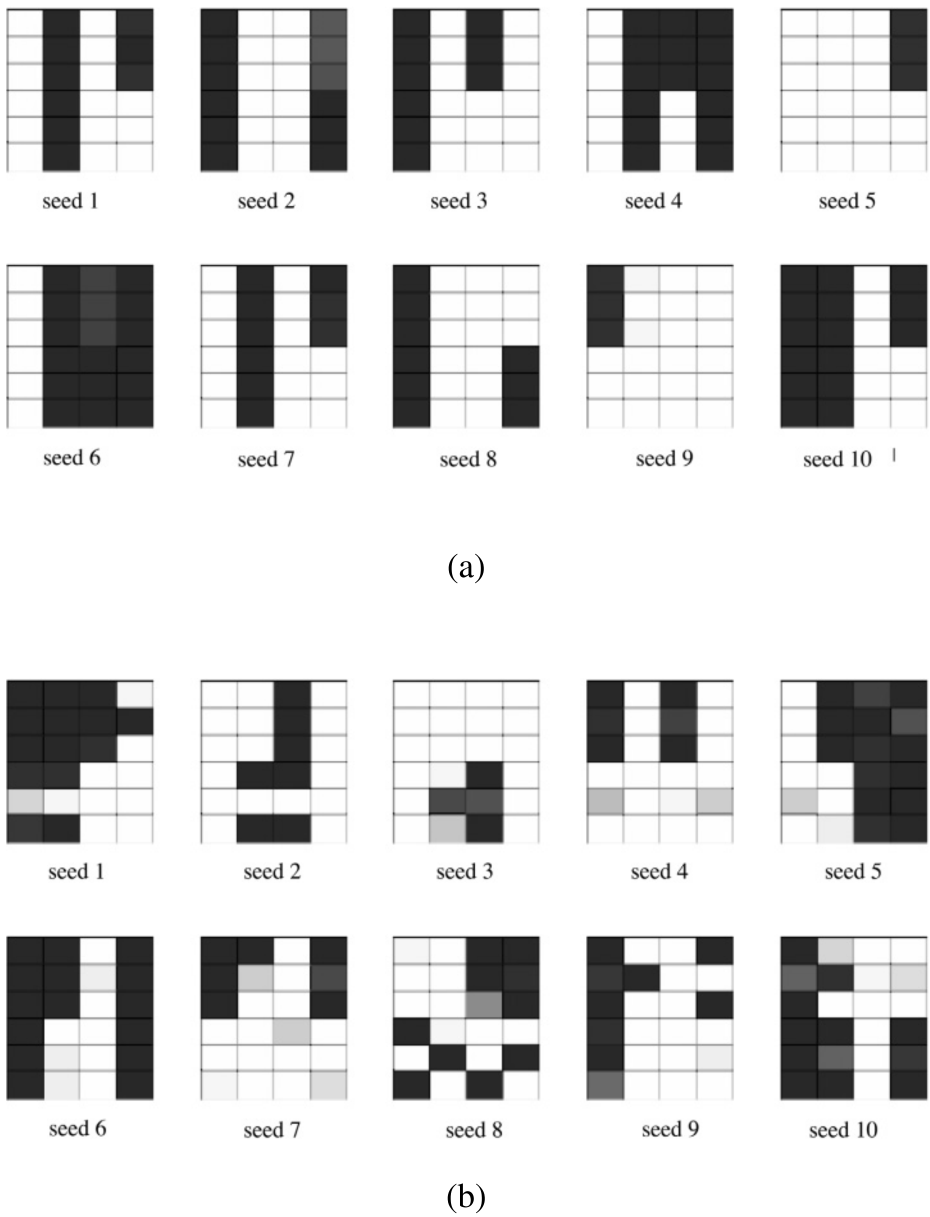

Figure 13. Activation level of the 4 units of the higher internal layer (a) and of the 4 units of the lower internal level (b) in response to each of the 6 visual input patterns, for the architecture in Figure 12a.

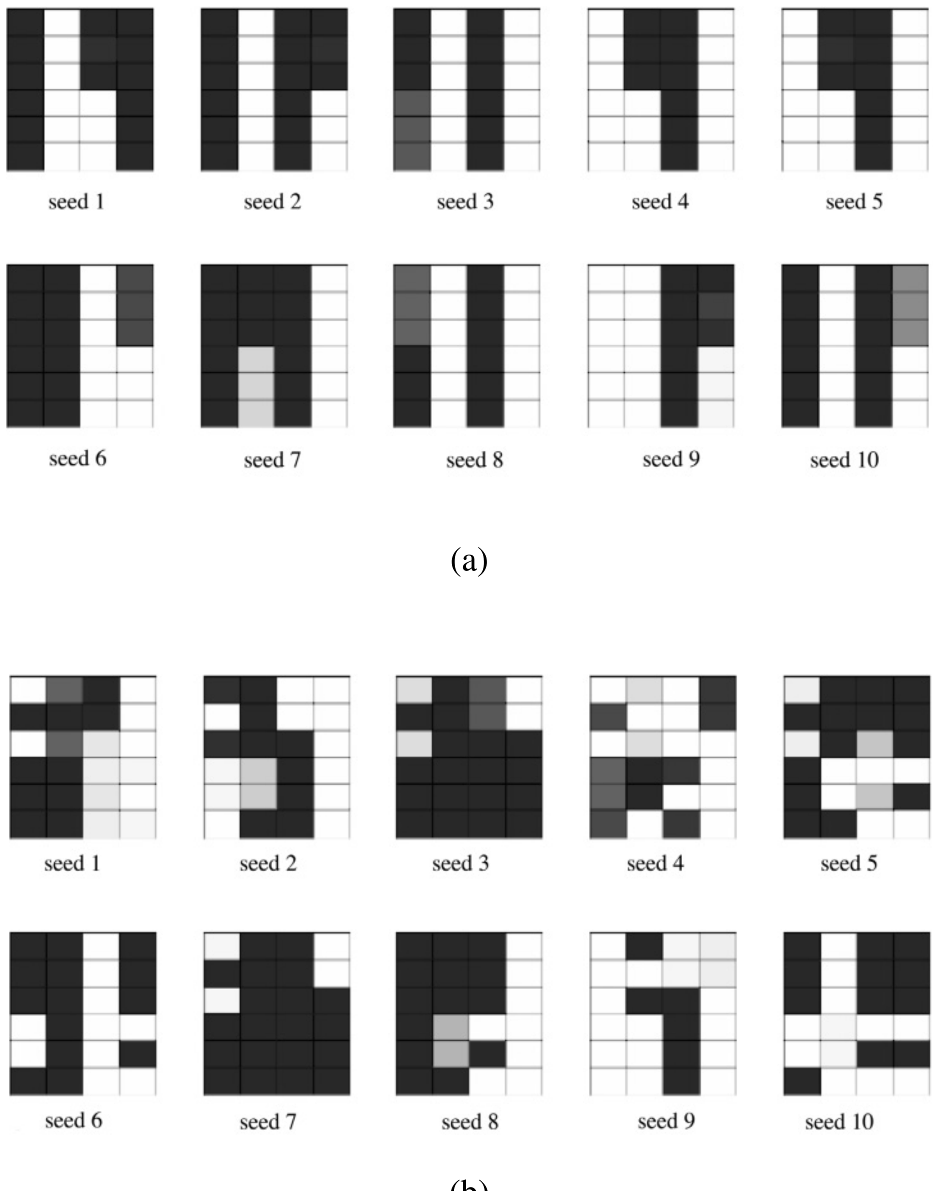

Figure 14. Activation level of the 4 units of the higher internal layer (a) and of the 2+2=4 units of the lower internal level (b) in response to each of the 6 visual input patterns, for the architecture in Figure 12b.

The two figures show that the higher level of internal units, the “pre-motor” units, encode macro-actions in the same way as the single layer of internal units in the previous simulation. As in the previous simulation, there are three types of internal units, those with (almost) zero activation in response to all 6 input patterns, those with a constant activation of (almost) 1 for all input patterns, and those with an activation of (almost) 0 encoding the macroaction of “moving the arm toward the left (or right) portion of the visual space” and an activation level of (almost) 1 encoding the macroaction of “moving the arm toward the right (left) portion of the visual space”.

However, when we turn our attention to the internal units of the lower layer, which is closer to the visual input, things look differently. Although there are some units with either an almost constant 0 or 1 activation level, most units are selective, i.e., their activation level varies but it does not vary with the motor output. In fact, there is no encoding of macro-actions at this level of internal units. As a consequence, when we lesion the units of this lower layer there is no predictable and systematic behavioral damage.

What do the internal units of the lower layer encode, then? This question can be answered more clearly if we look at the lower layer of the second architecture, which is divided up into two separate sets of 2 units each (Figure 12a). In the other architecture in which the entire retina projects to all the 4 internal units of the lower layer (Figure 12b), it is difficult to interpret what is encoded in these units because each internal unit receives information from both the left and the right visual fields and the information which must be extracted from the entire retina is rather complex. On the other hand, when we examine the architecture in which the lower internal layer is divided up into two separate sets of 2 units which receive information from the left field and from the right field, respectively, it becomes clear that the lower layer of internal units encodes what is present in the retina. At the input level each portion of the retina may contain either an A object or a B object or nothing. These three different possibilities tend to be encoded in the internal units each with a distinct distributed activation pattern. This information is then fed to the higher layer of internal units which integrates the information from both fields and, as we know, generates an encoding in terms of macro-actions.

These simulations seem to imply that the layers of internal units which are closer to the visual input tend to produce internal representations which reflect the properties of the visual input as such, independently from the actions with which the neural network will respond to the visual input, and it is only the internal representations in the layers which are closer to the motor output which reflect the properties of the motor output. In the next Section we describe a third set of simulations that show that the situation is more complex and that <u>all</u> processing of the visual input inside the neural network reflects and prepares the motor output.

4. INTERNAL REPRESENTATIONS REFLECT THE CONTEXT THAT DICTATES WITH WHICH ACTION TO RESPOND TO THE SENSORY INPUT

In the simulations described sofar the macro-action with which the organism must respond to the visual input depends exclusively on the visual input. Given some particular visual input the network always responds with the same macro-action. But consider a somewhat more complex situation in which the context also has some role in determining the organism's response (for experiments, cf. Barsalou, 1983; 1991; for simulations, cf. Borghi, Di Ferdinando, and Parisi, in press; Di Ferdinando, Borghi, and Parisi, 2002). The context can be some internal motivational state of the organism or an external verbal command. Given the same visual input the organism responds with one macro-action if the context is X and with a different macro-action if the context is Y. What we call context is an additional input which arrives to the network and interacts with the visual input to determine the organism's behavior. What are the consequences of this contextual information for the neural network's internal representations?

4.1 SIMULATIONS

As in the preceding simulations the organism has a total visual field of 2x4=8 cells divided into a left portion and a right portion of 2x2=4 cells each. However, unlike the preceding simulations the organism always sees two objects at the same time, either two objects of the same shape (two A objects or two B objects) or one A object and one B object, with the A object in the left portion of the visual field and the B object in the right portion, or viceversa. Hence, the organism's retina can encode one of the 4 visual scenes represented in Figure 15.

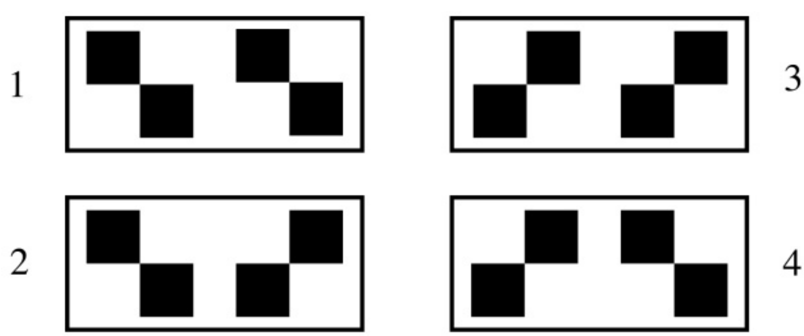

Figure 15. The content of the organism's visual field

The organism has the same two-segment arm which it uses for reaching one of the two objects. However, in the new simulations the object to be reached in any given trial depends on a command the organism receives from outside which can tell the organism to reach either the object on the left or the object on the right.

The neural network has three separate sets of input units. The first set (8 units) encodes the current content of the visual field and the second set (2 units) encodes the proprioceptive input which specifies the arm's current position. This is like the previous simulations. The third set (2 units) is new. These units are called "command units" and they encode two possible commands to the organism to reach with its arm either the object in the left portion of the visual field (encoded as the activation pattern 10) or in the right portion (encoded as 01). The internal architecture of the organism's neural network is composed of two successive layers of internal units. The lower layer of 4 internal units is divided up into two sets of 2 units each separately encoding the content of the left and right portion of the organism's visual field. All 4 units of the lower layer project to all 4 units of the higher layer. Therefore, the visual information from both the left and right half of the visual field, separately elaborated by the lower layer of internal units, is put together at this higher layer of internal units. Finally, all the 4 units of the higher layer project to the 2 output units which encode the arm's micro-actions.

We have applied two experimental manipulations. As in the preceding simulations, the 2 proprioceptive input units completely by-pass the internal units and are directly connected with the output units. On the contrary, for the 2 input units encoding the two commands "Reach the object on the left" and "Reach the object on the right", we have adopted two different network architectures in two separate simulations. In one simulation the command units project to the lower layer of internal units (Low Command, Figure 16a) and in the other simulation they project to the higher layer of internal units (High Command, Figure 16b). This implies that in the Low Command condition the lower layer of internal units elaborates the visual input already knowing the action to be executed in response to the visual input, whereas in the High Command condition the command information becomes available to the neural network only at the higher internal layer and the lower internal layer must process the visual input while still ignoring what to do with respect to it. The two network architectures are shown in Figure 16.

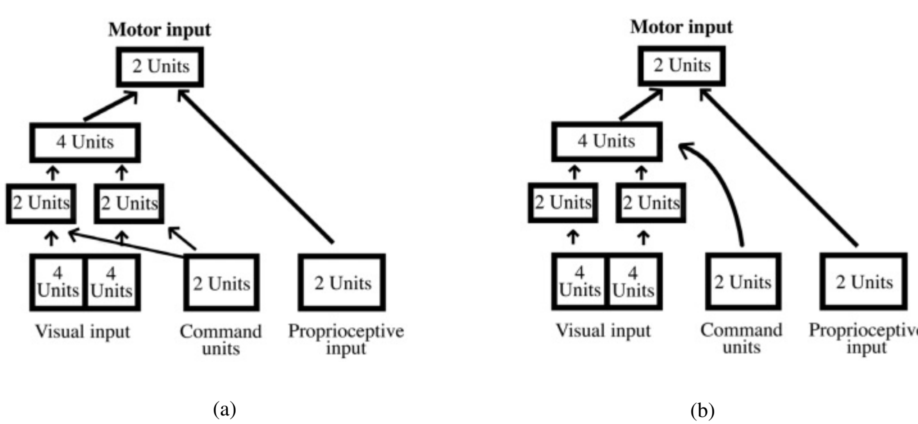

Figure 16. The two network architectures used in separate simulations

The second experimental manipulation concerns the manner in which the organism has to grasp the two objects, object A and object B, after reaching them. We assume that the objects have an "handle" which is used for grasping them. In one condition (Same Handle) both objects have the same handle, which is located where the two filled cells that represent an object touch each other. Hence, both A and B objects are grasped in the same manner, that is, by putting the arm's endpoint (the "hand") on the point of contact between the two filled cells. In the other condition (Different Handle) the objects have handles located differently: A objects are grasped by putting the "hand" on the higher cell and B objects are grasped by putting the "hand" on the lower cell (Figure 17).

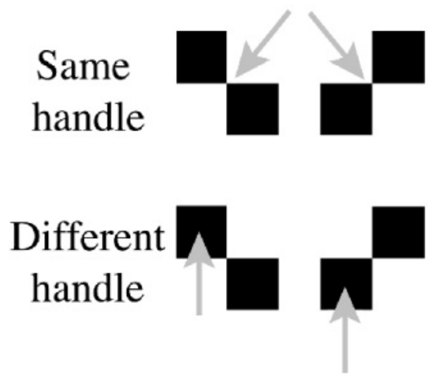

Figure 17. In the Same Handle condition both objects are grasped by putting the "hand" on the point of contact between the filled cells whereas in the Different Handle condition objects of type A are grasped by putting the hand on the higher cell and objects of type B are grasped by putting the hand on the lower cell.

One consequence of this experimental manipulation is that in the Same Handle condition the organism's neural network can ignore the difference in shape between the two types of objects since the different shape does not affect the nature of the motor action the organism must execute in grasping the objects. On the contrary, in the Different Handle condition the shape of the object must be recorded and elaborated by the organism's neural network since objects of shape A require a different type of motor action (grasping) from objects of shape B.

In total we have four experimental conditions: (1) Low Command/Same Handle; (2) Low Command/Different Handle; (3) High Command/Same Handle; (4) High Command/Different Handle. For each of the four experimental conditions we have run 10 replications of the simulation using the genetic algorithm with different initial conditions. In all 4 experimental conditions our organisms acquire the ability to respond appropriately. At the end of the simulation the best organisms are able to respond by reaching for the appropriate object and grasping it in the required manner. We now examine the internal organization of our organisms' neural networks.

4.2 ANALYSIS OF THE INTERNAL ORGANIZATION OF THE NEURAL NETWORKS

For each of the four experimental conditions, we examine the activation level of the 2 layers of internal units of the best individual of the last generation in each of the 10 replications. For the higher layer, which is closer to the motor output, the results are identical to what we have found in the previous simulations: this layer encodes the macro-actions to be produced by the organism. In the two Same Handle Conditions there are only two macro-actions: “reach and grasp the object on the left” and “reach and grasp the object on the right”. On the contrary, in the two Different Handle Conditions there are four different macro-actions: “reach and grasp the A object on the left”, “reach and grasp the B object on the left”, “reach and grasp the A object on the right” and “reach and grasp the B object on the right”. Correspondingly, we find two different activation patterns in the higher internal layer in the two Same Handle Conditions and four different activation patterns in the two Different Handle Conditions.

Let’s now turn to the lower layer of internal units, which is closer to the visual input and which therefore should reflect the properties of the visual input rather than those of the motor output. Figure 18 shows the results of our analysis for 4 individuals, one for each experimental condition.

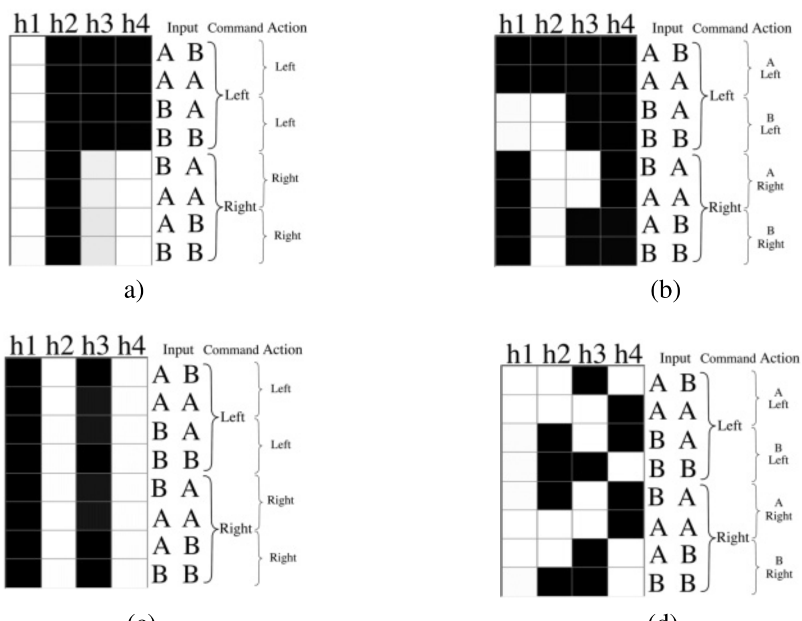

Figure 18. Activation patterns observed in the lower layer of internal units in response to the four visual stimuli when the command is “Reach and grasp the object on the left” (first four rows) and when the command is “Reach and grasp the object on the right” (last four rows), in the Low Command/Same Handle Condition (a), in the Low Command/Different Handle Condition (b), in the High Command/Same Handle Condition (c), and in the High Command/Different Handle Condition (d). The results are from 4 individuals, one for each condition.

What Figure 18 tells us is that the properties of the visual input are represented in different ways in the four conditions. In particular, two main results emerge:

(1) The internal representations of the two objects A and B are different only in the Different Handle Conditions. In the Same Handle Conditions, both when the command arrives to the lower layer (Low Command) and when the command arrives to the higher layer (High Command), there is no difference between objects A and B.

2) However, even in the two Different Handle Conditions, when the command arrives to the lower layer (Low Command) there is a difference between objects A and B only in the two internal units connected to the portion of the visual field which contains the object to be reached and grasped, while when the command arrives to the higher layer (High Command) the difference appears in both pairs of units.

Let us try to explain these results. In the two Same Handle Conditions the two objects have to be grasped in the same way and therefore there is no need for the neural network to know whether an object is A or B (Figure 18a and 18c). On the contrary, in the two Different Handle conditions, the two objects have to be grasped in different ways and therefore the neural network produces different internal representations for objects A and for objects B (Figure 18b and 18d).

The correctness of this analysis can be demonstrated by comparing, for the best individual of each of the 10 replications of the simulation, the normalized distance between the internal representations of the two objects A and B in the lower layer for the High Command/Same Handle Condition and for the High Command/Different Handle Condition (Figure 19). Similar results are found if we compare the Low Command/Same Handle Condition and Low Command/Different Handle Condition.

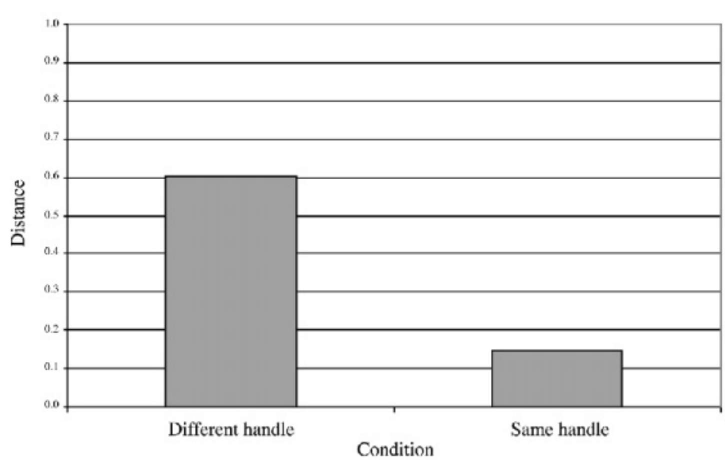

Figure 19. Normalized distance between the internal representations of the two objects A and B in the lower layer for the High Command/Same Handle Condition and for the High Command/Different Handle Condition

These results show that the network represents the two objects A and B with much more different activation patterns in the Different Handle Condition than in the Same Handle Condition. Thus, also in internal layers which are close to the visual input the internal representations tend to reflect the properties of the motor output rather than those of the visual input.

However, in the Different Handle Conditions the two objects A and B are represented in different ways depending on the internal layer to which the command arrives. In the Low Command/Different Handle Condition the lower layer of internal units already knows what macro-action must be executed and therefore we expect that the network's internal representations will reflect already at this level of processing of the visual input the requirements of the macro-action to be executed. If the command is "Reach and grasp the object on the left" A and B objects will be represented in different ways in the internal units connected to the left half of the visual field but they will be represented in the same way in the internal units connected to the right half, and viceversa if the command is "Reach and grasp the object on the right". This is exactly what we observe in Figure 18b. On the contrary, in the High Command/Different Handle Condition the information on whether the object on the left or the object on the right is to be reached and grasped arrives later on to the neural network, i.e., at the level of the second (higher) layer of internal units. Therefore, in the lower layer of internal units, since at this stage of processing of the visual input the network still ignores which action is to be executed in response to the visual input, the internal representations will reflect only the properties of the visual input and ignore the requirements of the actions. This is what we observe in Figure 18d.

The robustness of this analysis can be demonstrated by comparing in the Low Command/Different Handle Condition, for the best individual of each of the 10 replications

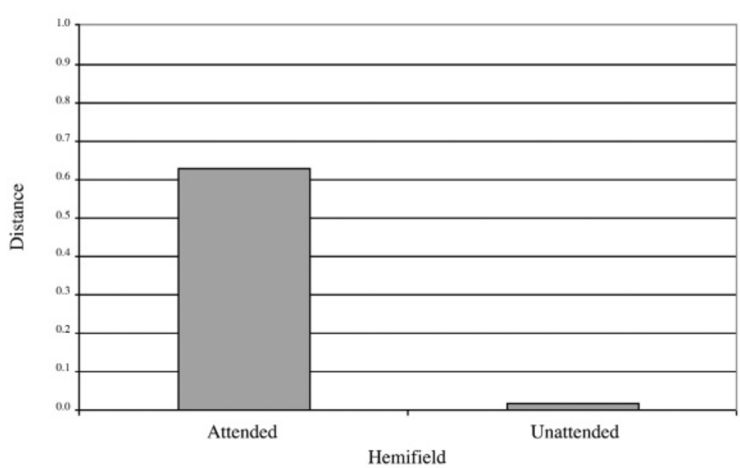

Figure 20. Normalized distance between internal representations of the two objects in the lower layer, in the two internal units that are connected to the portion of the retina which contains the object to be reached (attended), and in the two internal units connected to the other portion (unattended), in the Low Command/Different Handle Condition.

of the simulation, the normalized distance between the internal representations of the two objects in the lower layer, separately for the two units that are connected to the portion of the visual field where the object to be reached is found and for the two units connected to the other portion (Figure 20).

This analysis shows that the network represents the two objects A and B with much more different activation patterns in the two internal units that are connected to the portion of the visual field which contains the object to be reached (attended hemifield) than in the two units connected to the other portion (unattended hemifield). Thus, not only the manner in which neural networks represent the visual input depends on the requirements of the actions with which the organism must respond to the visual input rather than on the visual input's intrinsic properties, but these representations can vary in the same neural network according to the particular task in which the organism is currently involved.

5. CONCLUSIONS

This chapter has been concerned with the question of what internal representations underlie the ability of organisms to respond to sensory input with the appropriate motor output. The behavior of organisms is controlled by their nervous system and we have modelled the nervous system using artificial neural networks. Neural networks can be trained to exhibit desired behaviors (sensory input/motor output mappings) and then one can examine their internal representations, i.e., the activation patterns caused by the sensory input in the network's internal units.

We have described a series of computer simulations that allow us to draw the following conclusions.

Both macro-actions (meaningful sequences of micro-movements that achieve some goal for the organism) and micro-actions (the micro-movements that make up a macro-action) are real for neural networks. Micro-actions are the activation patterns observed in each input/output cycle in the network's motor output units. Macro-actions are observed internal representations which remain the same during an entire succession of input/output cycles and which, together with the constantly changing proprioceptive feedback, control the succession of micro-actions until a goal has been reached.

Internal representations tend to reflect the properties of the motor output (macro-actions) with which the organism respond to some particular visual input rather than those of the visual input itself. It is the visual input that is the immediate cause of the internal representations but it is the motor output that, through the adaptive history of the organisms, dictates the form of these representations. The properties of the visual input are retained in the internal representations only in so far as they are relevant for the action to be executed in response to the visual input.

Of course, if the neural network includes a succession of internal layers the early layers which are closer to the visual input will tend to reflect more the properties of the visual input and the later layers which are closer to the motor output the properties of the motor output. However, the internal representations of the visual input even in layers which are very close to the visual input will preserve of the visual input only those properties that are relevant for the action with which the organism must respond to the input. If the same visual input may elicit different actions depending on the context, the internal representations of the visual input will vary as a function of the current context and therefore of the contextually appropriate action which has to be generated. The critical result is the result for the two High Command conditions in the simulations of Section 4. Even if at the lower level the network still does not know what action must be generated in response to the visual input, the internal representation of the visual input at this lower level does not simply reflect the properties of the visual input. If the adaptive pattern of the organism, which is the current result of the adaptive history of the organism (genetic algorithm), requires the organism to respond in the same way to A and B objects (Same Handle condition), A and B objects will be represented in the same way at this lower level (Figure 18 (c)). But if the adaptive pattern requires the organism to respond in two different ways (Different Handle condition), A and B objects will be represented in two different ways (Figure 18 (d)). We conclude that there is no neutral representation of the sensory input in sensory-motor networks, a representation which simply reflects the properties of the sensory input, but all representations of sensory input, at all levels, are informed by the requirements of the action with which the organism must respond to the sensory input. The visual input is not internally represented in a fixed way which only reflects its intrinsic properties but it appears to be flexible and adaptable to the current needs of the organisms and to the specific action with which the organism must respond to the input given the particular context.

6. CODA

Gaetano Kanizsa wanted to keep perception separate from cognition because, at the time he was writing, there was a real danger that perception could be assimilated to cognition as symbol manipulation, whereas Kanizsa thought that perception is a (physical?) dynamical system. Our approach assimilates perception not to cognition but to action and it interprets everything which takes place in the mind (i.e., in the brain) as a physical dynamical system.

Prof. Andrea Di Ferdinando
Istituto di Psicologia
CNR
V. K. Marx
00137, Roma, Italy

Prof. Domenico Parisi
Istituto di Psicologia
CNR
V. K. Marx
00137, Roma, Italy

REFERENCES

BARSALOU, L.W. (1983). Ad hoc categories.*Memory and Cognition*, 11, 211-217.

BARSALOU, L.W. (1991). Deriving categories to achieve goals. In *The psychology of Learning and Motivation*, ed. by G.H. Bower. New York, Academic Press, pp. 1-64.

BORGHI, A.M., DI FERDINANDO, A., and PARISI, D. (in press). The role of perception and action in object categorization. In J.A. Bullinaria and W. Lowe (eds), *Connectionist Models of Cognition and Perception*. Singapore: World Scientific.

DI FERDINANDO, A., BORGHI, A.M. and PARISI, D. (2002) The role of action in object categorization.. In *Proceedings of Workshop on Categorization and Concept Representation*, Florida Artificial Intelligence Research Society Conference. Pensacola, FL.

FODOR, J.A. (1981). *Representations. Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, Mass., MIT Press.

GIBSON, J.J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Erlbaum.

HOLLAND J.H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: The University of Michigan Press (Second edition: Cambridge, MA: MIT Press, 1992).

MIGLINO, O. AND WALKER, R. (2002). Genetic redundancy in evolving populations of simulated robots. *Artificial Life*, 8, 265-277.

O'REGAN, J.K., NOE, A. (in press). A sensorimotor account of vision and visual consciusness. *Behavioral and Brain Sciences*, in press.

RUMELHART D. & MCCLELLAND J. (1986*). Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

This page intentionally left blank

LUIS GONCALVES, ENRICO DI BERNARDO AND PIETRO PERONA

MOVEMES FOR MODELING BIOLOGICAL MOTION PERCEPTION

1. INTRODUCTION

In 1973 Gunnar Johannson [18] discovered the surprising ability of our visual system in perceiving *biological motion*, i.e. the motion of the human body, even from highly empoverished and ambiguous stimuli. In attempting to develop models for explaining the perception of biological motion we examine models developed for different purposes in two areas of engineering, computer vision and computer graphics.

Perceiving human motion, actions and activities is as important to machines as it is to humans. People are the most important component of a machine's environment. Endowing machines with biologically-inspired senses, such as vision, audition, touch and olfaction appears to be the best way to build user-friendly and effective interfaces. Vision systems which can observe human motion and, more importantly, understand human actions and activities, with minimal user cooperation are an area of particular importance.

Humans relate naturally with other humans, therefore animated characters with a human appearance and human behavior (as well as, of course, other creatures inspired to animals) are an excellent vehicle for communication between machines and humans. Accurate rendering techniques developed in computer graphics and ever increasing computer performance make it possible to render scenes and bodies with great speed and realism. The next frontier in computer graphics is animating automatically and interactively characters to populate screentops, web pages and videogames.

In sum: the next generation of both animation and vision systems will need to bridge between the low-level data (rendered bodies, observed images) and the high-level descriptions that are convenient as either input (to graphics) or output (from computer vision) for automatic interaction with humans. Models developed within this engineering effort shed light on biological motion perception since it shares the same computational underpinning. We argue that *models of how humans move are the key ingredient* to this quantum step. We start by discussing the inadequacies of the current models (sections 2 and 3) and propose an alternative style of modeling human motion based on *movemes*, a felicitous term which we adopt from Bregler and Malik [10] and further extend from meaning "elementary stereotypical motions" to "elementary motions parameterized by goal and style" (section 4). We explore the practical aspects of movemes in sec. 5 and discuss six case studies in sections 6, 8, 9 and 10. The issues of styles and composition are discussed in sections 8 and 11. The details of the models are discussed in the appendix.

A. Carsetti (ed.), Seeing, Thinking and Knowing, 143-170.

2. MODELS IN MACHINE VISION

While it is easy to agree that machines should "look" at people in order to better interact with them, it is not immediately obvious which measurements should a machine perform on a given image sequence, and what information should be extracted from the human body. There are two classes of applications: "metric" applications where the position of the body has to be reconstructed in detail in space-time (e.g. used as input for positioning an object in a virtual space), and "semantic" applications where the meaning of an action (e.g. "she is slashing through Rembrandt's painting") is required. The task of the vision scientist/engineer is to define and measure "visual primitives" that are potentially useful for a large number of applications. These primitives would be the basis for the design of perceptual user interfaces [28, 14] substituting mouse motions and clicks, keystrokes etc. in existing applications, and perhaps enabling entirely new applications.

Which measurements should we take? It is intuitive that if one could reconstruct frame-by-frame the 3D position of each part of the body one would have and excellent set of visual primitives. One could use such measurements directly for metric applications, and feed them to applications. This avenue, which we shall call *puppet-tracking*, has been pursued by a number of researchers. The most successful attempts start from the premise that the overall kinematic structure of the human body is known, and its pose may be described synthetically by the degrees of freedom of the main joints (e.g. three in the shoulder, one in the elbow, two in the wrist etc) so that the whole body is represented in 3D by a stick-figure not dissimilar from the wooden articulated puppets that art students use for training in figure-drawing. Precise knowledge of the body's kinematics, together with some approximation of the dynamics of the body, allows one to invert the perspective map if each limb of the body is clearly visible throughout the sequence. One may thus track the body both on the image plane and in 3D with come success and accuracy even using a single camera. For example: Rheg and Kanade [22] were able to track the movements of the fingers of a hand, even when the hand underwent complete rotation around the wrist. Goncalves et al. [16, 3] demostrated their monocular 3D arm-tracking system in real-time and used it as an interface to a rudimentary 3D vitual desktop. Bregler and Malik [11] showed that they could track accurately walking human figures in the Muybridge sequences. In the first two studies the dynamics of the body was approximated by a random walk. In Bregler and Malik the model is more sophisticated: four second-order random walks with different statistics governed by a four-state Markov process.

These initial results are encouraging: they demonstrate that successful 3D tracking of the main degrees of freedom of the body may be achieved from monocular observations of a subject who is not wearing special clothing nor markers. However, they also point to some fundamental difficulties of the puppet-tracking approach. First of all: *these trackers need to be initialized near the current state of the system*, for example: in Goncalves' case the user has to start from a standard pose and hit a button on the keyboard to initiate tracking. Moreover, the tracker needs frequent re-initializations (every 1000 frames or so in Goncalves et al., not reported by the other studies). This is not satisfactory. Moreover: all *these systems depend crucially on precise knowledge of the mutual position of key features* and/or of the kinematics of the body, and therefore require an initial user-specific calibration; moreover, the user has to wear tight-fitting clothes, or, more entertainingly, no clothes at all in order to allow sufficient accuracy in measuring the position of body features frame-by-frame. Furthermore, occlusion has devastating effects. This is a far cry from the performance of the human visual system which can make sense of 3D human motion from monocular signals (e.g. TV) even when the actors wear loose clothing, such as overcoats, tunics and skirts, and when the signal is rather poor (e.g. a person seen at 100m spans fewer than 100 “pixels” in our retina).

What is wrong with the puppet-tracking approach, and what may be the secret of the human visual system?

First of all: the issue of automatic detection of humans and automatic initialization of human trackers has been so far ignored and needs to be addressed. We will not discuss this topics any further here; initial encouraging results are reported in [27, 26].

Second: currente models are too “brittle” for being practical. Unrealistically accurate and detailed kinematic modeling is required, clothing is a nuisance, occlusion is a problem. One could say that the puppet-tracking approach tries to estimate too much and knows too little. On one hand, one must realize that for the “semantic” applications is should not be necessary to measure the position of all limbs frame-by-frame in order to interpret human activities which span many hundreds of frames On the other hand, random walks cannot be the ultimate model of human motion. Humans do not move at random.

In the following we will argue that one has to reverse the current approach and take a counterintuitive point of view. Rather than reconstructing the 3D pose directly from the image stream *and then* feed such signal into a recognition box for semantic analysis, one should first recognize the motions of the body at some discrete high-level *and then* (if necessary) reconstruct the frame-by-frame pose from these. We will argue that this approach has several advantages: robustness, parsimony, convenience for high-level processing.

3. MODELS IN ANIMATION

It is increasingly common to see computer-generated human characters in motion pictures that are so life-like and convincing that one may be led into believing that the problem of automatic animation has been satisfactorily solved. This is far from the truth. One minute’s worth of animation can take weeks to produce and is the result of the manual labor of talented artists: *current characters are puppets rather than actors*. Those animated characters whose motion is synthesized automatically (surch as for a character on a webpage, or in a computer game) do not move realistically and have highly restrictive repertoires; Animating automatically realistic human characters is still an unsolved problem.

The ultimate technique for automatic animation would require high-level directions much like a director controls an actor ("walk towards the door expectantly", "pick up the glass in a hurry") rather than a puppetteer's step-by-step and joint-by-joint control. It would produce motion patterns that would be indistinguishable from the ones obtained from real people.

The automatic generation of realistic human motion has been a topic of research within the computer graphics and robotics communities for many years. Three techniques are used: dynamic simulation, optimization of contrastrints, and editing of motion capture data.

Dynamic simulation models start from the physics: a 3-D model of a body built out of rigid segments with associated masses and moments of inertia is moved by torques to the various joints simulating muscle actions. Hodgins et al. [17] use this approach to animate human athletes. Hand-crafted controllers place the feet, rotate the joints, and impose the phase of arm and leg swing. Humans running, spring board vaulting and diving have been generated this way.

This method has two shortcomings. The animations it generates look "robotic"; moreover, success appears to be restricted to athletic motions, since these are dynamically constrained and highly optimized: there are not many ways to run/jump competitively. However, for ordinary non-athletic motions such as strolling or picking up a light object the motion can be performed in many ways: each person has a particular, identificable walking style and a good actor can imitate various styles. In this case it is the brain that defines the motion and the dynamics of motion is virtually irrelevant. Thus for everyday actions dynamic simulation methods will fail to generate realistic motion unless the brain's motor control signals can be properly modeled and imitated.

Another approach to motion synthesis is to use the robotics-based techniques of inverse kinematics [21] (see also a similar technique called "Space-time constraints" developed by Witkin et al. [29]). When applied to a robot, these techniques allow the computation of the robot's configuration (set of joint angles) that will place the end-effector at a certain position and orientation in space (subject possibly to additional constraints imposed by the environment or other considerations). When combined with energy or torque minimizing principles, these methods can be used to produce efficient and well-behaved robot motions. Badler et al. [1] have used such solutions in the development of their Virtual Jack project. Virtual Jack has a complex virtual skeleton with over 100 joints, and was designed to simulate the posing (and to some extent the motion) of the human body in order to provide accurate ergonometric information for ergonomic studies. However, Jack is known to have a stiff back, stand in awkward poses, and move in a robotic fashion. The reason for this is that these robotic approaches have no notion of the naturalness of a pose, or of the intricate, characteristic phases between the motions of different body parts.

A third class of methods for generating realistic motion is by manipulating motion capture data (3-D recording of people's movements). Bruderlin [12] uses the technique of Motion Signal Processing, whereby multi-resolution filtering of the motion signals (changing the gain in different frequency bands) can produce different styles of motion. For examples, increasing the high frequency components produces a more nervous-looking motion. Gleicher [15] introduced the method of motion editing with space-time constraints. Here, a motion is adapted to satisfy some new constraints, such as jumping higher, or further. The new motion is solved for by minimizing the required joint displacements over the entire motion, and this way attempting to preserve the characteristics of the original motion. In can be thought of as a generalization of the inverse-kinematics techniques, where instead of computing the pose to satisfy the constraint only during a single frame, the modification of pose in done through time. However, since there is no notion of how realistic (or human-like) a modification is, the method can be used only to generate small changes - otherwise, the laws of physics appear to be defied.

4. WHY MOVEMES AND WHICH MOVEMES?

In 1973 Gunnar Johannson [18] discovered the surprising ability of our visual system in perceiving *biological motion*, i.e. the motion of the human body. He filmed people wearing light bulbs strapped to their joints while performing everyday tasks in a darkened room. Any frame of Johannson's movies is a black field containing an apparently meaningless cloud of bright dots. However: as soon as the movie is animated one has the vivid impression of people walking, turning the pages of a book, climbing steps etc. We formulate the hypothesis that in order to solve this apparently impossible perceptual task *people must move in stereotypical ways, and our visual system must have an exquisite model of how people typically move.* If we could understand the form and the parameters of such model we would have a powerful tool both for animation and for vision.

In looking for a model of human motion one must understand the constraints to such motion. First of all: our motions are constrained both by the kinematics and by the dynamics of our body. Our elbows are revolute joints with one degree of freedom (DOF), our shoulders are ball joints with three DOF etc.. Moreover, our muscles have limited force, and our limbs have limited acceleration. Knowledge of the mechanical properties of our bodies is helpful in constraining the space of solutions of biological motion perception. However, we postulate that there is a much more important constraint: *the motion of our body is governed by our brain.* Apart from rare moments, when we are either competing in sport or escaping an impending danger, our movements are determined by the stereotypical trajectories generated by our brain [4]; the dynamics of our body at most acts as a low-pass filter. For example, our handwriting is almost identical whether we write on a piece of paper or on a board – despite the fact that in one case we use fingers and wrist and in the other we use elbow and shoulder with completely different kinematics and dynamics at play (this principle of *motor equivalence* was discovered by Bernstein and collaborators in the first half of the 1900s [2]).

Why would our body move in stereotypical ways? Our brain moves our body in order to achieve goals, such as picking up objects, turning to look at the source of a noise, writing. The trajectories that are generated could, in principle, be different every time and rather complex. However: generating trajectories is a complex computational task requiring the inversion of the kinematics of the body in order to generate muscle control signals. Rather than synthesizing them from scratch every time the brain might take a shortcut, and concatenate a number of memorized pre-made component trajectories into a complete motion. Neurophysiological evidence suggests that indeed the nervous system may encode complex motions as discrete sequences of elementary trajectories [6, 8]. Moreover these trajectories appear to be parameterized in terms of Cartesian 'goal' parameters, which is not surprising given the fact that most motor tasks are specified in terms of objects whose position is available to the senses in Cartesian space.

This suggests a new computational approach to biological motion perception and to animation. One could define a set of elementary motions or *movemes* which would roughly correspond to the 'elementary units of motion' used by the brain. One could represent complex motions by concatenating and combining appropriate movemes. These movemes would be parameterized by 'goal' parameters in Cartesian space. This finds analogies in other human behaviors: the "phonemes' are the elementary units both in speech perception and produciton; in handwriting one thinks of 'strokes' as the elementary units.

How realistic is this approach? Which are the natural movemes and how many are they? How should one parameterize a moveme? We address these questions in the next sections.

In comparing the 'frame-by-frame puppet-tracking' approach to a moveme-based approach one notes that the second has the potential to reduce dramatically the number of parameters to be estimated, thus conferring great robustness on a vision system that knows about movemes. Moreover, a moveme-based approach transforms the description of human motion from continuous time trajectories to sequences of discrete tokens. The latter description is a better starting point for high-level interpretation of human motion.

5. MOVEMES IN PRACTICE

For a moveme-based approach to exist one must specify a set of movemes. How might one define a moveme, and how might one discover this set? As we mentioned above, our working hypothesis is that movemes are building blocks used by the brain in costructing the trajectories that our body should follow in order to achieve its goals. Therefore it is not easy to measure movemes directly; we may have to settle with observing movemes phenomenologically and indirectly. The following five criteria summarize our intuition on what makes a good moveme, and guide our search. A moveme should be

Simple - Few parameters should be necessary to describe a moveme. There should be no natural way to further decompose a moveme into sub-movemes.

Compositional - A moveme should be a good 'building block' to generate complex motions. One should avoid movemes that are not necessary.

Sufficient - The set of all movemes should suffice to represent all common human actions and activities with accuracy.

Semantic - Motion is most often goal-oriented; movemes should correspond to simple goals, which provide a natural parameterization for the moveme. Roughly speaking, a good moveme should correspond to a verb in language [23] and it should be parameterized by a meaningful goal in ambient space. The descriptors for a moveme are thus at least two sets of parameters: the identity of the moveme, and the goal of the moveme (we will see later that at times additional 'style' parameters are needed).

Segmentable - It should be easy to parse the continuous pattern of motion that a body describes over several minutes, or hours, into individual movemes. This is the case if the beginning and ending of a moveme are stationary or otherwise stereotypical. This makes estimating moveme models easier; more importantly, it makes it easier to compose complex motions from movemes (simple boundary conditions between individual movemes) and cheaper to recognise movemes automatically in a computer vision system (easy criteria for bottom-up segmentation).

This last property is not strictly necessary, but it is covenient and thus desirable.

A 'complete set' of movemes may be defined as the set of all the movemes that a human will use. Alternatively, it may be defined as the minimal set the movemes combining and concatenating which all the motions of a human may be described with sufficient accuracy.

The task of enumerating a complete set of movemes goes beyond the scope of this paper. If one takes the analogy of phonemes this effort is likely to take many years of accumulated experience and insight. The goal of this paper is more modest: beyond proposing an argument for the study of movemes, which we have just done, we wish to examine a small set of 'case studies' that will allow us to exemplify our intuition and to explore such practical issues as the complexity of a moveme, the accuracy with which we may expect a moveme to model human motion, how to compose movemes into complex motions.

6. METHODS

Guided by the principles listed above, we have chosen six movemes for our analysis: step, step over, look, reach, draw, run. These movemes are a sufficient set to test a number of modeling strategies, and represent a range of moveme complexity.

To model each moveme, we take a phenomenological or 'perceptual' approach. Rather than attempting to build models based on physical or optimality principles, we build models from observations of people performing movements. Thus our models will be empirically derived functions mapping moveme parameters to the motion of the entire body. For example: in the case of the moveme 'reach' we parameterize the moveme with the three coordinates of the point to be reached. Thus we assume that the entire motion of the body for the 1-2 seconds that the motion lasts (i.e. approximately 10^4 parameters) is determined by just three numbers, the Cartesian coordinates of the target.

In order to satisfy the compositional criterion of a moveme, a moveme must take into account the initial pose the body is in at the beginning of the movement. This can be done by using *state* parameters along with the goal parameters. For instance, in placing a step, it is necessary to know where the foot starts off, and this can be encoded in a state parameter for the initial foot position.

Collecting the data to build a moveme model involves the following steps:

• Capture: A 3-D optical motion capture system is used to record the motion of a person. The systems records motion at 60Hz with an accuracy of 1mm. 18 markers are placed on the body at the main joints. The subject acts out several samples of a moveme by, for example, walking back and forth to provide samples of taking a step, or repeatedly reaching to various locations to provide samples of reaching.

• Segmentation: The motion capture data is segmented into individual moveme sample trajectories by detecting the start and stop of each sample trajectory. To facilitate analysis, all the trajectories are re-sampled in time to be of same duration.

• Representation: The motion capture data is converted to 'bone-marker' data; rather than using the coordinates of the 18 markers directly, a 3-D skeletal model of the actor is used to compute the corresponding motion of the joints. Then virtual mocap data is generated, placing 21 bone-markers at the major joint centers. This procedure ensures a standard representation of body motion, and eliminates inaccuracies due to inexact placement of markers on the body.

• Labeling: Each moveme sample trajectory is labeled with the appropriate values of the goal and state parameters. For instance, reach moveme samples are labeled with the 3-D coordinate of the final hand position during the reach.

The set of labeled moveme sample trajectories form a set of input-output pairs (the goal and state parameters, and the resulting body trajectories) that can be used to build models, as discussed in the next section. One decision that remains, however, is how to represent the motion of the entire body. There are two natural representations to chose from. The body motion can be represented in terms of the angular coordinates of all the body joints, or it could be represented in terms of the motion of the 21 bone-markers. The former representation has the possible advantage of implicitly incorporating some knowledge of the kinematic structure of the human body. However, it is a highly non-linear space, as compared to the 3-D Euclidean space of marker coordinates, which may make it more difficult to learn an accurate model. Furthermore, errors in angular coordinates affect the body pose in a cumulative fashion as one progresses through the kinematic chain of the spine and limbs, whereas errors in marker coordinates do not affect each other. Finally, Bizzi [7] has put forward the case that the brain represents and controls trajectories in Cartesian coordinates. For these reasons, we represent body motion with the trajectories of the 21 bone-markers in Cartesian space.

7. MODELS

Let $\mathbf{m}$ be a moveme, and $\mathbf{y}_{\mathbf{m}}(x)$ represent the sample trajectory of moveme $\mathbf{m}$ with label (goal and state parameters) x. Then modeling of movemes can be viewed as a function approximation problem; given a set of input-output pairs {**x, y**}, find a function **f** which approximates the output **y** given the input **x**. If we found estimate **f** from our training data, then animation can be generated from a very high level; by choosing the moveme and the goal parameters, the motion of the entire body is generated. Likewise, for recognition and perception, one needs to find which of many possible moveme functions (and associated values of the goal and state parameters) best fit an observed motion.

We experimented with several different model types. Namely, linear models, higher-order polynomial models, radial basis function networks, and feed-forward networks with sigmoidal hidden units (see appendix A for details of each model type). Surprisingly, the simplest of models (the linear model) usually performs quite well, as will be shown in section 10, where we discuss the issue of performance evaluation.

8. STYLES

Thus far we have proposed to parameterize movemes by action type (walk, reach, look, etc.) and by goal. Johansson noticed that, from his displays, one can tell the age, sex, etc. of the person performing a motion. Therefore we need to add a new set of parameters, which we call "style" parameters. Now, every time new parameters are added the number of samples necessary to estimate a moveme increases. We postulate that *the style and goal parameters are orthogonal*, i.e., that we may estimate them independently. More precisely:

Suppose $F : \mathrm{x} \rightarrow \mathrm{y}$ is a function which maps a given label x to a moveme trajectory y. Suppose also that samples $\{\mathrm{x}_i, \mathrm{y}_i\}$ of the moveme performed with a new style are available. Then we can calculate the residual motions due to the new style as $\{\mathrm{r}_i \equiv \mathrm{y}_i - F(\mathrm{x}_i)\}$.

Using the residual motion and the sample labels, $\{\mathrm{x}_i, \mathrm{r}_i\}$, a residual function $R : \mathrm{x} \rightarrow \mathrm{y}$ can be learned. Then, to synthesize a new trajectory withe the new mood or style, the original function F and the residual function R can be combined. In fact, a modulating parameter, α, can be used to control how much of the new mood to use:

$$\mathrm{y}_{\mathrm{new-mood}} = F(\mathrm{x}) + \alpha R(\mathrm{x}) \qquad (1)$$

This is similar to the recent paper by Blanz and Vetter for face morphing [9] where linear combinations of sample faces are combined to produce new faces, except here entire trajectories are morphed.

Experimentally, it was verified that this technique works well; with fewer examples, it is possible to learn a new style. Intuitively, one can argue that a mood or style typically can be viewed as a 'lowpass' perturbation of the overall motion. While the nominal function needs to be learned with many examples to ensure that a proper moveme trajectory is generated throughout the input label's domain, the new mood's residual is a small signal that is not as complicated (does not change as much as a function of the label). Thus, fewer examples are needed to encode it. Furthermore, perceptually, given that the overall moveme is correct, 'errors' in mood or style are difficult to perceive.

9. DETAILS OF MOVEMES

In this section we describe in detail the movemes that were analyzed. We describe the moveme, the datased acquired, and the choice of goal parameterization.

9.1 2-D REACH

Dataset: Figure 1a shows some snapshots of a reaching moveme end-poses. The data was acquired with a 3-D mocap system using a single camera. The actor stood facing the camera, with arms by his sides, and then reached to a location around him, as if picking an apple from a tree (or from the ground). In order to make the dataset consistent, the actor always stood in the same initial position at the onset of each sample, and returned to that pose at the end of each sample. 91 samples of reach movemes were collected, the goal locations are shown in figure 1b. The actor reached towards 12 regularly spaced directions, near the limit of his reach range, and half way. The duration of each reach moveme sample trajectory varied from 90 to 117 frames (almost 2 seconds), and they were all uniformly re-sampled to be 117 frames long.

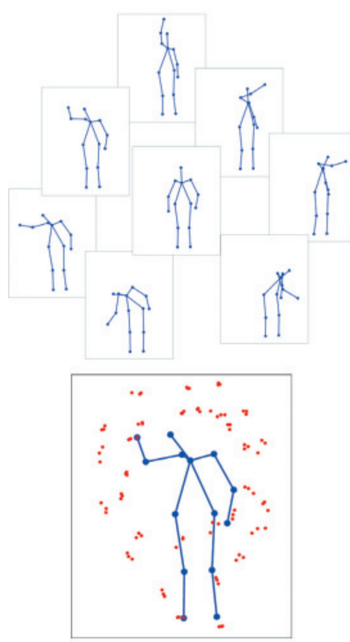

Figure 1. (a): Sample reach poses for learning a 'reach moveme'. Starting from the the rest position (center), the subject reached with his right hand to various locations around him. In (b), the 91 sample reach locations are plotted. The motion captured trajectories are used to learn a model of the elementary reach 'moveme', parameterized by the desired reach location (the 'goal' of the moveme).

Moveme parameterization: In the reach dataset, since the actor always started from the same initial condition and ended in the same final position, no state parameters are needed. The goal parameters are the 2-D screen coordinates of the position reached. Figure 1b shows the locations of the 91 sample reaches.

9.2 3-D REACH

Dataset: This dataset is similar to that of the 2-D reach, except that the data was captured in 3-D. The actor started from the rest position (standing facing forwards, arms at side) and reached to various locations in 3-D, returning to the rest position after each reach.

Moveme parameterization: As in the 2-D case, since the actor always started from the same initial pose, no state parameters are used, and the goal parameters are the coordinates of the 3-D reach location.

9.3 2-D DRAW

Dataset: Figure 2a shows some examples of 2-D drawing data. To perform this moveme, the actor stood and faced the camera. Starting at a random location, the actor performed simple strokes, as if drawing on a chalk board, consisting of either straight lines or curved arcs (but never curves with inflection of curvature). In all, 378 samples of the draw moveme were collected, with the strokes varying in position in space, size, and amount of curvature. The draw moveme was chosen to be analyzed because it is a good candidate for exploring how to compose movemes to create more complicated actions, as described in section 11 on concatenation. For example, writing on the board can be decomposed as a sequence of elemental draw movemes.

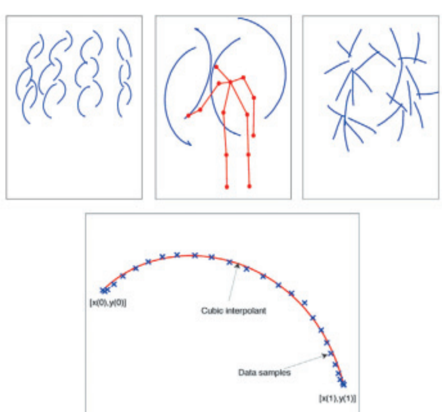

Figure 2. (a): Sample drawing strokes of a ‘draw’ moveme. Both straight and curved strokes at various scales were drawn, for a total of 378 samples. The motion of the entire body labeled with a parameterized encoding of the shape of the stroke is used to learn a draw moveme model. (b): A typical sampled stroke, and corresponding cubic interpolation defined by an 8-dimensional stroke label. Since the sample strokes were constrained to not have inflection of curvature, a simple cubic interpolant represents the dataset trajectories very well.

Moveme parameterization: The goal of the 2-D draw moveme is to draw a simple stroke. Thus, the goal parameters need to describe the path of the drawing hand. To represent the path in a compact form, it is represented by the coefficients of a cubic interpolant:

$$x(t) = \sum_{i=0}^{3} a_i t^i \tag{2}$$

$$y(t) = \sum_{i=0}^{3} b_i t^i \tag{3}$$

where time has been rescaled to be $t = 0$ at the beginning of motion and $t = 1$ at the end. Finding the coefficients to solve the above equations is a minimization problem, and does not guarantee a perfect fit. However, since we want to have an accurate control/estimation of the initial and final state of the body for the purpose of composing movemes, it is important to represent the starting and ending position accurately (i.e., equations 2 must be hard constraints for $t = 0$ and $t = 1$, and soft constraints for $0 < t_i < ... < t_n < 1$). Let

$$P(t) = [\ 1 \ \ t \ \ t^2 \ \ t^3\ ] \tag{4}$$

and let C^* be a solution of

$$\begin{bmatrix} P(0) \\ P(1) \end{bmatrix} C^* = \begin{bmatrix} x(0) \\ x(1) \end{bmatrix} \tag{5}$$

Also let C^{n_1}, C^{n_2} be a basis to the null space of $[P(0)^T P(1)^T X\ S]^T C^* = 0$. Then a solution satisfying the hard constraints and minimizing the error in the soft constraints can be found by solving for w_1 and w_2 which solve the least squares problem:

$$X - PC^* = [PC^{n_1}\ PC^{n_2}] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{6}$$

where $X = [x(t_1);...;x(t_n)]$ and $P = [P(t_1);...;P(t_n)]$

Figure 2b shows a sample fit of a stroke (it is difficult to visualize the resulting coefficients themselves since they lie in an 8-dimensional space).

9.4 3-D WALK

Dataset: The subject walked back and forth along straight and curved paths, starting and stopping, with a consistent style of walking. The aim was to capture samples of walking with as much variability as possible in terms of lenght of step, and curvature of path. Due to a limitation of the mocap system (3x4 m capture space and only 4 cameras), the actor had to always face the same general direction (+/- 30 degrees) and was not free to walk around naturally, which would have been ideal. Instead, short sequences of walking along a straight or curvy path for 5 steps, followed by walking back to the starting position, were repeated over and over. After segmentation and labeling, 124 samples of stepping with the left foot, and 119 samples with the right foot were obtained, to be used to learn two separate movemes (stepping with each foot). The samples were all resampled in time to be 45 frames in duration (0.75 seconds).

Moveme parameterization: the 3-D walking dataset is more complicated than the 2-D reach or draw datasets, in the sense that it requires a parameterization of the state of the actor at the beginning of the step. Also, it is not immediately obvious how to parameterize a step.

One possible parameterization is the position of the feet at the start and end of the step, and postulate that the entire body motion follows from that. Specifically, the reference frame of each step is taken as the (initial) position of the stationary (support) foot, with the x-axis along the foot direction. The state of the character is defined by the initial position of the moving foot, and the goal is defined by the final position. Figure 3 shows three sample start and end poses for stepping with the left foot, and figure 4 shows the labels for all the examples of left and right footed steps.

9.5 3-D HAPPY WALK

Dataset: Examples of walking in a different style - ‘happy’ (swaying back and forth, with a little bounce in the step) - were acquired. The aim was to use this dataset to learn an incremental model of a new style of walking based on the original walking moveme (as described in section 8 on learning styles). Thus, only 16 samples of happy walking (16 steps with left and right foot each) were acquired.

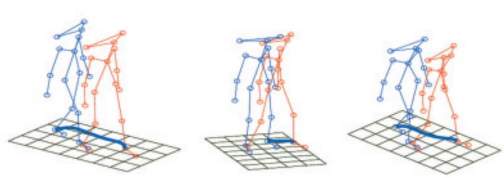

Figure 3. Three sample start and end poses for stepping with the left foot, with left ankle trajectory traced out in 3-D. A total of 124 samples were used to learn the ‘walk’ moveme.

Moveme parameterization: In this dataset, the labeling of the samples is done exactly as in the 3-D walking case. The level of happiness is assumed to be constant (the actor needs to perform the moveme samples consistently). During synthesis (after the appropriate models have been learned), the happy mood can be combined with nominal walking. Figure 5 shows the labels for the dataset.

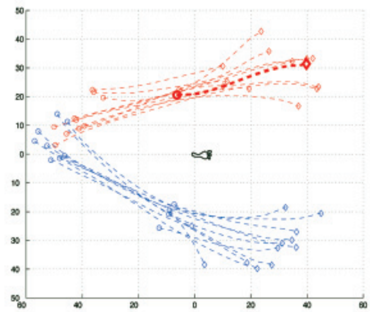

Figure 4. Parameterization of the ‘walk’ moveme involves defining the initial and final position of the moving foot with respect to the reference frame of the stationary pivot foot. The pivot foot (marked by the foot icon) has the ankle marker at the coordinate origin, with the foot-tip marker aligned along the x-axis. Circles denote sample start positions of the stepping foot, and diamonds denote the end positions. Trajectories on top (red) are those of the left foot ankle marker (left-footed steps), and those on bottom (blue) are those for the right-foot. Units in cm.

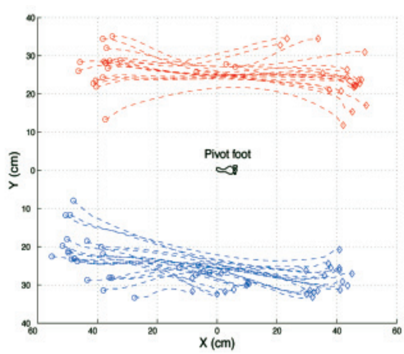

Figure 5. Plot of foot start and stop positions (in cm) and trajectories for walking data acquired with a new style: happy walking. Learning of styles is done based on the original moveme model and requires fewer samples (16 for stepping with each foot). Plot conventions same as figure 4.

9.6 3-D SAD WALK, 3-D SEXY WALK

Dataset: Two additional styles of walking were acquired. In one, the actor walked as if he were very sad; head hung low, dragging his feet, and with little arm movement. In the other, the actor walked ‘sexy’, with exagerated hip swing and arm motion.

Moveme parameterization: As with the happy style, the motion is parameterized with the same type of parameters as the original 3-D walk moveme data.

9.7 3-D STEP-OVER

Dataset: Examples of stepping-over an object were recorded. The variability of the samples included the size of the object (roughly three heights, 15 cm, 30 cm, and 50 cm were

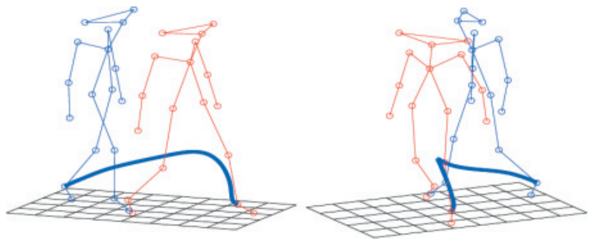

Figure 6. Two representatives of the stepping-over moveme, with the trajectory of the stepping ankle traced out in 3-D.

used). Because an actual physical object would interfere with the motion capture process, an imaginary object of variable height and length was used. Also, the angle of attack, and curvature of the walk path during the stepping-over was varied. In all, 29 steps with the left foot and 34 with the right foot were captured. Figure 6 shows the start and end pose of two samples of stepping-over with the left foot, displaying the 3-D trajectory of the left ankle.

Moveme parameterization: For the 'stepping-over' moveme, not only does the initial and final position of the stepping foot have to be specified, but also the height of the object (actually the height to raise the foot to) is part of the goal parameters. This value was extracted from the examples as the maximum height that the moving foot achieved. See figure 7 for the labels of all the step-over examples. Note that for stepping-over an object, there are four possible movemes involved; the lead step with the left or the right foot, and the follow-through step with the other foot. The dynamics of the movement of the lead foot is quite different from that of the follow-through foot, and so those movemes should be learned separately. Due to a limitation in the mo-cap system (the light-bulb suit limited the range of motion of the knees), it was not possible to capture examples of the follow-through moveme.

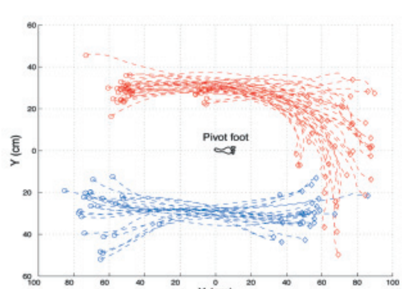

Figure 7. Plot of foot start and stop positions (in cm) and trajectories for the stepping-over moveme. Note that these steps are much longer than during normal or happy walking. The moveme is parameterized not only with the foot start and stop positions, but also by the height of the object being stepped over. Plot conventions same as for figure 4.

9.8 3-D LOOK

Dataset: Starting from the same initial position, looking straight ahead, the actor turned his head, neck and torso (entire body, in fact) to look in various directions; looking straight up, down to the floor, to the left, to the right. The visual hemi-field was sampled approximately every 20 degrees (vertically and horizontally) to generate the samples. In total, 34 samples were acquired.

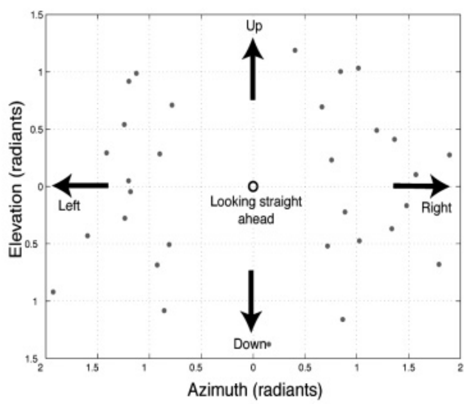

Figure 8. The labels for the 34 acquired samples of the look moveme. X-axis is azimuth of head, Y-axis is elevation (angles in radians), where origin (open circle) corresponds to the subject looking straight ahead.

Moveme parameterization: In the 3-D look dataset, the actor always started and ended from the same initial position, thus it is not necessary to encode any state parameters. The goal parameter is the gaze direction (azimuth and elevation), computed as the azimuth and elevation of the vector from the centroid of the two ear markers to the forehead marker. Figure 8 shows the labels for the samples.

9.9 3-D RUN

Dataset: Data to learn a run moveme was captured by placing an actor on a treadmill and having him run at various speeds, from standstill to all-out sprint (or as close to a sprint as was safe). Ideally, one would capture data of a subject running along various straight and curved paths, but this was not possible due to the small capture volume of our motion capture system. Like the walk moveme, data was separated into left-footed and right-footed strides to learn two (half-cycle) run movemes.

Moveme parameterization: Since the actor was facing forward on the treadmill at all times, the only goal parameter definable is the speed of the running, measured as the average speed of the treadmill during each stride (with the lenght of the stride varying as a function of speed).

10. PERFORMANCE

In this section we study the quality of the different moveme models using three diagnostics. First, the root mean square error can be used as a numerical measure of quality. Second, a perceptual evaluation can be made by using the models to synthesize movemes and having naive observers rate them. Finally, the quality of the models when they are used to synthesize movemes outside of the normal input range is studied by qualitative visual comparison.

10.1 RMS PERFORMANCE

We use the normalized root mean square error to quantify the quality of a moveme model. The normalized root mean square error is the RMS error of the model divided by RMS error of the simplest of all models, the constant mean value model. A value of 0 indicates a perfect model with no error, and a value of 1 indicates the model is as inaccurate as the mean value model.

The first two tables (Tables 10.1, 10.2) show the normalized error levels of the linear, quadratic, and radial basis function models for the 2-D reach and 2-D draw moveme, respectively. The first column shows the error when the entire dataset was used to learn the model. The second column shows the cross-validation error, where some of the data samples were not used for building the model, but instead were used only to compute the model error. For the reach moveme, increased model order improves the accuracy of representation, and only minimal over-fitting is occuring. Note that for the same size of model the polynomial models are significantly more accurate than the RBF models. With the draw moveme, since the input (parameter) space is large (8-dimensional) and not many samples are available (378), the effects of over-training are clearly visible; the cross-validation error of the quadratic model is almost double that of the linear model. Also the radial basis function model has an error more than 10 times as large as the linear model, since it is very difficult to cover a large dimensional space.

From the results above we conclude that polynomial models outperform radial basis function models, and so concentrate further analysis on those models only. In table 10.3 we indicate the representational accuracy of linear and quadratic models for the different 3-D movemes. Note that in many instances the quadratic model does not improve significantly over the linear. For styles, only the result for linear models are shown since there are not enough data samples to create a quadratic model.

10.2 PERCEPTUAL EVALUATION OF MOVEME MODELS

The perceptual quality of the moveme synthesis method was evaluated with formal and informal tests with various subjects.

In the formal tests, subjects were presented a two alternative forced choice paradigm where they were asked to distinguish between original and re-synthesized movemes. Two actions were tested, 2-D reaching, and 2-D drawing. The movemes were presented

Table 10.1. Comparison of different learning models for the reach moveme. Linear, quadratic, and cubic interpolants, as well as Radial Basis Function (RBF) networks with 5, 9, and 20 basis functions were used to create models mapping the 2-D reach location to the motion of the entire body (14 makers in 117 frames). Increasing the polynomial degree increased the accuracy of model, seeminghly without over-training. For a similar model size, RBFs generalized less well. Cross-validation values represent normalized RMS error mean ± standard deviation for 1000 independent iterations of training and testing.

Method		RMS Error (all data set)	RMS Error (2/3 learn, 1/3 test)	Model size
Polynomial	Degree=1	0.1357	0.1533 ± 0.0163	3276x3
	Degree=2	0.0677	0.0842 ± 0.0108	3276x6
	Degree=3	0.0460	0.0663 ± 0.0100	3276x10
RBF	N = 5	0.0921	0.1335 ± 0.0145	3276x6
	N = 9	0.0461	0.1011 ± 0.0122	3276x10
	N = 20	0.0314	0.0945 ± 0.0109	3276x21

Table 10.2. Comparison of different learning models for the draw moveme. Due to the large dimension of the input space (an 8 dimensional vector is used to parameterize the shape of the stroke to draw), the RBF models are approximately three times less accurate than polynomial interpolants, given the same number of model parameters. Although the errors of the polynomial interpolants are larger than for the reach moveme models, trajectories synthesized with these models still retain a realistic appearance.

Method		RMS Error (all data set)	RMS Error (2/3 learn, 1/3 test)	Model size
Polynomial	Degree=1	0.260	0.258 ± 0.10	1680x9
	Degree=2	0.215	0.344 ± 0.08	1680x17
RBF	N = 8	0.749	0.872 ± 0.04	1680x9
	N = 16	0.608	0.831 ± 0.04	1680x17

as moving light displays [18], and each subject viewed 30 stimuli pairs. For each stimuli pair, an original moveme was randomly chosen from samples in the motion capture dataset, and the corresponding moveme label was used to create a synthetic moveme. For the reaching action, a 3rd order polynomial model was used, while for the drawing action a linear model was used. After presenting the stimuli pair (with random order of appearance between the original and synthesized moveme), the subject chose which appeared more realistic. In case the subject was unsure, he was instructed to guess. If the true and re-synthesized movemes were completely indistinguishable, subjects should perform the task at chance level (50% correct responses). The results in table 10.4 show that indeed our subjects were unable to distinguish between real and synthetic movemes.

Various informal tests were also conducted. During the development of the different moveme model techniques, perceptual tests were always used to assess model quality. Perhaps the most significant perceptual tests were those where an interactive demo that used the moveme models to synthesize animation in real-time was shown to profession-

Table 10.3. Summary of the size of 3-D moveme data and models. The first column list the number of sample motions acquired for each moveme. The second column is the number of goal and/or state parameters used to parameterize the moveme. The 'RMS Error' columns denote the normalized RMS error when linear and quadratic models are used to represent the movemes. For learning of styles, only the linear model can be applied due to the small number of sample motions. 'Compression' indicates the ratio of the size of the raw data of a sample trajectory divided by the number of label parameters that specify the movement.

Moveme	**N. Sample Motions**	**N. Labels (Parameters)**	**RMS Error**		**Compression**
			Linear	**Quadratic**	
3-D Walk - Left Step	124	4	0.332	0.314	600
3-D Walk - Right Step	119	4	0.281	0.267	600
3-D Look	32	2	0.531	0.385	3160
3-D Run	56	1	0.675	0.627	2400
3-D Reach	33	3	0.393	0.270	6480
Step-Over - Left Step	29	5	0.427	-	480
Step-Over - Right Step	34	5	0.451	-	480
Happy - Left Step	16	4	0.754	-	600
Happy - Right Step	16	4	0.726	-	600
Sad - Left Step	19	4	0.743	-	600
Sad - Right Step	18	4	0.735	-	600
Sexy - Left Step	22	4	0.709	-	600
Sexy - Right Step	25	4	0.697	-	600

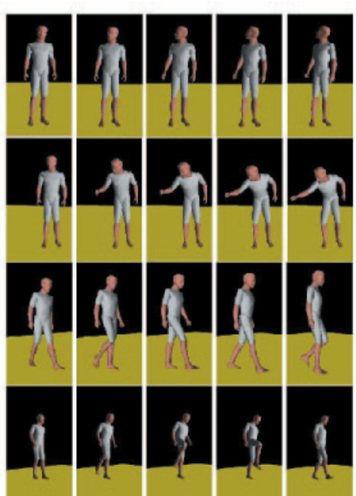

Figure 9. A real-time, interactive synthetic character performing look, reach, walk, and step-over movemes based on the high-level control of a user. The resulting motion is not only extremely versatile due to the use of goal and style parameters, but is also very realistic and fluid.

al game developers at a Game Developer's Conference (September 98) and in private meetings (August 99). They were unanimously impressed by the quality of the animation, many claiming that they had never seen such high quality real-time animation. Figure 9 shows some screen shots of a synthetic character performing various movemes.

10.3 EVALUATION OF MODEL EXTRAPOLATION

When synthesizing a new moveme for animation purposes (and even during recognition and perception), it is not always guaranteed that the moveme parameters will fall within the convex hull of the samples used to learn the model. To see how different models fare, the same (extrapolating) parameters are applied to the various models, and the resulting moveme are inspected visually. As expected, the linear model is much better behaved than the quadratic model and the radial basis function model. The radial basis function model simply fails to represent motion output beyond the convex hull, resulting in poor synthesis. The quadratic model "explodes" due to the large effect of the quadratic terms beyond the convex hull.

Thus the empirical conclusion from all the different evluations of the moveme models is that linear models are the best compromise in terms of accuracy of representation, requiring fewer examples, and extrapolation ability.

Table 10.4. Perceptual discriminability between original motions and reconstructions for the reach and the draw movemes. Subjects were presented 30 pairs of real and synthetic motions in random pair-wise order and were asked to determine whether the first or the second was the original motion. The mean and standard deviation match quite closely to the theoretical values for 30 i.i.d. coin tosses (mean of 50%, standard deviation of 9.13%), indicating that subjects found it difficult to dinstinguish between the two.

Subject	**Reach % Correct**	**Draw % Correct**
B.P.	36.7	50.0
L.G.	60.0	50.0
J.B.	33.3	53.3
Y.S.	46.7	-
Y.K.	63.3	-
P.M.	60.0	-
P.P.	46.6	-
M.M.	-	46.7
E.D.	-	46.7
Mean	49.51	48.28
St. dev.	11.93	3.59

11. CONCATENATION

If movemes are to be used as elemental motions, it should be possible to concatenate them together to form more complex actions. This is possible with our definition of movemes through careful design. To ensure good concatenation, the body pose at the end of one moveme needs to match the initial pose of the next moveme. For the look, and reach movemes, the initial (and final) pose is the 'rest' pose; standing looking forward, with arms at the side. These movemes are also cyclical, in the sense that the full motion is represented; both the forward moving part of a reach (or look) and the return motion are represented. Thus they can easily be concatenated. The step moveme is also cyclical, so that it can be concatenated with itself. However, in a typical stride the body pose will not match up with the rest pose of the reach and look movemes. To ensure that concatenation of the walk moveme with the reach and/or look moveme is possible, the set of sample trajectories of the walk moveme included examples of starting to walk, and stopping, where (respectively) the subject started from the rest pose and ended in the rest pose.

In the limit, one can hypothesize that the models for all movemes have state parameters that encode the initial pose of the body (or at least the pose of the relevant body parts) so that a properly concatenated moveme can always be generated.

12. DISCUSSION AND CONCLUSIONS

We have proposed to model human motion by concatenating discrete elements, the 'movemes', which are learned by observing people move. Our argument in favor of such model stems from the observation that common human motions are stereotypical and determined by somewhat arbitrary patterns that are generated by the brain, rather than being determined by the mechanics of the body. Suggestions in this direction come both from the motor control literature and from the physiology and psychology of perception.

We have explored six movemes and highlighted their properties. We demonstrated that movemes greatly compress motion information: just a few goal and style parameters encode human motions which are characterized by thousands of position parameters. All the movemes that we studied had a natural parameterization in terms of the goal of a specific simple action that is associated with the moveme. Simple linear or quadratic maps were found to represent each moveme very well, to the point that human subjects are unable to tell motions produced by human actors apart from motions produced by our models. Such maps are specified by few parameters and thus require few training examples to be learned.

More questions are opened rather than settled by postulating the existence of movemes. First of all: how many movemes might there be and how might one go about building a catalogue. We do not have an answer to this question: we may only take educated guesses by comparing with phonemes and by counting action verbs in the dictionary; the estimates we get range around 100 movemes (order of magnitude). A related practical question is whether it would be possible to define subsets of movemes that are sufficient for visual perception and animation of human motion in restricted situations: e.g. urban street scenes, mountain hiking, playing hide-and-seek in woods.

A second class of questions regards the practical aspects of using moveme-based models for animation and vision. We have successfully animated characters using movemes and combinations of movemes, and we have attributed different 'styles' to these characters as well. Rose et al [24] also describe an animation system based on 'verbs' and 'adverbs' which is similar to our movemes with goal and style parameters. It is still unclear, however, whether superimposing and concatenating new types of movemes will always be natural and possible, or whether there will be a combinatorial explosion of boundary conditions to be handled in combining different movemes.

Where does all this place us in interpreting biological motion perception? Movemes appear to be a natural and rich representation which the brain might employ in perceiving biological motion. However, more work needs to be done to build a complete moveme-based human motion tracker/recognizer that might be compared with psychophysical measurements [18]. A germ of such a system may be recognized in Bregler and Malik's motion tracker [13] which switches between four linear time invariant dynamical systems representing different phases of a walking step. Black et al [25] also describe a movemelike approach where parameterized 'temporal primitives' of cyclical motions are estimated from motion capture data, and the models are then used to track the motion of a subject in 3-D. Finally, Mataric [20] has developed a perception-action model based on a compact set of 'basis behaviors' that enables a humanoid robot to both perceive human action and imitate it.

A third class of questions has to do with the high-level aspects of the analysis of human motion. The final aim in many applications is to 'understand' human activity from images, or synthesize motion from high-level descriptions. This is done at a more abstract level than movemes: we are ultimately interested in descriptions such as 'she is cautiously pruning her rose-bushes' rather than '... steps forward, then reaches for stem while looking at stem ...'. How well will we be able to map sequences of movemes to high-level descriptions of activities and vice-versa?

A. MOTION MODELS

A.1 Linear models

The simplest model one can use is a global linear model. If $\{x^s\}$ is the set of all label vectors (written as column vectors) for a particular dataset with N_s samples, we can form an augmented input matrix, X: $X = [1...1; x^1 ... x^{N_s}]$. X is the matrix of sample inputs, augmented by a row of 1's. This row is used to learn the constant bias term in the model.

Likewise, if $\{y^s\}$ represents all the sample movemes, where for sample s, the column vector y^s consists of the stacked up coordinates of all the markers throughout all the frames of the moveme, then we can form the sample output matrix Y:$Y = [y^1 ... y^{N_s}]$.

Now the best linear model can be found by solving for L in the least squares problem $Y = LX$, and this is easily done as $L = YX^t(XX^t)^{-1}$.

Then our map is $y = f(x) = YX^t(XX^t)^{-1}x$.

A.2 Higher-order polynomial models

The next obvious choice for a model, then, is to include higher order terms in the multidimensional polynomial interpolant. One can learn a global quadratic model by adding additional rows to the input matrix X corresponding to the pairwise products of individual label parameters. For a 2-dimensional label, three such products can be formed; with a 4-dimensional label there are 10.

The process can be formalized by defining N_b polynomial basis functions, with n^{th} function $\Phi_n(x)$ defined as

$$\Phi_n(x) = \prod_{i=1}^{N_D} x_i^{c_{i,n}} \tag{7}$$

where x is a sample moveme label, x_i is the i^{th} component of the label vector, and $c_{i,n}$ is the power to which the i^{th} label component is raised in the n^{th} basis function. For the constant basis function, all c's would be zero; for linear basis functions, only one c is 1 and all others zero; and for quadratic basis functions the sum of the exponents has to equal 2 (either one is 2 and the others zero, or two of them are 1).

If we denote $\Phi(X)$ the matrix generated by applying all the basis functions to all the sample moveme labels.

$$\Phi(X) = \begin{bmatrix} \Phi_1(x^1) & \dots & \Phi_1(x^{N_S}) \\ \dots & \dots & \dots \\ \Phi_{N_b}(x^1) & \dots & \Phi_{N_b}(x^{N_S}) \end{bmatrix} \tag{8}$$

then the best global polynomial model can be found by solving for the coefficient matrix W in $Y = W\Phi(X)$, which is also solved by the least squares pseudo-inverse method.

In principle, with basis functions of higher and higher polynomial degree the function approximation can be more and more accurate (if the basis includes all possible polynomial terms up to a certain degree, the model is a multi-dimensional Taylor expansion around the zero vector). However, the model can quickly become unwieldy, as the number of basis functions (and size of the coefficient matrix W) grows exponentially with the degree of the approximation (and this in turn demands the use of more training data to prevent over-fitting).

Although it will be shown in the experimental section that global quadratic models improve the fidelity of the re-synthesis of the sample movemes, it will also be shown that the synthesized outputs do not degrade gracefully when the input label goes outside the range of the training examples. This is the same behavior as in the simple scalar input and output case, where it is known that the higher the polynomial order, the worse the extrapolation error.

A.3 Radial basis functions

A further generalization of the method of learning weights for a set of basis functions is to use other sets of functions other than polynomial functions for the set of basis. One set of functions that has been widely studied is that of radial basis functions [5]. The basic idea behind radial basis functions is that instead of using global polynomials of high degree to learn a highly non-linear function, one can use many basis functions with local support, and each one encodes the local behavior of the system; The n^{th} radial basis function is defined as:

$$\Phi_n(x) = exp((x - \mu_n)^t \Sigma_n^{-1} (x - \mu_n)) \tag{9}$$

Thus the n^{th} radial basis function is centered at μ_n in the input space, has value 1 when $x = \mu_n$, and the value decays to 0 as x moves away from μ_n at a rate controlled by the covariance 'spread' matrix Σ_n. As explained in [5], if we associate a radial basis function with each sample input such that for Φ_n we let $\mu_n = x^n$ (where x^n is the n^{th} sample input), then a matrix of weights can be learnt that will interpolate through the samples exactly. Just as in the global polynomial models, we solve for W in $Y = W\Phi(X)$, where now the matrix $\Phi(X)$ consists of the values of all the radial basis function for all the sample inputs. Furthermore, for an appropriate value of the spread parameters Σ_n, the model can be made to interpolate smoothly between samples.

However, one may not want an exact interpolation through all the samples because a) one knows that there is noise in the samples, and/or b) it would result in too large a model (if there were many samples). In this case, one uses many less radial basis functions, and the basis centers, μ_n, also have to be learnt, along with the coefficient matrix W and the spread matrices Σ_n.

There are many different learning algorithms for optimizing the model parameters. Because of the highly non-linear nature of the model, it is very computation intensive to optimize over all the parameters. One simplification is to replace the spread matrices with a constant, an a priori specified value that is the same for all basis functions. This is the form of the model that we experimented with.

Although radial basis function models can provide good results in some instances, it suffers from three drawbacks that make it mostly unsuitable for learning models of human motion. First, the number of basis functions required to 'fill the input space' grows exponentially with the number of dimensions of the moveme labels. Second, the basis functions are placed where the data lies in the input space, but if the input space is not sampled uniformly, there may be gaps, and then the model is not guaranteed to interpolate properly within large gaps. Finally, because of the local extent of each basis function, the model cannot extrapolate very well.

A.4 Feed-forward networks

The widely known method of feed-forward neural networks trained with the back-propagation scheme [19] can also be used to learn moveme models. A network with two layers could be used functions $\phi_i(x)$ of the input label x, and the output units compute linear transformations of those hidden variables, $y_j = \Sigma_i W_{ji} \phi_i(x)$. The structure of the network is identical to that of the global polynomial interpolators, or the radial basis function networks; the only thing that changes is the functional form of the nonlinearities $\phi_i(x)$. In this type of network, the nonlinearities are sigmoidal activation functions:

$$\phi_i(x) = g(w_i^T x + b_i) \tag{10}$$

where

$$g(\alpha) = \frac{1}{1+e(-\alpha)} \tag{11}$$

Figure 10(a) shows the activation function of $g(\alpha)$. Near zero, g is linear, until it saturates at a value of 1 above, and 0 below. In equation 10, w_i and b_i define a separating hyper-plane in the input space (depicted for a 2-dimensional input space in the figure 10(b)). The width of the linear region depends on the magnitude of w_i; the smaller the magnitude, the larger the region.

When a FFNN is used for pattern classification, any particular output of the network is 'high' (value 1, say) when the input vector belongs to that class. For this type of application, the network produces useful computations because of the saturating property of the hidden units. Each hidden unit determines whether an input vector is on one side or the other of a hyper-plane boundary in the input space. The combined outputs of several hidden units can be used to define intricate boundaries between different regions of the input space, each region representing a particular class of inputs.

When a FFNN is used to learn a continuous function, the behavior is very different. In this application, it is the linear region of each hidden unit that is important. Over the linear region, a hidden unit provides a linear approximation of the function being learned. In the saturation regions, it provides only a constant bias. A function is properly approximated only if the linear regions of each hidden unit cover the entire range of the input space. Because fo the fact that each hidden unit saturates above and below, FFNNs inherently have difficulty extrapolating; beyond the input range specified by the training examples, it is likely that all hidden units become saturated.

There are ways to overcome this difficulty. For example, the input range over which extrapolation is desired can be specified a priori, and during the network training procedure the hidden unit weights w_i can be constrained to be small enough to guarantee that the linear region of each hidden unit is at least as wide as the desired extrapolation range. Another solution might be to use a new type of nonlinearity, which saturates only on one side. With such a nonlinearity, each hidden unit becomes a linear interpolant which 'turns on' on just one side of a hyper-plane that divides the input space in two halves.

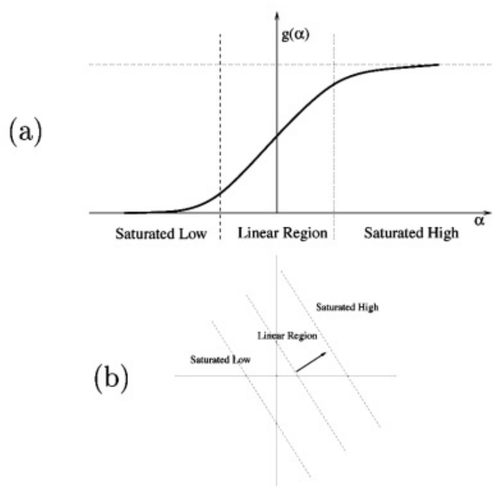

Figure 10. (a) The nonlinear sigmoidal activation function that generates hidden layer outputs $\phi_i(x)$ (b): the partitioning of a 2-D input space into a linear region and two saturation regions. The width of the linear region is determined by the magnitude of the weight vector w.

Besides the extrapolation deficiency, FFNNs have other characteristics that make them less desirable for motion learning. They are very slow to train, require the use of validation data to determine when to stop training the network to prevent over-fitting, and the trained network needs to have its weights regularized to guarantee that marker outputs are continuous through time. For these reasons, the use of FFNNs was not further explored. Nevertheless, the analysis of FFNNs was fruitful, in the sense that it provides a clearer idea of the functional properties a good moveme model should have. Namely, the input space should be separated into regions, with each region activating a local (linear) approximator. Some of the regions must be unbounded, so that the system extrapolates properly.

Prof. Luis Goncalves
California Institute of Technology
136-93,Pasadena
CA, 91125, U.S.A.

Prof. Enrico Di Bernardo
California Institute of Technology
136-93,Pasadena
CA, 91125, U.S.A.

Prof. Pietro Perona
California Institute of Technology
136-93,Pasadena
CA, 91125, U.S.A.

REFERENCES

[1] N. BADLER, C.B. PHILLIPS, and B.L. WEBBER. *Simulating Humans.* Oxford University Press, 1993.

[2] P.V. BASSIN, N.A. BERNSTEIN, and L.P. LATASH. On the problem of the relation between structure and function in the brain froma contemporary point of view. *Motor control*, 3(4): 332-342, 1999.

[3] E. DI BERNARDO, L. GONCALVES, and P. PERONA. Monocular tracking of the human arm in 3d: Real-time impementation and experiments. In *International Conference on Pattern Recognition*, August 1996.

[4] N.A. BERNSTEIN. Movements. *Motor control*, 3(2): 107-130, 1999.

[5] C.M. BISHOP. *Neural Networks for Pattern Recognition.* Oxford Univ. Press, 1995.

[6] E. BIZZI, S.F. GISZTER, E. LOEB, F.A. MUSSA-IVALDI, and P. SALTIEL. Modual organization of motor behavior in the frog's spinal cord. *Trends in Neurosciences*, 18(10): 442-446, 1995.

[7] E. BIZZI, N. HOGAN, F.A. MUSSAIVALDI, and S. GISZTER. Does the nervous-system use equilibrium-point control to guide single and multiple joint movements. *Behavioral and Brain Sciences*, 15(4): 603-613, 1992.

[8] E. BIZZI and F.A. MUSSA-IVALDI. Neural basis of motor control and its cognitive implications. *Trends in Cognitive Sciences*, 2(3): 97-102, 1998.

[9] V. BLANZ and T. VETTER. A morphable model for the synthesis of 3-d faces. In *SIGGRAPH 99 Conference Proceedings,* pages 187-1894, August 1999.

[10] C. BREGLER and J. MALIK. Learning and recognizing human dynamics in video sequences. in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.,* pages 568-574, Puerto Rico, 1997.

[11] C. BREGLER and J. MALIK. Tracking people with twists and exponential maps. In *Proc. IEEE CVPR*, 1998.

[12] ARMIN BRUDERLIN and LANCE WILLIAMS. Motion Signal Processing. In *Int. Conf. on Computer Ghaphics (SIGGGRAPH) '95*, pages 97-104, August 1995.

[13] J. MALIK C. BREGLER. Video motion capure. Technical Report UCB-CSD-97-973, CS Dept., University of California at Berkeley, 1997.

[14] W.T. FREEMAN, D.B. ANDERSON, P.A. BEARDSLEY, C.N. DODGE, M. ROTH, C.D. WEISSMAN, W.S. YERAZUNIS, H. DAGE, K. KYUMA, Y. MIYAKE, and K. TANAKA. Computer vision for interactive computer graphics. *IEEE Computer Graphics and Applications,* 18(3): 42-53, 1998.

[15] M. GLEICHER. Motion editing with space-time constraints. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics,* 1997.

[16] L. GONCALVES, E. DI BERNARDO, E. URSELLA, and P. PERONA. Monocular tracking of the human arm in 3d. In *Proc. 5th Int. Conf. on Computer Vision*, pages 764-770, Cambridge, Mass, June 1995.

[17] JESSICA K. HODGINS, WAYNE L. WOOTEN, DAVID C. BROGAN, and JAMES F. O'BRIEN. Animating Human Athletics. In *Computer Graphics Proceedings, Annual Conference Series,* pages 71-78, 1995.

[18] G. JOHANSSON. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics,* 14: 201-211, 1973.

[19] A. KROGH, R. PALMER, and J. HERTZ. *Introduction to the Theory of Neural Networks.* Persues Press, 1991.

[20] MAJA MATARIC. Getting humanoids to move and imitate. *IEEE Intelligent Systems,* pages 19-24, July 2000.

[21] R.M. MURRAY, Z. LI, and S.S. SASTRY. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[22] J. REHG and T. KANADE. Model-based tracking of self-occluding articulated objects. In ICCV95, pages 612-617, 1995.

[23] G. RIZZOLATTI and M. ARBIB. Language within our grasp. *Trends in Neurosciences*, 5(21): 188-194, 1998.

[24] C. ROSE, M. COHEN, and B. BODENHEIMER. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications,* pages 32-40, September/October 1998.

[25] H. SIDENBLADH, M.J. BLACK, and D.J. FLEET. Stochastic tracking of 3d human figures using 2d image motion, pages 702-718, Dublin, June 2000.
[26] Y. SONG, X. FENG, and P. PERONA. Towards detection of human motion. In *Proc. IEEE CVPR*, volume 1, pages 810-817, June 2000.
[27] Y. SONG, L. GONCALVES, and P. PERONA. Monocular perception of biological motion - clutter and partial occlusion. In *Proc. ECCV*, volume 2, pages 719-733, June/July 2000.
[28] M. TURK, editor. *Workshop on Perceptual User Interfaces (PUI'97)*. Banff, Canada, 1997.
[29] A. WITKIN and M. KASS. Spacetime constraints. *Computer Graphics,* 22(4): 159-168, 1998.

JEAN LORENCEAU

FORM CONSTRAINTS IN MOTION INTEGRATION, SEGMENTATION AND SELECTION

INTRODUCTION

Perception is a process by which living organisms extract regularities from the physical fluxes of varying physical characteristics in the external world in order to construct the stable representations that are needed for recognition, memory formation and the organisation of action. The exact nature of the process is still not well understood as the type of regularities that are indeed used by sensory systems can be manifold. However, perception is not a process by which living organisms would reproduce the physical fluxes such as to build an internal representations identical to its physical counterpart.

One issue then is to understand the kind of physical regularities that are relevant for perceiving and recognising events in the outside world. One way to address this question is to consider what we do not perceive. For instance, in vision, we are perceptually blind to individual photons and cannot see how many of them stroke the retinal receptors, or determine their exact wavelength, despite the fact that photons are the primary inputs that activate the visual brain. In contrast, we do perceive objects of all kind, even though they are seen for the first time, are partially hidden by other objects, or are seen under different illumination levels and altered by shadows. Thus, despite huge modifications of the photon flux, we are able to perceive and recognise the same objects.

THE GESTALTIST APPROACH TO INVARIANT

Early in the 20th century, Wertheimer (1912) analysed in details the theoretical consequences of the fact that two static flashes presented in succession with appropriate temporal and spatial separations elicit a perception of visual apparent motion. This observation was taken as a strong indication that the visual system does more than simply registering external events, suggesting that the "whole is more than the sum of its parts". From this seminal analysis of apparent motion emerged a scientific attempt to define the general principles underlying that claim, known as the Gestalt psychology.

The gestalt school, with the impulsion of Wertheimer (1912), Koffka (1935) and Köhler (1929), and later in the century Metzger (1934), Kanisza (1976) and a number of others, developed experimental paradigms to define and isolate the general rules underlying perceptual organisation. Using simple visual or auditory stimuli some principles involved in perceptual organisation could be qualitatively assessed. In vision, figure/ground segregation and perceptual grouping of individual tokens into a "whole" appeared to strongly rely on several rules such as good continuity, proximity, closure, symmetry, common fate, synchronicity etc. Most importantly, these principle define

A. Carsetti (ed.), Seeing, Thinking and Knowing, 171-189.

spatial and temporal relationships between "tokens", whatever the exact nature of these tokens: dots, segments, colour, contours, motion, etc. Implicitly, the general model underlying the Gestaltist approach is a geometrical one, stressing the spatial relations between parts rather than concerned with the intrinsic processing of the parts themselves. However, the attempt of the gestalt school to offer a plausible neuronal perspective that could explain their observations on perceptual organisation failed, as the Gestaltists were thinking in term of an isomorphism between external and internal geometrical rules whereby spatial relationships between neurones would mimic the geometry of the stimulus. Electrophysiological and anatomical studies did not revealed such an isomorphism. Yet, in this paper, we consider recent neurophysiological findings that suggest how geometrical principles may be implemented in the brain and discuss hypotheses about the physiological mechanisms that may underlie perceptual grouping.

VISION THROUGH SPATIAL FREQUENCY FILTERS

In contrast with the Gestaltist program of research, and following the psychophysical approach defined by Fechner at the end of the 19th century, another view of how the visual system processes its inputs developed. The goal of this approach was to establish the quantitative relationships between the physical stimulus inputs and the perceptual outputs of sensory systems, in order to define and measure the elementary sensations processed by the human brain. In this atomist approach, less emphasis is put on perceptual organisation and grouping and more on the early processing performed by the central nervous system, using threshold measurements as a tool to probe sensory systems. This approach requires to define a model of the stimulus, a model of the perceptual and neural processes at work and a model of the decisional processes needed to generate an observer's response. Some powerful laws of perception could be demonstrated in this methodological framework, among which the Bouguer-Weber's law on discrimination thresholds or the so-called Fechner's law according to which sensation grows as a function of the logarithm of the intensity of the physical inputs. This approach soon benefited from the tremendous progresses in electrophysiological techniques in the mid of the 20th century that allowed the recording of the responses of cortical cells in response to visual stimulation. The discovery that retinal ganglion cells respond optimally to a restricted portion of the visual space, and are selectively activated by localised distributions of luminance or chromatic contrast gave birth to the notion of a spatially limited receptive field that processes locally specific characteristics of the visual inputs (Hartline, 1940; Hubel & Wiesel, 1968). Several classes of ganglion cells with different functional and morphological properties were identified, among which the magnocellular and parvocellular cells project through parallel pathways to the lateral geniculate nucleus and from there to primary visual cortex and higher areas. These new findings led psychophysicists to look for the perceptual consequences of the existence of these cells, as the spatio-temporal structure of their receptive fields provided strong experimental evidence to determine a model of the stimulus relevant to biological vision. In their seminal study, Campbell & Robson (1968) offered that the centre-surround receptive field structure is well suited to process the incoming retinal inputs in a way analogous to that proposed by the French mathematician Fourier to analyse periodic events. Fourier's theorem states that any complex event can be mathematically decomposed into a simple sum of sinusoidal waves of different frequency, amplitude and phase. According to Campbell & Robson, the ganglion cells in the retina would decompose visual inputs into a spectrum of spatial frequency bands, each being selectively processed by a sub-population of neurones. The idea that the visual system transforms a spatial distribution of light into a set of different spatial frequency bands had a huge impact on subsequent research. Since then, the model of the stimulus considered as relevant to study the visual system was based on the linear Fourier decomposition. Consequently, the most elementary stimulus, represented by a single point in the frequency domain, is a sinusoidal distribution of luminance contrast, consisting of a simple extended oriented grating. One of the great advantage of this model was that the Fourier transform is a linear operation, and therefore appeared as a powerful tool to determine whether the visual system itself behaves as a linear spatial frequency analyser.

A Fourier decomposition of a two-dimensional (2D) image result in both an energy spectrum, describing the distribution of amplitude in different spatial frequency bands at different orientations, and a phase spectrum, that contains information about the absolute phase of different spatial frequencies and orientations, and that represents the position of different spatial frequencies relative to an arbitrary origin (figure 1).

In practice, both psychophysicists and electrophysiologists used extended gratings of different spatial frequencies, orientation and contrast to probe the visual system, as they were mostly concerned with the effects of the energy spectrum on contrast sensitivity and neuronal selectivity but were less concerned with the phase spectrum. However, simple cells in primary visual cortex were found to respond to the spatial phase, the position of a grating within their receptive field, and psychophysical studies showed that observers rely heavily on the phase spectrum to recognise objects and scenes, and to a lesser extent on the energy spectrum. For instance, when blending, through image synthesis, an energy spectrum of an image A with the phase spectrum of an image B, image B is more easily recognised than image A (Piotrowski & Campbell, 1982).

One issue with the representation of images in the Fourier domain is that the position and geometrical relationships between different parts of an image, although they are somehow embedded in the phase spectrum, are difficult to visualise and analyse. This is mainly because the phase of each spatial frequency component is expressed relative to an arbitrary origin, but does not represent directly the relative phase between different spatial frequencies that would give information about their spatial relationships. Consequently, researchers have often discarded the phase spectrum, and thus the geometry, when analysing their stimuli. More recently, the development of multiscale analysis and wavelet transform that use filters well localised both in space and in the frequency domain provided new tools to describe the morphological properties of images while modelling more accurately the response profile of cortical cells. However, the extraction of

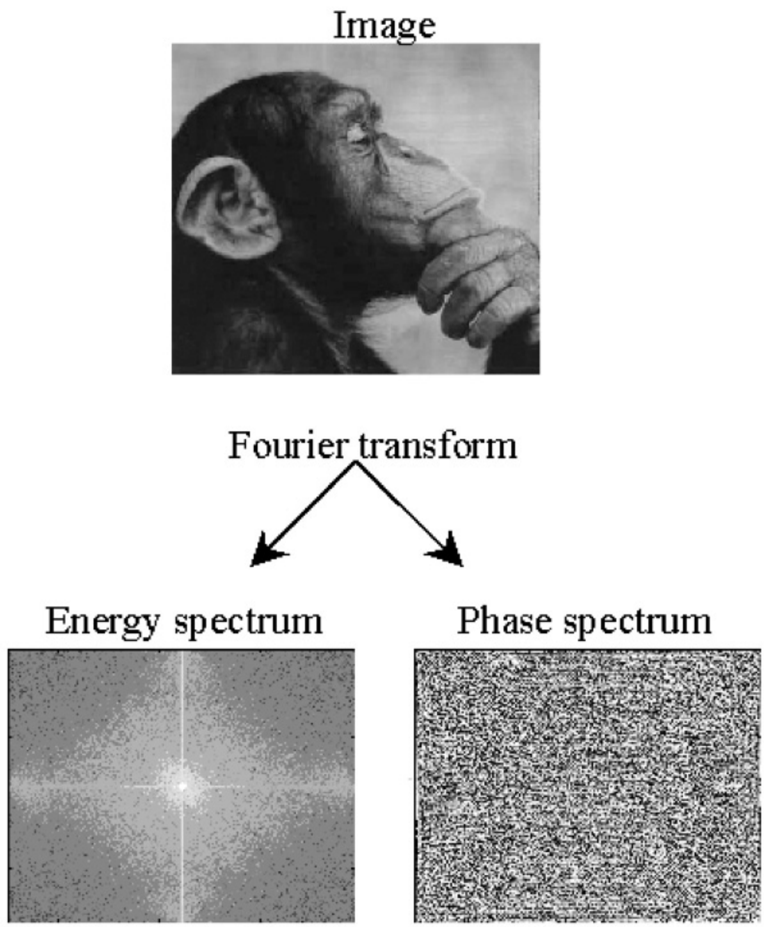

Figure 1. An image and its Fourier transform that describe the amplitude and orientation of the frequency spectrum (left) and the phase spectrum (right).

the geometrical properties of images with the wavelet transform remains a difficult problem (Mallat, personal communication).

The strong electrophysiological evidence suggesting that visual neurones with spatially restricted receptive fields indeed respond selectively to spatial frequency led to the idea that the visual cortex consist in a set of spatial frequency analysers working independently and in parallel (De Valois & De Valois, 1988). Accordingly, a complex input image activates a large population of cells distributed across the cortical surface, each analysing a different region of the space and processing a specific spatial scale of the stimulus input. Given this distributed representation of an image in the visual cortex, important issues arise: how does the brain aggregate the numerous neuronal responses to a complex object in order to segregate it from its background? How and under what conditions does the visual system bind together these distributed responses while avoiding spurious combinations? What are the rules and physiological mechanisms involved to account for this perceptual organisation?

In the following, we very briefly describe the organisation of the visual brain and present some of the challenging issues that emerge. We then present and analyse experimental results related to this issue using a simple experimental paradigm that proved useful to uncover the role of geometrical constraints in perceptual grouping and to understand how the brain builds up perceptual moving objects from their moving parts. Finally, we discuss some physiological mechanisms that may account for the experimental findings.

The recent progress in our knowledge of both the anatomy and physiology of the visual brain indicates today that it consists of an elaborately interconnected network of over 30 visual areas with highly specialised functional roles (Van Essen & De Yoe, 1995). Their organisation is generally thought of as a roughly hierarchical structure in which neural computations become increasingly complex. According to this view, neurones at the lower levels of the hierarchy would process elementary characteristics from relatively small regions of visual space. The results of these analyses are then passed onto and processed by higher level units that integrate information across larger spatial and, possibly, temporal extents. Anatomical and physiological evidence support this convergence scheme. For instance, the responses of rods and cones to distribution of light intensities are combined both spatially and temporally through convergent projections to build up the centre-surround receptive field of ganglion cells tuned to spatial frequency. Further combination of the outputs from lower level detectors would then explain the processing of orientation, colour, movement etc. (Hubel & Wiesel, 1968). According to this view, at each cortical level, detectors respond to those features to which they are preferentially tuned, within a fixed location in retinal space. Moreover, neurones tuned to different dimensions such as motion, colour or form are located in distinct areas distributed along two parallel pathways, often viewed as the expression of a "what and where" or "perception/action" dichotomy. This stems from a number of electrophysiological recordings of cells in areas distributed along these pathways and from psychophysical and neurological studies, showing that motion processing and oculomotor control were specific to the dorsal pathway, while colour and form analysis would be mostly performed in the ventral pathway.

CONTOUR INTEGRATION AND LONG-RANGE HORIZONTAL CONNECTIONS

In the primary visual cortex, the receptive fields of visual neurones are organised in a retinotopic fashion, such that neighbouring neurones analyse near regions of the visual field. Neurones in a single column perpendicular to the cortical surface are selective to the same orientation while orientation selectivity changes smoothly from one column to the next resulting in hypercolumns where neurones are processing neighbouring regions of the visual space. This pattern suffers exceptions however, as it was found the cortical surface present singularities, called pin-wheels, where neurones change rapidly their orientation selectivity as well as their positions in visual space.

This organisation suggested that the brain processes its visual inputs in parallel through spatially limited receptive fields insensitive to remote influences. This view has recently been challenged by physiological studies showing that the responses of V1 neurones to oriented stimuli presented within their receptive field can be markedly modulated by stimuli falling in surrounding regions, which by themselves fail to activate the cell. These influences could be mediated by cortico-cortical feedback projections from higher cortical areas as well as by long-range horizontal connections found within V1. Indeed, these horizontal connections link regions over distances of up to 6-8 mm of cortex, tend to connect cells with similar orientation preferences and more specifically cells whose receptive fields are topographically aligned along an axis of colinearity. Thus, this circuitry –feedback and long range connections within a single area- provide a possible physiological substratum to compute some of the geometrical properties of the incoming image.

In support of a functional link between neurones through horizontal connections in primary visual cortex, a number of recent psychophysical studies uncovered strong contrast dependent centre-surround interactions, either facilitatory or suppressive, that occur when one or several oriented test stimuli are analysed in the presence of surrounding oriented stimuli. For instance, contrast sensitivity is improved by similar flankers, collinear and aligned with the test stimulus. Changing the relative distance, orientation, spatial frequency or contrast of the flankers modulates the change in sensitivity, allowing the analysis of the architecture of these spatial interactions (Polat & Sagi, 1994). In addition, the ability to detect the presence of a specific configuration of oriented bars immersed within a surrounding textures of randomly oriented elements with similar characteristics is better for configurations of collinear and aligned elements than for parallel configurations. Field & collaborators (1993) proposed that perceptual association fields are involved in this contour integration process, and suggested that the architecture of horizontal connections may underlie these effects. This notion of association field is supported by studies showing that these interactions are decreased or suppressed in amblyopic patients who suffer from a disorganisation of long-range connectivity. Overall, these studies are compatible with the view that long-range connections play a functional role in perceptual contour integration, and further suggest that they may constitute a physiological substrate that implement some of the gestalt rules at an early processing stage.

INFLUENCE OF SINGULARITIES IN MOTION INTEGRATION, SEGMENTATION AND SELECTION

It has long been known that object motion or self motion can elicit a perception of a form –structure from motion- that would not be recognised if the retinal image was static, as is the case with biological motion (Johansson, 1950) or rotating three dimensional clouds of dots. However, less is known on the influence of form on motion perception.

We now briefly describe the results of psychophysical experiments concerned with the influence of perceptual interactions between form and motion processing on motion integration, segmentation and selection. We present evidence that motion grouping relies heavily on the processing of local singularities such as junctions and line-ends, and on more global properties of objects such as collinearity, closure and surface formation, i.e. geometrical properties of the stimulus. In addition, experimental evidence suggests that form /motion interactions do not result from a convergence in late visual areas of motion and form information conveyed by the dorsal and ventral pathway, but already occurs at an early processing stage.

An object's motion is analysed in primary visual cortex by direction selective neurones with oriented receptive fields of limited spatial extent. It is easy to show both theoretically and experimentally that these neurones are unable to accurately signal the physical direction of an oriented contour that crosses a neurone's receptive field (Henry & al., 1972; Fennema & Thompson, 1979). The inaccuracy of these motion selective neurones occurs because they cannot process the motion component parallel to the contour, which by itself does not produces any change in the input to the cell, and can therefore only respond to the component perpendicular to the preferred cell orientation, a limitation known as the "aperture problem". There are several ways to overcome this problem. One is to rely on the richer information available at contour extremities, where the existence of singularities provides sufficient cues to solve the aperture problem in a small region of the visual field. Another possibility is to combine the ambiguous neuronal response elicited *across space* by moving contours with different orientations and to compute the physical direction according to some combination rule (Adelson & Movshon, 1982; Wilson & Kim, 1994). Note that the law of common fate proposed by Gestaltists, in which components moving in the same direction with the same speed are bounded together and interpreted as belonging to the same object (Koffka, 1935) cannot account for motion grouping, as this simple rule is insufficient to constrain a unique interpretation of visual motion. Indeed, the common fate principle implicitly assumes that visual neurons analyse 2D motion whereas most cortical neurons signal only one-dimensional (1D) motion (see above: the aperture problem). In addition motion in a three-dimensional (3D) space projects on a two-dimensional (2D) retinal space. Thus, identical retinal motion may correspond to different trajectories or conversely movements in different directions with different speeds may correspond to a unique motion of a single object.

To study how local motion signals are integrated into a global object's motion and to determine the contribution of form information in solving the aperture problem, Lorenceau

& Shiffrar (1992) designed a simple paradigm in which simple geometrical shapes were presented behind rectangular windows while moving smoothly along a circular trajectory (see figure 2a).

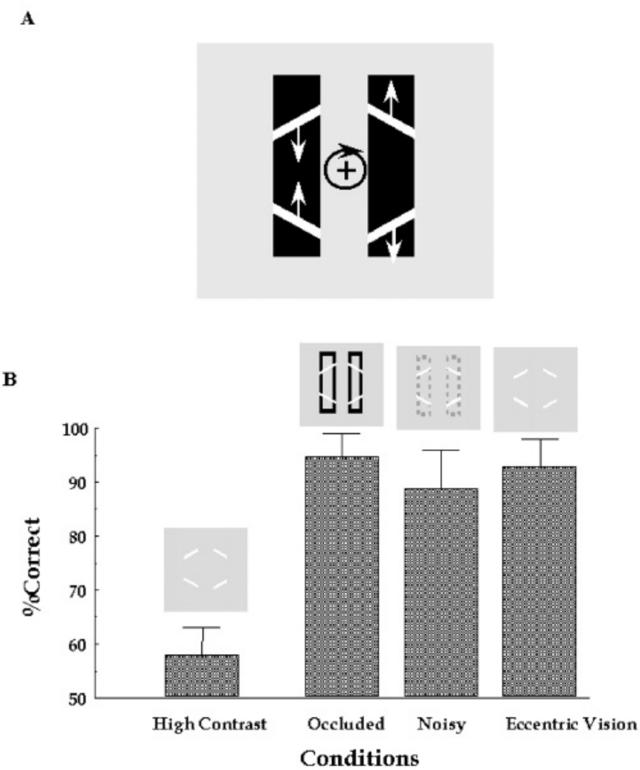

Figure 2. **A**: Stimulus used in the experiments: an outlined diamond is seen behind windows that conceal the vertices at all time. Moving the diamond along a circular path in the plane (large arrow in the centre) results in a vertical translation of component segments within each window (small arrows). Since each segment does not provide rotational information, integration across space and time is necessary to recover object's motion. **B**: Percentage of the trials where observers successfully recovered the clockwise or anticlockwise direction of motion, as a function of the different conditions tested (see text for details).

Under these conditions, a single contour segment visible in each window does not provide enough information to determine the global direction of object motion: each segment appears to move back and forth within each window with no rotational component. In order to recover the global circular object's motion it is necessary to group and combine the different segment motions. This stimulus thus offers a simple tool to test whether human observer can or not combine segment motion across space and time, in situations where singularities such as vertices do not provide a direct relevant information. Altering this occluded stimulus by changing the geometry between its constituent segments or by altering the information provided by occlusion points at the border between the window and the moving form, offer a way to assess the role of these features on perceptual grouping.

When looking at a revolving diamond under these "aperture viewing" conditions, a striking observation is that whenever the apertures are visible, either because they have a different luminance from the background or because they are outlined, the diamond appears rigid and moving coherently as a whole along a circular path whose direction can easily be determined. Decreasing the contrast between the apertures and the background decreases the perceived coherence and observers have trouble discriminating the diamond's direction. When the apertures and the background have the same hue and luminance, observers report seeing a jumbled mess of four moving segments. Clear perceptual transitions between a moving whole and its moving parts can be induced by this contrast manipulation, although at a given contrast level, such perceptual transitions also occur spontaneously over time. To get insights into this phenomenon and test different potential explanations, we modified the salience of line ends, either by using jagged apertures, that alter the salience of line-ends due to symmetrical and rapid changes in contour length during the motion, or by changing the luminance distribution along the contour (i.e. high luminance at the centre and low luminance at the line ends or the reverse). As a general rule, we found that motion coherence and discrimination performance improve as terminator salience decreases. Similar improvement in performance is observed when the overall contrast of the segments decreases, suggesting the existence of a threshold above which singularities are resolved and the diamond segmented into parts. These observations show that singularities – junctions, end points, vertices – exert a strong control on perceptual integration of component motion over space and time, and are used to segment objects into parts.

Eccentric viewing conditions produce dramatically different results. Whatever the aperture visibility, the diamond always appears as a rigid object whose direction is effortlessly seen. This effect is not easily explained by an increase of receptive field sizes with eccentricity, since we found that reducing the size of the stimuli has little influence on perceived coherence in central vision. Rather, the effect of eccentric viewing conditions could reflect the relative inability of peripheral vision to resolve local discontinuities. A summary of these different results is presented in figure 2b. Performance in a forced choice direction discrimination task is plotted as a function of the different conditions tested.

At this point several hypotheses that could be invoked to account for these phenomena can be discarded. For example, the idea that motion integration is facilitated with visible as compared to invisible apertures because the former, but not the later, provides a static frame of reference cannot explain why low contrast stimuli are perceptually coherent in central vision when the windows are invisible. Also, neither the idea that human observers use a constraint of rigidity to recover object motion, nor a significant role of attention in binding is supported by our results: Prior knowledge that a rigid diamond is moving does not help to determine its direction and attentional efforts to glue the otherwise incoherent segments into a whole coherent perception are useless. A complementary demonstration with a stereoscopic diamond stimulus support the view that early parsing of the image relies on 2D discontinuities and depth ordering: if a high contrast diamond has a positive disparity relative to the apertures and thus appears in front, its motion appears incoherent, whereas negative disparities, inducing a perception of a diamond moving behind the apertures, produce a highly coherent perception of a rigidly moving object. Thus, despite the fact that the monocular image is identical in both conditions, the perceptual outcome is dramatically different. This effect brings additional support to the hypothesis that changes in terminator classification and depth ordering regulates the transitions from motion integration to motion segmentation.

GLOBAL FORM CONSTRAINTS IN MOTION INTEGRATION

Although the experiments described above indicate that the processing of singularities provide strong constraint on motion grouping, they do not directly address the role of more global geometrical properties in motion integration. To answer this question, it is necessary to modify the spatial relationships between the constituent of a shape, without modifying the singularities or the total energy spectrum of the stimulus. In this way one can ascertain that the potential effects of form on motion integration is not caused by differences in the processing of end-points or in the distribution of Fourier energy in different spatial frequency bands. One way to do this is to permute the positions of the object's parts without modifying the apertures or the segment characteristics. This was done in a series of experiments using outlines of a variety of simple geometrical shapes, shown in figure 3a, such as a diamond, a cross or a chevron, etc. Eight different shapes were used, all constructed with the same component segments, but with different spatial distributions. Note that the energy spectrum of these different shapes is highly similar, the only important differences lying in the phase spectrum.

Thus, any difference in the ability to recover the coherent global motion of these different occluded shapes should be due to differences between their phase spectra. We then ask observers to indicate the clockwise versus anti-clockwise motion of these shapes when seen behind windows that occlude their vertices. Surprisingly, the performance of human observers in the global motion discrimination task strongly depends on which shapes is shown (figure 3b). As a general rule, "closed" figures made of relatable segments (see Kellman & Shipley, 1991), for instance the diamond, yield much better performance than "open" figures, such as a cross or a chevron, for which observers hardly recover the global direction of motion. In addition, observers report the closed figures as being highly coherent shapes moving as a whole, whereas open shapes appear as non rigid sets of line segments moving incoherently.

These results strengthen the view that contour and motion binding depends mainly on the phase spectrum of these stimuli but little on their energy spectrum. What is it about

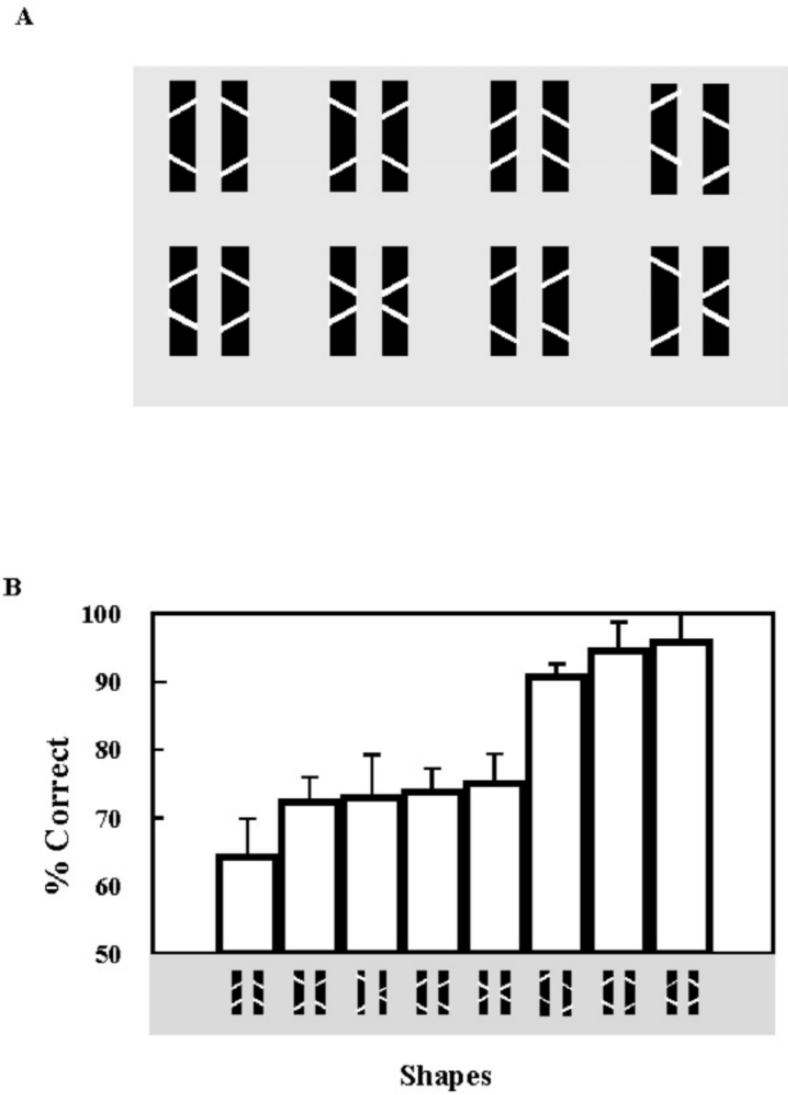

Figure 3. **A**: Stimulus used in to uncover the role of form on motion integration. Different shapes made up of identical segments with different spatial distributions are used. Note that the energy spectra of these stimuli are highly similar whereas the phase spectra differ. **B**: Percentage of the trials where observers successfully recovered the clockwise or anticlockwise direction of motion, as a function of the different shapes tested. The results show that different shapes with identical segment motions yield different performance.

the ‘difficult’ shapes that makes motion integration so troublesome? To test whether this difficulty results from a lack of familiarity with the occluded stimuli, we conducted additional experiments where observers practice the task during a number of sessions or are presented fully visible static shapes for one second before each test trial. Results show that training and knowledge of the shapes in advance does not facilitate global rotation discrimination for difficult shapes (Lorenceau & Alais, 2001). Since performance seems immune to the influence of cognitive strategies arising from prior knowledge of the stimulus, these finding suggest that the origin of the limiting factor rendering motion integration more difficult for these shapes lies at a low-level in the visual system.

It was noted above that motion and form are processed in parallel within the parvocellular and magnocellular pathways. In addition to their different specialisation for form and motion, these pathways also respond differentially to several stimulus variables, particularly to contrast and speed. By varying these parameters it is thus possible to alter the relative contributions of the two pathways to visual processing, and assess their respective contributions. In particular, the poor sensitivity of the ventral “form” pathway to luminance contrast and speed permits to create stimuli which would favour the dorsal “motion” path at the expense of form processing. This was done by reducing the luminance of the contours which define the stimuli and by doubling the speed of rotation. Given that the difference in global motion discrimination between the shapes seems simply to be a matter of geometrical form, we expected that this would reduce the difference in performance between ‘easy’ and ‘difficult’ shapes. Reducing contour luminance resulted in a dramatic improvement in performance on the global rotation task for the difficult shapes, with good performance for the cross and chevron. Importantly, speed also interacted with stimulus shape, such that performance for the cross and chevron was noticeably better at the higher speed. Thus, not only does overall performance improve as stimulus conditions increasingly favour the dorsal “motion” pathway, but the distinction previously seen between easy and difficult shapes is progressively attenuated.

These findings show that reducing the contribution of the form pathway reduces the differences between easy and difficult shapes. More specifically, it is performance for the difficult shapes which improves most, rising toward the near-perfect level of the diamond shape. This confirms that the difference between the shapes really is simply a matter of geometrical form, since, once the influence of form information is reduced, all of the shapes are essentially identical in terms of their spatiotemporal content and thus produce the same global motion solution. This points to a strong interaction between form and motion processing, whereby the form processing path can exert a strong suppressive influence on the motion pathway, determining whether or not local motions are integrated into coherent global motions. This influence of form on motion could result from late interactions between the “form” and “motion” pathways, as neuronal selectivity to shapes such as diamonds, crosses and chevrons is found in the ventral pathway, whereas neurones detecting the kind of rotary motions used here are found in the dorsal pathway. However, several aspects of the present data, such as the strong contrast dependence, the absence of priming effects, and the lack of learning for difficult shapes, suggest an early form/motion interaction. In addition, we propose that there is something about the difficult shapes which actually impedes motion integration. Our data suggest that the role of form information is to regulate whether motion integration should go ahead or not: contours forming convex, closed forms (good gestalts) would favour motion integration, while open, convex forms would trigger a veto from the form system which would prevent motion integration. Giving the form pathway a right of veto over motion integration would help prevent the motion system from integrating local motions which do not belong together, which is especially a risk when they are spatially separated or partially occluded. Integration in occluded regions must be sensibly constrained to prevent spurious motion integration, and a form-based veto could do this. A decision to veto motion integration would need to be made early and would explain the observation that performance with ‘difficult’ shapes remained poor despite extended practice or priming with complete shapes. If motion integration were vetoed early for ‘difficult’ shapes, no learning could take place in higher-level form areas. This suggests that the influence of form on motion could already take place between the magno and parvo streams that are known to interact as early as V1.

FORM AND MOTION SELECTION

In the experiments presented so far, the perception of a moving “whole” is contrasted with the perception of its parts. Although this design permits to shed light on the processes involved in the integration and segmentation of component motions, it is not well suited to address the problem of selection, by which the visual system should decide which and when local motions must be bound with others. Consider instead the stimulus shown in figure 4a, which consists in two overlapping shapes moving in different directions. If these stimuli are seen behind small apertures, such that only straight segments are visible, the activity elicited in orientation and direction selective cells in the primary visual cortex – but this would also be true for any local motion sensor – might resemble the pat-

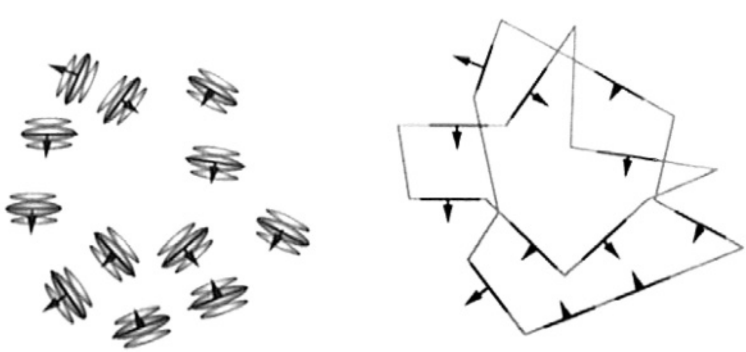

Figure 4. Illustration of the problem of selection: two partially occluded overlapping figures moving in different directions (right) elicit a response from a collection of neurones in primary visual cortex. The visual system must select and combine the responses to one object while discarding the responses to the second object and avoid spurious combination so as to correctly recover the motion of objects in the scene.

tern presented in figure 4b, in which a population of neurones face the aperture problem, i.e. responds to the motion component motion orthogonal to the preferred orientation.

The question, then, is not only to determine whether the neural responses to individual component motions should be bound together or not, but also to determine which responses should be bound together while avoiding spurious associations between the responses elicited by the contours of a different object. This, in principle, could be done by determining which component motions are mutually consistent. However, when observers are shown this stimulus and asked how many objects are present or in what direction they move, they have difficulty to answer. The perception if that of a single non-rigid flow grossly moving in the average component direction. This observation suggests that observers cannot select component motions that belong to the same rigid object on the sole basis of the mutual consistency of the component directions, so as to segment these motions from the remaining inconsistent moving contours. Presumably, other constraints -or prior assumptions- must be used to solve this binding problem. Amongst them, the constraints related to form information appear to play a critical role. This possibility was tested using two moving diamonds, partially visible behind windows that concealed their vertices at all times. This two-diamond stimulus may help uncover the constraints involved in motion selection as it is inherently ambiguous, so the different perceptions it may elicit can reveal the characteristics of motion selection processes. When this stimulus is static, it can be decomposed into a small diamond surrounded by a large one or into two overlapping diamonds of the same size (Figure 5). However, ob-

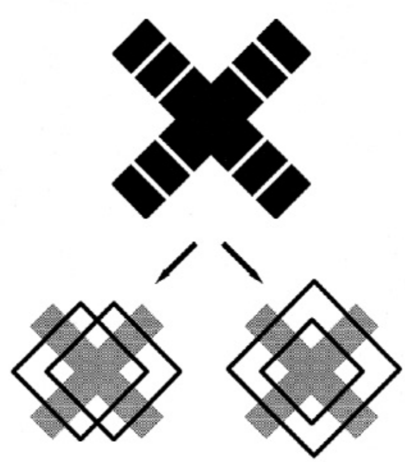

Figure 5. Stimuli used in to uncover the role of form on motion selection. **A**. When static this stimulus yield two distinct perceptual organisation: two overlapping diamonds of the same size or a small diamond embedded in a large one. Observers spontaneously choose the later solution. **B**: when both diamonds are moving in opposite directions, different motion combination are possible: depending on which motion signals available in each aperture are selected one can see incoherent motion of individual segments (no integration), coherent motion in depth or coherent translation in the plane. See text for details. The results indicate that observers favour the grouping of segments forming closed figures, whatever the motion percept implied by this grouping strategy.

servers spontaneously perceive two diamonds of different sizes, rather than two identical diamonds. This is expected however, because proximity, good continuation and closure, known to be powerful cues for grouping static contours, favour this interpretation.

To test whether these rules are also used to drive motion selection, we designed moving versions of this two diamond stimulus. If two identical diamonds oscillate back and forth in opposite direction behind windows (Figure 6a), several interpretations are possible, depending on how the motion signals available through the apertures are combined. The perception of a small expanding diamond surrounded by a contracting diamond could emerge if component motions in the centre were bound together, as in the static version of this stimulus, and segmented from the outer component motions that would on their own yield the perception of a large contracting diamond. Alternatively, the component motions of diamonds with identical size could be grouped by similarity, yielding the perception of two overlapping diamonds translating in opposite directions in the plane. Other possibilities – absence of grouping, selection by proximity within a window – also exist and can elicit different interpretations. Simple experiments were done to determine what is the dominant perceptual organisation of motion. Observers were asked to report whether they saw two unequal diamonds moving in depth – expanding and contracting – or two identical diamonds translating back and forth in the

A

B

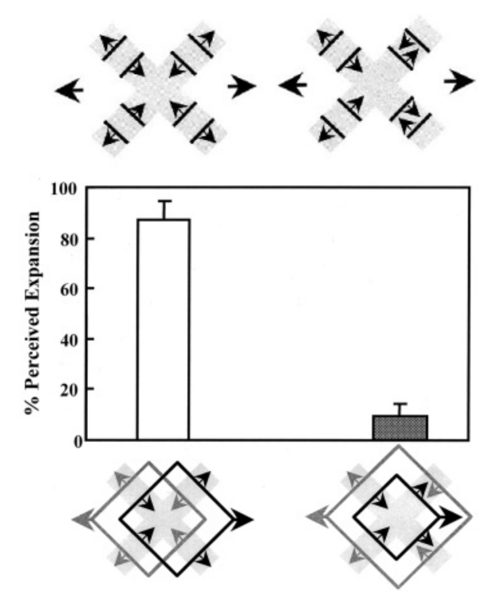

Figure 6. Results of a forced choice experiment in which observers were required to indicate whether they perceived two diamonds expanding or contracting over time, or two diamonds translating back and forth in the fronto-parallel plane. Observers' choice depends on which segments are grouped. The results indicate that observers based their choice on the perceived form as they always group the four central segments yielding a small and a large diamond. This spatial configuration is then either seen as expansion and contraction or as two translations in opposite directions.

fronto- parallel plane – to the left and to the right –. The results (figure 6b) clearly show that observers systematically report seeing a small and a large diamond moving in depth, but rarely perceive two translating diamonds or other motion combination. If one uses instead a large diamond surrounding a small one, so as to display the same spatial configuration of eight static segments and then apply the same horizontal oscillation, the perception of motion changes dramatically: observers no longer see motion in depth but report seeing a small and a large diamonds translating in opposite directions in the same plane, although the perception of two diamonds of equal size moving in depth would be an equally possible interpretation. Thus, the same sets of four segments were selected in both configurations –same size or different diamond sizes –, resulting in very different perception of motion. This suggests that motion signals are not selected on the basis of the sole motion information and that observers are not biased toward a specific interpretation, for instance because motion in depth would be more relevant for the organism. One explanation is that aspects of static forms, such as collinearity, alignment and closure strongly determine which signals are selected to drive the motion integration/segmentation process.

Altogether, these results powerfully demonstrate the critical role played by geometrical information in global motion computation. Local singularities such as vertices, junctions or line-ends appears to exert a strong control on the balance between integration and segmentation as salient contour terminators appear to be used to parse the image into parts. Global geometrical image properties also appear to provide strong constraints on the integration process, as integrating moving contours into a global motion is a simple task for some configurations (diamonds) while it is difficult for others (crosses and chevrons). The observation that extrapolation of the contours of these different shapes also produces two distinct classes of stimuli provides insights to account for this dichotomy. In the case of the diamond, contour extrapolation produces a closed, convex shape, whereas open concave shapes are produced in the case of the cross or the chevron. The closure inherent in the diamond's form may provide a reason for its superiority over the other shapes. Closure of the diamond by amodal completion (Kanisza, 1979), together with the filling-in of its interior this may engenders, would serve effectively the segregation of the diamond from its background. Consequently, judging the diamond's direction of rotation would be much easier than for open shapes which generate poorer responses at the level of object representation. The available neural evidence suggests that these processes of completion, filling-in, and figure/ground segregation are initiated early in visual processing. Cells in V1 have been shown to respond to contours rendered discontinuous by occlusion (Sugita, 1999), and V1 cells are also capable of responding to filled-in areas and not just to their borders (Komatsu et al., 1996). Brain imaging has also revealed figure/ground segregation as early as V1 (Skiera et al., 2000). Moreover, neural correlates of the boundary-initiated surface formation process described above have been observed in V1 (Lamme, 1995).

The effects of boundary completion, filling-in and figure/ground segregation, can all be considered broadly under the rubric of form processing. Our data suggest that the role of form information is to regulate whether motion integration should go ahead or not.

It is likely that cooperative interactions between neighbouring contours observed in area V1 (Gilbert, 1992; Kapadia et al., 1995) can provide the cortical substrate to explain the influence of form on motion grouping, possibly by conducting a pre-shaping of elements into proto-forms. Evidence from physiology and psychophysics points to low-level mechanisms being extensively involved in contour completion, filling-in and figure/ground segregation.

A simple model that synthesise the present findings is shown in figure 7. Depending on the salience of singularities and on their spatial configurations, contours would be grouped through horizontal connections in primary visual cortex (V1), area V2 would further classify the depth relationships at occlusion points and provide modulating inputs to the MT/MST complex in the dorsal pathway which could in turn and depending on the results of this initial processing, integrate or segment the selected motion signals.

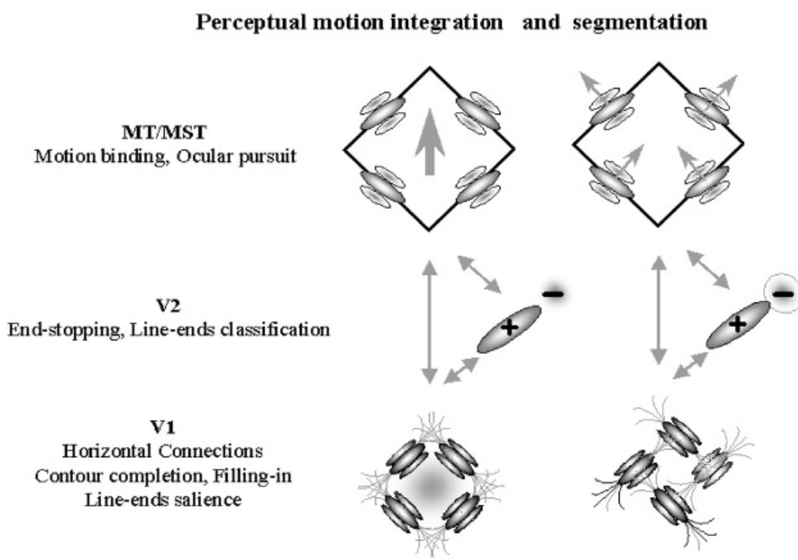

Figure 7. Hypothetical model of form and motion binding. Long-range horizontal connections in V1 would link relatable contour segments depending on the salience of their end-points and on the collinearity and alignment between them. V2 would further process and classify singularities –vertex, junction, occlusion point- in the image. MT/MST would compute partial solution to the aperture problem, depending on the inputs from V1 and V2 that would "tag" the signals selected for further motion analysis. Feed-back from MT could help maintain a viable solution.

CONCLUSION

The present paper summarized some of the numerous studies that converge to support the idea that geometrical relationships between visual elements or "tokens", as initially stated by the Gestaltists, play a fundamental role in the perceptual organisation of form and motion. Recent developments in neuroscience and the available anatomical and physiological evidence suggest that the neuronal circuitry described in the primary visual cortex possesses some of the properties needed to process the geometrical characteristics of the retinal inputs. This is certainly not the whole story, however: many other aspects of form and motion, that are selectively processed in areas distributed along the dorsal and ventral pathways, may also play a role. In addition, attention and prior knowledge could modulate perceptual grouping, although the present experiments failed to demonstrate such influence. Finally, the fact that motion can by itself provide sufficient information to segregate and recognise the form of objects indicates that interactions between form and motion are bi-directional. Future studies will with no doubt shed light on the intricate relationships between the processing of motion and form.

Prof. Jean Lorenceau
UNIC-CNRS
Avenue de la Terrasse
91198, .Gif-sur-Yvette, France

REFERENCES

ADELSON, E. & MOVSHON, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523-525.

CAMPBELL F. W. & ROBSON J.G., (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551-566.

DE VALOIS, R.L. & DE VALOIS, K.K. (1988). *Spatial vision*. Oxford University Press.

FENNEMA, C.L. & THOMPSON, W.B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 9, 301-315.

FIELD DJ, HAYES A, HESS RF (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Research*, 33, 173-93.

GALLANT, J. L., CONNOR, C. E., RAKSHIT, S., LEWIS, J. W. & VAN ESSEN, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76, 2718-2739.

GILBERT CD (1992). Horizontal integration and cortical dynamics. *Neuron*, 9, 1-13.

HENRY, G.H., BISHOP, P.O. TUPPER, R.M. & DREHER, B. (1974). Orientation axis and direction as stimulus parameters for striate cells. *Vision Research,* 14, 767-777.

HARTLINE, H.K. (1940). The receptive field of optic nerve fibers. *Am. J. Physiol.*, 130, 690-699.

HUBEL, D. & WIESEL, T. (1968). Receptive fields and functional architecture of the monkey striate cortex. *Journal of Physiology*, 195, 215-243.

JOHANSSON G. (1950). *Configurations in event perception*. Uppsala, Almqvist & Wiksell.

KANIZSA G. (1976) Subjective contours. *Scientific American* 234(4):48-52.

KAPADIA MK, ITO M, GILBERT CD, WESTHEIMER G (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15, 843-56.

KELLMAN PJ, SHIPLEY TF (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23, 141-221.

KOFFKA, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt Brace.

KOHLER, W. (1929). *Gestalt Psychology*, New-York, Liveright.

KOMATSU H, MURAKAMI I, KINOSHITA M. (1996). Surface representation in the visual system. *Brain Res Cogn Brain Res*., 5, 97-104.

KOVACS I, JULESZ B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, 370, 644-6.

LAMME VA (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of Neuroscience*, 15, 1605-15.

LORENCEAU, J. & ALAIS, D. (2001). Form constraints in motion binding. *Nature Neuroscience*, 4, 7, 745-751.

LORENCEAU, J. & SHIFFRAR, M., (1992) The influence of terminators of motion integration across space. *Vision Research*, 32,2, 263-273.

METZGER W. (1934). Betrachtungen uber phenomenal identitat. *Psychologische Forschung*, 19, 1-60.

PIOTROWSKI L.N. & CAMPBELL, F., W. (1982). A demonstration of the importance and flexibility of spatial-frequency amplitude and phase. *Perception*, 11, 337-346.

POLAT U, SAGI D (1994). The architecture of perceptual spatial interactions. *Vision Research*, 34. 73-8.

SKIERA G, PETERSEN D, SKALEJ M & FAHLE M (2000). Correlates of figure-ground segregation in fMRI. *Vision Research*, 40, 2047-56.

SUGITA Y (1999). Grouping of image fragments in primary visual cortex. *Nature*, 401,269-272.

VAN ESSEN, D. C. & DEYOE, E. A. (1995). *Concurrent processing in primate visual cortex*. In M. Gazzaniga (Ed.), The Cognitive Neurosciences (pp. 383-400). MIT Press.

WERTHEIMER, M. (1912). Experimentelle stuidien uber das Sehen von Beuegung. *Zeitschrift fuer Psychologie*, 61, 161-265.

WILSON H.R. & KIM J. (1994) A model for motion coherence and transparency. *Visual Neuroscience, 11*, 1205-1220.

This page intentionally left blank

SCINTILLATIONS, EXTINCTIONS, AND OTHER NEW VISUAL EFFECTS

1. INFLUENCES

After completing studies in mathematics and engineering, I turned to molecular biology, and became interested in the problem of how molecules recognize each other. In many molecular processes, there are crucial stages in which a molecule has to select a cognate partner, among many non-cognate ones. Sometimes errors are made, but on the whole, the selection procedures are remarkably accurate. For instance, the error-rate in the reproduction of DNA can be as low as $2x10^{-10}$ in some cells, per monomer incorporated. In the 1970's there were a number of puzzling observations on the patterns of errors made in mutant cells exhibiting either higher accuracy, or lower accuracy than standard cells. The prevalent doctrine was that the errors were due to error-prone processes, which operated in parallel with the normal error-free process. Inspired, in part, from readings in psychology and psychoanalysis, I was inclined to consider errors as products of the normal process, signatures which revealed its inner workings. I thus showed, by a simple mathematical analysis, that the error-patterns could be interpreted in this way [1], and developped a body of ideas on how accuracy could be achieved in molecular processes [2]. In this field, my name is associated to the name of John Hopfield [3], a physicist now famous in cognitive psychology for his contributions to neural network theory [4].

While working in a molecular biology laboratory, I was reading books and articles on vision and the brain. Three books, by Karl von Frisch [5], Bela Julesz [6] and Richard Gregory [7] made a lasting impression on me. From von Frisch, I learnt not to take an experiment at face value: bees do not discriminate between red and black, yet they have color vision. They distinguish two whites, identical to our eyes, on the basis of their ultraviolet content. (Later, I found that a similar result had been established, much earlier, by Lubbock on ants). From Julesz, I learnt that one could do experiments probing the inner workings of the brain, using carefully designed images. From Gregory, I learnt all about constancies in vision, and how much of a paradox stable vision was. I was also impressed by his style, as a scientific writer.

However, the book which gave me a real opportunity to join the field, was a more academic one, a synthesis on visual illusions by Robinson [8]. This book contained an exhaustive description of the known geometrical visual illusions, and a summary of most, if not all theories put forward to explain them. None of these theories satisfied me, and I thought there would be room for a fresh attack, in the line of my work on accuracy in molecular biology. Geometrical illusions, far from being the result of error-prone processes in the brain, would, on the contrary, be the signature of intelligent procedures to represent shape and spatial relationships. A map of a portion of the earth may look dis-

A. Carsetti (ed.), Seeing, Thinking and Knowing, 191-201.

torted, yet it may have been constructed according to a rigorous mathematical algorithm. The distorsions come from the need to accomodate the constraints (in the case of the geographic map, one has to represent a spherical surface on a planar one), and not from the inadequacy of the mapping procedure. I therefore dwelled upon the geometrical problems of vision, and worked both on geometrical visual illusions, and stereoscopic vision, first theoretically [9, 10] then experimentally (e.g., [11,12]).

Ultimately, I became a professional in the field, and attended the European Congresses on Visual Perception (ECVP). At these congresses, there was a predominance of talks by scientists from English-speaking countries, and the Italians were often relegated to the posters (the situation has improved, since), but it was often there that I found the most creative visual stimuli, there that I took the measure of the Italian contributions to the field. There was a strong trend, among the ruling psychophysicists, to describe tedious experiments, made on boring visual stimuli that involved just three points, or three line segments, or two gratings, or even worse, two Gabor patches. I considered that vision has to deal with a 3d world, which is seized with mobile eyes attached to a mobile head. Visual spatial analysis then requires, to perform correctly, inputs with some minimal complexity. Seven or eight anchoring points seem to be a strict minimum for 3d spatial analysis (see, [9, 13]). With stimuli lacking complexity, the normal visual algorithms may not work properly, and what one studies then is perhaps a "default" setting of the visual system. I thus became increasingly attentive to the astute visual stimuli designed by scientists from the Italian school (see, e.g., the collection of contributions in [14, 15]). I began to develop ties with several of them, and I realized how much this school owed to Gaetano Kanizsa. Last, but not least, I became familiar with Kanizsa's work in its globality [16, 17], and took the measure of the depth of his thinking. Above all I appreciated his way of embodying his conceptions into striking visual examples. Mathematicians, dealing with a conjecture, the proof of which appears beyond reach, occasionally defeat it by the discovery of a single counterexample. Kanizsa had the art of constructing with maestria, the right counter example to defeat the too simplistic explanations of the phenomena he was interested in.

Rather than embarking into wordly discussions of, say the interrelationship between top-down and bottom-up streams in visual analysis, I will more modestly, introduce, with minimal comments, a few of my favourite images: Images which would have perhaps elicited inspired comments from Kanizsa.

2. TEXTURES, AND SUBJECTIVE CONTOURS

At the beginning, I was greatly influenced by Julesz, and was an admirer of his camouflaging textures involving random lattices of black and white squares, used in stereoscopic stimuli. However, real-life scenes contain edges at all orientations, and I sought to design camouflaging textures rich in orientations. I thus produced "random-curve stereograms" [18]. There, a 3d surface is represented by a computer-generated random curve, or by a distorted lattice. In spite of the low density of lines on the surface, it is perceived as essentially complete, and one can perhaps classify it as a kind of subjective surface.

Later, I tried to produce camouflaging textures manually, and some of my efforts are reminding of Kanizsa's *biotessitures* (reproduced in [19], part III). In Fig. 1, I show a stereoscopic image of a mask, covered with a hand-made texture. Although the details of the shape are concealed in monocular vision, some information can be retrieved under the conditions of 'monocular stereoscopy'. When symmetry is introduced, in the manual, quasi-random textures, visually-rich patterns build up, expanding outwards from the axis of symmetry (Fig. 2).

In my view one of the most exciting development in the domain of subjective contours is the extension of the phenomenon to 3d surfaces, first in stereo vision [20, 21] then to drawings involving both masking and interposition clues [22]. Kanizsa was not the first to notice subjective contours, but he made penetrating analyses connecting this domain to the domain of the perception of transparency. One point which intrigues me is why, in figures in which black and white play symmetrical roles (e.g., Fig 3) we call "subjective" the white surface, and "real" the equivalent black surface ?

3. SUBTLE DIFFERENCES

One way of settling a point, in visual perception, is to design a couple of images which differ, in their construction, by a hardly noticeable feature, and yet produce strikingly different effects. For instance Kanizsa showed that the perception of subjective letters, represented by their shadows failed at first when the letters were not displayed in their usual orientation, and was recovered once the anomaly was recognized. I have used the strategy of the subtle difference to study the role of orientation disparity in stereo vision [23].

Here, I show a couple of figures in which the Fraser spiral illusion works, or does not work, depending on a subtle detail (Fig. 4). In my opinion, this couple of images establishes a bridge between the Fraser illusion and the gestalt principle of segregation by contrast (a corolloray to the principle of association by grey-level proximity).

4. ALTERNATIVE 3D INTERPRETATIONS

With minimal changes, the drawing of a flat figure can evoke a 3d object. Kanizsa showed interest in the nature of the low-level cues which contributed to global 3d interpretation. There are also figures which elicit both 2d and 3d interpretations. In Fig. *5* the two trapeziums are at first interpreted as flat shapes. After a certain time, a 3d interpretation develops, in which the trapeziums are not even planar! They appear like twisted ribbons [24]. Once the 3d interpretation is acquired, it is difficult to see the trapeziums as planar again. Incidentally, this may be taken as a (rare) counterexample to the genericity principle [25]. For other examples of switches in 3d interpretation, see [26, 27].

5. CONTRAST EFFECTS

The domain of visual contrast effects is still producing novelties, see review in [28]. While contrast effects were not, for me, a central preoccupation, I was driven into the field by my interest for geometrical questions. Take the well-known Hermann grid effect. It is usually presented as a two-dimensional area of black squares, separated by horizontal and vertical rectilinear arrays. Is this very peculiar geometry necessary to the illusion ? Orientation, at least, must be respected, for if a Hermann grid is rotated by 45 degrees, a different contrast effect becomes manifest. Perhaps, then, the Hermann grid is providing cues relative to the geometrical layout of the neurons that are performing. in some area of the brain, the local contrast calculations. I thus teamed with a geometrically-minded visual scientist, Kent Stevens, to investigate the geometrical requirements of the Hermann grid illusion. Hundreds of geometric variants of the grid were generated, including variants giving the scintillation effect [29, 30] (see Fig. 6). The original question was not settled, but out of the stack of variants, new visual phenomena emerged [31-33].

The most spectacular one is the extinction effect [31] (Fig. 7). There, only a few disks are seen at a time, in clusters which move with the fixation point. It is as though, outisde the fixation point, a feature which is above the spatial threshold for detection, needs also to be above some contrast threshold, with respect to background, in order to be brought to attention.

Distorting the squares of the Hermann grid, one can observe an effect in which illusory lines are seen to pulsate [32] (Fig.8). The orientations of these lines are unusual. They correspond to the directions of knight's moves on a chessboard. These lines go through both black and white regions (see Fig. 5 in [32]) and could be testimonies of a cooperation between neurons having aligned receptive fields of opposite contrasts.

In the display of Fig. 9, the diamond-oriented domains appear differently contrasted, depending on whether they contain near horizontal or near vertical stripes [33]. For some people, the domains with near vertical stripes appear highly contrasted, while those with near horizontal stripes appear faded. For other subjects, it is the opposite. Once the effect is noticed, it can be detected in many variants. This type of pattern combines easily with many other visual effects. For instance, when straight lines are superimposed on the patterns with stripes at different orientations, striking Zollner-type distorsions are observed (Fig. l0).

After having produced theories, then psychophysical data, I find more and more satisfaction, as Kanizsa did, in elaborating striking images. Whereas, in his case, the images must have been the outcome of a completely rational line of thinking, in my case they came by surprise. They were - at least for Fig. 7 and 8, the unexpected reward of a very systematic work of variations in the geometry of the stimuli.

Prof. Jacques Ninio
Ecole Normale Supérieure
24 rue Lhomomd
75231, Paris, France

REFERENCES

[1] J. Ninio, *J. Mol. Biol.* 84, 287 (1974)
[2] J. Ninio, *Biochimie 57,* 587 (1975)
[3] J. J. Hopfield, *Proc. Nat. Acad. Sci. USA* 71**,** 4135 (1974)
[4] J. J. Hopfield, *Proc. Nat. Acad. Sci. USA* 79**,** *2554* (1982)
[5] K. von Frisch, *The Dance Language and Orientation of Bees,* Bellknap Harvard University Press, Cambridge, USA (1965)
[6] B. Julesz, *Foundations of Cyclopean Perception.* Chicago Univeristy Press, Chicago, USA (1971).
[7] R.L. Gregory, *The Intelligent Eye.* Weidenfeld and Nicolson, London, U.K. (1970).
[8] J.O. Robinson, *The Psychology of Visual Illusion.* Hutchinson University Press, London, U.K. (1972)
[9] J. Ninio, *Perception 6,* 627 (1977)
[10] J. Ninio, *J. Theoret. Biol.* 79, 167 (1979)
[11] P. Herbomel and J. Ninio, *Vision Research* 33, 1813 (1993)
[12] J. Ninio and J.K. O'Regan, *Perception 25,* 77 (1996)
[13] H.C. Longuet-Higgins, *Nature* 293, 133 (1981)
[14] F. Purghé, N. Stucchi and A. Olivero (eds), *La percezione visiva.* UTET Università, Torino (1999)
[15] M. Zanforlin and L. Tommasi (eds), *Research in Perception.* Logos, Saonara, Italy (1999)
[16] G. Kanizsa, *Grammatica del Vedere.* Il Mulino, Bologna, Italy (1980)
[17] G. Kanizsa, *Vedere e Pensare.* Il Mulino, Bologna, Italy (1991)
[18] J. Ninio, *Perception* 10, 403 (1981)
[19] S. Sumi and K. Noguchi. *New Horizons in the Study of Gestalt Perception. In Memory of the Late Professor Gaetano Kanizsa.* Keio University, Yokohama, Japan (1996)
[20] J.P. Harris and R.L. Gregory, *Perception* 2, 235 (1973)
[21] Idesawa, M. and Zhang, Q., SPIE Proceedings 3077, 770 (1977)
[22] P.U. Tse, *Perception* 27, 977 (1998)
[23] J. Ninio, *Perception* 14, 305 (1985)
[24] J. Ninio, *L'empreinte des Sens,* 3rd ed., Odile Jacob, Paris, France (1996)
[25] J.J. Koenderink and A.J. van Doorn, *Biol. Cybernet.* 32, 211 (1979)
[26] J. Ninio, *Stéreomagie,* Seuil, Paris, France (1994)
[27] J. Ninio, *The Science of Illusions,* Cornell University Press, Ithaca, USA (2001)
[28] P. Bressan, *Perception* 30, 1031 (2001)
[29] J.R. Bergen, *Investigative Ophtalmology & Visual Science, Supplement* 26, 280 (1985)
[30] M. Schrauf, B. Lingelbach, and E.R. Wist, *Vision Research* 37, 1033 (1997)
[31] J. Ninio and K. A. Stevens, *Perception* 29, 1209 (2000)
[32] J. Ninio, *Perception* 30, 253 (2001)
[33] J. Ninio, *Perception*.

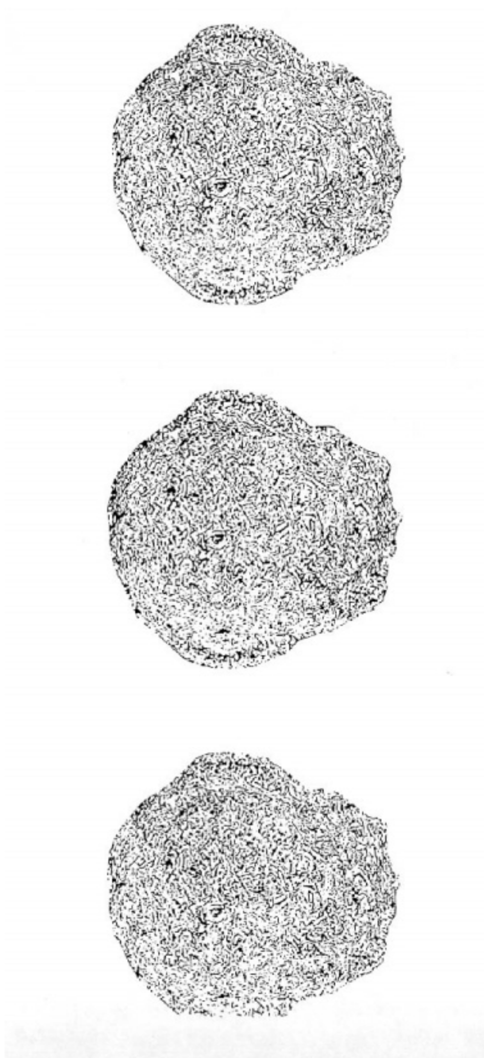

Figure 1. Camouflaged stereogram. A mask of the face of a monkey, covered with hand-made texture, was photographed from two viewpoints to generate this stereogram. Use the central image with the left image for convergent viewing, or with the right image for parallel viewing. Some depth information may be retrieved by looking at a single image through a narrow tube - for example, with the hand against the eye, the fingers folded to create a cylindrical aperture.

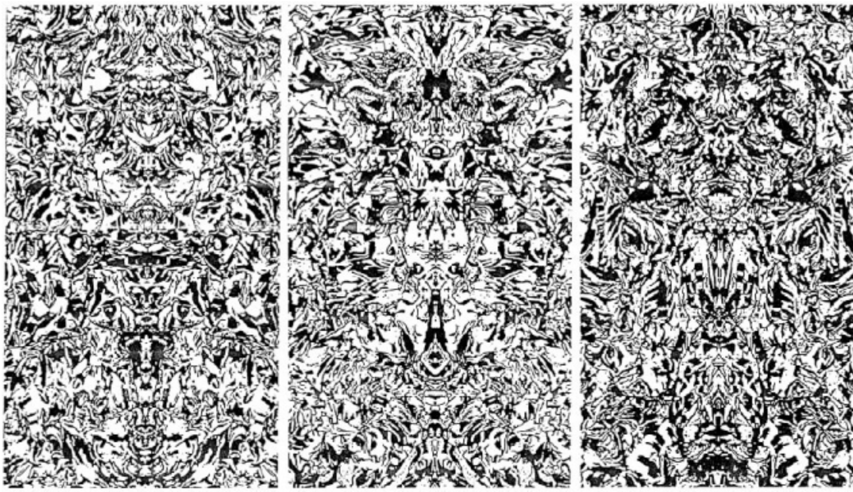

Figure 2. Visually-rich patterns obtained by introducing symmetry in a hand-made texture. Near the symmetry axes, the appearance of the texture is lost in favor of the patterns. The texture is better appreciated on the sides, or when rotating the figure by ninety degrees.

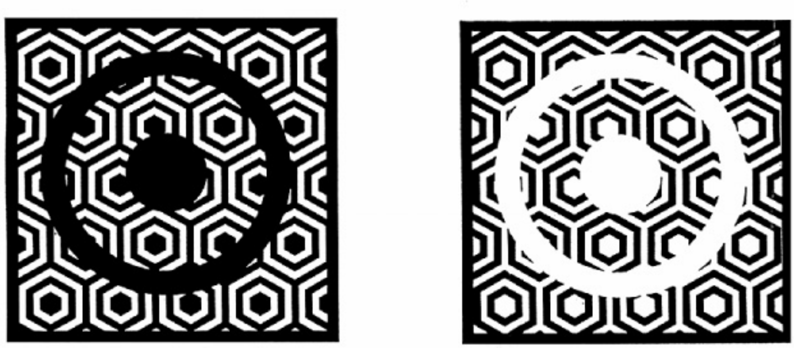

Figure 3. Subjective contours. In both figures, one sees a regular ring over a background of spaced hexagons. One is tempted to say that the ring on the left is real, and the ring on the right is subjective. However -ignoring the frames- the left and right figures differ by a mere inversion of black and white. Note also that, in a sense, each ring must be “cut away” from lines of its own colour.

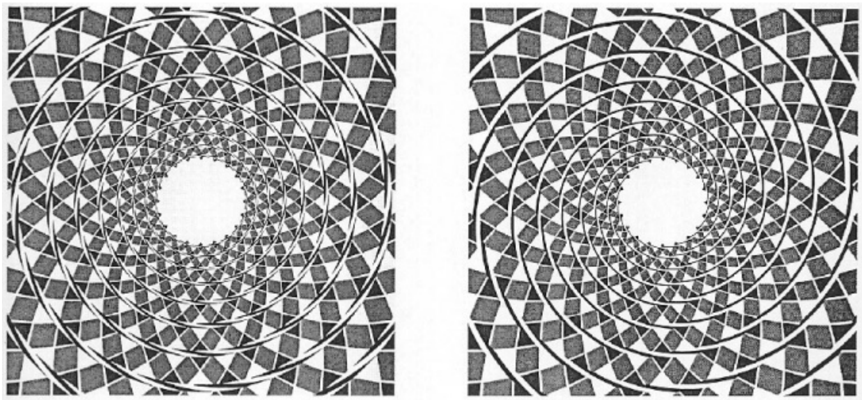

Figure 4. A subtle difference. The Fraser's spiral illusion works well in the left image, but not in the right image. The two images differ by a slight rotation of the concentric rings, making their borders real in the right image, and subjective in the left image. The orientations of the black or white arcs of a ring, taken separately, are typical of a real spiral field.

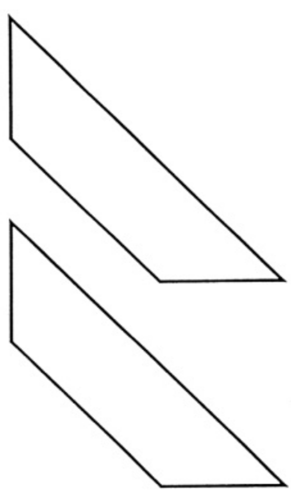

Figure 5. Twisted trapeziums. It is possible to see the trapeziums protruding in 3d as twisted ribbons, the horizontal sides being at the front, and the vertical sides at the back.

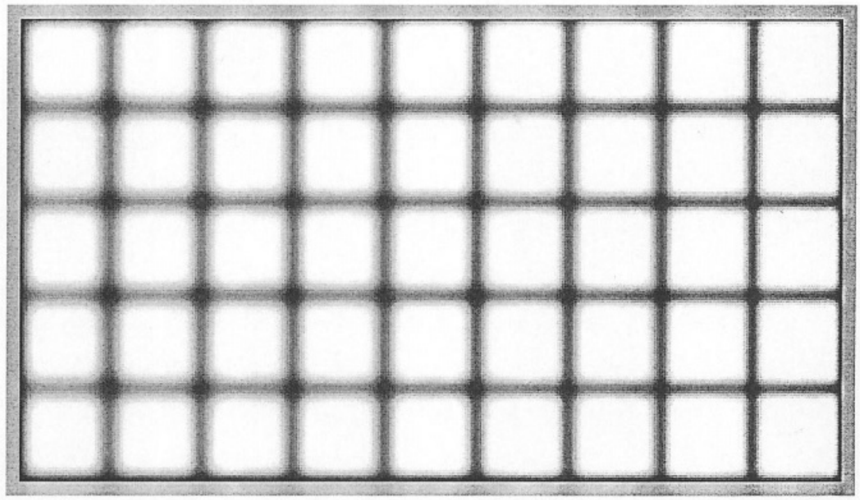

Figure 6. Scintillation effect (adapted from Bergen [29]). Brilliant spots appear to flash at the crossings of the dark alleys.

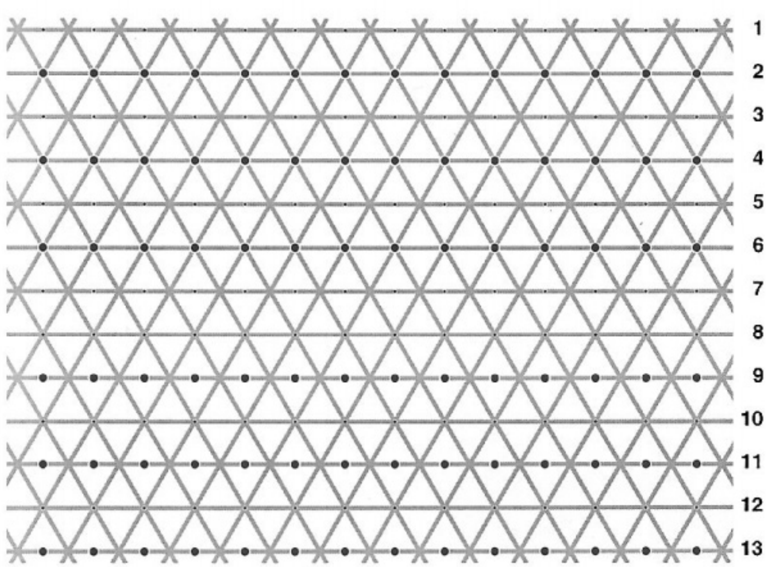

Figure 7. Extinction effect. On lines 9, 11 and 13, containing large disks half-way from alley-crossings, all disks are seen, while many of the large disks situated at the crossings, on lines 2, 4 and 6, are seen erratically.

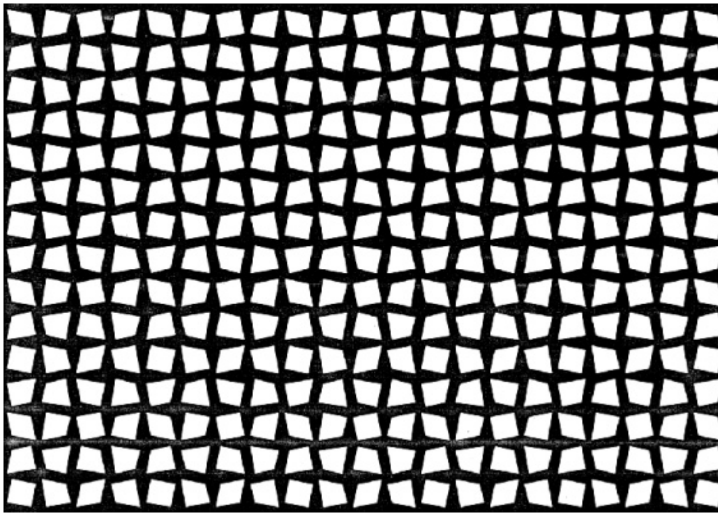

Figure 8. Flashing lines. Two sets of bright lines are seen pulsating at about 300 and 1200 with respect to the horizontal.

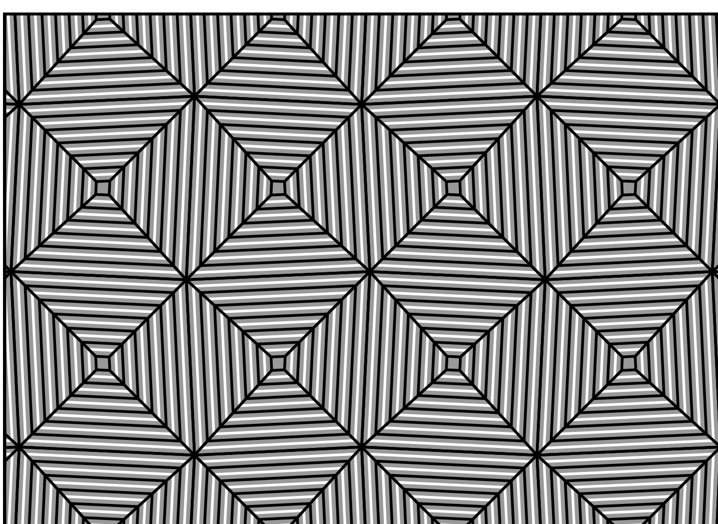

Figure 9. Orientation-dependent contrast. To most observers, the grey-level range appears narrower, either in the domains with horizontal stripes, or the domains with vertical stripes. Domains of one kind appear well contrasted, and domains the other kind appears toned down, although the stripes in it are seen with normal resolution.

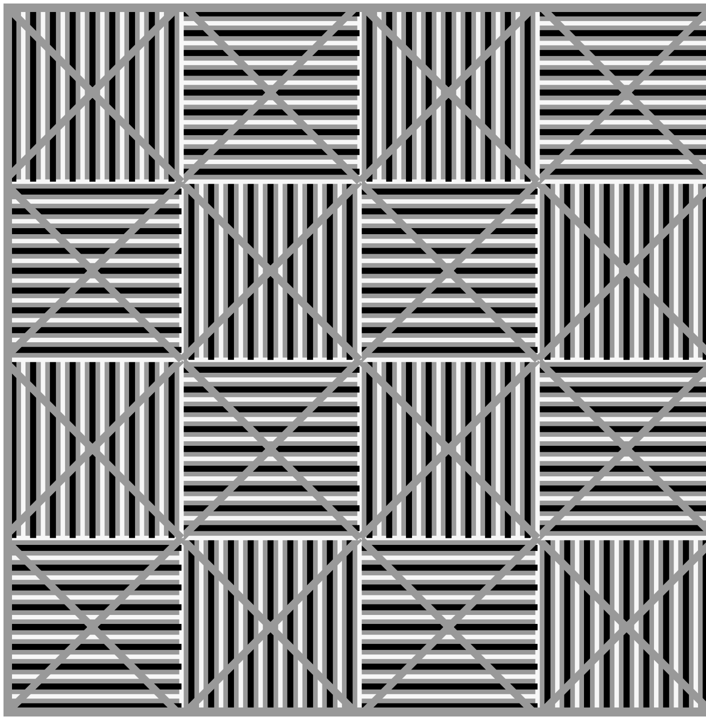

Figure 10. Distorted diagonals. The distorted appearance of the diagonal lines is reduced or cancelled when these lines are oriented horizontally or vertically. The effect is also observed with black or white diagonals.

This page intentionally left blank

M. OLIVETTI, R. DI MATTEO, C. DEL GRATTA, A. DE NICOLA, A. FERRETTI, G.L. ROMANI

COMMONALITIES BETWEEN VISUAL IMAGERY AND IMAGERY IN OTHER MODALITIES: AN INVESTIGATION BY MEANS OF FMRI

INTRODUCTION

The attempt to shadow the differences between seeing and thinking by stressing their similarities is not an epistemologically correct operation, because by using principles related to another domain (as thinking) to explain vision may induce a pre-packed explication. Instead, stressing differences may lead to the discovery of new rules governing only one of the two processes under investigation (Kanizsa, 1991). We report this provoking statement of Kanizsa, while approaching our research on mental imagery for two main reasons: 1) the main part of the psychological research on imagery is devoted to visual imagery, implicitly assuming that imagery derived from other sensory modalities will present characteristics that are similar to those of visual imagery; 2) a lot of studies on visual imagery are devoted to assess whether primary perceptual circuits are implied also in imagery and, therefore to assess how much seeing is similar to imaging. In this study we accepted Kanizsa's suggestion by trying to assess differences between visual and other-senses imagery in order to detect their peculiarities and the grade of their overlap.

Mental imagery has recently gained a renewed interest thanks to the advent of brain mapping of cognitive functioning by means of new non-invasive techniques (fMRI, functional Magnetic Resonance Imaging, and PET, Positron Emission Tomography). This new approach permits the recording and visualization of different parameters reflecting brain activity, with a high temporal and spatial resolution.

The neuroimaging approach to mental imagery was mainly focused on mapping brain correlates of well-established behavioral data in order to clarify the status (epiphenomenal vs. autonomous) of the processes underlying mental imagery. In particular, classical experiments on mental manipulation and image generation have been replicated showing the involvement of several brain areas in mental imagery.

The first question raised in this debate is linked to the extent of the involvement of the primary visual areas, if at all, in visual imagery. This idea is supported by evidence showing that focal brain damaged patients exhibit similar impairments in visual perception and imagery (for a review see Farah, 1995), and by neuroimaging data showing activation in the occipital lobe in various visual imagery tasks (Chen, Kato, Zhu, Ogawa, Tank & Ugurbil, 1998; Kosslyn & Thompson, 2000; Klein, Paradis, Poline, Kosslyn & Le Bihan, 2000). The hypothesis of the involvement of the primary visual areas in imagery is based on the assumption that visual imagery is depictive in nature (Kosslyn, 1994) and should share the same neural substrate of visual perception (Kosslyn, Alpert, Thompson, Maljkovic, Weise, Chabris, Hamilton, Rauch & Buonomano, 1993). This idea rests on the hypothesis that mental imagery activates backward projections from

A. Carsetti (ed.), Seeing, Thinking and Knowing, 203-218.

'high-level' to 'low-level' areas of the visual system (Kosslyn, Maljkovic, Hamilton, Horwitz & Thompson, 1995) due to the retrieval of stored information in order to reconstruct spatial patterns in topographically organized cortical areas.

However, other studies did not find any evidence of the involvement of primary visual areas in visual imagery (De Volder, Toyama, Kimura, Kiyosawa, Nahano, Vanlierde, Wanet-Defalque, Mishina, Oda, Ishiwata et al., 2001; Mellet, Tzourio Mazoyer, Bricogne, Mazoyer, Kosslyn & Denis, 2000; Cocude, Mellet & Denis, 1999; Mellet, Tzourio Mazoyer, Denis & Mazoyer, 1998; D'Esposito, Detre, Aguirre, Stallcup, Alsop, Tippet & Farah, 1997), or found selective impairment in either imagery or perception following focal brain damage (Bartolomeo, Bachoud-Levi, De Gelder, Denes, Dalla Barba, Brugieres & Degos, 1998).

It should be noted that between these two groups of studies there are many differences in techniques, procedures and experimental tasks.

Regarding other brain areas, the middle-inferior temporal region, especially on the left hemisphere, has been repeatedly found to be active in various image generation (D'Esposito et al., 1997) and mental rotation (Iwaki, Ueno, Imada & Tonoike, 1999; Barnes, Howard, Senior, Brammer, Bullmore, Simmons, Woodruff & David, 2000; Jordan, Heinze, Lutz, Kanowski & Länche, 2001) tasks. These data support the idea that modality specific processes underlie mental imagery because activation in this area has been found also in visual object recognition (Stewart, Meyer, Frith & Rothwell, 2001) and in tasks requiring the recovery of visual features (Thompson-Schill, Aguirre, D'Esposito & Farah, 1999). However, by reviewing previous data on imagery tasks, Wise, Howard, Mummery, Fletcher, Leff, Büchel & Scott (2000) suggest that the core of this activation should be situated in the hetero-modal associative temporal cortex. According to these authors, this area mediates access to the amodal/non-linguistic internal representations of word meanings, and this role would be more coherent with the results obtained by means of such a wide range of cognitive tasks. In this view, its role in mental imagery would be independent from modality specific representations.

Activation of associative areas in the parietal lobe has also been found but its role in mental imagery is somewhat controversial. Some authors suggest that these regions contribute to the processing of spatial attributes of imaged objects (Diwadkar, Carpenter & Just, 2000), others outline their role in the construction of a supra-modal representation by binding together modality specific information (Lamm, Windischberger, Leodolter, Moser & Bauer, 2001; Richter, Somorjai, Summers, Jarmasz, Menon, Gati, Georgopoulos, Tegeler & Kim, 2000). Finally, Carey (1998) suggests that this area should be considered a key component of a third visual stream (besides the ventral and the dorsal pathways) having perceptual, attentional and motor-related functions.

In the context of mental imagery, the role of the prefrontal cortex known to be responsible for working memory operations, has been somewhat neglected, perhaps due to the heterogeneous pattern of activation emerging from different studies. As reported by several authors, spatial working memory tends to activate the right prefrontal cortex, whereas verbal tasks involve mainly the left or bilateral prefrontal cortex (Burbaud, Camus, Guehl, Bioulac, Caillé & Allard, 2000; Bosch, Mecklinger & Friederici, 2001).

Studies examining the relationship between imagery and processes related to modalities other than vision are very rare. Among them Zatorre, Halpern, Perry, Meyer & Evans (1996), by comparing the PET data from auditory perception to those derived from auditory imagery, conclude that the same brain regions were activated in the two tasks. Höllinger, Beisteiner, Lang, Lindinger & Berthoz (1999) compared slow potentials accompanying the execution of movements to the response accompanying their imagination, and found again that similar regions were at work in both cases.

In summary, although data support the idea that perception and imagery share a common neural substrate, findings are not univocal, suggesting that this common substrate does not involve early perceptual stages.

Neuroimaging techniques offer the possibility to investigate another interesting aspect of mental imagery, i.e. the distinctive features of intermodal mental imagery.

From a behavioral point of view the interest for this topic is rather old as testified by the construction of quantitative instruments aimed at evaluating mental imagery, not only in the visual modality, but also in the auditory, haptic, kinesthetic, gustatory, olfactory and organic ones (Betts, 1909; Sheehan, 1967; White, Ashton & Brown, 1977). Some of these studies investigate the relationships between visual and auditory imagery (Gissurarson, 1992), visual and kinesthetic imagery (Farthing, Venturino & Brown, 1983), and visual imagery and olfactory stimulation (Wolpin & Weinstein, 1983; Gilbert, Crouch & Kemp, 1998). Other studies are concerned with the reported vividness of experienced imagery (see for example Campos & Perez, 1988; Isaac, Marks & Russell, 1986). Overall, these studies contributed to the imagery debate by legitimizing and encouraging further investigations in this field.

From a neurophysiological perspective, an increasing number of researchers has recently adopted different psycho-physiological and neuroimaging techniques in order to investigate intermodal connections (see for example Fallgatter, Mueller & Strik, 1997; Farah, Weisberg, Monheit & Peronnet, 1990; Del Gratta, Di Matteo, De Nicola, Ferretti, Tartaro, Bonomo, Romani & Olivetti Belardinelli, 2001; De Volder et al., 2001).

However, until now little is known about how we imagine either an odor or the taste of our favorite dishes, or how we mentally reproduce the typical sound of everyday events. At least two specific aspects should be investigated both from a behavioral point of view, and from a neuro-physiological perspective: first, the specificity of mental imagery linked to each sensory modality; second, the degree of overlap between visual imagery and other types of imageries. Both questions would allow us to clarify the nature of mental imagery: the former by studying the imagery process on a more extensive set of perceptual-like objects, the latter by studying how much imagery according to various sensory modalities is tied to the processing of visual features.

The present study is devoted to the second question by trying to identify the common substrate of visual images and images generated according to other sensory modalities. It consists of a fMRI block design while participants were requested to generate mental images cued by short sentences describing different perceptual object (shapes, sounds, odors, flavors, self-perceived movements and internal sensations). Imagery cues were presented visually and were contrasted with sentences describing abstract concepts, since differences in activation during visual imagery and abstract thoughts were often assessed in literature (Lehman, Kochi, Koenig, Koykkou, Michel & Strik, 1994; Goldenberg, Podreka, Steiner & Willmes, 1987; Petsche, Lacroix, Lindner, Rappelsberger & Schmidt, 1992; Wise, Howard, Mummery, Fletcher, Leff, Büchel & Scott, 2000).

EXPERIMENT

METHODS

PARTICIPANTS

Fifteen healthy volunteers, after signing an informed consent waiver, participated in this study, which was approved by the local ethics committee. Eight of them were females and seven of them were males, and their age ranged between 19 and 20. All of them were right handed as well as their parents.

DESIGN

The experimental task required subjects to generate mental images cued by visually presented stimuli. Each experimental session of a single subject consisted of three functional studies and a morphological MRI. In each functional study, stimuli belonging to one experimental condition (regarding one of the seven sensory modalities) were delivered, together with stimuli belonging to the control condition. The first experimental condition always belonged to the visual modality, while those in the other two were evenly chosen among the remaining six modalities. Overall, the visual modality was studied fifteen times, while the other six modalities were studied five times. The number of modalities studied for each subject was limited to three in order to avoid lengthy recording sessions. The visual modality was always included and used as a reference.

Functional studies were performed according to a block paradigm, in which 12 volumes acquired during mental imagery – i.e. during experimental stimulus delivery – were alternated three times with 12 volumes acquired during baseline – i.e. during control sentence delivery. Experimental and control stimuli were presented at the start of the first volume, and then at every fourth one, so that three different experimental stimuli or three different control stimuli, were presented in each block. Each stimulus, or control sentence, remained visible until it was replaced by the following. Thus, the subject could see every stimulus for the whole time interval corresponding to the acquisition of four volumes, i.e. 24 seconds. The duration of a block was therefore 72 seconds, and the total duration of a study was 7 minutes 12 seconds. Overall, nine different experimental stimuli and nine different control sentences were presented in each study.

STIMULUS MATERIAL

Eight different sets of sentences referring to either concrete or abstract objects were used as mental image generation cues. Seven sets were used in the experimental condition and the remaining one was used in the control condition as a baseline. Each experimental set consisted of nine sentences, whereas the control set consisted of 27 sentences, and each sentence was composed by three or four words identifying either a definite perceptual object or an abstract concept.

The experimental sets contained sentences referring respectively to the visual, auditory, tactile, kinesthetic, gustatory, olfactory, and organic modalities. The control set contained sentences referring to abstract concepts. The English translation of an example included in each set is given: *Seeing a coin* (visual), *Hearing a rumble* (auditory), *Touching a soft material* (tactile), *The act of walking* (kinesthetic), *Smelling wet paint* (olfactory), *Tasting a salty food* (gustatory), *Feeling tired* (organic), *Admitting a misdeed* (abstract).

The entire set of stimuli was presented to the participant of the present experiment after the end of the experimental session in order to obtain data on the effectiveness of the material. Results revealed that participants classified 96% of visual items, 85% of auditory items, 88% of tactile items, 74% of kinesthetic items, 90% of olfactory items, 90% of gustatory items, 47% of organic items, and 55% of abstract items respectively as visual, auditory, tactile, kinesthetic, gustatory, olfactory, organic, and none of the previous categories. Chi-square comparison for each modality between observed and expected frequencies reveals that participants' responses match the item classification ($p<0.001$). Moreover, the rating of the power to evoke mental images (on a scale range from 1 to 7) revealed that modality specific items obtained an average score of 5.30 (s.d. 0.67) while abstract items achieved an average value of 2.10 (s.d. 1.22) ($t=11.98$, $p<0.0001$). The result was confirmed also for each single modality vs. abstract items comparison ($p<0.001$).

PROCEDURE

Subjects were interviewed in order to verify the lack of contra-indications at participating in the experiment and were acquainted with the experimental apparatus. They were then informed that they would be presented a set of sentences and were instructed to mentally read these sentences, without moving their lips, to concentrate on them, and to try to imagine their content.

Experimental and control sentences were projected on a translucent glass placed on the back of the scanner bore by means of an LCD projector and two perpendicular mirrors. An additional mirror fixed to the head coil inside the magnet bore allowed the subject to see the translucent glass. The LCD projector was driven by a PC placed at the scanner console and connected to it via a VGA cable through a hole in the shielded room. The PC was manually controlled by an operator, according to the volume acquisition timing. The stimuli and control sentences were administered by means of a slide presentation software, and were printed in yellow on a blue background. No artifacts due to the projector or the VGA cable were visible in the functional as well as in the morphological images.

APPARATUS

Functional MRI was performed with a SIEMENS VISION 1.5T scanner endowed with EPI (Echo Planar Imaging) capability. Each functional volume was acquired by means of an EPI FID (Free Induction Decay) sequence with the following parameters: 30 bicommisural transaxial slices 3 mm thickness; no gap, matrix 64 x 64; FOV (Field Of View) 192; 3 mm x 3 mm in-plane voxel size; flip angle 90°; TR 6 s; TE 60 ms. Scan time for one volume was three seconds. The image volume covered the whole brain.

In addition to functional images, a high resolution, morphological MRI was acquired at the end of each session, by means of a 3D-MPRAGE (Magnetization Prepared Rapid Gradient Echo) sequence. The parameters characterizing this acquisition were: 240 axial slices, 1 mm thickness, no gap, matrix 256 x 256, FOV 256 mm, in-plane voxel size 1 mm x 1 mm, flip angle 12°, TR = 9.7 ms, TE = 4 ms.

DATA ANALYSIS

Functional data were analyzed using MEDx software by Sensor Systems. First, all volumes in a study were realigned, in order to correct for physiological subject movement, with the software AIR included in the MEDx software package. Then, data were grouped according to the various sensory modalities. Three different groups were formed for the visual modality so that comparison between the latter and other modalities was performed within the same group of subjects. All functional volumes were transformed into Talairach space and, within a single modality, or within a single group of subjects in the visual modality, were merged to form a larger block paradigm, consisting of 360 volumes. All volumes in such a group were normalized to the same baseline level. A spatial gaussian filter 4 mm FWHM was applied. Voxel time courses were high pass filtered. The volumes in each modality group were divided into subgroups corresponding to volume acquired during the presentation of modality specific stimuli, and during the presentation of control sentences respectively. Then a voxel-by-voxel Student t-test was performed, and the corresponding Z-score maps were calculated and thresholded at $Z=2.5$ corresponding to a null probability $p<0.006$ (uncorrected). Subsequently the clustering algorithm of the MEDx package was run on these maps, thus selecting only the clusters of activation with a probability larger than 0.5. Finally the clustered Z-score maps were superimposed on a high resolution, Talairach transformed, morphological image.

Finally in each map we looked for activation areas common to the visual on one hand, and the remaining modalities on the other hand. To this end we compared the thresholded Z-score maps of a pair of modalities and selected the voxels that were significantly activated in both. This yielded maps of voxels activated in both modalities, which were then classified according to their neuroanatomical location by means of the Talairach atlas.

RESULTS

In all modalities a number of activated areas was clearly observed above the threshold. In Table 1 group data are listed. For each activation area the Talairach coordinates of the centroids of clusters of activation are indicated, together with the corresponding Brodmann area, where applicable.

In the visual modality the most prominent areas of activation are distributed bilaterally even though the right hemisphere, overall, appears to be more activated than the left one. However, activation in the temporal area was more intense on the left. Other prominent activation is observed in several orbital frontal areas, mainly on the right hemisphere.

In the auditory modality the main areas of activation were, bilaterally, in the middle temporal area, in the middle and superior pre-frontal area, and, unilaterally, in the left insula-precentral gyrus.

In the tactile modality the activation pattern is quite asymmetrical, with the most prominent activation in the left hemisphere. In the left hemisphere areas of activation are observed in the middle-inferior temporal, inferior frontal, inferior parietal areas. One symmetrical activation is observed in the insula, which is however much larger and more intense in the left hemisphere. Another symmetrical activation is in the post-central gyrus, here too, much more intense in the left hemisphere.

In the olfactory modality, bilateral areas of activation are observed in the middle frontal gyrus. In the left hemisphere, prominent activated areas are observed in the inferior-middle temporal gyrus, in the parietal area and in the middle prefrontal gyrus. Overall, the left hemisphere appears to be more activated than the right one.

The gustatory modality shows a rough symmetry regarding the location of the active areas, but a strong asymmetry regarding their extension, the activation in the left hemisphere being much larger. Activated areas in both hemispheres are in the parietal region, in the post-central gyrus, in the insula, and in prefrontal areas.

The organic modality shows a bilateral compound symmetrical activation pattern around the superior temporal area, with a maximum in the insula, in the pre-central operculum, and in the post-central gyrus, and a bilateral activation in the middle-superior frontal areas. Activation was also observed in the left parietal area, and in the left inferior temporal gyrus.

The kinesthetic modality shows rather symmetrical activation in the cingulate gyrus, in the middle and inferior temporal areas. In the left hemisphere, activation is observed in the precuneus while in the right hemisphere activation is observed in the fusiform gyrus.

The maps of voxels that were significantly activated both in the visual modality and in each of the other sensory modalities compared with it yielded the following results.

Visual and auditory modalities share a bilateral activation in the middle-inferior temporal area, although the maximum of communality is limited to a little portion only.

Modality	VISUAL		AUDITORY		TACTILE		OLFACTORY		GUSTATORY		ORGANIC		KINESTHETIC	
Hemisphere	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
Area	x y z	x y z	x y z	x y z	x y z	x y z	x y z	xx y z	x y z	xx y z	x y z	x y z	x y z	x y z
Fusiform/Hippocampal gyrus (BA37)		-36, -26, -14	-					30, -34, -16					45, -36, -14	
Middle-inferior temporal gyri (BA37)	55, -49, -6	-53, -57, -8	61, -43, -8	-49, -53, -9		-53, -61, -9		-48, -53, -8				-56, -54, -6	46, -15, -2	-49, -47, -8
Superior temporal gyrus (BA37)													46, -57, 8	-36, -62, 8
Insula				-36, -1, 15	32, -1, 15	-38, -5, 7			40, -6, -7	-38, 1, 2	39, -3, -1	-39, -3, -6		
Post-central gyrus (BA2)											58, -29, 39	-60, -27, 40		
Post-central gyrus (BA43)					51, -19, 16	-59, -22, 16			37, -2, 15	-38, -12, 16	54, -11, 19	-59, -21, 19		
Precuneus (BA7)														-12, -62, 45
Inferior parietal lobule (BA40)	53, -37, 34	-49, -37, 35				-53, -39, 47		-55, -26, 30	59, -35, 31	-53, -39, 37		-56, -41, 40		
Middle-inferior frontal gyri (BA6)	29, 9, 55								58, -2, 31	-57, -5, 34	50, -1, 9	-56, -4, 9		
Middle frontal gyrus (BA9/10)	27, 36, -4		28, 35, -3	-28, 35, -3				-22, -37, -4			32, 59, 4	-35, 40, -2		
Superior frontal gyri (BA9/10)	34, 51, 3		10, 59, 8	-10, 59, 8					42, 38, 19	-39, 39, 13				
Middle-inferior frontal gyri (BA44/46/47)	42, 38, 12					-56, 4, 25	50, 39, 17	-44, 34, 17			47, -31, 16	-51, 15, -2		
Posterior cingulate gyrus (BA31)	5, -40, 27												16, -19, 37	-12, -28, 35

Table 1. Talairach coordinates for the activated areas in the different modalities.

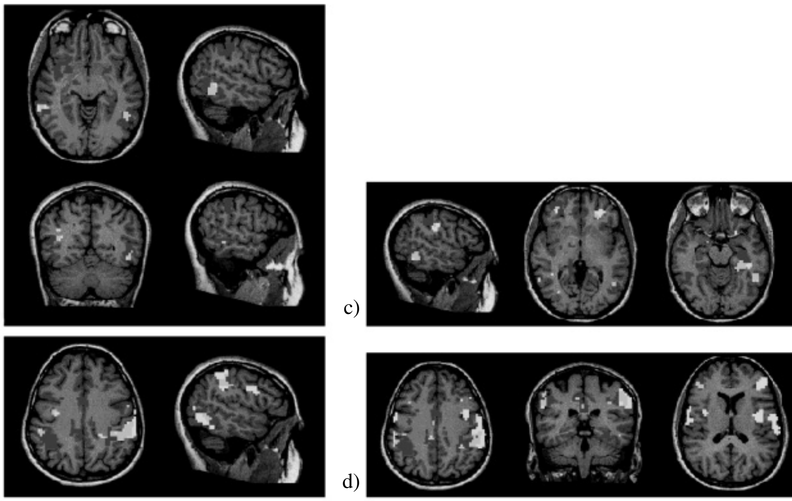

Figure 1. a) Visual and auditory modalities exhibit common activation in middle temporal areas. The upper right panel shows a left sagittal view, while the lower right shows a right sagittal view. In this, and the following figures, colored voxels indicate Z>2.5 corresponding to a null probability p<0.006. Blue voxels represent visual imagery activation, yellow voxels represent compared modality activation, and green voxels indicate common areas of activation. Note that the left side of each image represents the right hemisphere and viceversa.
b) Visual and tactile modalities share common bilateral parietal areas (left panel): note the different degree of asymmetry in the extension of the activation. Other common activated areas in the parietal and middle temporal regions are shown in a left sagittal view (right panel).
c) Visual and olfactory modalities reveal common activated areas in the parietal and middle temporal region (left panel showing a left sagittal view), in the left middle temporal and middle frontal region; note the reversed asymmetry of the two patterns of activation (central panel), and in the left fusiform gyri (right panel).
d) Visual and gustatory modalities show common activated areas in the inferior parietal region bilaterally (left and central panel) and in the right middle-inferior frontal areas (right panel).

In visual and tactile modalities activated areas are seen in the parietal lobe, with a different degree of asymmetry: in the left hemisphere activation is about equal, while in the right hemisphere the visual modality produces a more extended activation. Other common areas are in the left middle-inferior temporal area.

Visual and olfactory modalities show quite different activation patterns, with activated areas mainly in the left hemisphere, namely in the middle-inferior temporal region, and the parietal area. Middle frontal areas are activated bilaterally in both modalities, but with a reversed pattern of asymmetry. Indeed, in the left hemisphere the olfactory modality shows a more extended activation, while the reverse is true in the right hemisphere.

In visual and gustatory modalities, an overlap of activation in the inferior parietal area is observed bilaterally but with a different degree of symmetry. In addition, overlaps are seen in the right middle-inferior frontal areas.

No significant communalities were found between visual and organic imagery and between visual and kinesthetic imagery. However, both the organic and the kinesthetic modalities show a consistent activation in the left middle-inferior temporal area, even if it does not overlap with the corresponding area in the visual condition. In addition, activation in the parietal lobe was found for the organic modality but it does not share a common area with the corresponding activated region for the visual modality.

DISCUSSION

This study indicates that each type of mental imagery exhibits a different degree of overlap with visual imagery for what concerns their cerebral correlates as revealed by fMRI.

In general, visual imagery activates mainly the right hemisphere, while the tactile, olfactory, and gustatory imageries elicit prominently left activation. Finally, auditory, kinesthetic, and organic imageries involve equally both hemispheres.

The most consistent region of overlap is the middle-inferior temporal area, especially on the left hemisphere. In fact, auditory, tactile, and olfactory imageries all show common activated areas in this region. In addition, organic and kinesthetic modalities also show activation in this region even though without any overlap.

The parietal associative areas also exhibit a certain degree of consistency, because common areas of activation with visual imagery were found for tactile, olfactory and gustatory modalities. Here again, the organic modality shows an activated area in this region but it does not overlap with the corresponding area in the visual condition.

The prefrontal areas show a less consistent pattern of activation as they reveal common areas of activation only for the visual-gustatory comparison. However, activation in prefrontal areas was also found in the visual-olfactory comparison, although the intensity pattern in the two hemispheres was reversed across modalities.

In some cases, different sensory imageries activate the same area, in other cases the areas of activation are close to each other in the same neuroanatomical area, indicating that the region is involved in both modalities but perhaps in a non-perfectly coincident way. Both cases could be explained, according to Calvert, Brammer & Iversen (1998), by the fact that the hetero-modal cortex either contains neurons responding to more than one modality or has closely interspersed populations of modality specific neurons, which are responsive to different modalities.

Another tentative explanation could be derived from the proposal put forward by Singer (2000) regarding the coexistence in the mammalian brain of complementary strategies for the representation of mental contents: a strategy for items that occur very frequently and/or are of particular behavioral importance, and a second one reserved for items which are infrequent, novel, or of high complexity. Reviewing experimental data on vision, audition, motion and olfaction, Singer suggests that, in the latter case, items are coded by dynamically associated assemblies of feature-tuned cells formed by rapid and transient synchronization of the associated neurons. According to this hypothesis, the partial mismatch between activated areas in visual imagery and activated areas in other sensory modalities could be due to different degrees of either salience or complexity of the requested images.

In this study, the inferior temporal region is often activated bilaterally, although the most consistent activation is found on the left side, especially in the auditory, tactile, olfactory modalities compared with the visual one. Regarding the visual modality while D'Esposito et al. (1997) reported left fMRI activation in this site, Mellet et al. (1998 and 2000) reported a bilateral PET activation. They maintain that the right side of the inferior temporal area is responsible for the processing of complex shapes, while the left side seems to be engaged in the processing of verbalizable shapes. In this view the presence or the absence of right temporal cortex activation would depend on the complex or simple nature of the visual image processing, while the left activation would depend on being or not being verbalizable. In our study, all the images were cued by verbal items and the task requirements were held constant across conditions, therefore there is no reason to suppose complexity differences in the processing of different images. In light of previous data, the left side of this area might have a role in connecting the verbal encoding of a word with its deeper representation (Wise et al., 2000). However, the lack of any overlapping temporal activation for the gustatory modality and a non perfectly coincident overlap for the organic and kinesthetic modalities suggest also that this area may reflect the segregation of semantic knowledge into anatomically discrete, but highly interactive, modality specific regions (Thompson-Schill et al., 1999).

The activation of the parietal region in imagery processes is often reported in literature, frequently in association with task related to spatial processing (Iwaki et al., 1999; Banati, Goerres, Tjoa, Aggleton & Grasby, 2000; Barnes et al., 2000; Diwadkar, Carpenter & Just, 2000; Jordan et al., 2001). However, although in our task spatial processing was not requested, a consistent common activation, albeit not including all the modalities, was found in the parietal region. Jordan et al. (2001) suggest that this region may be responsible for the transformation of shapes into a supra-modal form, thus enabling the cognitive system to process visuo-spatial features in a way that is independent from sensory features. According to these authors, the network underlying this transformation may be involved in low-level attentional processes, working for many types of cognitive processes (Coull & Frith, 1998; Coull & Nobre, 1998). In this view, the common area of activation we found in the parietal region may reflect supra-modal transformations mediated by low-level attentional process.

Regarding the prefrontal areas, some data support the idea that this region may be related to the memory retrieval of mental images. Our data is consistent with previous data indicating a hemispheric domain-specificity of the prefrontal cortex (right-sided for spatial WM, bilateral or left-sided for verbal WM). In our study the visual modality show a right-sided activation in the prefrontal areas, while other modalities show a composite pattern distributed either bilaterally or on the left hemisphere, yielding only an overlap in the right middle prefrontal area between the visual and the gustatory modalities. This pattern of activation may suggest that visual imagery (and perhaps also the gustatory one) relies on spatial processes, while other modalities rely more upon verbal processes.

The reverse pattern of activation found in prefrontal areas for visual and olfactory modalities is coherent with data reported by Zald & Pardo (2000) asserting a mainly left prefrontal activation in odorants hedonic judgments. However, this data refers to a perceptual task and can supply only an indirect indication to olfactory imagery.

The lack of any consistent activation in primary sensory areas could be due to the kind of task used in this study. As suggested by Thompson, Kosslyn, Sukel & Alpert (2001), the primary visual cortex is activated more often when participants are requested to use the image in some way. In our study, in order to minimize differences among the conditions, apart from those related to the imagery modality, participants were simply requested to mentally represent the target item, i.e. they were requested to perform an image generation task. According to Behrmann (2000), image generation is a process more specific to imagery than image manipulation, because it involves the active reconstruction of a long-term mental representation. Moreover, in our opinion, image manipulation involves some kind of on-line processing that might be more dependent on the specific content of the image to be manipulated. As our study was aimed at identifying the common substrate of different imagery modalities, the image generation task seems to imply processes supposed to be less variable across modalities.

An alternative explanation for the lack of activation of primary visual areas may be due to the visual presentation of the items. However, studies that contrasted concrete items vs. abstract items by using an auditory presentation (De Volder et al., 2001; Mellet et al., 1998; D'Esposito et al., 1997) found substantially the same pattern of results for the visual modality.

Whether common areas indicate either the involvement of amodal functional circuits in mental imagery, or the presence of a visual imagery component also in different types of mental images, should be the object of further investigations. However, the first hypothesis is a little more coherent with the results of the ratings of vividness of the material used in this study, which show a clear cut among different types of images.

From this study, we can derive three key findings. First, common brain areas were found to be active in both visual imagery and imagery based on other sensory modalities. These common areas are supposed to reflect either the verbal retrieval of long-term representations or the segregation of long-term representations into highly interactive modality specific regions.

Second, each imagery modality activates also distinct brain areas, suggesting that high-level cognitive processes imply modality-specific operations. This result is coherent with the domain-specific hypothesis proposed for the functioning of the fronto-parietal associative stream (Rushworth & Owen, 1998; Miller, 2000).

Third, primary areas were never found to be active, suggesting that different, though interactive, neural circuits underlie low-level and high-level processes. Although this claim is only indicative, as in this study, no direct comparisons were made between imagery and perceptual/motor processes, it outlines the lack of primary cortex activation for imagery in those modalities that were not accompanied by any corresponding sensory stimulation due to either the visual presentation of the stimuli or to the noisy apparatus. Further investigations will be essential to extensively clarify this claim.

Prof. Marta Olivetti Belardinelli
Dipartimento di Psicologia
Università di Roma "La Sapienza"
V. dei Marsi 78
00185, Roma, Italy

Prof. Rosalia di Matteo
Dipartimento di Psicologia
Università di Roma "La Sapienza"
V. dei Marsi 78
00185, Roma, Italy

Prof. Cosimo Del Gratta
INFM, L'Aquila and
Dipartimento di Scienze Cliniche e Bio-immagini
Università di Chieti "G. D'Annunzio"
V. dei Vestini 33
66013, Chieti, Italy

Prof. Andrea De Nicola
INFM, L'Aquila and
ITAB
Università di Chieti "G. D'Annunzio"
V. dei Vestini 33
66013, Chieti, Italy

Prof. Antonio Ferretti
INFM, L'Aquila and
ITAB
Università di Chieti "G. D'Annunzio"
V. dei Vestini 33
66013, Chieti, Italy

Prof. Gian Luca Romani
INFM, l'Aquila and
Dipartimento di Scienze Cliniche e Bio-immagini
Università di Chieti "G. D'Annunzio"
V. dei Vestini 33
66013, Chieti, Italy

REFERENCES

BANATI, R.B., GOERRES, G.W., TJOA, C., AGGLETON, J.P., & GRASBY, P. (2000). The functional anatomy of visual-tactile integration in man: A study using positron emission tomography. *Neuropsychologia*, 38, 115-124.

BARNES, J., HOWARD, R.J., SENIOR, C., BRAMMER, M., BULLMORE, E.T., SIMMONS, A., WOODRUFF, P., & DAVID, A.S. (2000). Cortical activity during rotational and linear transformations. *Neuropsychologia*, *38,* 1148-1156.

BARTOLOMEO, P., BACHOUD-LEVI, A.-C., DE GELDER, B., DENES, G., DALLA BARBA, G., BRUGIERES, P., & DEGOS, J.D. (1998). Multiple-domain dissociation between impaired visual perception and preserved mental imagery in a patient with bilateral extrastriate lesions. *Neuropsychologia*, *36,* 239-249.

BEHRMANN, M. (2000). The mind's eye mapped onto the brain's matter. *Current Directions in Psychological Science*, *9,* 50-54.

BETTS, G.H. (1909). *The distribution and functions of mental imagery*. New York: New York Teachers College, Columbia University.

BOSCH, V., MECKLINGER, A., & FRIEDERICI, A.D. (2001). Slow cortical potentials during retention of object, spatial, and verbal information. *Cognitive Brain Research*, *10,* 219-237.

BURBAUD, P., CAMUS, O., GUEHL, D., BIOULAC, B., CAILLÉ, J.-M., & ALLARD, M. (2000). Influence of cognitive strategies on the pattern of cortical activation during mental subtraction. A functional imaging study in human subjects. *Neuroscience Letters*, *287,* 76-80.

CALVERT, G.A., BRAMMER, M.J., & IVERSEN, S.D. (1998). Crossmodal identification. *Trends in Cognitive Science*, *2,* 247-253.

CAMPOS, A., & PEREZ, M.J. (1988). Vividness of Movement Imagery Questionnaire: Relations with other measures of mental imagery. *Perceptual and Motor Skills*, *67,* 607-610.

CAREY, D.P. (1998). Action, perception, cognition, and the inferior parietal cortex. *Trends in Cognitive Science*, *2,* 162-164.

CHEN, W., KATO, T., ZHU, X.-H., OGAWA, S., TANK, D.W., & UGURBIL, K. (1998). Human primary visual cortex and lateral geniculate nucleus activation during visual imagery. *Neuroreport*, *9,* 3669-3674.

COCUDE, M., MELLET, E., & DENIS, M. (1999). Visual and mental exploration of visuo-spatial configurations: Behavioral and neuroimaging approaches. *Psychological Research*, *62,* 93-106.

COULL, J.T., & FRITH, C.D. (1998). Differential activation of right superior parietal cortex and intraparietal sulcus by spatial and nonspatial attention. *NeuroImage*, *8,* 176-187.

COULL, J.T., & NOBRE, A.C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, *18,* 7426-7435.

DE VOLDER, A.G., TOYAMA, H., KIMURA, Y., KIYOSAWA, M., NAHANO, H., VANLIERDE, A., WANET-DEFALQUE, M.-C., MISHINA, M., ODA, K., ISHIWATA, K., & SENDA, M. (2001). Auditory triggered mental imagery of shape involves visual association areas in early blind humans. *NeuroImage*, *14,* 129-139.

DEL GRATTA, C., DI MATTEO, R., DE NICOLA, A., FERRETTI, A., TARTARO, A., BONOMO, L., ROMANI, G.L., & OLIVETTI BELARDINELLI, M. (2001). Sensory image generation: A comparison between different sensory modalities with fMRI. *NeuroImage*, *6* (special issue)*,* S394.

D'ESPOSITO, M., DETRE, J.A., AGUIRRE, G.K., STALLCUP, M., ALSOP, D.C., TIPPET, L.J., & FARAH, M.J. (1997). A functional MRI study of mental image generation. *Neuropsychologia*, *35,* 725-730.

DIWADKAR, V.A., CARPENTER, P.A., & JUST, M.A. (2000). Collaborative activity between parietal and dorsolateral prefrontal cortex in dynamic spatial working memory revealed by fMRI. *NeuroImage*, *12,* 85-99.

FALLGATTER, A.J., MUELLER, T.J., & STRIK, W.K. (1997). Neurophysiological correlates of mental imagery in different sensory modalities. *International Journal of Psychophysiology*, *25,* 145-153.

FARAH, M.J. (1995). Current issue in the neuropsychology of image generation. *Neuropsychologia*, *33,* 1455-1471.

FARAH, M.J., WEISBERG, L.L., MONHEIT, M., & PERONNET, F. (1990). Brain activity underlying mental imagery: Event-related potentials during mental image generation. *Journal of Cognitive Neuroscience*, *1,* 302-316.

FARTHING, C.W., VENTURINO, M., & BROWN, S.W. (1983). Relationship between two different types of imagery vividness questionnaire items and three hypnotic susceptibility scale factors: A brief communication. *International Journal of Clinical and Experimental Hypnosis*, *31,* 8-13.

GILBERT, A.N., CROUCH, M., & KEMP, S.E. (1998). Olfactory and visual mental imagery. *Journal of Mental Imagery*, *22,* 137-146.

GISSURARSON, L.R. (1992). Reported auditory imagery and its relationship with visual imagery. *Journal of Mental Imagery*, *16,* 117-122.

GOLDENBERG, G., PODREKA, I., STEINER, M., & WILLMES, K. (1987). Patterns of regional cerebral blood flow related to meaningfulness and imaginability of words - An emission computer tomography study. *Neuropsychologia*, *25,* 473-486.

HÖLLINGER, P., BEISTEINER, R., LANG, W., LINDINGER, G., & BERTHOZ, A. (1999). Mental representations of movements. Brain potentials associated with imagination of eye movements. *Clinical Neurophysiology*, *110,* 799-805.

ISAAC, A., MARKS, D.F., & RUSSELL, D.G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire. *Journal of Mental Imagery*, *10,* 23-30.

IWAKI, S., UENO, S., IMADA, T., & TONOIKE, M. (1999). Dynamic cortical activation in mental image processing revealed by biomagnetic measurement. *Neuroreport*, *10,* 1793-1797.

JORDAN, K., HEINZE, H.-J., LUTZ, K., KANOWSKI, M., & LÄNCHE, L. (2001). Cortical activations during the mental rotation of different visual objects. *NeuroImage*, *13,* 143-152.

KANIZSA, G. (1991). Vedere e pensare. Roma: Il Mulino.

KLEIN, I., PARADIS, A.-L., POLINE, J.-B., KOSSLYN, S.M., & LE BIHAN, D. (2000). Transient activity in the human calcarine cortex during visual mental imagery: An Event-Related fMRI study. *Journal of Cognitive Neuroscience*, *12,* 15-23.

KOSSLYN, S.M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

KOSSLYN, S.M., & THOMPSON, W.L. (2000). Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In M.S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 975-985). Cambridge, MA: MIT Press.

KOSSLYN, S.M., ALPERT, N.M., THOMPSON, W.L., MALJKOVIC, V., WEISE, S.B., CHABRIS, C.F., HAMILTON, S.E., RAUCH, S.L., & BUONOMANO, F.S. (1993). Visual mental imagery activates topographically organized visual areas: PET investigations. *Journal of Cognitive Neuroscience*, *5,* 263-287.

KOSSLYN, S.M., MALJKOVIC, V., HAMILTON, S.E., HORWITZ, G., & THOMPSON, W.L. (1995). Two types of image generation: Evidence for left and right hemisphere processes. *Neuropsychologia*, *33,* 1485-1510.

LAMM, C., WINDISCHBERGER, C., LEODOLTER, U., MOSER, E., & BAUER, H. (2001). Evidence for premotor cortex activity during dynamic visuospatial imagery from single-trial functional magnetic resonance imaging and event-related slow cortical potentials. *NeuroImage*, *14,* 268-283.

LEHMAN, D., KOCHI, K., KOENIG, T., KOYKKOU, M., MICHEL, C.M., & STRIK, W.K. (1994). Microstates of the brain electric field and momentary mind states. In M. Eiselt, U. Zwiener and H. Witte (Eds.), *Quantitative and topological EEG and MEG analysis*. Jena: Universitätsverlag Mayer.

MELLET, E., TZOURIO MAZOYER, N., BRICOGNE, S., MAZOYER, B.M., KOSSLYN, S.M., & DENIS, M. (2000). Functional anatomy of high-resolution visual mental imagery. *Journal of Cognitive Neuroscience*, *12,* 98-109.

MELLET, E., TZOURIO MAZOYER, N., DENIS, M., & MAZOYER, B.M. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport*, *9,* 803-808.

MILLER, E.K. (2000). The prefrontal cortex: No simple matter. *NeuroImage*, *11,* 447-450.

PETSCHE, H., LACROIX, D., LINDNER, K., RAPPELSBERGER, P., & SCHMIDT, H.E. (1992). Thinking with images or thinking with language: A pilot EEG probability mapping study. *International Journal of Psychophysiology*, *12,* 31-39.

RICHTER, W., SOMORJAI, R., SUMMERS, R., JARMASZ, M., MENON, R.S., GATI, J.S., GEORGOPOULOS, A.P., TEGELER, C.U.K., & KIM, S.-G. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience*, *12,* 310-320.

RUSHWORTH, M.F.S., & OWEN, A.M. (1998). The functional organization of the lateral frontal cortex: Conjecture or Conjuncture in the electrophysiology literature? *Trends in Cognitive Science*, *2,* 46-53.

SHEEHAN, P.W. (1967). A shortened form of Betts' questionnaire upon mental imagery. *Journal of Clinical Psychology*, *23,* 386-389.

SINGER, W. (2000). Response synchronization: A universal coding strategy for the definition of relations. In M.S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 325-338). Cambridge, MA: MIT Press.

STEWART, L., MEYER, B.-U., FRITH, U., & ROTHWELL, J. (2001). Left posterior BA37 is involved in object recognition: A TMS study. *Neuropsychologia*, *39,* 1-6.

THOMPSON, W.L., KOSSLYN, S.M., SUKEL, K.E., & ALPERT, N.M. (2001). Mental imagery of high- and low-resolution gratings activates area 17. *NeuroImage*, *14,* 454-464.

THOMPSON-SCHILL, S.L., AGUIRRE, G.K., D'ESPOSITO, M., & FARAH, M.J. (1999). A neural basis for category and modality specificity of semantic knowledge. *Neuropsychologia*, *37,* 671-676.

WHITE, K.D., ASHTON, R., & BROWN, R.M.D. (1977). The measurement of imagery vividness: Normative data and their relationship to sex, age, and modality differences. *British Journal of Psychology*, *68,* 203-211.

WISE, R.J.S., HOWARD, D., MUMMERY, C.J., FLETCHER, P., LEFF, A., BÜCHEL, C., & SCOTT, S.K. (2000). Noun imageability and the temporal lobes. *Neuropsychologia*, *38,* 958-994.

WOLPIN, M., & WEINSTEIN, C. (1983). Visual imagery and olfactory stimulation. *Journal of Mental Imagery*, *7,* 63-74.

ZALD, D.H., & PARDO, J.V. (2000). Functional neuroimaging of the olfactory system in humans. *International Journal of Psychophysiology*, *36,* 165-181.

ZATORRE, R.J., HALPERN, A.R., PERRY, D.W., MEYER, E., & EVANS, A.C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, *8,* 29-46.

PART II

FORMS AND SCHEMES OF PERCEPTUAL AND COGNITIVE SELF-ORGANISATION

This page intentionally left blank

VICTOR ROSENTHAL

MICROGENESIS, IMMEDIATE EXPERIENCE AND VISUAL PROCESSES IN READING

INTRODUCTION

The concept of *microgenesis* refers to the *development* on a brief *present-time* scale of a percept, a thought, an object of imagination, or an expression. It defines the occurrence of immediate experience as *dynamic unfolding* and *differentiation* in which the 'germ' of the final experience is already embodied in the early stages of its development. Immediate experience typically concerns the focal experience of an object that is thematized as a 'figure' in the global field of consciousness; this can involve a percept, thought, object of imagination, or expression (verbal and/or gestural). Yet, whatever its modality or content, focal experience is postulated to develop and stabilize through dynamic differentiation and unfolding. Such a microgenetic description of immediate experience substantiates a *phenomenological* and *genetic* theory of cognition where any process of perception, thought, expression or imagination is primarily a process of genetic differentiation and development, rather than one of detection (of a stimulus array or information), transformation, and integration (of multiple primitive components) as theories of cognitivist kind have contended.

The term *microgenesis* was first coined by Heinz Werner (1956) as a means of providing a genetic characterization of the structure and temporal dynamics of immediate experience, and, more generally, of any psychological process (Werner, 1957; Werner & Kaplan, 1956; Werner & Kaplan, 1963). But the genetic framework to which this term referred actually emerged in the mid-1920s in the context of Werner's work at the University of Hamburg and, to a certain extent, of the work of the *Ganzheitspsychologie* group in Leipzig led by Friedrich Sander. For Werner, *microgenesis* had not only a substantive (as a psychological theory) but also a methodological meaning. As a method, it either referred to *genetic realization* (*Aktualgenese*) which sought to provide the means of externalizing the course of brief perceptual, or other cognitive processes by artificially eliciting 'primitive' (*i.e.* developmentally early) responses that are normally occulted by the final experience (see in this respect Sander, 1930; Werner, 1956). Or it referred to *experimental psychogenesis* which aimed to construct small-scale, *living models* of large-scale developmental processes in such a way as to 'miniaturize' (*i.e.* accelerate and/or telescope) the course of a given process and bring it under experimental control. Experimental psychogenesis, devised by Werner in the 1920s, played afterwards an important role in the work of Vygotsky and Luria who further extended its field of application and gave it a historical dimension (Catán, 1986; Vygotsky, 1978; Werner, 1957, first German edition published in 1926; Werner & Kaplan, 1956). As a theoretical framework, microgenesis constituted a rectification of Gestalt theory especially in regard to its overly structural and agenetic character[1].

A. Carsetti (ed.), Seeing, Thinking and Knowing, 221-243.

Yet, together with the latter, microgenesis gave psychology its first cognitive paradigm. In its modern version, microgenesis offers a genetic, phenomenological alternative to the information-processing metaphor, an alternative that reunites mind and nature and restores to cognition its cultural and hermeneutic dimensions.

My purpose in this essay is to provide an overview of the main constructs of microgenetic theory, to outline its potential avenues of future development in the field of cognitive science, and to illustrate an application of the theory to research, using visual processes in reading as an example. In my overview, I shall not dwell on the history of microgenesis (the reader may find the relevant sources in Catán, 1986; Conrad, 1954; Sander, 1930; Werner, 1956; Werner, 1957; Werner & Kaplan, 1956) but rather describe its main constructs from a contemporary perspective.

MICROGENETIC DEVELOPMENT

Microgenetic development concerns the psychogenetic dynamics of a process that can take from a few seconds (as in the case of perception and speech) up to several hours or even weeks (as in the case of reading, problem solving or skill acquisition). It is a living process that dynamically creates a structured coupling between a living being and its environment and sustains a knowledge relationship between that being and its world of life (*Umwelt*). This knowledge relationship is protensively embodied in a readiness for further action, and thereby has practical meaning and value. Microgenetic development is thus an essential form of cognitive process: it is a dynamic process that brings about readiness for action[2]. Microgenesis takes place in relation to a thematic field which, however unstable and poorly differentiated it might be, is always given from the outset. To this field, it brings stabilized, differentiated structure and thematic focalization, thereby conferring value and meaning to it. Figure/ground organizations are an illustration of a typical microgenetic development. Yet, one should bear in mind that however irresistible an organization might appear, it is never predetermined but admits of alternative solutions, that a ‘figure’ embodies a focal theme, and that a ‘ground’ is never phenomenologically or semantically empty. *Thematic field* denotes here a definite field of consciousness, and has both phenomenological *and* semantic meaning (see Gurwitsch, 1957). Focal thematic embodiment of microgenetic development thus differs from unfocussed, heterogeneous, and heterochronic ontogenesis, which spans a considerable portion of life and requires organic maturation and growth (see Werner, 1957, for a discussion of differences between microgenetic development and ontogenesis; Werner & Kaplan, 1956).

MEANING AND FORM

It should be noted that form, meaning and value are not considered separate or independent entities. According to microgenetic theory, whatever acquires the phenomenological status of individuated form acquires, *ipso facto*, value and meaning.

This may not necessarily be a focally attended meaning, as the definite experience of meaning depends on whether a form is given a focal thematic status. Yet, regardless of the status of the meaning experience, microgenetic theory postulates that form has of necessity semantic and axiological extensions. Incidentally, this point highlights the radical opposition between microgenesis and the standard cognitivist stance, where form, meaning and value are deemed independent and mobilize processes that are intrinsically alien to one another. If meaning and value are acknowledged to affect perception, as the seminal experiment of Bruner and Goodman (1947) revealed by showing that the size of a coin is seen as bigger when it is highly valued, it is assumed that this influence obtains via the interaction of processes. Yet no precise explanation has been supplied as to how structurally different processes, which deal with incommensurable factors, can ever interact with one another[3]. Werner and Wapner (1952) observed many years ago, that theories which separate sensory, semantic, motivational and emotional processes, and view perception as a construction of abstract forms out of meaningless features (only to discover later their identity and meaning), face in this respect insurmountable paradoxes. If semantics postdates morphology, then it cannot affect form reconstruction, and if semantics is concomitant with form reconstruction, how can it influence morphological processing prior to 'knowing' what the latter is about? Finally, since morphological and semantic processes are viewed as incommensurable, how can they be brought to cooperate together without recourse to yet another, higher-order process? Invoking such a process would either amount to conjuring up a sentient device of the homunculus variety or would stand in contradiction to the very postulate of the distinctness and independence of meaning and form.

CATEGORIZATION

According to the present account, no such interaction is to be sought because meaning and form are not separate or independent entities; on the contrary, perception is directly meaning and value-laden, with actual meaning developing along the global-to-local (indefinite/general-to-definite/specific) dynamics of microgenesis. The gradual differentiation of a meaning, percept or concept involves a *global-to-local* course of development, where meaning and value go hand-in-hand with perceptual or cognitive organization, developing from vague and general to definite and specific. Note, though, that no direct holistic principle can be viable if it does not rely on a process of categorization. Immediate categorization represents another essential feature of microgenetic development: it provides the dynamic link between holistic differentiation, meaning and readiness for action. Consider, indeed, that even the most basic categorization has meaning – meaning is thus not the end product of perception but rather part and parcel of the perceptual process[4]. The psychological literature gives ample evidence of the fact that subjects carry out basic categorization instantaneously (e.g., discrimination of relevant from irrelevant stimuli), without first making a more complete identification of the stimuli, and that preliminary categorization improves the rate of definite identification (Brand, 1971; Ingling, 1972). It should also be emphasized that categorization necessarily delineates a horizon of action, a horizon that comprises a range of relevant acts that the subject may potentially be compelled to enact.

Perception can then be said to act under the assumption of the consistency and meaningfulness of the world in which we live: the perceptual system 'assumes' that whatever it encounters has structure and meaning. It therefore anticipates and actively seeks meaningful structures (objects) and immediately categorizes them on a global dynamic basis[5]. Microgenetic theory contains here a hermeneutic principle: in order to be meaningful, perception must consist in dynamic categorization evolving from general to specific, from vague and global to precise and local (see Rastier, 1997). Incidentally, this explains why the 'germ' of the final percept is already embodied in early stages of the perceptual process. Immediate categorization allows for the categorical continuity of forms throughout the entire perceptual process giving cohesiveness and stability to the perceived world (see Cadiot & Visetti, 2001; Gurwitsch, 1966, chap. 1). This primary categorization may be insufficient for the precise and overt perceptual identification of objects – as required by standard psychological experiments – and may then need to be completed by a process of local *discrimination*. This complementary discrimination, necessary for the focal thematization of a 'figure', is greatly constrained by the former process; it operates within a restricted categorical domain, and can thus bear selectively on the properties of the percept that are distinctive. From a phenomenological viewpoint, discrimination is what brings about the overt identification of a percept.

BREAKING UP THE HOLISTIC FABRIC OF REALITY

The segmentation of the perceptual field into individual objects is thus the result of perceptual differentiation, and not the objective state of affairs that perception would merely seek to detect and acknowledge. In this sense, microgenesis is the process that breaks up the holistic fabric of reality into variably differentiated yet *meaningful* objects, beings and relations. From Aristotle to Poincaré and Thom, scores of philosophers and mathematicians have speculated about the ontological precedence of continuum over discrete structures, and suggested that individuated forms are created by breaking up the continuous fabric of reality, and not the other way around. From the microgenetic viewpoint, we may invoke *genetic*[6] precedence of continuum over discrete structures, where categorization and dynamic thematization act as organizing principles in breaking up the continuous fabric of reality into individuated forms.

The idea of the genetic precedence of holistic fabric over individuated forms in the course of perceptual differentiation runs counter to standard cognitivist theories where form perception is basically viewed as a reconstruction from components (or elementary features), followed by the projection of the reconstructed object onto the internal screen of the mind (*i.e.* representation). It bears noting that the idea of a projection onto a mental screen is phenomenologically vacuous (*i.e.* provides no explanation of perceptual *ex-*

perience) and smacks of a homunculus (it takes a homunculus to contemplate the screen). Yet, since phenomenological issues seldom preoccupy the proponents of cognitivism, and they bluntly dismiss any alternative perspective – confident as they are that their theoretical stance will ultimately be corroborated by neurophysiological and psychological evidence – we may provisionally embrace their concerns and examine the issue of elementary features. Because any real form can be decomposed into a countless number of features, and what makes a useful feature in one case may have little utility in another, one may wonder how the perceptual system is able to pick in advance the useful features of an as yet unreconstructed form. A way out of this problem might be to suggest the existence of a finite set of generic features that could be made use of in the (re)construction of any possible form. But then there would be tremendous differences with respect to the ease with which various forms are reconstructed, and the task may even turn out to be impracticable in the absence of an organizing principle, which, again, would have to be known in advance. Clearly, the proposition that perception is based on reconstruction from elementary components raises more problems than it may be expected to solve.

PRESENT-TIME EXPERIENCE

Optics, acoustics, chemistry, topology, as well as technological metaphors of photography, motion pictures, television or recording devices have, during the past century, greatly inspired scientific theorizing on perception. In their physicalistic fervor, generations of psychologists and neuroscientists alike somehow lost sight of the very phenomenological character of reality, let alone the necessity of explaining why present-time experience has continuity and depth. Why is it that what occurs in present-time is not infinitely brief, that experience does not consist of a kaleidoscopic succession of disconnected instants but has consistency and duration? Bergson, Husserl, and Merleau-Ponty, to mention the most outstanding authors, have penetratingly described and analyzed the issue of a non-evanescent present, of which my own description would be a pale rendition. Let me underscore, nevertheless, that perception critically involves an enduring and consistent *presence in experience*. This presence signifies that there is a continuous structure to experience, or more properly, a continuous forward-oriented dynamics, so that the present-time is neither infinitely brief nor evanescent, but has depth (or thickness) and consistency stretching dynamically from its immediate predecessor to its anticipated successor. To use Husserl's terminology, the now has retentions and protentions. Perception theorists who keep on brushing aside this continuous forward-oriented dynamics of present-time experience can be likened to conscientious parents who throw their baby out with the bathwater. Even if one were to regard the perceptual field as a kind of external memory – to quote a recent theory (O'Regan, 1992) – where any part of the field is kept available for further inspection – this very availability critically depends on the continuous dynamics of a forward-oriented stretch of present-time. Were this not so, the issue of availability for further inspection would be pointless as, at each and every instant, the perceptual process would have to start anew. Whatever is present in experience is so by virtue of a process that dynamically extends in time. This presence in experience is by no means illusory, if by illusory one implies something unreal, because the only reality available to us is the one we experience.

DUAL DYNAMICS OF TIME: UNFOLDING AND DEPLOYMENT

It is my suggestion that microgenetic theory provides an adequate framework for the explanation of this dynamic process. To show how, I shall first introduce the concept of *autochrony* which refers to the internal, unidirectional (*i.e.* forward-oriented), self-generated time characteristic of the living process (Rosenthal, 1993). This self-generated, internal time, which determines the flow of the living process and is proper to each living species, has a phenomenological and biological meaning. It confers on temporal dynamics its intrinsic direction as well as the periodicity specific to each species[7]. It is at the very heart of the autonomy of action and provides the latter with its driving impulse. Note, indeed, that in order to be autonomous, rather than merely reactive, an action has to be self-generated. Yet, if we are to account for the continuous forward-oriented dynamics of present-time experience, what is further required is the idea of a dual dynamics of microgenetic development, one of *unfolding* and one of *deployment*[8]. Experience has consistency and duration because it has a developmental history, a history that diachronically deploys and unfolds. Unfolding refers to the developmental succession of intermediate phases of ongoing experience, whereas deployment designates the fact that a figure has temporal extension, the time it takes to deploy in experience. This dual dynamics of microgenetic development, whereby experience gradually unfolds through differentiation and the deployment of intermediate figures, and where successive deployments tend to occult their predecessors but not the very sense of developmental history, confers on present-time experience its temporal depth and consistency. There is thus depth and consistency in the present-time because we sense the developmental history of ongoing experience without being able, at least usually[9], to evoke its intermediate deployments, as they are occulted by the current occurrence of the present.

The cohesiveness of gradually unfolding present-time experience depends also on the *anticipatory* and *categorial* character of microgenetic development. Categorization allows for the continuity of form identity throughout perceptual development, giving it cohesiveness and stability. Anticipation, which should not be mistaken for the effective expectation of definite objects or states of affairs, designates a protensive readiness for action: we actively anticipate and seek meaningful structures and immediately categorize them in view of prospective action.

GRADUAL CHARACTER OF IMMEDIATE EXPERIENCE

The hidden, gradual character of immediate experience attracted considerable attention on the part of the founders of microgenetic theory. The method of *genetic realization*

(*Aktualgenese*) was actually developed by Sander in order to externalize the course of microgenetic development by artificially eliciting 'primitive' (i.e. developmentally early) responses which are normally occulted by the final experience (Sander, 1930; Werner, 1956). In the field of visual perception, this was accomplished by repeatedly presenting very brief, poorly lit or miniature stimuli, and gradually increasing exposure time, improving lighting or letting the stimulus grow to 'normal' size. The subjects or, more appropriately, the *observers* were invited to describe what they perceived and felt as the experiment unfolded. Sander provided minute descriptions of these 'primitive' responses, observing that "the emergent perceptual constructs are by no means mere imperfect or vague versions of the final figure (…) but characteristic metamorphoses with qualitative individuality, 'preformulations' (*Vorgestalten*)" (ibid, p. 193). He noted that in the course of an unfolding perception, the development does not amount to a steady, progressive improvement whereby each successive deployment is a more elaborate version of its predecessor that comes closer to the final percept. Rather, the development observed in *Aktualgenese* exhibits the characteristic structural dynamics at work in perception. "The formation of the successive stages, which usually emanate one from the other by sudden jerks, has a certain shading of non-finality; the intermediaries lack the relative stability and composure of the final forms; they are restless, agitated, and full of tensions, as though in a plastic state of becoming." Moreover, "this structural dynamics, which (…)[is] one of the determining factors in the process of perception itself, enters our immediate experience in the form of certain dynamic qualities of the total 'state of mind', in emotive qualitative tonalities" (p. 194).

The structural dynamics at work in an unfolding perception generates intense emotional involvement on the part of the experiencer. The perceptual development, artificially externalized by the method of *Aktualgenese*, is not something the observer follows with cool objectivity and detachment, but "all metamorphoses are engulfed in a[n]… emotional process of pronouncedly impulsive and tensor nature, and take place through an intense participation of the whole human organism" (p. 194). There is an 'inner urge' for 'formation of the ill formed' and for meaningfulness. The intermediate deployments are thus experienced with a 'peculiar feeling-tone' correlated with the instability and non-finality of a given occurrence and are animated by the dynamics of what Sander's gestaltist counterparts called *Prägnanz* (the 'urge' for symmetry, regularity, homogeneity, simplicity, stability…). The emotional involvement observed in *genetic realization*, which could be viewed as excessive in regard to an unremarkable object of actual perception, can nevertheless be experienced under 'normal' conditions. A picture hanging crooked on the wall can become unbearable and can literally shriek to be set straight.

Werner gave a markedly similar description of these structural dynamics at work in microgenetic development and of the intermediate deployments that are occulted by the final experience, placing an emphasis on total bodily feeling, emotional-kinesthetic dynamics and action-like inner gestures. But he was more concerned with the semantic aspects of microgenesis and specified the characteristics of meaning stabilization and differentiation (Werner, 1930; Werner, 1956; Werner & Kaplan, 1956; Werner & Kaplan, 1963). In particular, he noted an early emergence of the general sphere of meaning, and described the developmental dynamics of meaning structure as characterized by sphere-like deployments, where gradual differentiation is not necessarily accomplished by contracting the semantic sphere but also involves shifts of the 'center of gravity'. Thus a target 'cigar' may at one point elicit the "primitive" response 'smoke', at another, 'cancer'. Some of the most interesting observations he made stemmed from neuropsychology where pathological behavior due to brain damage was described as an arrest of the microgenetic process at an earlier stage of development so that patient's responses took the form of unfinished 'products' which would normally undergo further development (see Conrad, 1954; Semenza, Bisiacchi, & Rosenthal, 1988; Werner, 1956).

THEMATIC ORGANIZATION OF CONSCIOUS EXPERIENCE

The foregoing descriptions of microgenetic experiments shed light on the non-unitary and gradual character of conscious experience. For one thing, the thematic organization of conscious experience does not amount to mere contrastive juxtaposition where the theme (focal figure) is granted focal awareness and the ground is phenomenologically empty. Background objects are not speechless; they hang together with the theme as a sort of supportive frame, yet each brings in its intentional horizon and thereby constitutes a potential landmark for alternative thematic organizations. Moreover, thematic organization is not inherent to the field and is largely dependent on the subject's engagement in action; accordingly, access to phenomenal sensations depends on this engagement in action. Thus, for instance, physically the same perceptual context can give rise to different reports depending on the type of action in which the subject is involved (see e.g. Marcel, 1993). Finally, although the history of a microgenetic development is usually obscured by the final deployment, both the *Aktualgenese* experiments and the elusive fading impressions of intermediate deployments we sometimes have (and which are not necessarily imperfect versions of the final figure) suggest that conscious experience develops gradually and that the organization of the thematic field undergoes successive adjustments. These dynamic characteristics of conscious experience bear witness to the importance of the concept of structural instability for the theory of immediate experience.

PHYSIOGNOMIC CHARACTER OF PERCEPTION

The overall dynamic structure of microgenetic development may also account for the *physiognomic* character of perceptual experience. *Physiognomic* means here that we perceive objects as "directly expressing an inner form of life" (Werner, 1957, p. 69), that is, in the same manner in which we experience physiognomies, facial expressions, gestures, or, more generally, acts of living beings. Following this line, perceived forms are not static morphological configurations but dynamic deployments, where the overall 'dynamic tone' is part and parcel of the experienced percept[10]. Accordingly, all perceived objects, whatever their nature, partake of physiognomic qualities. The physiognomic character of perception has been extensively discussed by Werner and also by Köhler (1938; 1947) and Arnheim (1954; 1969), but the idea of physiognomic perception also prompted a series of misunderstandings, most notably on the part of Gibson (1979) and Fodor (1964). It should be observed that two aspects of physiognomic perception must be taken into account. The first concerns the *expressive* character of percepts, and the second the *conative* dimension of perception whereby the readiness for action imbedded in perceptual experience 'urges' us to act upon, or use, perceived objects (see also the concept of gerundival perception in Lambie & Marcel, 2002). Gibson's concept of *affordance*, an anglicized version of Kurt Lewin's *Aufforderungscharakter* (invitation character), is partly grounded on this latter idea, though it doesn't convey the sense of an urge to act but merely invokes an invitation. This urge to action is most readily observed in the behavior of children, in so-called 'primitive peoples', and under the influence of certain drugs (Werner, 1957). Brain pathology gives an amazing example of the conative character of perception in so-called *utilization behavior* where the patient cannot but use whatever object he or she happens to come across (Lhermitte, 1983; Shallice, Burgess, Schon, & Baxter, 1989). As Lhermitte observed, for the patient, the perception of an object implies the order to grasp and use the object. As the above example suggests, this pathological behavior is by no means an aberrant creation of pathology, but an expression of the readiness for action imbedded in perceptual dynamics. In the social context of Western Societies, this readiness for action does not necessarily prompt effective object manipulation, at least in adult behavior, but in the context of certain brain lesions enactment may become irresistible.

The expressive character of perception is obviously no less imbedded in perceptual dynamics. As Köhler and Arnheim cogently argued, the expressivity of the perceived world is *directly* experienced by the perceivers and does not result from empathic projection or from perceived analogy with their own past expressions and feelings. For one thing, we cannot simultaneously be external observers and the experiencers of our own interiority and exteriority. How could we then acquire the dual knowledge that would serve as the basis for an analogy? Second, the analogy could only hold between comparable entities or configurations; yet when we perceive a *sad tree*, a *cheerful landscape*, or the *lovely* face of Dorothée, this can hardly be due to the knowledge of our own expressions of sadness and cheerfulness, or, for the present writer, of his own loveliness. Clearly, an indirect principle, whereby perceived morphologies or dynamic configurations (*e.g.* facial expressions, gestures) are subsequently interpreted by analogy or empathic projection, can hardly count as a satisfactory explanation of the expressivity of the perceived world. On the contrary, expressivity constitutes a forceful illustration of the dynamic principle at work in perceptual development whereby even the morphology of static forms is grounded in the configural dynamics of the deployment of the percept[11].

The acknowledgement of the physiognomic character of perception shouldn't be naively interpreted to suggest that our everyday perception is overflowing with an expressive world where objects and landscapes are animated by inner life. As adult members of Western Societies we certainly do not find ourselves overwhelmed by the expressivity of the perceived world – we generally barely pay attention to it – and many people will even be reluctant to admit that their perceptual experience has an expressive flavor. The expressivity of the perceived world is clearly at odds with the matter-of-fact style of our social world. In a labor-oriented society – where activity is governed by non-immediate goals and where, as Benny Shanon noted[12], *voluntary ignorance* of a good deal of what we are otherwise able to perceive but what falls outside the tacitly agreed upon terms of relationships, is one of the founding components of interpersonal relations – physiognomic impressions normally recede to the background and form, at best, an elusive feeling tone. The perception of inanimate objects is no less affected by the style of our social world, for we belong to social world even when alone. Yet, reports of physiognomic perception abound in child psychology, ethnopsychology and clinical psychology. Children, so-called 'primitive peoples' and, for instance, certain schizophrenics manifest in their behavior clearly identifiable reactions to the perceived expressive character of objects. Moreover, the ease and the naturalness with which we are receptive to expressivity in literature, painting or music would remain inexplicable were we not to assume that this receptiveness builds upon a disposition that was 'already there'. Clearly, these observations testify to expressivity in perception. The acknowledgement of the physiognomic character of perception brings us closer to a scientific explanation of the origin of *esthetic* and *ethical* attitudes. Although for many students of cognition this issue is secondary or falls beyond the scope of a scientific endeavor, I submit that the inability of cognitivist theories to account for the origin of esthetic and ethical attitudes, their failure to even perceive the fundamental status of esthetics and ethics in regard to human cognition, constitute some of their major shortcomings. It is certainly not irrelevant that Gaetano Kanizsa, whose work inspired this collection of essays and whose phenomenological orientation resolutely opposed cognitivist approaches to perception, was deeply concerned with perception's esthetic character as well as being an accomplished painter.

GENETIC PHENOMENOLOGICAL SCIENCE OF COGNITION

The basic constructs of microgenetic theory outlined so far may be viewed as landmarks for a genetic phenomenological science of cognition. A reader familiar with Gestalt theory will easily recognize in this overview the legacy of Wolfgang Köhler, in particular his idea of stabilization in dynamic system and his concepts of value and expressivity in perception. These ideas are, however, reformulated to take into account temporal dynamics so as to be able to define cognitive process in terms of a dynamic development characterized by gradual differentiation and deployments, variable stabilization as well as unfolding and thematic focalization. This micro-development has an anticipatory and categorial character, to which, strangely enough, the gestaltists paid little attention.

It should be stressed that the phenomenological character of microgenetic theory does not prevent it from being amenable to evaluation by the methods of natural science. Much as the original theory of Werner has built on ample experimental evidence, various specific postulates of the present theory may be subjected to experimental evaluation in psychology and neuroscience where, incidentally, a variety of tools to probe brain dynamics have recently been devised. Note that experimental evaluation *does not* require that phenomenology be naturalized. The very idea of an *experimental phenomenology* is precisely to bring a naïve openness on the part of the subject, uncontaminated by formal knowledge, to the experimental situation, with its precise physical measures and control of experimental variables. Similarly, a comparable naïve openness is required on the part of the scientist whose genuine questioning free of conceptual prejudice is the only way to 'get in touch' with the original reality he seeks to describe, and, as such, is the necessary counterpart of his otherwise naturalistic stance (see Bozzi, 1989; Rosenthal & Visetti, 1999; Vicario, 1993). In the next section, I shall briefly review neurophysiological, neuroanatomical, and eye-movement data and suggest that the bulk of available evidence is basically consistent with the notion of a global-to-local developmental dynamics in visual perception and the idea of the gradual differentiation of the percept. However, before I turn to this data, I should like to point out that some of the most promising developments for microgenetic theory in the field of cognitive science might be sought in the use of the modern mathematical and physical concepts of instability, and in the application of the theory of complex systems in *modeling* the dynamics of microgenetic differentiation (see also Visetti, this volume).

EARLY STRUCTURE IN VISION

There is a predilection among many vision scientists for the traditional atomistic explanation of visual perception according to which the putative percept is reconstructed at the level of the higher cortical structures from unstructured mosaic of elementary sensations that are produced on the retina and dispatched via retinofugal pathways to these cortical structures[13]. I shall argue, to the contrary, that the anatomical and physiological studies of the retinofugal pathways in primates support the proposition that considerable structure emerges already at the lowest levels of visual processes, and that these studies lend credence to the idea of holistic precedence, as well as, indirectly, to the overall schema of global-to-local structure of visual processes involving early categorization.

Retinal projections to the cerebral cortex are dominated by two major pathways, the magnocellular (M) and parvocellular (P) systems, which are relayed by the magnocellular and parvocellular subdivisions of the lateral geniculate nucleus (see Merigan & Maunsell, 1993; Shapley & Perry, 1986). The M ganglion cells have large soma, with extensive dendritic trees and large axons, whereas the P ganglion cells have smaller soma, small dendritic arbors and medium-size axons (see Leventhal, Rodieck, & Dreher, 1981). It is important to note that the conduction velocity of M cells is greater than that of P cells due to the larger axonal diameter of M cells. Moreover, the M cells have large receptive fields, rapid temporal dynamics, and are more sensitive to low spatial frequencies. This system is sensitive to the coarse spatial distribution essential to the differentiation of basic form and for figure-ground segregation. The P system, which has smaller receptive fields and is more sensitive to higher spatial frequencies, samples the retinal image with higher resolution that is relevant to local spatial detail and color. There seems to be a division of labor between the two systems such that the M system quickly processes coarse form and the P system subsequently specializes in fine detail and color. The two systems thus sense different but overlapping portions of visible spatial and temporal frequencies (Livingstone & Hubel, 1988). In a word, retinal 'sensations' are processed twice in a nonredundant fashion and each time using 'data' in a different format. Arguably, the M system provides a quick primal glimpse of the visual field, supplying sufficient structural information about gross spatial discontinuities and their position in the field to guide the processing of the P system. This enables the oculomotor system[14] to adjust gaze position and generates dynamic displacement thereby creating spatio-temporal discontinuities to which the M system is also sensitive (see Lehmkuhle, 1993; Merigan & Maunsell, 1993). Two important points emerge from this description. (1) Since the magnocellular system has temporal precedence, the earliest processes in vision are necessarily global and coarse-grained. (2) As the M ganglion cells have large receptive fields and are sensitive to coarse spatial distribution, displacement and temporal dynamics, there are reasons to believe that the M system segments the visual field on the basis of gross spatial and temporal discontinuities and of their joint displacement[15]. These observations strongly favor the proposition that the 'stimulus' brought by the magnocellular projection in V1 (striate cortex) already has considerable structure.

The notions of a global-to-local developmental dynamics in visual perception and of an early categorization of visual forms will, however, best be evaluated by combining the foregoing anatomical and physiological considerations with evidence from eye movement studies. It should be noted that the magnocellular system is mostly involved in extrafoveal (both parafoveal and peripheral) vision whose definition is insufficient for local detail, that it presumably exerts control on eye movements, and that foveal fixations (necessary for the exploration of local detail) are highly selective and cover only a small part of the visual field, mainly the *figure* (see O'Regan & Noë, 2002; Underwood, 1998; Yarbus, 1967). This selectivity is obviously inconsistent with the 'mosaic theory', at least as far as the whole visual field is concerned, for how can the visual system reconstruct the whole field when over 80% of its 'elementary components' are unavailable. But selectivity is also interesting for other reasons. In order to act selectively a system has to have prior 'knowledge' on which to base the selection. In this case, the system has to spot the figure first and, then, adjust the gaze so as to fixate parts of this figure. Now, several characteristics of what makes up a figure need to be recalled: (a) it is a form, (b) it is necessarily meaningful, and (c) it has thematic prominence with respect to the rest of the field. But how can this come about were we first to construct abstract forms out of meaningless features, only to discover later their identity and meaning? Obviously, such a form could not first be (re)constructed out of the mosaic of its local meaningless components, and then targeted for central fixation, because local components can only be explored when fixated in central vision. Incidentally, this observation lends further support to the above proposal that basic form emerges in early coarse vision. But since the form in question is also a figure standing in the field, it comes accompanied by its semantic and thematic extensions. One can hardly explain how a figure can be spotted in early coarse vision if visual perception did not have a directly categorial and anticipatory character. Moreover, since early vision can at best provide a *raw sketch* of the figure, which may then be further explored in central vision, there are reasons to assume that the postulated schema of global-to-local coarse-grained-to-fine-grained differentiation in visual perception rests on firm grounds.

SUBJECTIVE FIGURES

Various phenomena of perceptual completion, whether figures, surfaces or regions, provide an interesting illustration of microgenetic dynamics at work in perception. Consider the famous example of the Kanizsa square where a collinear arrangement of edges of four white 'pacmen' (inducers) on a black background gives rise to the perception of a black square whose area appears slightly darker than the background. In addition, the surface of the square appears to the observer to be in front of four disks that it partly occludes. Since the square is perceived in spite of the absence of corresponding luminance changes (*i.e*. forming complete boundaries), and thus does not reflect any real distal object, it can only be created by the visual system which purportedly completes, closes, and fills in the surfaces between 'fragments', so as to make the resulting 'subjective' region emerge as figure standing in the ground. Yet, as Kanizsa (1976; 1979) aptly showed, this and other examples of so-called subjective contours demonstrate the basic validity of Gestalt principles of field organization, in particular of its figure/ground structure and of *Prägnanz*, whereby incomplete fragments are, upon completion, transformed into simpler, stable and regular figures. Although this phenomenon is often described in terms of contour completion, it clearly demonstrates a *figural* effect, whereby the visual system imposes a figural organization of the field (and hence figure completion), and where *the contour results from perceiving a surface*, not the other way around, again as Kanizsa suggested. Moreover, these *subjective figures* illustrate the categorial and anticipatory character of microgenetic development, such that the perceptual system anticipates and actively seeks meaningful structures and immediately categorizes them on a global dynamic basis[16]. The crucial role of meaningfulness is demonstrated by the fact that no subjective figures arise in perception when the spatial arrangement of inducers does not approximate a 'sensible form' or when the inducers are themselves meaningful (*viz*. complete) forms[17].

What makes these subjective figures even more valuable for the present discussion is that they may be viewed as an instantiation of *early structure* and of *holistic precedence* in visual development. In recent years, there has been considerable debate in vision science concerning the neural mechanism underlying perceptual filling-in and several researchers have claimed to have identified subpopulations of cortical cells specialized in various aspects of perceptual completion (see e.g. Lesher, 1995; and Pessoa, Thompson, & Noë, 1998, for a review and critical discussion). One problem with these postulates of low-level filling-in neural mechanisms is the frequent confusion between what pertains to receptive field dynamics related to blind spot or scotomata and what pertains to the perception of genuine subjective figures. Another problem is that finding neurons whose behavior correlates with the perception of subjective figures does not imply that these neurons are actually responsible or even used for perceptual completion. The next problem is that no subpopulation of specialized cells can account for the fact that subjective figures are always sensible meaningful forms. Finally, many studies have tended to overstress the importance of contour (which as Kanizsa showed is secondary to surface perception) and thus assumed a critical role for collinear alignment of edge inducers when actually such alignments are not a necessary condition (as the Sambin/Kanizsa cross examples demonstrate, see Figure 1 below, and Kanizsa, 1976). It is important to note in this respect that there is presently a considerable body of neurophysiological and neuropsychological evidence supporting the idea that surface formation and completion, involving context-dependent figure/ground segregation, occurs very *early* in the course of vision and on *global* basis (Davis & Driver, 1994; Lamme, 1995; Mattingley, Davis, & Driver, 1997). This evidence confirms Kanizsa's results and further corroborates the microgenetic postulates of the dynamic, directly categorial (*viz*. meaning-laden) and anticipatory character of field organization. Although many scientists among the neuroscience intelligentsia continue to favor a modern version of the helmholtzian doctrine according to which the percept (here the subjective figure) is reconstructed at the level of 'sentient' higher cortical structures from an unstructured mosaic of elementary sensations processed by specialized local detectors, I submit that the above examples and discussion provide powerful arguments in support of the microgenetic theory of perceptual development outlined in this essay.

THE MICROGENESIS OF VISUAL PROCESSES IN READING

I shall turn now to a specific illustration of certain principles of microgenetic theory in the field of reading. I have chosen reading because it is a peculiar skill. It takes both language and perception to become a reader, yet language and perception won't suffice; some people never become proficient readers, and a brain lesion can disrupt reading skills in subjects otherwise showing no defect in object perception and spoken language. The persistence of oral civilizations and of nonliterate societies further teaches us that not any form of social world is appropriate for the advent of literacy. Moreover, the passage from nonliterate to literate society deeply alters syntax, vocabulary and language use, as well as the mnemonic and cognitive practices of society members, and, ultimately, the society itself. At the same time, reading is interesting for our purpose as it handily lends itself to the evaluation of the postulates of immediate categorization and meaningfulness, selectivity, and the global-to-local structure of perceptual development.

Language and perception are unsettling accomplices of literacy. Their relationship involves a kind of co-determinism where it is difficult to regard written language as a sim-

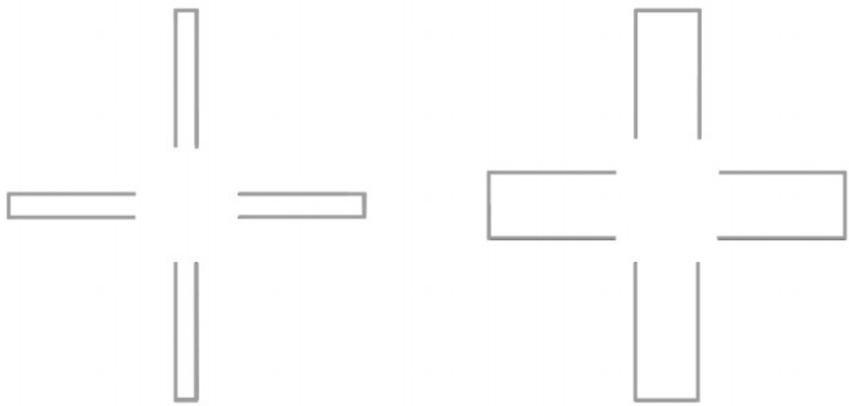

Figure 1. Sambin-Kanizsa cross examples. Note that the subjective surface in the center of the thin cross (left) has the form of a circle whereas the subjective surface in the center of the thick cross has the form of a square.

ple externalization of information, lying ‘out there’ and waiting to be identified and reinternalized. Wertheimer once observed that for a proficient literate individual it is neither necessary to identify each individual letter nor to overtly recognize every word while reading a text. In 1913, he and Pötzl reported on an alexic patient who had “lost the ability to perceive words as gestalts”. In spite of his preserved capacity to recognize individual letters, the patient was practically unable to read words; his preserved ability to identify component letters (elementary segments) along with the inability to recognize words (functional wholes) – which, incidentally, proved to be unresponsive to training – constituted in Wertheimer’s view an illustration of a gestalt organization in reading.

Although Wertheimer did not elaborate any further on this organization, and his observations remained quite general, he clearly alluded to the difficulty that would confront a theory of perceptual processes in reading stated in terms of (mechanical) unit identification and conversion. On the one hand, typical silent reading (in orthographic writing systems) can neither be characterized as literal (not all letters are identified) nor as purely holistic (letters still matter for reading, and not all words are overtly recognized). On the other hand, many letters and words are typically left unidentified in the course of proficient reading[18]. It is thus patent that the metaphors of unit (whether letters or whole words) identification and conversion, which fed the century-long debate between the proponents of letter-by-letter or direct whole word recognition in reading, are unenlightening[19]. For the contents of the *perceptual experience* that underlies reading are not ‘out there’ on a sheet of paper waiting to be detected and internalized (in the form of mental representation). Since, on the one hand, the contents of experience in reading are not ‘out there’ waiting identification, internalization or conversion, and, on the other hand, a text represents a highly elaborated yet very compact material for experience, reading may serve as a living small-scale model of immediate experience illustrating the various issues relevant to microgenesis reviewed earlier in this essay.

An interesting source of insights into reading comes from so-called 'deep dyslexia', a reading pathology due to brain damage where patients are characteristically unable to identify letters and read aloud pronounceable nonwords. They preserve nevertheless the ability to read words, though to a variable degree: nouns, verbs and adjectives are read best whereas conjunctions and articles are seldom read aloud. In reading content words, they quite often make semantic paralexias (e.g. reading 'priest' for 'church') and sometimes, but rather infrequently, visual (e.g. 'deep' for 'deer') or derivational (e.g. 'registered' for 'register') errors (Coltheart, Patterson, & Marshall, 1980). One thing that is striking about observations on deep dyslexia is that they illustrate a condition in which perceptual-morphological, syntactic and semantic aspects of reading are all interwoven. Note, for instance, that in order to read 'priest' for 'church', words that visually have almost nothing in common, some perceptual-morphological processing of the printed word 'church' is necessary. This processing must, however, be insufficient for the overt identification of the target word, yet it must be sufficient to hit upon the sphere of meaning relevant to 'church' so as to allow the patient to respond using the word 'priest'. How could this occur where meaning and form alien to one another? On the other hand, it should also be borne in mind that, were the target a function word, chances are that a deep dyslexic patient would not be able to read it out loud. But how can he know that the target is a function word in so far as he is unable to overtly identify it? Clearly, in this example, form, meaning and function cannot be independent and mobilize processes that are intrinsically alien to one another.

The above example is also remarkably reminiscent of observations described by Werner (1956) and Conrad (1954) in which pathological behavior due to brain damage was presented as an arrest of the microgenetic process at an early stage of development, thereby letting occur unfinished 'products' in patients' behavior, that would normally undergo further development. Moreover, an examination of patients' semantic 'errors' produced for the same target word, whether in the same reading session or in different sessions, shows the same character of instability, sphere-like deployments and shifts of 'center of gravity' as those described by Werner with respect to *Aktualgenese* experiments conducted with normal subjects.

In a series of experiments undertaken recently in my laboratory we sought to further evaluate the postulates of immediate categorization and meaningfulness, selectivity, and global-to-local structure of microgenetic development in reading. These experiments were mainly intended to probe the structure of visual processes in reading but since reading normally applies to meaningful texts, other issues related to text interpretation, gradual development of meaning, and meaning and form relationship arose as well. In particular, we sought to evaluate the general hypothesis of a *global-to-local structure* of visual processes in reading by picking a specific instantiation of this hypothesis in terms of the *selective* processing of component letters depending on their orthographically discriminative character. The basic idea underlying this was the following: the selective processing of component letters that depends on their orthographically discriminative character presupposes prior identification of letter slots which are ambiguous and thus need discrimination. However, in order to determine which letter slots are ambiguous, it is first necessary to gather at least some 'knowledge' about the class of word shapes to which the target word belongs. Only a global word-shape-based process can bring about such 'knowledge'. Moreover, the very existence of letter processing after a prior word-shape preview implies that the latter global process is too coarse-grained or otherwise insufficient for word identification[20]. In this sense, the idea of the selective processing of component letters depending on their orthographic discriminativity lets us evaluate the underlying hypothesis of the global-to-local structure of perceptual differentiation in reading.

A few words of clarification may be needed here. What defines the ambiguous or critical *letter slots* in a word is the existence of its *orthographic shapemates*, *i.e.* other words sharing the same *global shape* (global word-form irrespective of internal letter features) but which differ locally with respect to component letters that occupy these slots. Of course, the letters in question are of similar stature (ascender to ascender, descender to descender, etc…); otherwise the words would not share the same global shape or be similar in regard to global word-forms. Because the primary global process is assumed to be coarse-grained, *orthographic similarity* is defined by the similarity of global word-forms at the level of spatial resolution which is insensitive to internal letter features. It is only in a later phase, and if the reader seeks local discrimination, that the visual system becomes sensitive to fine-level internal features. Thus, for instance, the fifth letter slot (*r*) in the French word 'effarer' is ambiguous due to the existence of another word 'effacer' which shares with the former the same global shape (it is its shapemate) and the two words therefore can only be distinguished from one another by checking locally the fifth letter slot (*r* vs. *c*). On the other hand, there is no ambiguous letter slot in the French word 'migraine' because no other French word shares its global shape. It can then be said that the fifth letter 'r' in 'effarer' is discriminative and hence critical for the identification of this word, whereas the 'r' in 'migraine' is noncritical because 'migraine' has no shapemates from which it would have to be distinguished.

The experiments which were conducted in order to evaluate the hypothesis of the selective processing of component letters were based on the letter cancellation technique or on the analyses of eye movements.

The letter cancellation technique (Corcoran, 1966; Healy, 1994) requires subjects to cross or circle each instance of a specific letter while reading a text for comprehension. It has been used to study the issue of perceptual units in reading (Drewnowski & Healy, 1977; Hadley & Healy, 1991; Healy, 1976) in relation to the effect of linguistic function (e.g. content vs. function words) on letter detectability (Greenberg & Koriat, 1991; Koriat & Greenberg, 1991; Koriat & Greenberg, 1994), and in relation to the phonological status (e.g. pronounced vs. silent) of component letters (Corcoran, 1966). Studies based on this technique have shown that subjects always miss a certain amount of target letters while reading real text and that the rate of omission depends on certain parameters. For instance, letters in function words are more often missed than letters in content words

(Greenberg & Koriat, 1991), silent letters remain undetected more often than pronounced letters (Corcoran, 1966) and word meaning may also influence letter detection (Moravcsik & Healy, 1995). Since the misdetections of target letters reported in these studies are quite systematic and take place in spite of efforts to detect all instances of these letters, it is assumed that they are indicative of the characteristics of the reading process.

In all experiments based on letter cancellation, the critical comparison was obtained by contrasting words in which the substitution of the target letter (most often 's' and 'r') creates at least one orthographic shapemate, and words where no substitution of the target letter makes an existing word. The main prediction of these experiments was that because discriminative letter slots are likely to be targeted for local verification and thereby come to the center of local attention, the detection rate of target letters, which are critical to shapemate word differentiation, should be substantially higher than that of non-critical target letters, or, alternatively, that subjects should miss many more non-critical targets than critical ones. The main result of these experiments was that component letters that differentiate orthographic shapemates are better detected than letters that are in unambiguous slots (which give rise to twice as many detection errors). This *critical-letter effect* was obtained on five different passages of prose, as well as on meaningless scrambled assemblies of words. Moreover, it was found to criterially depend on orthographic similarity: the effect did not occur when orthographically legal letter substitutions altered word shape (e.g. 'mé**r**ite' vs. 'mé**d**ite'). These findings unequivocally corroborate the general idea that local letter-level analyses are pretuned by an earlier global process and thus they lend support to the hypothesis of the global-to-local structure of perceptual differentiation in reading[21].

Experiments based on the analyses of the eye fixations of subjects reading various types of text provided independent evidence of the above critical-letter effect and brought additional insights into the structure of visual and interpretive processes in reading that are relevant here. First, the results substantiated the hypothesis of the orthographic determinants of fixation locations in words by showing a systematic relationship between the distribution of fixation locations and the presence or absence of orthographically discriminative letters: eye fixations tended to land on the area of discriminative letters in words that have orthographic shapemates and to spread over the body of words with unambiguous shapes. Second, these results showed that the presence or absence of orthographically discriminative information does not affect the probability of fixating a (content) word: readers fixated just as much words that have orthographic shapemates and words that have unambiguous shapes. Third, the results showed that while reading normal two-page texts, subjects centrally fixated only 44% of the words.

The finding that an orthographically ambiguous word (*i.e.* word having shapemates) will not necessarily be fixated shows that the 'decision' of whether to fixate a word is not governed merely by orthographic considerations (e.g. the search to explicitly identify words) but by the ongoing process of text comprehension (see also Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Rayner & Well, 1996). This conclusion, along with the consideration that less than 50% of words were centrally fixated in our experiments, suggests that parafoveally-gained information may be sufficiently meaningful for it to take part in text interpretation. This is consistent with the proposition that meaning goes hand-in-hand with perceptual categorization, developing along the same lines from general to specific, from relatively vague and global to articulate, precise, and local.

The finding that readers fixate not only words that have ambiguous shapes but also words that are unambiguous – whether because they have orthographically unique shapes and/or due to strong contextual evidence – suggests that definite word identification does not take place in parafoveal vision. We may thus ask the following question: if only words gaining foveal fixation are explicitly identified, how is it that subjects can skip more than 50% of words and still properly understand a text, as is shown by their ability to correctly answer questions about the content of what they have read? It is noteworthy that in our experiments this skipping very often concerned content words – one out of three 7-10 letter words (mainly content words) were not fixated by our subjects – and cannot therefore be attributed to a word class effect (e.g. certain function words being selectively skipped because they are highly predictable on syntactic grounds). Clearly, these results indicate that text comprehension in reading does not require explicit identification of *all* component words. This is not to suggest that the meaning of words that are not explicitly identified is simply ignored. Although parafoveal inspection does not allow for explicit word identification it does appear to feed the ongoing text comprehension process with adequate information (see also Lavigne, Vitu, & d'Ydewalle, 2000). This information may only be partial or incomplete from the point of view of a dictionary definition of word meaning, but it nevertheless appears to be contextually appropriate and sufficient for the comprehension of a given text[22]. In any case, if parafoveal inspection can both constrain word discrimination and inform the process of text comprehension, there are grounds in psychology for the concept of immediate coarse-grained categorization of (printed) forms that is directly meaning-laden (due in part to a form/meaning relationship). This is precisely what microgenetic theory stipulates.

The foregoing example is admittedly no more than a partial illustration of the application of microgenetic thinking to the research context of reading. Beyond its relevance for a theory of visual processes in reading, the primary purpose of this illustration was to show that the microgenetic theory offers a viable and productive research strategy. The subversive quality of the theory, which even on partial and fairly local application forces a deep revision of the field of reading research, shows that microgenesis is not a mere collection of local hypotheses, that it makes a coherent though as yet emergent framework for the study of lived experience. After all, the issue at stake is genetic phenomenological science of embodied cognition.

Prof. Victor Rosenthal
Centre Baul Broca
2ter rue d'Alésia
75014, Paris, France

Notes

[1] It should be underscored that microgenesis shared with the Berlin School of Gestalt theory some of its basic tenets (*e.g.* the concept of field, the idea of stabilization in a dynamic system) and its phenomenological orientation. However, it proceeded in directions neglected by the gestaltists: it focused on fine-grained temporal dynamics of psychological processes and on the categorial character of meaning and perception; postulated that perceptual experience is directly meaning-laden and intrinsically emotional, that forms are inherently semantic, and not merely morphological constructs. In contrast to the Berlin group, early work on microgenesis was highly concerned with language and language development, and with cognitive disorders due to brain damage.

[2] Action should be distinguished from mechanical reaction. What characterizes genuine action is that it implies the autonomy and spontaneity of an agent, and a knowledge of the environment in which the action will take place. Indeed, the very possibility of autonomous and spontaneous action is ipso facto a demonstration of the agent's of knowledge of the environment. To put it in terms of a phenomenology of action (and indeed also of living): doing is a basic form of knowing (see Arendt, 1958; Whitehead, 1983).

[3] The popular flowchart models (see e.g. Shallice, 1987; Shallice, 1988) where in order to acquire meaning, a semantically vacuous categorical percept has to access so called 'semantic memory', and where various 'semantic effects' are dealt with by invoking the concept of 'level of activation', do not provide a better solution to this problem. For, if semantics postdates morphology in the course of perception, and the latter is independent of the former, no room is left for the influence of meaning and value upon the size of perceived objects.

[4] In line with Gestalt tradition, the microgenetic theory assumes that perception generically instantiates the structure of cognition.

[5] Correlatively, it thus becomes understandable why cognitive and perceptual processes *are not infallible*. Although microgenesis is globally adequate for our conditions of living, its anticipatory and directly categorial character conditions its potential failures. Accordingly, the observation that cognitive, perceptual or language processes are intrinsically fallible becomes a source of insights into the structure of cognition (see Rosenthal & Bisiacchi, 1997). For instance, the obstinate resistance of 'perceptual errors' to contradictory evidence handily illustrates the 'cost' of the anticipatory and directly categorial character of microgenetic differentiation.

[6] *Genetic* refers here to the developmental dynamics of a process, not to a genome or to an adjectival use of the metaphor 'genetic program'.

[7] It goes without saying that it is not the real line that can formally represent autochronic time. Self-generation of time can only occur by fits and starts (or by pulsing) with variable periodicity.

[8] We are concerned here with the tentative explanation of the dynamics of processes in themselves. The reader should, however, be aware that the general proposal bears on the dual structure of autochronic time.

[9] We may sometimes have an elusive, fading impression of intermediate deployments which nevertheless escapes thematization however much we strive to bring it to conscious inspection.

[10] For instance, Werner noted that colors are experienced not only in terms of hue, brightness, and saturation but also in terms of being strong or weak, cool or warm; lines not only have extent and curvature, etc., but may be seen as gay or sad…

[11] Note that physiognomic perception further instantiates the value-laden character of perceptual experience which I discussed in the initial sections of this essay. The perceptual world is indeed directly invested with values by virtue of the same dynamic principles that confer 'interiority' on perceived objects and dynamic configurations and urge perceivers on to action. Accordingly, values are not indirectly associated with objects on the basis of past experience and/or rational evaluation (though of course in certain particular situations an object may be valued on the basis of rational evaluation) any more than expressive qualities are inferred by analogy.

[12] (Shanon, 1982).

[13] This idea is even presented unquestioned in recent handbooks (see e.g. Palmer, 1999). The logical difficulties with which this 'mosaic theory' is confronted are hardly mentioned.

[14] The magnocellular system appears to exert control on eye movement.

[15] Although the M cells are often described as detectors of movement, it should be borne in mind that spatial and temporal discontinuities induced by movement are the very condition for form perception. Indeed, self-induced (eye and/or head) movements are necessary for seeing, and the 'static retina', *i.e.* when eye movements are prevented or artificially compensated for, is blind (see Yarbus, 1967). Incidentally, this latter observation was anticipated by Husserl and Merleau-Ponty.

[16] The global character is obvious since the figure cannot be constructed from its components. The dynamics can be explained by the co-presence (or co-occurrence in time) of inducers (fragments) and their joint displacement upon self-induced movement (*e.g.* eye-movement).

[17] Note also that although subjective figures are often illustrated by geometrical forms, geometric regularity is unnecessary, and any sensible figure, even irregular, can arise under similar conditions. The phenomenal completion is thus not an effect of Euclidean principles encoded in the brain.

[18] Rosenthal, Parisse, and Chainay (2002) showed that subjects skip (*i.e.* do not fixate in central vision) more than 50% of words while reading regular texts.

[19] It bears noting that the so-called *interactive* solution (viz. interactive recognition of letters and whole words, see e.g. McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) is in this respect no solution at all as it presupposes an *a priori* segmentation into relevant units.

[20] The foregoing formulation should not be interpreted literally as suggesting a two-stage (first global, then local) theory of visual processes in reading, which, let it be said in passing, would be at odds with the microgenetic theory of gradual development. It simply intends to instantiate the idea of the temporal precedence of global coarse-grained over local selective and fine-grained differentiation.

[21] One may notice, on the other hand, that the necessity of differentiating words having the same global shape presupposes the prior occurrence of a process that categorizes words on the basis of their global shape. In this sense, these results corroborate the proposition that perceptual differentiation involves immediate categorization.

[22] Since overt identification of all words is not necessary for contextually appropriate text comprehension, the proportion of words being explicitly identified may vary depending on strategic attentional factors, the type of text being read, and the individual's interests and reading skills (see also the concept of the effective visual field in Marcel, 1974, and the discussion of the use of context).

REFERENCES

ARENDT, H. (1958). *The Human Condition*. Chicago: University of Chicago Press.

ARNHEIM, R. (1954). *Art and visual perception; a psychology of the creative eye*. Berkeley: University of California Press.

ARNHEIM, R. (1969). *Visual thinking*. Berkeley: University of California Press.

BALOTA, D. A., POLLATSEK, A., & RAYNER, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*(3), 364-90.

BOZZI, P. (1989). *Fenomenologia Sperimentale*. Bologna, Italy: Il Mulino.

BRAND, J. (1971). Classification without identification in visual search. *Quarterly Journal of Experimental Psychology, 23*, 178-186.

BRUNER, J. S., & GOODMAN, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal Social Psychology, 42*, 33-44.

CADIOT, P., & VISETTI, Y. M. (2001). *Pour une théorie des formes sémantiques*. Paris: Presses Universitaires de France.

CATÁN, L. (1986). The dynamic display of process: Historical development and contemporary uses of the microgenetic method. *Human Development, 29*, 252-263.

COLTHEART, M., PATTERSON, K., & MARSHALL, J. C. (Eds.). (1980). *Deep dyslexia*. London: Routledge & Kegan Paul.

CONRAD, K. (1954). New problems of aphasia. *Brain, 77*(4), 491-509.

CORCORAN, D. W. J. (1966). An acoustic factor in letter cancellation. *Nature, 210*, 658.

DAVIS, G., & DRIVER, J. (1994). Parallel detection of Kanizsa subjective figures in human visual system. *Nature, 371*, 791-793.

Drewnowski, A., & Healy, A. F. (1977). Detection errors on the and and: Evidence for reading units larger then the word. *Memory & Cognition, 5*, 636-647.

EHRLICH, S. F., & RAYNER, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641-655.

FODOR, J. A. (1964). Review of Symbol Formation by Heinz Werner and Bernard Kaplan. *Language, 40*, 566-579.

GIBSON, J. J. (1979). *The Ecological Approach to Visual Perception*. Hillsdale, N.J.: Lawrence Erlbaum.

GREENBERG, S. N., & KORIAT, A. (1991). The missing-letter effect for common function words depends on their linguistic function in the phrase. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(6), 1051-1061.

GURWITSCH, A. (1957). *Théorie du champ de la conscience*. Paris: Desclée de Brouver.

GURWITSCH, A. (1966). *Studies in Phenomenology and Psychology*. Evanston: Northwestern University Press.

HADLEY, J. A., & HEALY, A. F. (1991). When are reading units larger than the letter? Refinement of the unitization reading model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(6), 1062-1073.

HEALY, A. F. (1976). Detection errors on the word the: evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance, 2*(2), 235-242.

HEALY, A. F. (1994). Letter detection: a window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin & Review, 1*(3), 333-344.

INGLING, N. (1972). Categorization: a mechanism for rapid information processing. *Journal of Experimental Psychology, 94*, 239-243.

KANIZSA, G. (1976). Subjective contours. *Scientific American, 234*(4), 48-52.

KANIZSA, G. (1979). *Organization in Vision: Essays on Gestalt Perception*. New York: Praeger Press.

KÖHLER, W. (1938). *The Place of Value in a World of Facts*. New York: Liveright.

KÖHLER, W. (1947). *Gestalt Psychology*. (Revised Edition ed.). New York: Liveright.

KORIAT, A., & GREENBERG, S. N. (1991). Syntactic control of letter detection: Evidence from English and Hebrew non-words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(6), 1035-1050.

KORIAT, A., & GREENBERG, S. N. (1994). The extraction of phrase structure during reading: evidence from letter detection errors. *Psychonomic Bulletin & Review, 1*(3), 345-356.

LAMBIE, J. A., & MARCEL, A. J. (2002). Consciousness and the varieties of emotion experience: a theoretical framework. *Psychological Review, 109*(2), 219-259.

LAMME, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience, 15*, 1605-1615.

LAVIGNE, F., VITU, F., & D'YDEWALLE, G. (2000). The influence of semantic context on initial eye landing sites in words. *Acta Psychologica, 104*, 191-214.

LEHMKUHLE, S. (1993). Neurological basis of visual processes in reading. In D. M. Willows, R. S. Kruk, & E. Corcos (Eds.), *Visual processes in reading and reading disabilities* (pp. 77-94). Hillsdale, N.J.: Lawrence Erlbaum Associates.

LESHER, G. W. (1995). Illusory contours: Toward a neurally based perceptual theory. *Psychonomic Bulletin and Review, 2*, 279-321.

LEVENTHAL, A. G., RODIECK, R. W., & DREHER, B. (1981). Retinal ganglion cell classes in the Old World monkey: morphology and central projections. *Science, 213*(4512), 1139-42.

LHERMITTE, F. (1983). 'Utilization behaviour' and its relation to lesions of the frontal lobes. *Brain, 106*(Pt 2), 237-55.

LIVINGSTONE, M., & HUBEL, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science, 240*(4853), 740-9.

MARCEL, A. J. (1974). The effective visual field and the use of context in fast and slow readers of two ages. *British Journal of Psychology, 65*(4), 479-492.

MARCEL, A. J. (1993). Slippage in the unity of consciousness. *Ciba Foundation Symposium, 174*, 168-80.

MATTINGLEY, J. B., DAVIS, G., & DRIVER, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science, 275*, 671-674.

MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1 An account of basic findings. *Psychological Review, 88*, 375-407.

MERIGAN, W. H., & MAUNSELL, H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience, 16*, 369-402.

MORAVCSIK, J. E., & HEALY, A. F. (1995). Effect of meaning on letter detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 82-95.

O'REGAN, J. K. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology, 46*(3), 461-88.

O'REGAN, J. K., & NOË, A. (2002). A sensory-motor account of vision and visual consciousness. *Behavoral and Brain Sciences, 24*(5).

PALMER, S. E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA.: The MIT Press.

PESSOA, L., THOMPSON, E., & NOË, A. (1998). Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences, 21*, 723-802.

RASTIER, F. (1997). *Meaning and textuality*. Toronto: University of Toronto Press.

RAYNER, K., & WELL, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review, 3*, 504-509.

ROSENTHAL, V. (1993). Cognition, vie et... temps. *Intellectica, 16*, 175-207.

ROSENTHAL, V., & BISIACCHI, P. (1997). Tacit integration and referential structure in the language comprehension of aphasics and normals. *Journal of Psycholinguistic Research, 26*(5), 557-580.

ROSENTHAL, V., PARISSE, C., & CHAINAY, H. (2002). A critical-letter effect in reading. (submitted).

ROSENTHAL, V., & VISETTI, Y. M. (1999). Sens et temps de la Gestalt. *Intellectica, 28*, 147-227.

RUMELHART, D. E., & MCCLELLAND, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*(1), 60-94.

SANDER, F. (1930). Structures, totality of experience, and gestalt. In C. Murchinson (Ed.), *Psychologies of 1930* (pp. 188-204.). Worcester, MA.: Clark University Press.

SEMENZA, C., BISIACCHI, P., & ROSENTHAL, V. (1988). A function for cognitive neuropsychology. In G. Denes, P. Bisiacchi, & C. Semenza (Eds.), *Perspectives on cognitive neuropsychology* (pp. 3-30). London: Lawrence Erlbaum Associates.

SHALLICE, T. (1987). Impairments of semantic processing: multiple dissociations. In M. Coltheart, R. Job, & G. Sartori (Eds.), *The cognitive neuropsychology of language* (pp. 111-129). London: Lawrence Erlbaum Associates.

SHALLICE, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

SHALLICE, T., BURGESS, P. W., SCHON, F., & BAXTER, D. M. (1989). The origins of utilization behaviour. *Brain, 112*(Pt 6), 1587-98.

SHANON, B. (1982). *The Myth of the Box and the Package*. Jerusalem: The Rotman Center for Cognitive Science: The Hebrew University.

SHAPLEY, M., & PERRY, V. H. (1986). Cat and monkey retinal ganglion cells and their visual functional roles. *Trends in Neuroscience, 9*, 229-235.

UNDERWOOD, G. (Ed.). (1998). *Eye guidance in reading and scene perception*. Oxford: Elsevier Science Ltd.

VICARIO, G. B. (1993). On Experimental Phenomenology. In S. C. Masin (Ed.), *Foundations of Perceptual Theory*. Amsterdam, The Netherlands: Elsevier Science.

VYGOTSKY, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA.: Harvard University Press.

WERNER, H. (1930). Die Rolle der Sprachempfindung im Prozess der Gestaltung ausdrucksmässig erlebter Wörter. *Zeitschrift für Psychologie, 117*, 230-254.

WERNER, H. (1956). Microgenesis and aphasia. *Journal of Abnormal Social Psychology, 52*, 347-353.

WERNER, H. (1957). *Comparative psychology of mental development*. (Rev. ed.). New York: International Universities Press.

WERNER, H., & KAPLAN, B. (1956). The developmental approach to cognition: its relevance to the psychological interpretation of antropological and ethnolinguistic data. *American Anthropologist, 58*, 866-880.

WERNER, H., & KAPLAN, B. (1963). *Symbol formation: an organismic-developmental approach to language and the expression of thought*. New York: Wiley.

WERNER, H., & WAPNER, S. (1952). Toward a general theory of perception. *Psychological Review, 59*, 324-338.

WHITEHEAD, A. N. (1983). *Concept of Nature*. Cambridge: Cambridge University Press.

YARBUS, A. L. (1967). *Eye movements and vision* (L. A. Riggs, Trans.). New York: Plenum Press.

This page intentionally left blank

YVES-MARIE VISETTI

LANGUAGE, SPACE AND THE THEORY OF SEMANTIC FORMS

1. INTRODUCTION

Phenomenological and Gestalt perspectives have become increasingly important in linguistics, which should lead to better exchanges with semiotics and cognitive sciences. Cognitive linguistics, and to a certain extent what is known as *linguistique de l'énonciation*, have led the way[1]. They have each in their own way established something of a Kantian *schematism* at the center of their theoretical perspective, developing on this basis what we might call a theory of *semantic forms*. They have introduced genuine semantic topological spaces, and attempted to describe the dynamics of the instantiation and transformation of the linguistic schemes they postulate. It is thus possible, up to a certain point, to conceive the construction of meaning as a construction of *forms*, and in so doing, to analyze resemblances and differences between these various processes. As a result, the idea of grammar itself has been modified, and centered upon a universal *linguistic schematism*, which supposedly organizes the values of all units and constructions. At the same time a certain understanding of the phenomenon of *polysemy* has been obtained, at least as far as this grammatical level is concerned.

However, a closer analysis reveals a number of difficulties, which call for a better understanding of what a genuine phenomenological and Gestalt framework should be in semantics. First, if we agree with the fact that there is a privileged relation, or some kind of similar organization, between language and perception, we should make more precise the general theory of perception (and jointly of action!) which we take as a reference. Secondly, if we also agree with the idea of a *specifically linguistic schematism*, analog to, but different from, what is needed for 'external' perception-and-action, its realm of dimensions should be determined: but we note here that there is a real, important disagreement between the authors. Thirdly, if we view language activity as a construction of genuine, 'internal' semantic forms based on linguistic schemes, it is obvious that polysemic words should correspond to *transposable* and *plastic* schemes: but the works we have just evoked remain very vague on this point; most of the time they propose lists of cases rather than genuine transposition and/or transformation processes. As a matter of fact, very few authors consider polysemy as a fundamental property of language which should be taken into account by linguistics from the very beginning.

Furthermore, all these approaches acknowledge the importance of the spatial and/or physical uses of linguistic units, i.e. those uses which seem to be exclusively dedicated to qualify the topological, geometrical or physical structure of the tangible world. But now a question arises: what is the relationship between these uses, and all the other uses of the *same* units, which, depending on the context, can signify a great variety of meanings? For instance, what is the 'logic' connecting the different uses of the English preposition *ON*, like in *book on the table* (spatial use), *departure on Monday* (temporal),

A. Carsetti (ed.), Seeing, Thinking and Knowing, 245-275.

tax on income or *to count on one's friends* ('support' or 'foundation')? Should we consider that the spatial or physical values of *ON* are in a sense a basis for all the others? Are they more typical? Or should we put all the uses on the same footing, and derive the various meanings from a single more generic principle?

In this paper we will show how to escape these false dilemmas, and how to better assess the continuity between the perception of the tangible world, and the perception of the Semantic Forms upon which we intend to build a theory. Starting from the key question of prepositions and of the relation between their spatial and less- or non spatial uses, we shall try to put forth general semantic principles, applicable to all categories of words and constructions (section 2). After that (section 3), we shall come back very briefly to Gestalt and phenomenological theories of perception, stressing the fact that they are *semiotic* theories, and not only morphological or 'configurational' theories of perception. As an immediate application to semantics, we will show the interest of this kind of approach to clarify the meaning of other categories of polysemic words (e.g. nouns). We shall then propose (section 4) – but in a very sketchy way – some general postulates for a *microgenetic theory of Semantic Forms*, based upon the mathematical notion of *instability*. The theory postulates 3 layers of meaning (or 'phases' of stabilization), called *motifs, profiles*, and *themes*. Taken together, they shape linguistic structure and semantic activity. They apply in exactly the same way in lexical as well as in grammatical semantics. Actually, they are conceived in the perspective of being integrated more tightly into a global textual semantics, very akin to the one developed by F. Rastier (1987, 1989, 1994, 2000). Finally, we come back in conclusion to what should be the nature and place of grammar in a theory of Semantic Forms.

This paper motivates and sketches a *theory of Semantic Forms*, which is a joint work with P. Cadiot, arising from our common interest for semantics, Gestalt theory, phenomenology, and complex dynamical models (e.g. Visetti 1994, 2001; see also Rosenthal and Visetti 1999, 2003). Examples and their specific analyses – sometimes slightly reformulated – have been taken from P. Cadiot's previous works. We propose here a synthesis of several previous publications, with a special stress on the relation between language and space, and on the grammatical dimensions of meaning. The semantics of prepositions, and more generally grammatical semantics, should be considered as a very important starting point, and a first application of our theory. However our real purpose is much more global, and goes beyond that: we try to put from the very beginning – at least at a theoretical level – the whole semantics under the pressure of a fully dynamical, discursive, and diachronic perspective. The interested reader will find a much more detailed presentation in our recent book (Cadiot & Visetti 2001)[2].

2. FROM SCHEMES TO MOTIFS: THE CASE OF PREPOSITIONS

All the different trends in Cognitive Linguistics have placed the question of *grammar* in the foreground of their works, and have developed specific and original conceptions of it. As a matter of fact, they have severely criticized the autonomy of syntax postulated by generative linguistics in the line of Chomsky's work. But they have maintained a clear cut separation between *structure* and *content*: 'structure' refers to a central and universal schematic level of meaning, called *grammatical*, which extends to all units and constructions; 'content' refers to all the remaining dimensions (concepts, notions, domains…), specifically brought by the lexicon. Grammar is therefore a kind of *imagery*, a way of structuring, of giving 'configurations' to all semantic domains, and also to the 'scenes' evoked by speech. *Imagery* includes:

- structural organization of 'scenes' (space, time, movement, figure/ground or target/landmark organization, separation between entities and processes)
- perspective (point of view, ways of going over the scene)
- distribution of attention (focusing, stressing)
- and, for Talmy or Vandeloise (not for Langacker), some less configurational dimensions, like the system of *forces*, or dimensions like *control*, or *access*.

For all these authors, this kind of *schematism* is specific to language (e.g. topological, not metric), but has many common properties with perception of external space.

Most often there is a trend towards relying on a very general psychological prototype, according to which language, at its most fundamental level, encodes tangible and/or physical structures. Therefore, in order to describe all kinds of categories of words, linguistics should favor spatial and/or concrete uses, and even take them as a primary basis for all the other ones. This idea leads in cognitive semantics, and also in grammaticalization theories, to a hierarchy of meanings, which starts from spatial or physical values, taken as literal meanings, up to temporal or abstract meanings, which are supposed to be derived from the previous ones by some kind of metaphorical transfer process. However, authors like Lakoff, Langacker, Talmy or Vandeloise underline that these primary values proceed from specifically linguistic *schemes*, which should not be confused with perceptive 'external' structures: indeed they are far more schematic, and at the same time genuinely linguistic, since for example they shape space by introducing 'fictive' contours or 'fictive' motions (Talmy). But in spite of these very important addings, the primacy (and/or the prototypical status) of a certain kind of spatial and physical meanings is not really questioned. Furthermore, schematical relations between language and perception often rely on a very peculiar conception of the spatial and physical experience, which fails to appreciate the true nature of what the phenomenological tradition names the 'immediate experience' of subjects. It amounts to a reduction of this 'immediate experience' to a purely external space, and to a purely externalized physics of 'forces', both separated from their motor, intentional and intersubjective (even maybe social and cultural) sources. In this external space, language would identify relations between 'trajectors' and 'landmarks', conceived as independent, separate, individuals or places, entirely pre-existing to the relations they enter in.

We think that this type of analysis extends to semantics a very questionable conception of perception, which stems from ontological prejudices, and not from rigorous descriptions. As a consequence of this wrong starting point, some works in the field of grammar retain only a very poor and abstract schematism; while others, or even sometimes the same works, address only the spatial or physical uses, hoping that the thus created gap between these uses and all the others will be filled by an appeal to the magical notion of metaphor.

More precisely, concerning the type of those linguistic schemes currently postulated by LC, and their relation to our external, everyday perception, two main attitudes can be distinguished:

- sometimes (Langacker, particularly) the realm of dimensions prescribed by the historical Kantian framework is centered on purely abstract 'configurational' dimensions (abstract topology, abstract dynamics); those dimensions are supposed to be a permanent and obligatory basis of language in all semantic domains; on the contrary, dimensions like 'forces' (and *a fortiori* dimensions like interiority, animacy, agency,…) are considered as less grammatical, secondary dimensions, coming only from more or less prototypical uses (e.g. referring to the external perceived space); they can only add themselves to the configurational dimensions, and never 'neutralize' them
- sometimes the realm of dimensions is not reduced (Talmy, Vandeloise); but this realm is considered primarily as part of our experience of the external physical world; spatial uses are more than typical, they are the primary ones; and all other uses are considered to be derived by a kind of metaphorical process[3].

With the semantics of prepositions, we find in a particularly striking form the problem of the relation to space and to the physical world. We shall take this example as a fundamental illustration of the ideas we intend to put forth in this paper. Indeed, the approach we advocate is deeply different from those we have just evoked[4]. It aims at going beyond these kinds of schematism, while keeping some of their 'good' properties. The exact abstraction level as well as the interior diversity of each scheme are a first key matter. On the one hand, abstract topological and/or cinematic characterizations (call them 'configurational') are too poor. On the other hand, schemes weighted from the beginning by spatial or physical values are too specific, and furthermore rely on a very peculiar conception of spatial and physical experience. Actually, more 'intentional' or 'praxeologic' dimensions, intuitively related to 'interiority', 'animacy', 'expressiveness', 'appropriation', 'control', 'dependence', 'anticipation' etc. are needed. By entering in the process of discourse, all these dimensions – configurational or not – can be neatly put forward by speech, or alternately kept inside the dynamics of the construction of meaning as a more or less virtual aspect of what is thematized. In particular, configurational or morphological values are not a systematic basis: they may be pushed in the background, or even disappear, superseded by others, which are quite equally fundamental and grammatical.

More generally, these *motifs*, as we shall call them as from now, to distinguish them definitely from the problematics we criticize, appear deformed, reshaped, in various *profiles*, abstract as well as concrete. A *motif* is a unifying principle for this diversity of uses, which can only be understood if one takes into account from the very beginning dimensions of meaning which cannot be integrated into the narrow frame of a schematism – at least if by a 'schematism' we mean something (still predominant in cognitive linguistics) which can be traced to kantian philosophy (Kant [1781-1787]; for a discussion on this point, cf. Salanskis, 1994). Of course we have to consider all these fundamental dimensions at a very *generic* level, so as to assume that they are systematically put into play, and worked out by each use. But generic as they may be, our thesis is that these dimensions can be traced back to the immediate experience of perception, action and expression, if they are conveniently described in their social and cultural setting. This is why we decided to drop the designation of *scheme*, and to adopt the word *motif* to express the kind of 'germ of meaning' we wish to attribute to many linguistic units. Indeed, the word 'scheme' evokes a certain immanentism or inneism, a restricted *repertoire* of categories not constituted by culture and social practices, and a priviledge granted to a certain biased representation of the physical world. It is therefore a term not suitable for indicating an historical, cultural, 'transactional' unifying linguistic principle, whose function is to *motivate* the variety of uses of a grammatical or a lexical unit.

SOME SKETCHY CONSIDERATIONS ON FRENCH PREPOSITIONS[5]

There are great differences in the systems of prepositions in French and English, especially concerning so-called 'colourless' or only weakly depictable 'space prepositions' like EN or PAR. We will here present only short considerations about SUR, SOUS, CONTRE, EN, PAR, which evidently call for considerable developments, and should be in a systematic mood confronted to other languages. We hope at least that this will be understood as a way of challenging the routine frozen expression: "spatial preposition".

The case of SUR

A very sketchy analysis allows us to distinguish the following configurations.

A 'region SUR' constructed at the level of predication ETRE SUR ('to be on'), i.e. a construction of a site based on the connection [Preposition + Nominal], localization of the noun subject, and the contact enabled by the predicate:

(1) *Le livre est sur la table* ('The book is on the table')

In other cases, the 'region ON' is established by the context of the sentence, which allows for an adjustement or requalification of lexical and syntactic expectations.

(2) *Max s'est effondré dans le fauteuil* ('Max collapsed in the (arm) chair').

(3) *Max a posé timidement une fesse sur le fauteuil* ('Max timidly sat on the (arm) chair').

The motif 'contact' is permitted and enabled by the predicate. As opposed to a table or a sidewalk, an armchair is not a priori an acceptable object for the predicate ETRE SUR ('to be on'). The requalification is facilitated by the specific reference.

A zone established as a frame for what happens in the 'region SUR'. Compared with the previous examples, the possible fluctuations between contact and localization increase.

(4) *Les enfants jouent sur le trottoir* ('The children are playing on the sidewalk')

Still, there is a simple correlation between a topological notion and a uniquevocal localization in the thematic space.

However, this correlation is nullified, or made more complex, by many other uses with spatial implications. It may happen that the prepositional phrase does not localize the subject of the sentence.

(5) *Pierre joue avec sa poupée sur la table* ('Pierre plays with his doll on the table').

(6) *Pierre a vu un chat sur le balcon* ('Pierre saw a cat on the balcony').

Nothing indicates that the referent of 'Pierre' is localized by the 'region ON' (on the table, on the balcony). In fact, the contrary is noticably more likely.

The 'region ON' no longer has determined spatial limits at the thematic level. Following examples are quite particular to French, in which we can hypothesize that the motif is further developed.

(7) *Pierre travaille sur Paris* ('Pierre works *on/in Paris').

(8) *Pierre est représentant sur la région Nord* ('Pierre is a representative *on/for/in the north').

Here, the preposition SUR is used in the construction of "functional spaces" (zones specified only in the domain of the predication) and not of physical spaces, but the topological instruction of contact is preserved.

The motif of 'contact', which, based on the preceding examples, we might believe to be simply topological, can actually be easily requalified with new interpretative effects for which the spatial inferences are decreasingly concrete, proving itself to be inseparable from temporal and qualitative modulations (Dendale & De Mulder, 1997, whence the following examples):

- support (weight or imminence).

(9) *Une menace planait sur la ville* ('A threat hovered on?/over the town').

- foundation (assessment).

(10) *Juger les gens sur l'apparence* ('To judge people on?/by their appearance').

(11) *Il fut condamné sur de faux témoignages* ('He was convicted on false testimony').

- covering.

(12) *La couverture est sur la table* ('The tablecloth is on the table')

- objective (goal)

(13) *Marche sur Rome* ('March on Rome')

(14) *Fixer un oeil sur quelquechose* ('*pose/ *fix / *leave/feast one's eyes on something'.

-visibility, immediate access (as opposed to inclusion which would signify dependance, interposition of a border or a screen).

(15) *Il y a un trou sur ta manche* ('There is a hole *on/in your sleeve')

Semantic cues 'support' and/or 'foundation' can be extended easily to uses that are definitively 'non spatial' as in:

(16) *Impôt sur le revenu* ('tax on income')

(17) *agir sur ordre* ('act on orders')

(18) *Pierre a travaillé sur cette question depuis longtemps* ('Pierre has been working on this question for a long time').

Or even:

(19) *Sur cette question, Pierre n'a rien à dire* ('On this issue, Pierre has nothing to say').

Here the motif of contact is invested in a thematic zoning, which can be specified only in the domain opened by the predication or the introductor nominal argument.

Let us also remember the temporal uses differentially specifiable, which emerge from the motif of contact.

(20) *Sur ce, il disparut à jamais* ('*On/after this he disappeared for ever')

(21) *Pierre est sur le départ* ('Pierre is about to leave')

(22) *Il y a eu des gelées sur le matin* ('There was a frost this morning/on the morn' (archaic))

(23) *Il faut agir sur le champ* ('One must act at once').

In *compter sur ses amis* ('to count on ones friends'), *miser sur le bon cheval* ('bet on the right horse'), without entirely abandonning a certain value of 'to lean on', a modulation of the original motif, the preposition is requalified as a rectional marker.

These examples not only invalidate purely spatial and physical explanations of SUR. They also weaken explanations based on abstract topological schemas, which often seem artificial and demand further qualifications which call into doubt their validity. Above all, this type of schematics does not provide operable explanations, and as a result doesn't explain why only certain values and not others are called upon (by interaction with the surrounding lexical material, as we say). What's missing here is the possibility of recognizing the affinity and interrelation of these different values, which we would like to stabilize by way of lexico-grammatical motifs.

In this way, the topological instruction, even when purely configurational and despatialized (i.e. conceived independantly of the perceived space) seems to flag behind a richer, more open definition-delimitation of two 'segments' or 'phases' as they are construed during any type of contact. Compared to the image of 'surface' often invoked (geometrical notion), or to that of 'height' (Weinrich 1989), this motif of 'contact' would have the same statue as that of 'coalescence' for EN, or of 'means' in the case for PAR. Beyond its dynamic value it also offers a static characteristic which provides a border or a stabilized variation (localization, support) but it is fundamentally an aspectual motif, intentional in aim and in practice. At once a motif of exploitation and of valorisation of this contact by a type of immediate interaction (leaning, rebounding, perlaboration), giving the values of objective, imminence, achievement, effect, transition, cause and effect. Its configurational expression, once fully deployed, includes an axial orientation of momentum, another transversal orientation for the contact zone and the exteriority maintained between the two phases thus delimited. (if the contact zone is in fact the topological frontier of the access zone, it is still not appropriated as its border, but remains 'exterior').

Localization can certainly be explained in euclidean terms: surface, height, width, etc. But the diversity of possible instances of localization (the rich variety of contributing elements) calls for dimensions which are more dynamic (force, figure/background) compared to the more configurational ones. In the phrase *cup on the table*, we might emphasize the importance of [bearing-weight]. In *bandage on the arm*, *drawing on the wall*, *handle on the door*, *apple on the branch*, ON constitutes the sight as a [background], which guarantees a [detachability] for the figure, regardless of any more objective relations with the object/surface.

The case of SOUS

One can uncover five 'experiental types' (evidently a nice example of a family resemblance in the wittgensteinian sense):

- low position: *sous la table* ('under the table'); *sous les nuages* ('under the clouds');
- covering/protection: *sous la couette* ('under the covers'); (objet enfoui*) sous la neige* ('under the snow'); *sous une même rubrique* ('under/in the same rubric');
- exposition: *sous la pluie* (under*/in the rain); (marcher) *sous la neige* ((walk) 'in the snow'); *sous les regards* ('under the eyes of x'); *sous les bombes ('under fire')*; *sous la menace ('under the gun')*;
- inaccessibility: *sous terre ('underground'); sous le sceau du secret ('under heavy guard')*;
- depending from the external: *sous surveillance* ('under surveillance'); *sous influence* ('under the influence'); *sous la contrainte ('under pressure'); sous garantie* ('under warranty'); *sous arrestation* ('under arrest').

These uses involve a co-adjustement of the values selected from the NPs assigned by the preposition, and in some cases by the introductory element (see the example of snow). Together they evoke family resemblances of covering, protection, inaccessability, exposure to, and dependence upon, in varying degrees of explicitness?

Among the notions evoked above, certain seem more oriented to a topological schematic pole (surface constructed by the PP which establishes an interior space based on that boundary. The others closer to a more "instructional" pole (Cadiot 1999) which consists of the more dynamic values, aspectualised by a quasi praxeological perspective (no exit dynamic, opening blocked) indexed on the ambivalence of the situation (covering vs. exposed). Articulating these two poles of the boundary, which remains separate from the interior space, is just the configurational expression of this blocking and ambivalent. As in the case of SUR, this complex motif is diversely profiled and stabilized: by valorization, specification, or on the contrary inhibition, retreat, aspectualization of the different values it unifies.

The case of CONTRE

Let's note the following four 'experiential types':

- Proximity with contact: *s'appuyer contre le mur* ('leaning against a wall').
- Opposition (conflict): *être contre le mur de Berlin* ('be against the Berlin wall'); *contre toute attente* ('against all expectations').
- Exchange: *échanger sa vieille voiture contre un scooter* ('trade one's old car for a scooter').
- Proportion / comparison: *vingt mauvais films contre un bon* ('20 bad films *against/ for one good one').

For CONTRE we propose a motif instituting the affinity of opposition and reconciliation (force/counter-force, posing/opposing). This motif is sustainable, up to a certain point, in a schematic framework, which could be capable of reflecting relational categories like [force] in a plurality of spaces (not necessarily physical). But we insist again that this motif-schema must be modulated and specified in accordance with plausible profiles. As a result values such as 'counter-force' or 'dynamic coming together' can disappear almost completely from the profile. Even when so "virtualized" as in *Sofa against the wall* they remain as a motivation for the internal perspective or 'aspect' of the dynamic.

The case of EN

We will show two points:
- there is no clear-cut distinction between spatial and non-spatial uses or senses;
- the specifically linguistic meaning of it should be accessed in an immediate combination of schematic and intentional dimensions:

Let's have a look at following phrases:

(1)	*pommier en fleurs*	'apple tree in bloom'
(2)	*chien en chaleur*	'dog in heat'
(3)	*femme en cheveux*	'hair-dressed woman'
(4)	*propos en l'air*	'words up in the air'

The sense of these phrases can be paraphrazed by following intuitive formulations or characterizations: 'globally saturated physical image' (1), 'invasion' (2), 'emblematic access' (3), 'taken over from the inside/outside' (4).

They tend to show that space is only involved at a thematic level, and in some sort of continuous variation. The characterisations can be resumed in an unique notion, or motif, of coalescence, with no linguistically prescribed limits or 'bornage' (bordering), and assymetricaly oriented toward the referent of the second NP. The image of the first NP is, so to say, absorbed in the image of the second (fleurs, chaleur, cheveux, air).

But this motif is not only schematic or perceptual. It coalesces with a more instructional dimension: one has to associate the resulting image with its perspective, and with the intention through which or by which it was brought about. The scene is necessarily animated by the process which generated it. Otherwise other prepositions like DANS (with its bornage instruction) or even AVEC would be more appropriate.

A more direct evidence for this rather intuitive interpretation can be drawn from other data where space is not involved:

(5) *Max est en faute* ('Max is mistaken') / **Max est en erreur*
(6) *Max est en tort* ('Max is wrong') / **Max est en raison*
(7) *Max est en beauté* ('Max is handsome') / **Max est en laideur*
(8) *Max est en vie* ('Max is alive') / **Max est en mort*
(9) *Max est en difficulté* ('Max is in difficulties') / **Max est en facilité*.

There seems to be a rather regular paradigm of such cases, where only the 'resulting states' which can be associated with the intentional, subjective object-oriented path, or purpose that brought them about can be correctly introduced by EN. For example *Max est en vie* is pragmatically possible only in as much as one has reasons to believe that he could be dead (after some accident, presumably); *Max est en faute*, because he has done or said something which happened to be wrong or inappropriate; *Max est en beauté* means more than *Max est beau*: that he tried or at least, wished to be handsome...

The case of PAR

Even more evidently, it is impossible to differentiate spatial and not spatial uses in the case of PAR.

être emporté par le courant	'to get carried away by the current'
passer par le jardin	'to go through the garden'
prendre par la gauche	'to take a lefthand turn'
regarder par le trou de la serrure	'to look through the key-hole'
attraper par la cravate	'to grab by the tie'
tuer par balle	'to kill by bullets'
convaincre par son comportement	'to convince by one's behaviour'
impressionner par son intelligence	'to impress by/with one's intelligence'
passer par des moments difficiles	'to come through hard times'
renoncer par lassitude	'to give up from/because of lassitude'.

In English, BY works better with active referents and tends to internalize them in the scope of the schema, while with more external complements, THROUGH or even BECAUSE OF are better, and WITH seems at least to initiate a motion of externalization, or 'parallelization'. As is well known, PAR is typically used to express agentivity in passive constructions or in any type of constructions where a process is described from the point of view of its activation. So it expresses an inner activation principle. Being 'inner' corresponds to the schematic dimension, being 'agentive' to the intentional one. But both are intimately correlated and coactive in every instance, even when it corresponds to no specific local thematic or referential intuition.

We stop here this series of examples, and try now to draw some general conclusions. What is actually our own perspective? In summary, we advocate:

• No privilege for spatial or physical usage of words (as conceived by current trends in Cognitive Linguistics), and consequently no doctrine of metaphorical transfer of meaning, going from the spatial and/or physical uses towards more 'abstract' ones (as currently conceived by the same linguistics)

• Search for grammatical *motifs*, which are ways of giving/apprehending/displaying, immediately available in all semantic domains, without any analogical or metaphorical transfer stemming from more specific values, allegedly conceived as the primitive ones

• Rejection (most of the time) of purely configurational versions of those *motifs*: on the contrary, a *motif*, especially a grammatical one, is an unstable, and at the same time a strongly unitized, mean of building and accessing 'semantic forms'; it ties together, and defines a kind of transaction between many dimensions which cannot be dissociated at its level, but at the level of *profiling* inside specific semantic domains

• Rejection of an 'immanentist' explanation of the variety of uses, based upon an identification of the *motif* with some kind of 'autonomous' potential; indeed, depending on the specific use, some dimensions of the *motif* can be further specified, enriched with other dimensions, or on the contrary virtualized, even completey neutralized. The parameters controlling the profiling dynamics are not an internal property of the *motif*: the relation between the *motif* and a particular profile has to be considered as a *linguistic motivation*, because profiling a *motif* consists of recovering it within other dynamics, brought about by the co-text and the context, i.e. by an ongoing hermeneutic perspective

• A conception of the grammatical *motifs* (e.g. a motif of a preposition) as highly unstable 'forms' (or germs of forms) which can be stabilized only by interaction with the others constituents of surrounding syntagms, or even by more distant elements of the cotext: as we have said, this stabilization is not a 'simple' instantiation of the *motif*, but a recapture by other non immanent dynamics giving rise to the variety of its *profiles*.

Actually, this approach is very general, and applies both to grammar and to lexicon. It is strongly different from other approaches currently worked out by cognitive linguistics. We have already underlined some differences in the analysis of the grammatical expression of space, and in the assessment of its status relatively to the global functioning of the concerned units. But the situation is the same for grammar as a whole, and in particular regarding its difference with the lexical aspects of meaning. In short, we could say that cognitive linguistics tend to limit semantics to grammar, and grammar to a certain kind of 'schemes'. We have just criticized their schematism, as well as the conception of perception to which it is correlated. Indeed, concerning the type of the grammatical schemes, and their relation to our external, everyday perception, we have seen that two main attitudes can be distinguished:

• sometimes, the schemes are from the very beginning merged with a very peculiar conception of the physical world, in which the fundamental role of action, and of other kinds of anticipations, is underestimated (cf. Talmy, or Vandeloise 1991);

• sometimes they are abstract, and purely topological/configurational (Langacker).

The reason for this false alternative is simple: there is no generic diagrammatic representation of action, animacy, interiority, expressivity, intentionality and anticipation, as they are constituted by their cognitive, social, cultural and… linguistic modalities. So that whenever one tries to take some of these dimensions into account, the only way to recover some expressions of them is to resort to the physical experience – which is at the same time wrongly apprehended. Once again, such a conception of our 'immediate experience' not only provokes an impoverishment of the theory of grammar, it also introduces a gap between grammar and lexicon, as well as between the so-called litteral meaning and the figurative ones. Finally, so to speak, the only relation between grammar and lexicon, is… schematism ! And the only relation between the registered basic lexicon and the variety of uses is… a metaphoric relation to space ! In short, we think that cognitive linguistics have up to now too strongly dissociated 'structure' (identified to the schematical dimensions of meaning) from 'content'. Therefore the very foundation of semantics is still grammar, understood as a fairly autonomous device, in spite of whatever these authors may say about the continuity between grammar and lexicon. In the same way, there is a tendency to see grammaticalization as a pure bleaching process, which only retains values pertaining to a universal repertoire set once and for all.

We think, and actually numerous linguistic analyses show, that we need a richer theoretical apparatus, inspired by an integrated theory of perception, action and expression, really susceptible to be transposed into grammatical and lexical studies, which would then become more tightly unified if we view them in this perspective. We look therefore towards a fully intentional theory of perception, a semiotic and 'transactional' theory of immediate experience, constituted by the simultaneous grasp of practical (praxis), axiological (ethics and esthetics), and subjective values. In order to recover such a theory, we would have to read carefully the gestaltist writings, notably those of the Berlin School (Wertheimer, Koffka, Köhler), the message of which has been weakened by cognitive linguistics. Beyond that, we would have to come back to the phenomenological tradition (Husserl, Gurwitsch, Merleau-Ponty), to Cassirer's philosophy of symbolic forms, and also for example to Vygotsky's developmental psychology, which gives to *social practices* a constitutive role[6].

Once recovered in this way a much more relevant model of perception-and-action, we shall be in a position to transpose it into semantics, in order to provide for a more complex interplay between the dynamics of constitution and the constituted meanings, than anticipated by current schematisms. Language activity will be described as a process analogous to what is called a *complex system* in other disciplinary areas. Notably, the construction of 'semantic forms' will appear as a kind of microgenetic developmental process, with concurrent unstable and stabilization 'phases'. The description of the *linguistic motifs* as *unstable germs of forms* (in a gestalt sense of the word 'form', transposed to semantics) is thus fundamental in our perspective. This will result in three semantics 'modes' or 'phases' in the dynamics of the construction of meaning, which we shall call *motifs*, *profiles*, and *themes*.

3. TOWARDS A PHENOMENOLOGICAL AND GESTALT THEORY OF SEMANTIC FORMS

3.1 GESTALT, PHENOMENOLOGY, AND LANGUAGE ACTIVITY

Among the several fundamental references quoted at the end of the preceding section, we shall limit ourselves, and even then in a sketchy manner, to the gestaltist ones[7]. Gestalt psychology has often been reduced to its morphological and morphodynamical aspects (especially with the famous slogan 'the whole is more than the sum of its parts'). Actually, it describes a much richer and deeper unity between perception, action and expression. It is precisely this kind of unity that we want to put at the core of the construction of meaning, seen as a construction of 'semantic forms'. Under the expression 'semantic forms', we do not refer to a sensation conceived in isolation (even if the *theme* of the discourse resorts to our concrete, practical world), but to semiotic and multimodal 'forms' unfolding through language activity as units in all domains of thought and experience. We do not either take ordinary perception as a *foundation* for linguistics, but rather take it, when described according to the phenomenological style, as an essential *correlate*, and a particular *illustration* of the construction of meaning. Once again, the choice of a theoretical perspective on the perceptual experience is decisive for any linguistics which pretends to find here a model, and perhaps an origin.

For example, turning back to of our fundamental relationship with space, we find currently in linguistics three main conceptions of this reference space:

• physical, objective space, with a universal geometry, and objective, universal categories of 'objects'

• perceived, psychological space (still independent of culture and language diversity as a general framework – even if it is differently worked out by cultures and subjects)

• semiotic space, whose overall perception bears immediately upon social practices and cultural knowledge

Cognitive linguistics favor conception (b), with a very little touch of (c). We think more radically that:

• this approach of perception should be extended to include a broader *repertoire* of dimensions, which are unavoidably shaped by the social and cultural context. This *repertoire* cannot be defined on the basis of a purely pre-linguistic or extra-linguistic perspective. Each particular language defines its own realm of dimensions, including those that are closer to the sensible ones.

• perception, for what concerns its 'continuity' with semantics, is less a matter of encountering concrete, external things or places, than a matter of establishing *qualified relations* with things, space, and other perceiving agents; therefore another conception of subjective experience, as well as a more *intersubjective* perspective, are here fundamental; they put forth immediately intertwined attentional, modal, behavioral, axiological values, which cognitive linguistics treat only as secondary or derived, and at best in a very parsimonious way.

What seems to be related to language at its most profound level, in an intimate and *reciprocal* connection, is the social and cultural *Lebenswelt*, which includes centrally the socially and culturally constituted experience of the body, in its relationship to its practical environment and to others subjects. Spatialist and/or purely topologist approaches apprehend only certain wrongly isolated effects of this intimate connection. Furthermore, they tend to consider space as already constituted, and do not grasp it at the level where it is permanently reconstructed by our movements, and reshaped by our expectations. Quite differently, we want to insist on *the self centered bodily experience*, which is exemplified by qualitative terms, like: resistance/yielding, holding tight, rupture, softness, roughness, bury, block, insert, get rid of, drown, touch, etc. Consider also the 'motif' of *containment*, which is much richer than the relationship between the container and the contained. Think of the motif of *control*, which intertwines attentional, temporal, kinesthetic, modal, and even intersubjective aspects. Think equally of the English particle *up*, which is conceptually an aspectual marker for completion, and not essentially an indication of verticality, etc. Meaning is thus firstly anchored in anticipation, qualitative, often synesthesic feelings, and not in a directional grasping of "objects". But we insist (*contra* Lakoff) that this partial embodiment of semantics is only possible if the body in question is not considered as a pre-linguistic basis, but as a cultural construction, a truly 'fictive' body, constituted by social practices – and among them by language activity.

Precisely, the Gestalt and phenomenological tradition doesn't dissociate the grasping of forms and values; as we said, perception, action, and expression are here more tightly intertwined than in any other approach. 'Forms' in this sense:

• are to be simultaneously defined in all modalities (visual, auditive, tactile, motor and kinesthetic…), cf. the very important concept of *synesthesy* (objects, moves, changes that appear explicitly in one sensorial modality, are 'felt' in other sensorial and kinesthetic modalities as well)

• have *immediate* functional and agentive values (degree of spontaneity, distinction active/passive, differentiation of roles). Cf. Gibson's *affordances* (1979), which have been directly inspired by Lewin's *Aufforderungscharakter* and Kohler's *requiredness* (1938): e.g. artifacts like a hammer, a chair, are perceived immediately with their gestual, postural, functional values; seeing a mailbox immediately sketches, depending upon our attitude, parts of an integrated social scenario

• have also immediate esthetic and 'behavioral' values, with emotional resonance. Recall the examples of Köhler (1929, 1938): a wave, a musical crescendo. Cf. also Michotte's work (1946) on the perception of movements as behavioral styles (walking, running [away, after], swimming, flying…)

• include an immediate perception of forces or causes, of intentional moves (intersubjectivity, animacy, agency), and of expressive values (joy, fear, demand…).

Perception in this sense has to be considered as instantiating a *general* structure of cognition, and not only as resorting to a purely sensorial and peripheral organization. As a slogan, we could say that 'to perceive is from a single move to act and to express'. Perception already gives access to, and sketches, a meaning. It implies not only the presence of things, but a perspective of the subject, and a suggestion of acting. Perception in space is not grasping pure configurations or shapes, nor only a basis for other, subsequent 'associative' or 'metaphorical' interpretations: it is from the outset a dynamic encounter of 'figures' with no necessary dissociation between forms and values, apprehended in the course of actions, and deeply qualified by a specific mode of access or attitude. It is this notion of a *qualified relation* (which is a way of 'accessing', of 'giving', of 'apprehending'…) that we want to transpose into semantics, in order to view it as a kind of perception and/or construction of forms. At this level, any distinction between abstract or concrete, or between interior or exterior perception, is irrelevant.

In the same way as there is more than topology or geometry in our multiple relations to ambiant space, we can say that 'figures' are objective counterparts, phenomenological manifestations of the *relations* we have with them. Needless to say, the perceived relations are not prescribed by some kind of pre-existent exterior world: they are conditioned by a global perspective or purpose, which constitutes subjects and objects simultaneously. Any perceptive relation can thus be modulated towards its subjective side, or towards its objective one, in a way which is constitutive of the act of perceiving. As a *relation*, it can be transposed to multiple situations or referents. Language only radicalizes this: at its deepest level, it defines, differentiates, and records primarily *relations* – not the referents, which depend upon another, more thematic, linguistic and cognitive level (e.g. think to a contrast like *house*/*home*: possibly the same referent, but not the same relation to it). And as soon as language comes into play, *relations* are definitely *socially* constructed, as historically accumulated sediments. On the whole – and this is called *polysemy* – they are intrinsically transposable to a diversity of 'themes', in a variety of semantic domains correlated to a variety of social and cultural practices. Language activity appears, up to a certain degree, as a 'new' layer of *social perception*, made of intrinsically transposable, highly unstable germs of 'forms' (forms of relations), to be stabilized in a variety of domains: experiential (*qualia* and their evaluations), practical (actions and their domains), theoretical, mythical, etc…

Therefore, if the concept of Gestalt seems to be perfectly transposable to semantics, it is on condition that it be rethought so as to integrate the *socially* constituted nature of Semantic Forms, which are of a *linguistic and semiotic* nature, different from the more universal level of experience which has been studied in the visual modality[8].

3.2 AN INSIGHT INTO THE SEMANTICS OF NOUNS

In several recent works, we have applied to a set of strongly polysemic nouns of 'Basic French' a description principle, which takes into account on an equal footing all their uses[9]. We were thus moving away from the dominant lexicologic approach, which promotes a certain notion of 'litteral' meaning, supposedly combining tangible, concrete, reference and denominative function in a first primary layer. As for us, on the contrary, the meaning of the most frequent nouns can and must be devised long 'before' any logic of classification or of categorization of referents. As a matter of fact, nouns – at least the most frequent ones – are 'ways of access', or 'ways of establishing relationships', prior to being labels in a game of entities categorization and denomination. Their prior function is to be interpreted in terms of analogical generative potentials (or germ of forms), which we called *motifs*. These motifs may be intuitively presented as generic 'experiential bundles', and described, in the phenomenological and Gestalt style, according to different intertwined modalities: perception, action, *qualia* and evaluation. Of course, we do not intend to give full descriptions of them: such an enterprise would be endless. The only thing to do is simply to put forward some of their principal dimensions, which are already very enlightening for the question of polysemy and of the so called 'figurative meanings'. We shall give here very few examples, trying to choose them in such a way that their polysemic distribution in French be similar to the one of their usual translation into English. Other examples can be found in Cadiot and Visetti 2001, chap. 3; Visetti and Cadiot 2002, section 3.

Let us start with some *motifs* which seem to provoke a perception and/or a construction of forms of visual type. The words which correspond to them seem indeed to have as a basic signification a 'schematic' form, which is easily, almost mechanically, transposable from one domain to another.

• ARBRE ('tree'): *arbre fruitier* ('fruit tree'), *arbre généalogique* ('family tree'), *arbre syntaxique* ('syntactical tree'); also some uses considered as more figurative: *arbre de la Vie* ('Tree of Life'), *arbre de la Connaissance* ('Tree of Knowledge')

• VAGUE ('wave'): *vague d'enthousiasme* ('wave of enthusiasm'), *vague de chaleur* ('heat wave'), *Nouvelle Vague* ('New Wave')

These examples already show that motifs are not generally limited to configurational values (like a dynamical shape). Indeed, as in the gestaltist theory of visual perception, *motifs* unify a bundle of synesthetic values going far beyond purely morphological determinations. For example, the motif of ARBRE unifies a branching process with a specific coherence stemming from the root, and giving rise to a perspective of growth, generativity, support. Depending upon the specific use, some of these dimensions are salient, others are pushed into the background, or even vanish. The important point is that language offers the *possibility* to grasp simultaneously all these aspects, because they are put into transaction with each other, and blend together, giving rise to a kind of coalescence. At the same time, language offers the *possibility* of dissociating this same unity (up to a certain point), and of enriching it (if needed), in order to give rise to a variety of *profiles*.

Beyond the synesthetic values just exemplified, other nouns give direct access in their motif to dynamical-functional and practical (action-oriented) dimensions of meaning. Of course, this immediate *relation to praxis* makes increasingly more problematic the attribution of an original 'material' meaning ! Thus, for instance:

• BOUCHE ('mouth'): can be used in French as in English for a river ('Mouths of the Gange'), a volcano, etc. French also uses it for the subway's entrance (*bouche du métro*). One can see that the motif of BOUCHE includes dynamical-functional aspects, roughly evoking 'entry and exit'

• CLEF ('key'): *clef anglaise* ('adjustable spanner'), *clef de voûte* ('keystone'), *clef du succès* ('key of success'), *clef du mystère* ('key to the mystery'), *point-clef* ('keypoint'), *mot-clef* ('keyword'). One can propose that the motif of CLEF unifies 'exclusive access, (un)locking, and accuracy'. One can also see that the word CLEF can evolve according to a mainly perceptual and functional model (*clef anglaise, clef de voûte*), or according to a more explicitly intentional and practical model (*point-clef, mot-clef, clef du mystère*)

• MUR ('wall'): *mur de briques* ('brick wall'), *mur de Berlin* ('Berlin Wall'), *se cogner la tête à un mur* ('to hit one's head against a wall'), *se heurter à un mur d'incomprehension* ('to come up against a wall of incomprehension'). These examples show that MUR integrates in its motif 'to separate, to stand up, to surround, to protect, to hit…'. It is to be stressed that an agonistic dimension is already immediately present in this motif, and not subsequently inferred (but of course it is neutralized in many denominative uses).

Other words yet give access through their motif to a certain general 'quality of sensation', or to a certain 'norm of evaluation', which can be applied to an open set of entities, situations, states, etc., impossible to be determined a priori. These *linguistic qualia* have of course very important perceptual and emotional correlates, which are like their emblems; but being *linguistic*, these qualia are of course something else than these perceptible emblems: they are transposable to many kinds of experiences. Here are some examples, about which we shall not try to explicit any *motif* (except for the first example). We shall only underline that these conjectural motifs are neither concrete nor abstract, being totally entangled, as generic qualia, between physical, psychological, and axiological aspects:

• NUIT ('night'): the motif here tends to split into two sub-motifs, which nevertheless remain linked; the first evokes darkness: *la nuit tombe* ('night is falling'*), la nuit de l'ignorance* ('darkness of ignorance'), la nuit des temps *('the mists of time')*; the second evokes a period of rest: *passer une bonne nuit* ('to have a good night')

• BOUE ('mud'): *s'enfoncer dans la boue* ('to sink in the mud'), *traîner quelqu'un dans la boue* ('to drag someone's name in the mud')

• FOUILLIS ('mess'): *ta chambre/ ton article est un vrai fouillis* ('your room/paper is a real mess')

• NUAGE ('cloud'): rather than defining a motif, it is better to delineate it through the specific phraseology of the word (idiomatisms), of which it is a unifying principle. For example: *les nuages s'accumulent* ('clouds are gathering': in French, it applies to many kinds of situations where a threat is looming, like in English 'to be under a cloud'); *être dans les nuages* ('to be in the clouds'); *un nuage de tristesse passa sur son visage ('his face was clouded with sadness')*; and inversely, one can talk of *un bonheur sans nuages* (a happiness without clouds: 'a perfect bliss').

In this search for the *motifs*, the lexicalized figurative meanings play a very important role. Indeed, they do not function as heavily analogical mechanisms, but on the basis of an immediate *promotion* of the corresponding motif, which therefore appears as a general access principle, a qualitative relational index, immediately available in a variety of domains.

All these examples show that the notion of Gestalt can only be recast in semantics if it is taken in its widest diversity. Even less of course than for grammatical units, configurational or morpho-dynamical aspects do not suffice, since the motifs merge many other dimensions. As testified by polysemy, by the (so called) figurative meanings, and by their surrounding phraseology, nouns, at least the most frequent ones, register in their most internal kernel the coalescence of all these dimensions, much more than their dissimilation: this is why it is necessary to introduce motifs as unifying principles for the lexical diversity. On the other hand, this kind of unity does not define an invariant: on the contrary, motifs can be dissociated, and sorted out at the lexical level of *profiling*. Therefore, *profiling* do not consist in a 'simple' instantiation, but in a recapture of the motifs through more global dynamics: we contend that this process must be understood as a stabilization process, applied to unstable germs. And this leads us to the global theory sketched in the next and last section.

4. THE MICROGENESIS OF SEMANTIC FORMS: MOTIFS, PROFILES, THEMES

Our global theoretical perspective presents language activity as a construction, and/or a perception of semantic forms. That does not mean that we intend to reduce it to the perception or construction of simple 'external' entities. On the contrary, it means that we aim at describing the more specifically linguistic-semantic part of a global process giving rise to 'thematic forms', which are inextricably both linguistic and semiotic. These forms can be sensible, imaginary, or ideal; and their construction depends upon the subject's activity as well as upon the semiotic (social, cultural) ambient medium. This is why we have taken up concepts and principles inherited from the Gestalt and phenomenological traditions: indeed, they put into place, at least at the level of individual subjects, the appropriate setting for this kind of widening of perspective, and at the same time for its focalization on language.

In support of their approach of psychology, the gestaltists from the Berlin School (principally W. Köhler 1920) laid the basis of a general theory of Forms and organizations. Drawing upon their hypothesis of an *isomorphism* between the structures of the subjective immediate experience, on one side, and the functional dynamical organization of the brain, on the other side, they devised a theory *both phenomenological and physical*, inspired by field theory, statistical physics, and dynamical systems. But they considered it at that time as a speculative theory, or as a building metaphor, and not as a genuine model for the phenomenological mind and/or for the brain, hoping that future progress in neurosciences, in physics, in mathematics, and in the methodology of phenomenological descriptions, would confirm their insight. Since then, many works in various areas have pursued in the same direction, and actually gone far beyond, towards multiple theories of complex dynamical systems. Although we do not offer here any precise modelization project, we think that calling upon the most general principles of the gestaltist theory of Forms can help to stabilize our own theory, and to prepare its association with the important interdisciplinary field just evoked. As a reminder, here are some of the most fundamental features of this theory:[10]

- Relations between parts and wholes: synthesis by reciprocal determination of all dimensions of the field of forms
- continuous substrates, continuous modulations of forms, and at the same time delineation of forms by means of discontinuities
- figure/ground and trajector/landmark organization
- no form without an 'internal' time of constitution: time of integration and/or differentiation, identification of forms through the dynamical chaining of different profiles
- forms are intrinsically 'transposable' (transposition does not mean a two-step process, going from a field A to another field B: it refers to the *immediate availability* of an organizing 'scheme' in an open variety of domains)
- 'schemes' are not formal types, as in logical approaches, but 'potentials' to be actualized, evolving through practice.

Last, but not least, there appears in gestaltist writings, notably those by the so-called 'microgenetic' schools (Werner 1956; Flavel and Draguns 1957; cf. also Kanizsa 1991: 118; V. Rosenthal, this volume), that forms are to be considered as the result of dynamical *stabilization processes*, i.e. as units in an ongoing continuous flow, comprising more or less stable 'phases', depending on the moment and on the part of the flow. Of course, for lack of mathematics and physics, it was only possible to develop these concepts of stability/instability as from the 1960's. This more recent aspect of the theory is essential for the theory of semantic forms we want to build. Modern mathematical and physical concepts of instability, and recent advances in the theory of complex systems, allow us, not to modelize for the moment, but at least to conceive and to formulate a unified setting for language activity seen as *a construction of forms in a semantic field*[11]. Without taking into account such a notion of instability at the very heart of the linguistic theory, we would be obliged, either to drop the *immediate* link between language and action-perception (as logical approaches do), or to consider concrete, externally stabilized, referential uses as a first building layer (as cognitive linguistics mostly do). In all cases, this would imply the isolation of literal meanings, and the processing of all other uses by means of metaphor and metonymy (which strangely enough would admit at a later stage transformations such as mixing, deformations, etc. excluded from the first stage).

Let us see now how the dynamical principles we favor are redistributed in our theory.

4.1 MOTIFS

Let us first recall that we view *linguistic motifs* as unstable germs of semantic forms, which can be stabilized only by, and with: (i) the other constituents of surrounding syntagms, (ii) more distant elements of the co-text, and (iii) an ongoing context-and-topic. This stabilization process is not a 'simple' instantiation, but a recapture of the *motifs* by non immanent *profiling* dynamics, partly linked to specific semantic domains, partly constituted by generic grammatical means. All this process gives rise to the variety of lexical *profiles* (uses) of the words. Each *motif* blends, intertwines, different dimensions that can be dissociated only later (if ever) in the stabilization process, by inscription into a more specific semantic domain. Therefore a *motif* does not belong to a specific domain: on the contrary, it encompasses several ones (to the extent that 'semantic domains' can always be sharply distinguished from one another).

In a sense, *motifs* define the functional kernel of many linguistic units, whether monomorphemic or polymorphemic. Most importantly, these unstable 'germs' do not entirely control from the inside their own stabilization parameters, nor are they by themselves generative of the lexical values they *motivate*. Language activity has a polysystemic, multi-level organization, with strongly interacting and at the same time possibly uncoupled 'levels'. As a physical (thermodynamic) metaphor, this organization is not that of a homogeneous system, made of uniformly individuated and stabilized entities. It is that of a heterogeneous medium, with several coexisting more or less differentiated 'phases', ongoing phase transition, and diffusion/reaction processes.

More precisely, for our dynamical approach of the semantic reconstruction up to the level of text and discourse, we need:

• Coalescence and /or transaction between dimensions of meaning, the dissociation of which could only happen 'downstream' in a stabilization process in the co-text and context: this implies to introduce 'upstream', and constitutionally, a *structural instability* at the level of motifs (see a little lower in the text);

• Openness and immediate susceptibility of the linguistic motifs within the thematic and situational frame, allowing for a generalized form of indexicality (rooted in the themes of the discourse): because of this plasticity of the motif, and unpredictability of the exact part which is taken up at each occurrence, its internal organization has to be a complex, *chaotic* one (see lower in the text);

• Permanence of this type of organization through the traditional layers of integration (morphemes, words, phrases, texts).

In order to build such a theoretical linguistic concept, it is quite relevant to draw upon the various mathematical notions of instability. It even appears that we must go further than the Elementary Catastrophe Theory of R. Thom and E.C. Zeeman, from which are inspired the very few existing semantic models[12]. Pursuing the same lines, we can represent the participation of a given unit in the global construction of meaning (e.g. the contribution of this unit to the construction of an ongoing scene or scenario, or to a network of 'mental spaces') through a dynamical system operating in a certain semantic space, each state of which corresponds to a particular contribution of the unit. This dynamical system is coupled to certain parameters to be found in the co-text and in the context, and it controls in a reciprocal way some (other or the same) parameters in its semantic environment. If the analysis is situated at a microgenetic temporal scale, it is possible to postulate that the essential result of the construction is directed by the 'asymptotic', stabilized states of this dynamical system. In the right cases, the set of all these asymptotic states constitutes an *attractor set*, i.e. a region of the semantic space (a point, a cycle, or a more complex set, once called a 'strange attractor'), towards which converge all the trajectories, whatever their initial position in a wider region, called a *basin of attraction*. This attractor set represents a more or less complex state of the unit concerned, which may change according to the contextual parameters influencing the dynamics (and which also reciprocally influences these parameters). Thus, depending upon the contextual variations, a given attractor set can slightly move in the semantic space, without changing qualitatively its internal 'geometry' (structural stability). It can also change qualitatively, or even split up into several other different attractors ('structural instability', 'bifurcation'). In this way, a linguistic unit appears as a more or less unstable dynamical system, engaged in a reciprocal determination process with a certain part of the context. This 'deformation' process generally results in a more stabilized version of the initial dynamics, which drives the system into a certain attractor set, concentrating, so to speak, the resulting value, or *use*, of the unit. Hopefully, then, the modelizing process would consist in defining a *motif* as an unstable dynamical system, and in studying it relatively to a family of possible deformations (i.e. according to the different semantic fields and phrase constructions where the word appears in a corpus), so as to describe exhaustively the different cases of stabilization, as had been once done in other areas by the Elementary Catastrophe Theory.

Structural instability is one of the key concepts of the dynamical system theory. But there is another one, coming from the seminal work of D. Ruelle and F. Takens (1971), and which we have just alluded to (cf. Bergé and al. 1984; Dahan-Dolmenico and al. 1992; Ruelle 1993, 1996). Even if the ambient dynamic is stable, its asymptotic states can be very complex, because the corresponding attractor set itself has a very intricate topological structure, constituted by a bundle of dense, entangled trajectories, going through it in an unpredictable way ('strange attractor'). The attractor then represents a chaotic state, i.e. a global envelope of stabilization, which is accurately defined from an ideal geometrical point of view, but the trajectories of which cannot be known in their exact asymptotic evolution (unless the initial conditions are perfectly determined, and the computation 'infinite'). This important property, called 'sensitiveness to the initial conditions', defines a kind of 'stable turbulence', which is of a very high interest for our concept of *motif*, to the extent that a *motif* can be *promoted* as such by certain uses, in particular the 'figurative' or 'metaphoric' ones. In this model, the *promotion* of a motif corresponds to a chaotically organized state, which results in trajectories inside the semantic space the asymptotic evolution of which remains unpredictable.

We therefore see in which various meanings we need here to take up in semantics the mathematical concepts of instability:

First within the framework of a stable dynamics, comprising a chaotic attractor set, and consequently a kind of regional instability (allowing the promotion and the contextual elaboration of a motif, with fluctuating trajectories, and unpredictability of what is asymptotically integrated in this kind of use)

Secondly, in the framework of light global fluctuations of the global dynamical landscape, which do not imply important qualitative transformations (but only amplifications, or a kind of smoothing, inducing more simple or generic variants)

Thirdly, in the case of genuine structural deformations (structural instability), which modify the topology, and/or the number, of the attractor sets and their basins, and so reveal new principal contrasting dimensions, allowing a whole polysemic diversity of uses.

Let us underline that these phenomena can be simultaneously observed, depending on the dimensions on which the analysis is directed. Moreover, and this is most important, two dynamics can be topologically very similar, and even have exactly the same attractor sets, while strongly differing in their structural instability degree. When this dynamical setting is combined with a 'morphemic' conception of motifs (coalescence, transaction, between dimensions not yet dissociated at this level), several aspects of the construction of meaning, which are ordinarily presented as very distant ones, can be brought together without incoherence. Strange as it may appear, 'figurative' meanings appear very akin to the generic 'definitions' devised by lexical studies, and also to the generic 'potentials of meaning' brought out by linguistic theories. As a matter of fact, a generic *definition* of a motif promotes it through a global description of the topology of its attractor set, which reveals on its ground the intertwining of other linguistic motifs. While a *figurative* meaning promotes also the motif, not in a synoptic way, but rather by collecting some of its aspects along a largely unpredictable trajectory[13]. In both cases, the motif, as a dynamical chaotically organized unit, is perceived as such in the discourse – though in a more or less synoptic and global manner. What can be said, then, about the 'meaning potentials', which various linguistic theories postulate in order to introduce some kind of unity and generativity at the heart of a lexical unit ? In a dynamical setting like ours, such a 'meaning potential' is only another *structurally unstable* form of the motif, topologically very close to its chaotically stable ones (promoted by definition and/or figurative meanings). This structurally unstable form represents in our theory the generative potential of the corresponding linguistic unit, in as far as it is immediately available in an indefinite number of semantic lexical fields, through recapture and re-stabilization within their own dynamical frameworks. Each use then corresponds to a certain stabilization path. In this way, polysemy becomes a central and constitutive phenomenon in language organization and activity.

One sees therefore that it is possible to bring together in a unified setting deeply entrenched aspects of language activity, as well as more innovative ones: the key being to recognize at a theoretical level, and from the very beginning, a certain dynamical state, or semantic 'phase' (let us use here again the thermodynamic metaphor), which potentially combines the different forms of instability we have just mentioned. There remains now to see how this primordial instability is most of the time recaptured, and re-stabilized ('profiled') in order to construct the variety of semantic forms.

4.2 PROFILES

What do we then call *profiles*, or *profiling*?[14] Roughly speaking, *profiling*, which is of course context-and-situation dependent, occurs:

• by stabilization in lexical organizations (e.g. domains like music, cooking, sailing, architecture, business, law, mathematics; fields articulating several experiential domains and practices; denominative paradigms…)

• correlatively through grammatical units and constructions

• also through discourse organization (e.g. anaphors, comparisons).

From the point of view of the present theory of semantic forms, profiling implies:

• figure/ground repartition of the lexical content in semantic fields[15]

• possible dissociation of the involved motifs, through stabilization in co-text and context

• enrichment by new aspects, or on the contrary impoverishment of the involved motifs.

By this process, words (initially considered with all the 'morphemic openness' of their motif) become lexical units indexed on lexical classes, with more stabilized and individualized meanings[16]. Plasticity of the motifs through profiling is a key point in our theory. Depending on the reciprocal determination of the co-text and the context, some features can be completely neutralized, or on the contrary made salient. In many cases, some features are so to speak *virtualized*: they remain as a possible 'aspect' inside the dynamics of construction, without being explicitly integrated in the constructed forms. Nevertheless, they are as it were reserved, and can come back to the foreground if the discourse needs it afterwards. One of the reasons of these virtualization processes is that, by entering in a specific semantic domain to contribute to the formation of a lexical unit, a *motif* functions as a simple *motivation*: its proper contribution can be superseded by other afferent features, which are more important in this context. These features are either recorded in the lexicon, as a particular use of the word, or indexically integrated on the spot. But let us underline that even if these modulations of meaning are already registered in the lexicon, it is always the global stabilization dynamics in the current phrase, or in a larger co-text, and the peculiarities of the ongoing topic, which determine what exactly will be taken up from the lexical registration. Let us also underline that profiling is a *differential* process, which happens through contrasts and coordination between *several* inter-defining lexical units, which are the results of *reciprocal* stabilization paths.

At the level of a clause, lexical profiles stabilize through grammatical units and constructions, whose meanings stabilize correlatively at the same time. In this way, each statement appears as a *view* on the ongoing thematic organization, offering individuation, hierarchical structure, chaining, and grounding in the situation. In particular, a lexical profile can offer a certain view, or aspect, of a thematic unit. But this view is only a *characterization* of the unit: it cannot by itself decide what constitutes the thematic *identity* of the unit throughout the discourse[17].

The determination of a profile is not in the first place a matter of *type* instantiation, even if pre-recorded types can come into play. *Types*, in our view, are anticipations which pertain to the thematic level of language organization (like scenarios or 'actors'). The determination of profiles is performed, more fundamentally, by the mobilization of multiple frameworks which open the way to the thematization process Among the most current frameworks are the:

• modulation of specific differences of a lexical unit on the generic ground of a class: a lexical class appears as an area in a semantic space, where features, depending on the considered unit, circulate from the fore to the background (allowing, for example, metonymical shifts: *school* considered as a building, or as an institution)

• elaboration of functions and mereology, through lexicalization of parts and functions (a *gaming table* has *legs*, but a *table of contents* has not)

• exploration of the semantic neighborhood (synonyms, antonyms)

• fixation of an hyperonym, i.e. choice of a lexical unit bringing to the foreground some generic features of a given semantic class

• introduction of a scalar structure into a class (e.g. *few, many, too many*; *icy, cold, tepid, warm, hot*); more generally, introduction of a global 'geometrical' structure into a class (putting for example a week, generic value at the center, and a dense, emblematic *paragon* on the periphery[18])

• dissociation between processes, and roles or participants

• choice of a part of speech (nouns, verbs…)

• quantification, determination

• aspects, tenses, modalities

• constructions and grammatical functions.

As one can see, the problem of the construction of the lexicon, in its relation to the functional kernel of language (motifs and grammar), pertains indeed to the problematic of complex systems. First, the systemic variation is organized around unstable dynamics (here called motifs), which produce by being stabilized the diversity of profiles, whether new or registered. Secondly, there is a permanent adjustment of the system's categorization networks. The lexicon is not a set of labels, nor a nomenclature of concepts processed as such by the arrows of reference. It is the historical and heterogeneous result of a multitude of accesses to themes; these accesses are never registered alone, but in clusters, and at different depths of unification, stabilization, and exteriorization. The lexicon can only function because it is liable to establish in its own formats, and to register immediately, distinctions up to then original – which implies to weaken or 'virtualize' other already established distinctions, without loosing them. Lastly, the mobilization of motifs and profiles is aimed at the construction of thematic targets, which have their own structure. But language activity is not to be seen as a complete resorption of these semantic phases into a completely stabilized and/or externalized thematic level. It rather rests on the permanent co-existence of these different phases of meaning through the discourse.

4.3 THEMES

In order to complete the presentation of our theory of semantic forms, we must say at least a few words concerning the level of *thematic forms* and *thematic spaces* (recall that we take here 'thematic' in the full, literary sense of the word). At this level, the aggregation of profiles into thematic forms distributed throughout the text or the flow of speech (referring for example to narrative entities like actors, actions, and their transformations) is performed. In the same way as we have recalled in section 3 some principles from the phenomenological and Gestalt theory of perception and action, in order to transpose them into semantics, we should now come back to the phenomenological and semiotic theory of the *thematic field* (notably in A. Gurwitsch's work), in order to connect our theory of semantic forms, to the contemporary works on discourse, narratives, text semantics, etc. It would allow a criticism of the objectivist approaches, often correlated to referential semantics, and to the primacy of denominative uses. And it would also open the way to describing the new, original, motifs created by the discourse which elaborates, in a more or less innovative way, the pre-given linguistic motifs.

In the framework of the present paper, we shall limit ourselves to the following fundamental points:[19]

• In a situation of spontaneous speech, profiles are not perceived separately from the themes to which they give access, being nothing else than the transitory presentation of these accesses. The profiling dynamics cannot really enter into a stable state without a minimal thematic positioning, including the grasping of an ongoing topic. Profiling therefore depends constitutionally upon the global thematization movement.

• At the thematic level is carried out a global dynamics of construction and access to *themes* which are set as common objects of interest in the intersubjective field. Themes, in this sense, are partly externalized in our perceptive and practical world, as concrete objects or as effective actions. But this is only a partial aspect of their identity which is made, as already said, of an organized *history of profiled accesses* (e.g. an history developed in the structure of a scenario). Language opens on an exteriority which can be simultaneously sensible, imaginary, and ideal. We are here in a complete opposition to certain referential semantics, which pretend to favor concrete denominative uses, but actually have a very limited conception of what 'reference' means. To refer to a theme is not only to refer to its concrete facets, nor only to refer to its abstract, ideal, ones. Such conceptions do not allow to understand that language by nature addresses fiction as well as reality. Think for example of a chess game, and its pawns; the theme of the game is a synthesis of many different aspects; and necessary as they may be, pawns are a simple material substrate, invested by this whole thematic organization; or, more precisely, their visible and tangible configuration only defines a crucial perceptive facet of the ongoing theme – i.e. of the game.

• A thematic unit builds up its identity through a synthesis of successive profiles: an *actor*, for example, is identified by the open set of the participant profiles, which compose it from one clause to another, and define in this way its transformations and interactions with the other actors in an ongoing narrative (once schematized by the scripts and frames of the psychological semantics).

• The thematization activity can and must be understood at the semantic level of its linguistic accesses and effects, and without contradiction, as a global access to other of its textual, pragmatic, imaginary, conceptual, perceptive, and practical layers, which are less directly linguistic, but still semiotic (therefore cognitive and social at the same time).

The concept of *motif*, as we have seen, has allowed us to describe the functional kernel of language, and its unfolding in a permanently adjustable lexicon. But speech does not only stabilize, it also renews linguistic and lexical instability. Existing motifs are modulated, and new ones are sketched (even deeply elaborated), through discourses some of which are the starting point of an instituted modification, effectively registered in diachrony. It is therefore crucial that the dynamical structure of motifs (which is, so to speak, the most internal 'phase state' of language) allows an immediate interaction with the ongoing thematics. This kind of susceptibility makes it possible to index on an existing lexical unit a renewed motif, which condenses some essential dimensions of a new original theme, after having cut out part of its structure (e.g. its precise event structure). Of a prime importance are here the metaphoric innovations, and in a more commonplace manner, the uses mixing metonymic shifts and figurative operations[20].

This a complete reversal relatively to other theories, which start from an *ontology* conceived independently from language, or give primacy to the reference to a concrete, perceptible world, without asking what perception or practice consist of, when they are affected by language. *Ontologies* are complex thematic constructions, they are a very peculiar *result* of text, discourse, and other social practices, and not a universal starting point for semantics. Quite differently, we consider as a very important clue for the study of motifs in lexical semantics the figurative meanings, which precisely transgress ontological divisions. As we have said, we postulate that in many cases this kind of use *promotes* a linguistic motif, i.e. elaborates and puts it forward without absorbing it completely in a conventionalized lexical *profile*. We gave examples concerning nouns in section 2.

6. CONCLUSION: THE NATURE AND PLACE OF GRAMMAR IN A DYNAMICAL THEORY OF SEMANTIC FORMS

In this paper, we have systematically analyzed the principles according to which it is possible to build an analogy, and even a continuity, between language and perception. Starting from the case of prepositions, we identified several obstacles, or misleading choices currently made by cognitive linguistics. In particular:

• an erroneous model of perception, strangely disconnected from action, expression, and other essential dimensions of anticipation, leading to an inadequate separation between grammar and lexicon

• the non-taking into account of polysemy as a fundamental property of language

• an inability of the theory to allow the necessary interactions between the thematic developments in discourse, and the presumably most 'interior' level of language (the level of the so-called 'schemes' in cognitive linguistics).

In order to remedy all these deficiencies, we have introduced a more radically dynamical setting, which gives a fundamental role to the mathematical concepts of instability. On this basis, the construction of Semantic Forms can be distributed between three 'phases' named *motifs*, *profiles*, and *themes*. Indeed we claim that a theory of forms, suited to linguistics and susceptible to offer a coherent and global view on language activity, is possible only by introducing a diversity of concurrent semantic 'phase states', in a process made of structurally unstable or chaotic resources, and of partial stabilization dynamics (like in complex systems models).

In this way, we rejoin a Humboldtian conception of language, which considers it as an *energeia*, i.e. not as a finished product, but as a self-organized activity. This implies that we consider languages, not only as means to build (re)presentations, but also as capabilities of being immediately modulated, transformed, by their own activity. In order to better support this conception from a cognitive point of view, it appears necessary to come back to phenomenological and Gestalt theories of perception and action. In this way, the discussion is really opened on what can rightly be taken up again from them for semantics, while not forgetting the historical, social, and 'transactional' nature of what we have called linguistic motifs and lexical profiles.

We can now return to the question of the nature of grammar, and its relation to experience. We will stress the following points:

• Beware of the reduction of grammar to a universally pre-linguistically defined set of features. There should be NO prejudice concerning what the grammar of such or such language is: therefore NO 'a priori' or 'transcendental' approach, despite universal anthropological constraints. Think for example of the so-called 'classifiers' of many languages (Bantu, Amerindian, Australian), whose semantics and constructional properties do not fit well with the categories of the dominant Western-centered tradition. Let us not forget that grammars as well as languages are historical constructions, and that for a given language, grammatical routines are different depending upon types of discourses.

• Beware of the inadequate models of perception and/or schematism, and beware of an excessively focus on the relation to space. What matters first is the global framework of the perceptive and practical experience, apprehended from a point of view which, paradoxically, has to be bodily, subjective, *and* social. This subjective-and-social experience has also to be apprehended in its microgenetic structure.

• EXPERIENCE does not mean intuition + categories, like in a Kantian approach. It first means perception + action + expression. It is crucially made of anticipations, which are lived as such, and therefore are recursively anticipations of anticipations, etc.

• Grammar, being *the set of the most necessary and generic form-creating devices operating in language activity*, is:

– not a set of morphological axioms for a purely formal (logical or topological) intuition

– not a set of anticipations of an 'externalized' and 'stabilized' perception, taken in a narrow sense

• Grammar can be more safely compared to the set of the most necessary and generic linguistic anticipations of the subjective-and-social experience of speakers. Grammatical anticipations contribute to the stabilization of both the subjective and the objective sides of the utterance production dynamics, which constitutes at the same time the speaker/hearer and the current theme of interest.

• If experience and language activity are understood as a microgenesis of forms, we have to redistribute these generic anticipations among several concurrent 'phase states' of meaning arising in the thematization movement (i.e. in our theory, among *motifs, profiles, themes*). So the 'genericity' of grammar cannot be assigned to a unique level of stabilization and/or genericity: grammatical anticipations concern different microgenetic phases in the 'thematisation' dynamics.

• As a simplified working definition, grammar should be centered on the most generic 'gestalt-and-synoptic' aspects of the construction of Semantic Forms, i.e. on the generic *profile* of the Semantic Field (e.g. grammatical constructions). But actually, we have to take also into account the most generic dimensions of *motifs*, and the most generic *thematic* devices ('grounding' indexical markers, like deictic, determinant, etc.).

• Beyond the question of knowing what 'generic' and 'necessary' means, the grammatical/lexical distinction amounts to:

– the question of the proper appearance time of forms (synoptic vs developed in time)

– to an impossible clearcut distinction of strongly dissociated *units* (e.g. constituents) among the variety of semantic *forms*.

How is then the alleged unity of a word constituted – at least for those languages where the notion of 'word' is relevant? Our description makes it a *compromise* between three *concurrent* dynamic integration formats. At this level of the word (and even beyond in the case of compound lexical units), our theory puts in the center a 'phase state' of meaning, the instability of which (structural instability, instability in the sense of chaotic structures) can be described as *morphemic*. It makes possible the coalescence of dimensions which can be dissociated only further in a stabilization process, and thus radically differentiates *motifs* from what other theories call *types*. Motifs are generic in a specific sense, since they allow homogeneous thematic developments, as well as heterogeneous thematic dissociations or blendings, as in figurative meanings. From this point on, the question of polysemy can be redistributed in a new way among the three postulated meaning 'phases'. It is also possible then (see Cadiot and Visetti, 2001 ch. 4; also 2001b) to take into account the immediate interaction between the ongoing discourse, and the linguistic anticipations registered at these three levels, whether in the time of a conversation, or of a literary work, or also in the general evolution of language in diachrony.

Prof. Yves-Marie Visetti
Ecole Normale Supérieure
1 rue M. Arnoux
92120 Montrouge, France

NOTES

[1] Under the French heading *linguistique de l'énonciation*, we mean a linguistic current which can be traced back to K. Bühler (1934), through the work of E. Benveniste (1966/1974), and more recently, through the important contributions of A. Culioli (1990/1999), O. Ducrot (1984), J.C. Anscombre and O. Ducrot (1983).

[2] For a full presentation, see our book: *Pour une théorie des formes sémantiques – Motifs, Profils, Thèmes* (Cadiot & Visetti, 2001). See also Cadiot 1999a, 1999b, 2002; Cadiot & Visetti 2002a, b; Visetti & Cadiot 2000, 2002.

[3] As a matter of fact, this second trend is now developing towards a better acknowledgement of the role of action and its anticipations (cf. Vandeloise 2001, and other papers in the same Workshop). In Cognitive Linguistics, until recently, this kind of analysis remained limited to the theory of grammatical constructions and to the semantics of verbs (Fillmore's Case Grammar, Construction Grammar, cf. Goldberg 1995). At the present time, it has evolved to encompass other categories of words, like prepositions, by resorting to so-called 'functional' features.

[4] It draws on several recent works on prepositions (Cadiot 1991, 1997, 1999b).

[5] This section is taken from Cadiot 2002.

[6] Some among the most important references to the authors quoted in this paragraph are given in the References section. An excellent introduction to Husserl's phenomenology can be found in Salanskis (1998). For an introduction of some of Vygotsky's ideas in cognitive linguistics, cf. Sinha and Jensen de Lopez (2000).

[7] Readers interested in having more details on phenomenology and perception, in the perspective of a transposition in the field of semantics, may refer to our book (2001: particularly chap. 2).

[8] But even in dealing with vision, we should not forget that there is the 'seeing as' phenomenon: the way in which we *see* things depends on the way we *name* them. Cf. for instance several papers in *The 2nd Annual Language and Space Workshop*, University of Notre Dame, June 23-24 (L. Carlson, E. van der Zee, ed.): Smith; Richards & Coventry; Tversky & coll.

[9] Cf. Cadiot 1999a; Cadiot and Nemo 1997a,b,c; Nemo and Cadiot 1997; Cadiot and Tracy 1997; Visetti and Cadiot 2000; Cadiot and Visetti 2001b, 2001: ch. 3, section 3.1; see also Tracy 1997; Lebas 1999.

[10] For a reconstruction of Gestalt theory, and its assessment in the contemporary field of cognitive sciences, cf. Rosenthal and Visetti, 1999, 2003. For a presentation and illustration of a general dynamical paradigm in cognitive sciences, see Port and Van Gelder (1995), and most of all, J. Petitot's works quoted in the References section. See also Petitot, J., Varela, F., Pachoud, B. and Roy, J.-M. (eds) 1999.

[11] Far beyond the remarkable insights of the historical gestaltists, we see now mathematicians, physicists, biologists, computer scientists, modelizers in cognitive, social, ethological and ecological sciences, lay the foundations of a framework crossing their particular domains, and in which questions of stability and instability, invariant and variation, regulation and viability, can be deeply re-thought, and sometimes modelized. The following titles make it somewhat explicit: multiple spatial and temporal scales (at least two, micro- and macro-); importance of the topological, dynamical, and statistical characteristics; reciprocal determinations of local and global aspects; multiple dynamics for the formation of units (births, deaths, coalitions, etc.); co-existence of several dynamical 'phases'; adaptation, and active preservation of the internal and external viability domains; natural drift by coupling with a proper environment; behavioral repertoire organized around *unstable* dynamical processes, which constitute the system's functional kernel. On the whole,

all the system's characteristics are historically determined… Given the great variety of the fields and the models involved, we cannot do better than referring the interested reader to the site of the Santa Fe Institute (www.santafe.edu), and to the entire series of the SFI's Studies in the Sciences of Complexity. See also Weisbuch (1991). For a philosophical analysis of this paradigm shift, cf. Cilliers (1998).

[12] Examples of semantic models based upon Elementary Catastrophe Theory can be found in Thom (1974) or Zeeman (1977); in Brandt (1986), Petitot (1985, 1992, 1995), or Wildgen (1982); and more recently, with different theoretical orientations, in Piotrowski 1997), or Victorri and Fuchs (1996).

[13] This being said, the event of a figurative meaning does not only involve the level of motifs; it also implies processes at the thematic level: e.g. *blendings*, according to Fauconnier (1997) or Fauconnier and Turner (1999).

[14] We use the same term as Langacker (1987), but in a different theoretical framework. There is no theory of instability in Langacker's cognitive grammar. Furthermore, we have already criticized the strictly 'configurational' schematism he makes use of at the level of grammar. Lastly, we do not have the same conception of the 'thematic' level, nor of the alleged primacy or typicality of physical uses.

[15] The ground of a semantic field corresponds to its most generic features, and also to some more specific, but less relevant or salient ones, when the field is dynamically stabilized by the occurrence of a specific lexical profile (playing here the role of a figure).

[16] Not all words, however, possess a specific motif. Numerous technical terms are actually words indexed in a unique specific domain, which furthermore are very rarely used in a figurative meaning (examples chosen at random in a dictionary: *galvanoscope, gastritis, gasoline*). Of course, speech can always unlock the semantic game, and invent new meanings, which imply the creation of new (most of the time transitory) motifs. As an exercise, try for instance to say to your best friend: *You are a real gastritis*, or *You are my favorite gasoline*, and see what happens.

[17] Take for example a cooking recipe: the identity of the chicken (the central actor of the ongoing scenario) remains the same throughout. And nevertheless, its profiles change constantly, from the market up to the plate.

[18] A same word can possibly be placed in either position, e.g. the word *street* which functions according to the case as the generic term of the paradigm of urban ways (avenue, boulevard, lane, etc.), and as a kind of 'parangon' in denser (at the same time metonymical and somewhat figurative) meanings, like *to run about the streets, to find oneself out on the street, to come down into the street, man in the street…*

[19] For more details, see Cadiot and Visetti (2001: ch. 3, section 3.2.3); Visetti and Cadiot (2002, section 4.3).

[20] Cf. for example Fauconnier and Turner (1999), Coulson and Fauconnier (1999).

REFERENCES

ANSCOMBRE, J.C. and DUCROT, O. 1983. *L'argumentation dans la langue*. Liège/paris: Mardaga.

BENVENISTE, E. 1966/1974. *Problèmes de Linguistique générale*. (2 Vol). Paris: Gallimard. Partial Transltion 1973. *Problems in General Linguistics* (Miami Linguistics, No 8). Miami: Miami University Press.

BERGÉ, P., POMEAU, Y. and VIDAL, C. 1984. *Order within Chaos: Toward a Deterministic Approach to Turbulence*. NY: Wiley.

BRANDT, P.A. 1992 [1986]. *La charpente modale du sens: Pour une Sémiolinguistique Morphogénétique et Dynamique*. Amsterdam: John Benjamins.

BÜHLER, K. 1990 [1934]. *Theory of Language; The Representational Function of language*. Translated by D.F. Goodwin. Amsterdam: John Benjamins.

CADIOT, P. 1991. *De la grammaire à la cognition, la préposition POUR*. Paris: Editions du CNRS.

CADIOT, P. 1997. *Les prépositions abstraites en français*. Paris: Armand Colin.

CADIOT, P. 1999a. "Principe de conformité et génération analogique en sémantique nominale." *Verbum* 31: 383-407.

CADIOT, P. 1999b. "Espaces et prépositions." *Revue de Sémantique et pragmatique*, 6: 43-70.

CADIOT, P. 2002. "Schemas and Motifs in the Semantics of Prepositions." in *Prepositions in their syntactic, Semantic and Pragmatic Context*, S. Feigenbaum and D. Kurzon (eds): 41-57. Amsterdam, John Benjamins.

CADIOT, P. and NEMO, F. 1997a. "Propriétés extrinsèques en sémantique lexicale." *Journal of French Language Studies*, 7: 1-19.

CADIOT, P. and NEMO, F. 1997b. "Pour une sémiogenèse du nom." *Langue Française*, 113: 24-34.

CADIOT, P. and NEMO, F. 1997c. "Analytique des doubles caractérisations." *Sémiotiques*, 13: 123-145.

CADIOT, P. and TRACY, L. 1997. "On n'a pas tous les jours sa tête sur les épaules." *Sémiotiques*, 13: 105-122.

CADIOT, P. and VISETTI, Y.M. 2001. *Pour une théorie des formes sémantiques: motifs, profils, thèmes*. Paris: Presses Universitaires de France.

CADIOT, P., VISETTI, Y.M. 2002a. “Motifs, profils, Thèmes: une approche globale de la polysémie.” To appear in *Cahiers de Lexicologie*.

CADIOT, P., VISETTI, Y.M. 2002b. “Motifs linguistiques et construction des formes sémantiques. Schématicité, généricité, figuralité”. Actes du colloque *Representations du Sens linguistique*, LINCOM Europa, LINCOM Studies in Theoretical Linguistics.

CASSIRER, E. 1946 [1924]. *Language and Myth*. Translated by S. Langer. Reprint by Dower Publications.

CASSIRER, E. 1953 [1923]. *The Philosophy of Symbolic Forms. Vol 1: Language*. Translated by R. Manheim. New Haven: Yale University Press.

CILLIERS, P. 1998. *Complexity and Postmodernism. Understanding Complex Systems*. London: Routledge.

COULSON, S. and FAUCONNIER, G. 1999. “Fake Guns and Stone Lions: Conceptual blendings and Privative Adjectives.” In *Cognition and Function in Language*, B. A. Fox, D. Jurafsky and L. Michaelis (eds), 143-158. Stanford, California: Publications of the CSLI.

CULIOLI, A. 1990/1999. *Pour une linguistique de l'énonciation* (3 Vols). Paris: Ophrys.

DAHAN-DOLMENICO, A., CHABERT, J.-L., and CHEMLA K. (eds) 1992. *Chaos et déterminisme*. Paris: Seuil.

DUCROT, O. 1984. *Le dire et le dit*. Paris: Minuit.

FAUCONNIER, G. 1997. *Mappings in Thought and Language*. Cambridge, MA: Cambridge University Press.

FAUCONNIER, G., TURNER, M. 1999. “Metonymy and conceptual integration.”, in *Metonymy in Language and Thought*, K.U Panther and G. Radden (eds), 77-90. Amsterdam: John Benjamins.

FLAVEL, J. and DRAGUNS, J. 1957. “A Microgenetic approaach to perception and thought.” *Psychological Bulletin*, 54, 3: 197-216.

GIBSON, J.J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin. Reprint Hillsdale, N.J.: Lawrence Erlbaum Associates, 1986.

GOLDBERG, A. 1995. *Constructions. A Construction Grammar approach to Argument Structure*. Chicago: Chicago University Press.

GURWITSCH, A. 1964. *The Field of Consciousness*. Pittsburgh: Duquesnes University Press.

GURWITSCH, A. 1979. *Human Encounters in the Social World*. Pittsburgh: Duquesnes University Press.

HUSSERL, E. 1982 [1913]. *Ideas pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book: General Introduction to Pure Phenomenology*. Translated by F. Kersten. The Hague: M. Nijhoff.

HUSSERL, E. 1989 [1912-1928]. *Ideas pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Second Book: Studies in the Phenomenology of Constitution*. Translated by R. Rojcewicz and A. Schuwer. Dordrecht: Kluwer Academic Publishers.

HUSSERL, E. 1973 [1939]. *Experience and Judgement*. Translated by J.S. Churchill and K. Ameriks. London: Routledge and Kegan Paul.

KANIZSA, G. 1991. *Vedere e pensare*. Milan: Il Mulino.

KANT, I. 1999 [1781/87]. *Critique of Pure Reason*. P. Guyer and A.W. Wood (eds). Cambridge: Cambridge University Press.

KÖHLER, W. 1920. *Die physischen Gestalten in Ruhe und im stationären Zustand*. Berlin: Braunschweig. Excerpts in *A Source Book of Gestalt Psychology* 1938. W. D. Ellis, (ed.). London: Kegan Paul.

KOFFKA, K. 1935. *Principles of Gestalt Psychology*. New-York: Harcourt Brace.

KÖHLER, W. 1964 [1929]. *Gestalt psychology*. New York: Liveright.

KÖHLER, W. 1938. *The Place of Value in a World of Facts*. New-York: Liveright Pub.

KÖHLER, W. 1971. *Selected Papers*. Ed. by Mary Henle. New-York: Liveright.

LAKOFF, G. 1987. *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

LANGACKER, R. 1987, 1991. *Foundations of Cognitive Grammar* (2 vols). Stanford: Stanford University Press.

LANGACKER, R. 1999. *Grammar and Conceptualization*. Berlin/New-York: Mouton de Gruyter.

LEBAS F. 1999. *L'indexicalité du sens et l'opposition 'en intension' / 'en extension'*. PhD Thesis. Paris: Université Paris 8.

MERLEAU-PONTY, M. 1962 [1945]. *Phenomenology of Perception*. Translated by C. Smith. London: Routledge and Kegan Paul.

MERLEAU-PONTY, M. 1973 [1969]. *Prose of the World*. Translated by J. O'Neill. Northwestern University Press.

MICHOTTE, A. (1946). *La perception de la causalité*. Louvain: Études de Psychologie, Vol. 8.

NEMO, F. and CADIOT, P. 1997. “Un problème insoluble ?”. *Revue de Sémantique et Pragmatique*, 2: 9-40.

PETITOT, J. 1985. *Morphogenèse du sens*. Paris: Presses Universitaires de France. English version to appear 2001. Bern: Peter Lang.

PETITOT, J. 1992. *Physique du sens*. Paris: Editions du CNRS.

PETITOT, J. 1995. "Morphodynamics and Attractor Syntax. Dynamical and Morphological Models for Constituency in Visual Perception and Cognitive Grammar." In *Mind as Motion: Explorations in the Dynamics of Cognition*, R. Port and T. Van Gelder (eds), 227-281. Cambridge: MIT Press.

PETITOT, J., VARELA, F., PACHOUD, B. and ROY, J.-M. (eds) 1999. *Naturalizing Phenomenology. Issues in Contemporary Phenomenology and Cognitive Science*. Stanford: Stanford University Press.

PIOTROWSKI, D. 1997. *Dynamiques et structures en langue*. Paris: Editions du CNRS.

PORT R. and VAN GELDER, T. (eds) 1995. *Mind as Motion: Explorations in the Dynamics of Cognition*.

RASTIER, F. 1987. *Sémantique interprétative*. Paris: Presses Universitaires de France.

RASTIER 1997 [1989]. *Meaning and Textuality*. Translated by P. Perrone and F. Collins. Toronto: University of Toronto Press.

RASTIER, F. 2000. "On Signs and Texts: Cognitive Science and Interpretation." In *Semiotics as a Bridge between the Humanities and the Sciences*, P. Perron et al. (eds), 409-450. New-York/Toronto: Legas Press.

RASTIER, F. 2001 [1994]. *Semantics for Descriptions*. To appear. Chicago: University of Chicago Press.

ROSENTHAL, V. and VISETTI, Y.M. 1999. "Sens et temps de la Gestalt." *Intellectica*, 28: 147-227.

ROSENTHAL, V. and VISETTI, Y.M. 2003. *Köhler*. Paris: Les Belles Lettres.

RUELLE, D. and TAKENS, F. 1971. "On the nature of turbulence." *Communications in Mathematical Physics* 20: 167-192

RUELLE, D. 1993. *Chance and Chaos*. Princeton: Princeton University Press.

RUELLE, D. 1996. *Turbulence, Strange Attractors, and Chaos*. World Scientific Series on Nonlinear Science, Series A, Vol. 16. World Scientific Publishing Co.

SALANSKIS, J.-M. 1994. "Continuity, Cognition, and Linguistics.", In *Continuity in Semantics*, C. Fuchs and B. Victorri (eds), 127-153. Amsterdam: John Benjamins.

SALANSKIS, J.-M. 1998. *Husserl*. Paris: Les Belles Lettres.

SINHA, C. and JENSEN DE LOPEZ, K. 2000. "Language, Culture and the Embodiment of Spatial Cognition." Cognitive Linguistics 11-1/2:17-41.

TALMY, L. 2000. *Towards a Cognitive Semantics* (2 Vols). Boston: MIT Press.

THOM, R. 1980 [1974]. *Modèles mathématiques de la morphogenèse*. Paris: Christian Bourgois.

TRACY, L. 1997. "La clé du mystère: mettre le référent à sa place." *Langue Française*, 113: 66-78.

VANDELOISE, C. 1991 [1986]. *Spatial prepositions: A Case Study for French.* Translated by A. Bosch. Chicago: University of Chicago press.

VANDELOISE, C. 2001. "Force and function in the development of the preposition IN.". *The 2nd Annual Language and Space Workshop*, University of Notre Dame, June 23-24 (L. Carlson, E. van der Zee, ed.)

VICTORRI, B. and FUCHS, C. 1996. *La polysémie – construction dynamique du sens*. Paris: Hermès.

VISETTI, Y.-M. 1994. "Les modèles connexionnistes entre perception et sémantique." *Sémiotiques* 6-7: 15-48.

VISETTI, Y.-M. 2001. "Constructivismes, émergences: une analyse sémantique et thématique.". To appear in *Des lois de la pensée au constructivisme*, M.J. Durand (ed.). Paris: Editions de la Maison des Sciences de l'Homme.

VISETTI, Y.-M. and CADIOT, P. 2000. "Instabilité et théorie des formes en sémantique – pour une notion de motif linguistique." *TLE (Théorie, Littérature, Enseignement)*, 18: 137-169. Vincennes/Paris: Presses Universitaires de Vincennes.

VISETTI, Y.-M., CADIOT, P. 2002. "Instability and the Theory of Semantic Forms" in *Prepositions in their syntactic, Semantic and Pragmatic Context*, S. Feigenbaum and D. Kurzon (eds): 9-39. Amsterdam, John Benjamins.

VYGOTSKY, L. 1978 [1930]. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

VYGOTSKY, L. 1986 [1934]. *Thought and Language*. Cambridge, MA: Harvard University Press.

WEISBUCH, G. 1991. *Complex Systems Dynamics*. Santa Fe Institute Studies on the Sciences of Complexity. Addison Wesley.

WERNER, H. (1956). Microgenesis and aphasia. *Journal of Abnormal Social Psychology*, 52, 236-248.

WERTHEIMER, M. 1959 [1945]. *Productive thinking*. New York: Harper.

WILDGEN, W. 1982. *Catastrophe Theoretic Semantics. An Elaboration and Application of René Thom's Theory*. Amsterdam: John Benjamins.

ZEEMAN, E.C. 1977. *Catastrophe Theory: Selected Papers 1972-1977*. Mass: Addison-Weslay.

MANFRED WIMMER

EMOTION-COGNITION INTERACTION AND LANGUAGE

1. BASIC ASSUMPTIONS CONCERNING EMOTION - COGNITION INTERACTION

Analysing the triangle consisting of emotion - cognition - language from an evolutionary point of view it seems evident that emotion and cognition are much older than language. A lot of species act in a way, which we are accustomed to call cognitive or affective, without any indication of language.

Concerning the phylogenetic priority of emotions or cognitions a problem arises which highly depends on the definitions of emotion and cognition as well as on the related jargon of disciplines. For instance the so called cognitive theories of emotions (e.g. Schachter/Singer 1962; Schachter 1966; Lazarus 1982) are favouring the priority of cognitions. From this point of view each behavioral act is initiated by a cognitive process (conscious or unconscious) which can be followed by emotions and/or actions. Within these concepts the term cognition indicates each process of perception, evaluation, categorization etc., what makes the underlying concept of cognition very broad. These prior cognitive processes are followed by emotions. Opposed are these concepts by positions giving priority to emotions (e.g. Zajonc 1980). From this point of view organisms are always evaluating relevant aspects of the "Umwelt" within basic emotional classification schemes, leading to "affective judgements" which are followed by cognitive articulations and interpretations. The affective dimensions are often unconscious and have serious influence on the subsequent cognitive processes.

In this article a position is favoured, which postulates the unity of affect and cognition and tries to find a way between the priority debates. (see Wimmer 1995) Concerning the mentioned triangle of emotion - cognition - language I will start with the assumption that these three capacities are closely interrelated. The closest and oldest ties seem to exist between emotion and cognition - what will be the core issue of the first part of this article. Language is considered as being based upon these basic layers.

It will be one of the major goals to demonstrate the effects of the older layers of being (as there are emotions and cognitions) to the younger products of evolutionary processes (as there is language) as well as the large number of occuring reciprocal feedback processes. This means, that also the underlying and phylogenetically older cognitive and affective mechanisms seriously changed with appearing language capacities.

Concerning the forthcoming remarks and conceptual ideas one further very general point has to be underlined:

Taking into account evolution, means to take into account history, diachrony. One serious problem arising from this attempt is that sharp definitions - if they are seen within a time dimension - loose their clear borders. This will be demonstrated in relation to emotion, cognition as well as language. To prevent confusion, I will try to use a terminology always adequate to the level of analysis.

A. Carsetti (ed.), Seeing, Thinking and Knowing, 277-292.

In a first step I want to approach the core issues from a broad and basic perspective, focusing especially on cognition and the relations between emotions and cognition. This perspective is a general overview of affect-cognition interaction providing a solid ground for a more detailed analysis of language related issues. The evolutionary perspective on emotion-cognition interaction provides a view, where both appear as inseparably intertwined offsprings of one common root. (Wimmer 1995).

Forming a basic hypothesis and paraphrasing a citation by I. Kant:

"Affects without cognitions are blind and cognitions without affects are empty". (Wimmer 1995, p 41)

What does this mean and what contributes this hypopthesis to language related issues?

The core of the argument is the assumption, that from an evolutionary-phylogenetical viewpoint the distinction between affect and cognition seems to be artificially drawn leading to wrong conclusions. The sharp distinction between affect and cognition has deep roots in our cultural heritage, leading back to ancient greek philosophy. (comp. Averill 1996; Gardiner/Metcalf/Beebe-Center 1937) Beside these historical roots also recent neuroanatomical and neurophysiological research indicates a distinction between brain areas and mechanisms responsible for affective and cognitive processes. (Panksepp 1998; MacLean 1990) In contrast to these considerations an evolutionary approach leads to the assumption, that there exists one common root of emotion and cognition. A root, which in it's early and very basic form is very close to homoeostatic - regulatory mechanisms. The root metaphor is very helpful in proposing a picture of one common root, which branches off in different branches always remaining closely related to the basic root.

Even (in phylogenetical dimensions) the very young ability of language usage can be traced back to this basic root.

There are different attempts trying to find the roots of cognition and emotion. Some of the most prominent ones will be sketched out briefly.

1.1. COGNITION

The idea of grounding cognition in a broad evolutionary framework is quite old and K. Lorenz's statement, that "*...life itself can be considered as a knowledge gaining process"* is one of the basics of Evolutionary Epistemology. (Lorenz in Weiss 1971, p. 231)

These considerations also have their forerunners, especially E. Mach, L. Boltzmann, H. Spencer and even Ch. Darwin. (see Oeser 1993)

This means, that life processes to establish and maintain living structures have to take into account specific internal and external conditions (to get "informed" by these conditions, as Lorenz says), what leads to adequate reactions or adequate behavior. In general, organismic structures are considered as results of learning processes, especially from "genetic learning".

For a biological foundation of affect - cognition interaction it is important to mention, that almost all information processing mechanisms have a phylogenetic base the so called "mechanisms exploiting instant information" - as described by Lorenz (1981, p. 221f). These mechanisms are the core concept in Lorenz attempt of naturalising a priori cognitive structures. Their main functions can be found in processing information, without storing these information. (see Wimmer 1998, p. 125f).

Cognition in this sense enables an organism to perceive differences between different internal or external events within a genetically based frame. Lorenz's example for these mechanisms are regulatory circles with feedback loops. I am aware, that this usage of the term cognition is extremely broad, but for an evolutionary argumentation it is necessary to start with such a comprehensive definition which does not reduce cognition to symbolic, conscious processes, but finds its roots in organic, regulatory mechanisms.

The elements or the content of perception (the differences that are perceived) depends on the quality of the operating level (Oeser 1987, p. 21) and can be directed more toward 'internal' or 'external' events.

Beside this "grounding process" of cognition there is another very prominent and radical grounding process of cognition: Jean Piaget's Genetic Epistemology. Piaget's Genetic Epistemology is another type of naturalistic account considering "...life itself a self-regulatory process...". (Piaget 1967) Here also the roots of cognitions are found in basic organic processes.

Compared with Evolutionary Epistemology, Genetic Epistemology contains much more constructivistic, self-regulatory elements. Genetic learning in Piaget's perspective is less important putting major emphasis on self-regulatory processes. The development of cognitive structures is considered as a dialectical process of states in balance whose disturbances can lead to reorganisation processes leading to improved and more stabilised states. This concept of progressive growth is a result of Piaget's work on human ontogenetic development. (Piaget 1967; Piaget/Inhelder 1977)

1.2. AFFECTS

From my point of view you can trace back the roots of emotions as far as the roots of cognition. Probably this seems a bit ridiculous. Following the definition of emotion, as it is proposed by Kleinginna/Kleinginna (1981) the elementary emotions like rage, fear, pleasure etc. consist of the following constituents:

1) phenomenological component (subjective experience);
2) cognitive processes (evaluation, classification processes);
3) physiological processes (neurochemical pathways; neuroanatomical structures);
4) expressive component (mimics, gesticulation...).

Within a phylogenetic frame it seems obvious, that phenomenological as well as expressive components are quite "young", while cognitive and physiological processes appear as the older parts, which I would like to call *precursors of emotions.*

The basic physiological conditions form something like a "central - internal state" which has serious influence on all "higher" cognitive activities. (see Changeux 1984; Damasio 1994) Concerning perceptive processes, this view proposes that perception never is neutral and without any affective coloring, but always influenced by the basic physiological state which determines what in the surrounding conditions gets importance or relevance. So perception is always biased by the underlying internal state.

Anticipating some considerations which will be elaborated in more detail later, the affective part of behavior can be seen in the basic parameters of regulatory processes which have to be kept or - in the case of disturbances - be reestablished. The cognitive components constitute the perceptive and motoric aspects of behavior. Both descend from a common root, which gets differentiated within phylogeny, but which never gets separated. (see Wimmer/Ciompi 1996) Especially a phylogenetic reconstruction leads to the hypothesis, that emotion and cognition are inseparable components of one overlapping process leading to the above formulated hypothesis.

To make all these quite abstract considerations more clear, I will demonstrate the emotional and cognitive components in basic behavioral pattern as there are Kinesis and Taxis behavior.

1.3. KINESIS

Kinesis behavior - as a very simple form of behavior in unicellular organisms appears for the observer as increase of locomotoric activities of the organism coming to a region of worse environmental conditions. This behavior can be analysed quite well at a physiological level. Disturbance of the physiological homoeostasis (which may be perceived by specific receptor organs) leads automatically to increased locomotoric activities which get reduced if homoeostasis gets reestablished. If there is cognition, than just in this basic sense - as mentioned in relation to Lorenz and Piaget.

The organismic structure gets to *know* the changes of homoeostasis and compensatory mechanisms (e.g. locomotion) are induced. What the organism perceives is a change of the internal conditions. Something like an “outside world” does not exist.

Obuchowski (1982) describes this behavioral pattern as *“homoeostatic code”* characterised by the following elements:

1) information reaches the organism without the mediation of receptor organs;
2) behavior is modified by homoeostatic changes;
3) information is not identified by a specific apparatus;
4) the organism is totally involved in perception as well as in action (holistic reaction) (Obuchowski 1982, p. 236).

The organism exists within two possible internal states: homeostasis and disturbance. Each of these states is characterised by specific sensoric and motoric activities.

1.4. TAXIS

More complex appears Taxis behavior. Compared to Kinesis the major changes are the increasing number of internal states. As mentioned above Kinesis is characterized by the very global states balance and disturbance of balance -without contentspecific “markers”. Analysing Taxis behavior leads to the assumption, of an increasing number of internal states, which are characterized by specific physiological states and behavioral tendencies (e.g. hunger, thirst etc.) This appears for example as action toward or away from a specific stimulus (e.g.light/darkness).

A further important point is, that the internal state influences the general readiness for action and behavior makes a first step away from stereotypic, reflex-like action patter. In other words - the internal state “decides” if a behavioral program gets activated.

For example - if a turbellarian worm (Strudelwurm) smells food he does not react automatically because if he is satiated no reaction follows. The reflex-like stereotypic responses get more complex and the actual internal state (in this case the state “saturation”) determines perception in the way, that the relevant stimuli are even not perceived, or without any significance.

What these basic forms of behavior contribute to the question of emotion -cognition interaction is the following:

What gets perceived (even a disturbance of homoeostasis) just gets perceived in relation to an underlying element or internal state (what later will be called “centrating base”). At these levels of phylogeny these internal states (or centrating bases) are mainly physiological.

I myself propose to see affective and cognitive dimensions in centrating and decentrating components. Centration and decentration are terms coming from Piaget’s Genetic Epistemology, where centration means a dominance of subjective -individual perspectives. Decentration means the ability of reflecting on these privileged perspectives.

What I have in mind is to put decentration and centration within a broader phylogenetic frame, far away from conscious reflections. Centration and decentration are considered as *functional principles,* which remain the same at different levels of phylogenetic development. They are mainly descriptive terms and have the advantage not to *“substantialize”* or even *localize* emotion and cognition. They offer a perspective, where emotion and cognition appear as dynamic components of a more general process. They demonstrate the two different tendencies of organismic activities, one including the direction toward existing structures and their maintenance, the other tending toward changes and differentiation of these structures.

Kinesis demonstrates a clear dominance of centrating components, the organism being totally inclined in its own state and “being open” just for very few environmental influences.

Taxis behavior demonstrates increasing influence of decentrating tendencies leading to a much wider spectrum of environmental events that get relevance. For example having more specified sensory organs which allow a directed movement to (or away from) a specific stimulus.

In general the main function of the internal state can be considered as a mediator between sensory surface and behavioral (motoric) programs.

In more detail these functions are:

1) to motivate (affects as energizers or motors of cognitive activities);

2) to sensitize receptor organs;
3) to evaluate the input in relation to the internal needs and preferences;
4) storage effect - especially the hippocampus formation in the brain is essential for the so called affective memory;
5) organisation of cognitive elements: cognitive elements are stored depending on the actual affective background. This leads to so called integrated feeling - thinking - behaving programs. (Ciompi 1982; Ciompi 1996; Wimmer/Ciompi 1996).

I hope, this short excursion to the area of basic behavioral pattern was sufficient for clarifying the former mentioned hypothesis of affect - cognition interaction. Elaborating this base was necessary for understanding the preconditions of language.

2. EVOLUTIONARY APPROACH TO THE ORIGINS OF LANGUAGE

An evolutionary perspective tries to put language and it's precursors within a continuum of evolutionary processes. Following the main ideas of Cassirer (1923,1925), Langer (1965) and Buehler it seems plausible, to make a distinction between three major steps of language acquisition. These different stages can be interpreted in evolutionary terms - what will be explained in more detail below. For clarification it has to be emphasized, that in the following considerations I will use the term speech as the base of language. The term speech covers all the biologically given preconditions (sensorial, motorically, central processes etc.) providing the base of language production. Upon these preconditions language develops in the sense, that it enables the organisms to participate on a collectively used set of symbols and signs. (see Langer 1972, p. 325 f).

So language can be considered as an instrument of speech. (see Lachmann 2000) Following Cassirer three major steps will be distinguished:
1) speech as mimetic expression (sinnlicher Ausdruck);
2) speech as analogic expression (anschaulicher Ausdruck);
3) language as symbolic expression (begriffliches Denken) (Cassirer 1923, p. 124ff).

2.1. SPEECH AS MIMETIC EXPRESSION

This stage is characterised by a direct connection between the expressive phenomenon and the internal state. E.g. the uncontrolled cry in pain, or fear belongs to this stage.

At this stage all more abstract- symbolic representations are missing. Nevertheless these expressions can be used as a kind of signal, which gets interpreted by the conspecific, leading to adaptive behavior. It was especially Darwin who dealt with these aspects of speech, leading to the result that the expression of emotions is a weakened action. The intended action itself (for example biting) is no longer performed in the primordial way. It gets replaced by a weakened response, by just demonstrating the teeth. (Darwin 1872)

"…with mankind some expressions, such as the bristling of the hair under the influence of extreme terror, or the uncovering of the teeth under the furious rage, can hardly be understood, expect in the belief that man once exosted in a much lower and animal-like condition. The community of certain expressions in distinct though allied species as in the movements of the same facial muscles during laughter and by various monkeys, is rendered somewhat more intelligible, if we believe in their descent from a common progenitor". (Darwin 1872, p. 12)

This point is extremely important, because this expressive part of behavior marks the border between the directness and straightforwardness of basic sensomotorical behavior and more abstract forms of behavior, which during phylogeny became signals. (see Cassirer 1923, p. 127).

Signals expressing something (normally a specific internal state with relevance for the conspecific) are the first step away from the directness and immediateness of rigid, stereotype, reflex-like behavioral pattern[1].

Another example for this stage are onomatopoetic expressions with direct similarities between the expression and the expressed phenomenon or event.

2.2. SPEECH AS ANALOGIC EXPRESSION

The relation between the phonetic element and the object loses all kinds of direct similarities, without getting totally separated. What remains are analogic relations between phoneme and object.

E.g.: The Siam language (as other Asian languages) expresses different forms of meaning through modulation of tone pitch - a characteristic of analogic expression. Another example for analogic expression is reduplication for expressing plural terms.

As an example coming from ethology the alarm calls of vervet monkeys should be mentioned. (Cheney/Seyfarth 1985) These typical alarm calls indicate specific objects (e.g. specific predators), events etc., subsequently leading to appropriate behavior of conspecifics.

One of the main prerequisites of analogic expression are space and time categories.

Especially the mental representation of space (concept of space) improves analogic expression, in making clear distinctions and putting borders between the object and the surrounding space. This process contains a substantialisation and a related objectivation. This makes a major difference to the stage of mimetic expression, whose main characteristic is the immediate, direct involvement of the vocalizing organism, without any kind of objectivation of the object.

2.3. LANGUAGE AS SYMBOLIC EXPRESSION

At this level the direct connection between phoneme and object got lost. It is the level where theory construction, logic and science comes into play. Abstract concepts and categories can be used in a very flexible manner and provide the base for a new level of representation.

In relation to human ontogeny the main changes occuring with symbolic expression can be put as follows:

"The central theoretical point is that linguistic symbols embody the myriad ways of constructing the world intersubjectively that have accumulated in a culture over historical time, and the process of acquiring the conventional use of these symbolic artefacts, and so internalizing these construals, fundamentally transforms the nature of children's cognitive representations". (Tomasello 1999, p. 96)

Following S. Langer (1965) it is necessary to make a distinction between two modes of symbolic expression: presentational and discursive symbolic expression. The discursive mode of symbol usage comes close to the concept of scientific language, where the symbols have clear definitions, and their usage follows generally accepted grammatical rules. The main advocates of discursive symbolisms such as Carnap, Russell and Wittgenstein drew clear borders between scientific, discursive symbol usage and other ways of symbol usage such as metaphysics, art, subjectivity, feeling etc.

S. Langer's concept of presentational symbolism puts also arts, feelings etc. within a frame of articulated experiences and expressions, which are not irrational but just follow another kind of rationality. (see Meier-Seethaler 1997, p. 108f)

The main difference to discursive symbols is the fact, that the presentational symbols owe something like "implicite meaning" - what means that the semantic relations of this type of symbols is not stable and fixed (like e.g. in scientific language or mathematics), but more depending on the whole context as well as much closer to perceptive and emotional qualities. (Lachmann 2000, p. 73) Another important character of presentational symbols is their affinity to metaphors. This means "... the power of seeing one thing in another..." (Langer 1962, p. 153) -what is of major importance in relation to feelings and language.

In relation to affective expression the third point is dealing with the relations between the decrease of affective expression and symbolic reference.

3. DECREASE OF AFFECTIVE EXPRESSION AND SYMBOLIC REFERENCE

Having demonstrated different stages of speech or language usage I now come to some energetic - motivational considerations, which are important for understanding the relations between emotion - cognition and language. Looking at the transitions from mimetic - analogic to symbolic ways of expression from a motivational - energetic perspective a massive change appears. E.g. in comparing the energetic and dynamic dimensions of direct, mimetic expressions (like a cry of pain) with the energetics underlying controlled and rational language usage (as in a every day dialogue) the changes of expressed (emotional or more general: energetic) intensities seems obvious.

According to results from research on primate behavior as well as on human ontogenetic development there appears to be a close relation between the control of affective expression and related symbolic references. (Johnson 2000; Furth 1990) Following Vygotski (1978) early utterances in humans are closely related to the whole socio-emotional context within which they were learnt. So e.g. the term "car" is related not just to the object but to the whole set of emotional, social and physical activities related to car, like driving, sitting, fastening seat belts.... etc. So the usage of the word "car" is in the early stages of language acquisition always closely related to the whole context of activities around cars.

It is a important step, when the child starts to use the term car out of the actual context, what is combined with a flattening of the related affective background of the whole situation, which originally was combined with "car".

It is especially this temporal *detachment* of the utterance from the actual context (the decontextualisation), with the related and already mentioned emotional flattening, which provides the basis for a symbolic use of the word "car". This detachment enables the child to a much more flexible and creative usage of the symbol, often leading to very confused, phantastic and even surreal confabulations. It is especially - the later mentioned - societal frame which now comes into play for providing generally accepted limits and rules for these phantasies. (see Furth 1996)

These "decontextualisations" can first be seen in utterances which get produced beside the original context, where they were learnt and in applying them across a variety of different contexts. So the reinforcing contingencies of the original context are weakened and the range of possible applications of the symbol increases. (see Johnson 2000, p135) Here it is especially the previously mentioned presentational mode of symbolic expression which at this stage plays a major role.

Generally speaking this decrease of affective, energetic components as well as the related detachment are important prerequisites of symbolic expression. The neuroanatomical basis underlying these changes is the prefrontal cortex which is considerably enlarged in humans relative to other primates. The major functions can be found in inhibiting and regulating affective expression, as well as in planning activities. (Damasio 1994).

Brain lesions, with damage of the neurons and pathways connecting the prefrontal cortex with the deeper layers of the limbic system lead to uncontrolled affective expressions. (e.g. the famous case of Phineas Gage - see Bigelow 1850) Looking at the inhibiting effect from an energetic perspective it seems plausible, that the energies which get suppressed by prefrontal structures are redirected in the production of symbolic entities.

To put these considerations of a related decrease of affective expression and increase of symbolic capacities in a more general frame I come to point 4.

4. CONCRETE (SENSOMOTORIC) VERSUS ABSTRACT (SYMBOLIC) FORMS OF ACTION

Within the framework of cognitive psychology in the version of Piaget the developmental processes taking place between concrete sensomotoric actions and symbolic behavior are of major interest. Piaget's naturalistic theory of symbol formation provides a detailed framework for these extremely complex stages of human ontogenetic development, putting major emphasis on the idea that symbols arise from concrete actions. (Piaget 1967, 1981).

It is not space to deal with Piaget's complex hypothesis, but for language related issues it is necessary to take into account H.G. Furth's research, which is very close to Piaget. He tries to put the cognitivistic Piagetian view into a more motivational -energetic framework. (see Furth 1987, 1969; Wimmer 1998) Major emphasis is put on the transitions between concrete actions and related affectivity to emerging symbol systems and related behavioral changes. This is the link to the detachment problem, which was mentioned in the previous chapter.

In early pre-human as well as in most animal behavioral pattern the connection between motivations and related (motorical) action pattern is very close. Classical ethology has demonstrated in details, how the internal states (drives), the sensorial activities and the action pattern are related and activated. (Lorenz 1981; Tinbergen 1951) In most cases of animal behavior activated behavioral programs include specific motor pattern. Following Furth (1987) there is a essential difference between the concrete motor output of behavior as it can be seen in animal behavior, and especially human behavior, where concrete motorical action pattern are replaced by symbolic forms of behavior.

It is especially the increasing distance between the concrete behavioral acts from the underlying physiological basis that opens the door for symbolizing activities. This means, that with more complex mental - cognitive abilities the directness of behavioral acts gets lost. All the different stages of symbolic expressions and language related expressive processes demonstrate the growing distance from the basic processing mechanisms at the underlying physiological base. According to Piaget Furth considers "object permanence" as one major fact which provides the basis for all kinds of symbolizing processes. An object attains a permanent character, because

"...it is recognised as continuing to exist beyond the limits of the perceptual field, when it is no longer felt, seen or heard etc. ...". (Piaget 1953, p. 9)

Object permanence as a result of human ontogeny implies serious changes in affective as well as in cognitive dynamics. According to the arguments of Furth (who is quite close to S. Freud) the energy which normally underlies concrete behavior (sensomotoric actions) is now used within the symbolic domain for the generation and manipulation of symbols. This means, that impulses to actions which are no longer part of a concrete action are transformed in symbolic domains opening up a totally new field of human experience. As one of the main consequences the increasing range of releasing stimuli (which now are also effective in the mental domain) has to be underlined. Imagined situations (probably projected in the future) now seriously influence (and probably destabilize) the whole psychic domain.

This view of symbol production considers symbols as somewhat related to the underlying psychophysiological base. In contrast to basic assumptions of Al research and linguistically orientated philosophical positions symbols are not taken as neutral signs which are computed by specific rules governed by language. This very idealised view of symbol manipulation which is according just to the above mentioned discursive mode of symbol expression, which appears very late in human ontogeny and is in its most developed version something like an "ideal language", probably just realised in mathematics.

One further effect of the above mentioned object permanence is a serious change in the whole affective life, which can be called affect permanence. Before object permanence arises, affects were closely related to concrete sensorial impressions, concrete actions and concrete performed behavioral acts. With arising object permanence a stabilisation of the whole affective domain begins, because the mentally represented symbols as well as the related affective qualities can remain active - even if the relevant object is out of range for providing concrete input. Affective qualities coloring to mental entities remain active even when the concrete object gets out of sight, smell etc.

In general "object permanence" leads to a stabilisation and enlargement of internal, mental representations. If these representations get paired with phonetic cues stabilisation as well as storage qualities seem to improve seriously.

There is strong evidence that language arises after symbolization and the phonologic aspect appeared after the semantic. Semantic and phonetic dimensions together seem to have a reciprocal feedback resulting in the "explosion" of human language capacities.

The core of the argument is that this new symbolic dimension underlies two major influences: horizontal and vertical conditions.

The vertical dimension can be found in all the necessary cognitive, social and neuronal preconditions, providing the base for the ability to produce symbols. The horizontal dimension takes into account all the preexisting symbolic universes, into which the individual human mind gets socialized. (see e.g. Luckmann 1996; Assmann 2000).

Especially concepts coming from sociology and cultural sciences emphasize that there is something like a "cultural memory" (Assmann 2000, p. 11 f), a "sociohistoric a priori" (Knoblauch 1996, p. 16) reaching far beyond individual memory and forming a necessary frame for human ontogeny. This frame provides basic dimensions of "meaning" and "sense" for each individuum and also transcends individual memory. These symbolic systems have their internal coherence, their history and social foundation.

Language can be considered as an essential part of this frame, which contains specific "Weltanschauungen", contexts of meaning and versions of understanding. (Luckmann 1996, p. 91f)

4.1 FEEDBACK CIRCLES

Within the evolution of language and related changes in affectivity one very important element has to be underlined: Phonetic entities (like words) can - if they are expressed - have strong effects on the affective as well as the cognitive basis. A word (symbolic utterance), stored within a specific affective - cognitive context, can - as mentioned above - be reproduced beyond this primary context with weaker affective intensities. This leads to increasing flexibility and increasing abstract modes of symbol usage and representation.

But beside abstract usage of symbols, the same word - produced in a more or less neutral atmosphere - can evoke this primary, context-related "original" feelings and cognitions. This opens a new dimension of sociocultural development, because words can be representatives for specific events, persons, and situations and words can generate strong emotions (with related cognitions), which now are beyond the primordial, concrete context.

As examples for these complex relations I want to put rituals or ritualistic activities. In performing ritual activities a scenery is produced which induces shared feelings in the participants. The whole "setting" of ritual practices is quite complex including often repeated words (prayers, chants), specific body postures, rhythmic sounds etc. leading to specific sensations shared by the participants. (see Oubre 1996, p. 133f)[2].

Ritual can be seen as a way of regulated and articulated expression of emotions, as well as a way to generate emotions. (Langer 1965, p. 155) So in rituals emotions (or probably altered states of consciousness) get elicited by words within a more or less sophisticated context, demonstrating the ability to generate emotions artificially. Beside rituals there are a lot of other language related emotion - eliciting factors such as human artefacts and narratives (myths).

It is important to mention that all these rituals, art products etc. are expressing emotions and serve as emotion eliciting factors. This opens a dynamic I called "emotional hothouse effect" (emotionaler Treibhauseffekt) which plays a major role within cultural evolution. In expressing and eliciting emotions a dynamism arises which brings human emotionality far beyond the biological roots, without getting beyond the biological frame.

To make this clearer I will demonstrate early stages of usage of language, as it can be seen in the so called mythological thinking.

5. LANGUAGE WITHIN MYTHOLOGICAL THINKING

I just will rely on the most important points of this really complex field to demonstrate the close interrelations between language - cognition and emotion. Analysing language usage within mythological thinking offers the possibility to demonstrate early forms of language performance in humans. The importance of mythological thinking in general lies in the fact, that it dominated human history over long periods, and still plays a major role in all kinds of human affairs. Scientific thinking, if opposed to mythological thinking, can - within the timescale of human cultural evolution - be considered as very young with the first, very fine roots not older than 8000 years. (see Campbell 1991, Vol.1, p. 50).

So mythological thinking and related language are much closer to the biological inheritance. Looking at the formal structures of this kind of thinking, the dominant influence of the so called "ratiomorph mechanisms" (Riedl 1984) is obvious. Focusing on the formal structures of mythological thinking has the big advantage not to loose orientation, what can easily happen in a content related analysis of different myths[3].

A detailed and profound formal analysis of mythological thinking as a symbolic form and the inclined usage of language, as done e.g. by Cassirer (1923, 1925) shows the most essential roots of human language usage. The common separation between the sign and the designated object - which seems so evident for recent modes of thinking - at this stage simply does not exist. For this ancient modes of thinking and speaking a deep relation exists between the word (the sign or symbol) and the object designated by the sign.

"Name and essence are within an internal, necessary relation. The name not just designates, but the name is the essence itself - that is one of the main basic requirements of mythological thinking" (Cassirer 1925, p. 74. Transl. by M.W.).

This is also related to the human personality, whose core is closely connected with the name. For example the ancient Aegypt culture was characterized by the assumption, that beside the so called KA (as the more physical part of the person) the name is something like a double (Doppelgaenger) (Cassirer 1925, p. 117) Humans were seen as acting and having effects as long as the name is existing or is expressed. So beyond the "real" elemination of an enemy also the name (probably engraved in stone) had to be removed.

Taking into account the distinction between denotative and connotative meaning dimensions of words a further important characteristic of mythological language usage is the prevalence of connotative dimensions. Clear, stable and context-invariable definitions as in scientific language are missing at this stage.

This *variability of meaning* dimensions of words is strictly against the "principium identitatis" as a main principle of Aristotelian logic.

This can be demonstrated in the fact, that within mythological thinking a thing can change its "identity" if it is situated in a specific location. So the sorrounding space, the atmosphere, the context - to a great extent - influences the things identity. This concerns space and time dimensions in general, which are not considered as continuous and neutral dimensions but are characterized by inhomogenities and different qualities. Objects can change their main characters if they are situated in a specific location or enter a specific period. (see Eliade 1984)

A further general feature of mythological thinking and mythological language production is the so called "non reflected symbolism", being based on the dominance of expressive elements.

This means, that these words and these conceptions of the world ("Weltbilder") miss specific levels of reflexivity, containing the idea, that the used symbols and words refer to something "behind" them. This leads to a situation, where symbols and words are identified with the reference object what makes the emotional coloring of these words and symbols very intense.

Later periods of the development of modern human mind are characterized by the generation of further reflexive levels, bringing words from strong substantial entities and essences to more flexible signs just referring to other verbal entities. It seems very important to keep in mind these backgrounds and roots of human language origins, because even scientific language never gets totally rid of any substantialising effect.

Concerning emotions it is important to mention, that mythological thinking and language usage are closely related to high emotional intensities.

Two points have to be underlined:

1) with language a new and broad range of emotion eliciting stimuli appears, what means, that the range of emotion generating elements gets seriously enlarged. Words - evoking imaginations and activating memory can - especially if they are used within a specific context (e.g. rituals) generate strong emotions;

2) words and language about human emotions underlie severe historical changes. I just

want to mention two extremely different ways of speaking about emotions. For ancient greek people most of experienced emotions always were derived from external, supernatural entities like gods. (see Dodds 1951; Snell 1953)

"…that all departures from normal human behavior whose causes are not immediately perceived, whether by the subjects own consciousness or by the observation of others, are ascribed to a supernatural agency". (Dodds 1951, p. 13)

Obviously this way of dealing with and speaking about emotions presupposes a view of human existence, where emotional affairs are beyond autonomous self-control. The human subject is considered as being posessed and driven by external forces.

The opposing situation seems to be the recent way of dealing with emotions in focusing on internal psychic mechanisms causing all emotional experiences. The human subject considered as autonomous and self-controlling is the core causal point of all emotion generating activities. This leads to the concluding assumption, that humans are more or less fully responsible for their experienced emotions. These opposing examples just should demonstrate different ways of speaking about emotions.

Especially the social constructivistic theories of emotions emphasize the close connection between experienced emotions and related language. This includes the argument, that "...emotional experience ... requires that organisms possess a language of emotion". (Lewis/Saarni 1985, p. 8).

Similar ideas are formulated by Malatesta/Haviland:". ..the emotion words of a culture exert a powerful influence on the actual experience of emotion". (Malatesta/Haviland 1985, p. 110).

It was one of the major aims of this article to show that language and the required underlying levels of emotion and cognition appear as interacting phenomena, with a long history. None of all three functions can ever be isolated. There is neither a pure emotion, nor a pure cognition nor any kind of ideal language without any relations to the underlying levels of being.

Prof. Manfred Wimmer
Konrad Lorenz Institute
Adolf Lorenz Gasse 2
3422 Altenberg/Donau, Austria

NOTES

[1] Within emotion research especially the mimic changes occuring during emotional episodes are investigated in a very elaborated manner. The work of Hjortsjo (1969), Ekman/Friesen (1975), Eibl- Eibesfeldt (1986) and many others resulted in an immense ammount of empirical data. Other research strategies are dealing with voice and the influence of emotions on the different qualities (loudness, tone-pitch etc.) which occur in emotional states. (see Scherer 1986)In general they consider the biological function of the expressive components as communicative, that e.g. facial expression shows the conspecific the internal state leading to specific forms of behavior.

[2] Whitehead even goes further in postulating that the essential function of rites can be found in generating and experiencing emotions. (see Whitehead 1985).

[3] An overview of possible content related analysis can be found e.g. in Angherm 1996.

REFERENCES

ANGHERN E. (1996): *Die Ueberwindung des Chaos: Zur Philosophie des Mythos*. Suhrkamp: Frankfurt/Main.

ASSMANN J. (2000): *Religion und kulturelles Gedachtnis*. Beck'sche Reihe: Munchen.

AVERILL J.R. (1996): "An analysis of psychophysiological symbolism and its influence on theories of emotions". In: *The Emotions. Social, Cultural and Biological Dimensions.* Harrè R./Parrot W.G. (eds.) Sage Publications :London/New Delhi.

BIGELOW H.J. (1850): "Dr. Harlows case of recovery from the passage of an iron bar through the head". In: *American Journal of Medical Sciences* (19), (pp 13-22).

CAMPBELL J. (1991): *Die Masken Gottes*. Vol.1. Deutscher Taschenbuch Verlag: Muenchen. (Engl. Orig.: *The masks of God. Primitive mythology*. Viking Press: New York 1959).

CASSIRER E. (1923): *Philosophie der symbolischen Formen.* Bd. 1: Die Sprache. Zit. nach Cassirer (1997) *Philosophie der symbolischen Formen*. Bd. 1: Die Sprache. Primus Verlag: Darmstadt.

CASSIRER E. (1925): *Sprache und Mythos*. Studien der Bibliothek Warburg. Bd.6.

CHANGEUX J.P. (1984): *Der neuronale Mensch.* Reinbek bei Hamburg: Rowohlt.

CHENEY D.L.I./SEYFARTH R.M. (1985): "Vervet monkey alarm calls: manipulation through shared information?" In: *Behavior* 94 (150-166).

CIOMPI L. (1982): *Affektlogik: Uber die Struktur der Psyche und ihre Entwicklung*. Stuttgart: Klett-Cotta (Engl. version: *The psyche and schizophrenia. The bond between affect and logic*. Cambridge/Mass.: Harvard University Press).

CIOMPI L. (1997): *Die emotionalen Grundlagen des Denkens. Entwurf einer fraktalen Affektlogik.* Vandenhoeck&Ruprecht: Gottingen.

DAMASIO A.R. (1994): *Descartes error. Emotion, reason and the human brain*. New York: Putnams Sons.

DARWIN CH. (1872): *The expression of the emotions in man and animals.* Murray: London.

DODDS E.R. (1951): *The greeks and the irrational*. University of California Press:Berkeley/Los Angeles

EIBL-EIBESFELDT I. 1986: *Die Biologie des menschlichen* Verhaltens. Piper: Munchen.

EKMAN P./FRIESEN W. (1975): *Unmasking the face*. Prentice Hall: Englewood Cliffs/New York.

ELIADE M. (1984): *Das Heilige und das Profane. Vom Wesen des Religiosen*. Insel Verlag: Frankfurt/Main.

GARDINER H./METCALF R./BEEBE-CENTER J. (1937): *Feeling and emotion: A history of theories*. American Book Company: New York.

FURTH H. (1969): *Piaget and knowledge. Theoretical foundations*. New Jersey: Prentice-Hall, Inc.

FURTH H. (1987). *Knowledge as desire. An essay on Freud and Piaget*. Columbia University Press.

FURTH H. (1996): *Desire for society. Childrens knowledge as social immagination.* Plenum Press: London/New York.

HJORTSJO C.H. (1969): *Man's Face and Mimic Language.* Malmo.

JOHNSON CH.M. (2000): "The control of affective expression: The origins of symbolic reference". In: *The evolution of language. Proceedings of the 3rd International Conference 3 -6 April 2000*. Dessalles J.L. / Ghadakpour L. (eds.).

KLEINGINNA P.R./KLEINGINNA A.M. (1981): "A categorized list of emotion definitions, with suggestions for a consensual definition". In: *Motivation and Emotion 5 (345-355).*

KNOBLAUCH H. (1996): "Die Verfluechtigung der Religion ins Re1igioese". In: Luckmann 1996.

LACHMANN R. (2000) *Susanne K. Langer. Die lebendige Form menschlichen Fuehlens und Verstehens*. Fink Verlag: Dortmund.

LANGER S. (1962): *Philosophical sketches*. Baltimore: John Hopkins Press.

LANGER S. (1965): *Philosophie auf neuem Wege. Das symbol im denken, im Ritus und in der Kunst.* Fischer: Frankfurt/Main. (Engl. orig.: *Philosophy in a new key. A study in the symbolism of reason, rite and art.* (1942) Harvard University Press: Cambridge (Mass).

LANGER S. (1972): *Mind: An essay on human feelings*. Vol.2. Baltimore:Lohns Hopkins University Press.

LAZARUS R. (1982): "Thoughts on the relations between emotions and cognitions". In: *American Psychologist* 37 (1019-1024).

LEWIS M./SAARNI C. (1985): "Culture and emotions." In: *The socialisation of emotions*. Lewis M./Saarni C. (eds.) London: Plenum Press.

LORENZ K. (1981) *The foundations of ethology*. New York: Springer.

MEIER-SEETHALER C. (1997): *Gefuehl und Urteilskraft.Ein Plaedoyer fur die emotionale Vernunft.* Munchen: Beck'che Reihe.

LUCKMANN TH. (1996): *Die unsichtbare Religion*. Suhrkamp: Frankfurt/Main.

MACLEAN P. (1990): *The triune brain in evolution*. Plenum Press: New York.

MALATESTA C./HAVILAND J. (1985): "Signals, symbols and socialisation". In: *The socialisation of emotions.* Lewis M./Saarni C. (eds.) London: Plenum Press.

OBUCHOWSKI K. (1982): *Orientierung und Emotion*. Koln: Pah Rugenstein Verlag.

OESER E. (1993): "Die naturalistische Neubegruendung der Erkenntnistheorie durch Mach und Boltzmann im Wien der Jahrhundertwende". In Nautz/Vahrenkamp *Die Wiener Moderne: Einfluisse*, Umwelt, Wirkungen, 228-240.

OUBRE A.Y. (1996): *Instinct and Revelation. Reflections on the origins of numinous perception*. Amsterdam: Gordon and Breach Science Publishers.

PANKSEPP J. (1998): *Affective neuroscience. The foundations of human and animal emotions.* New York/Oxford: Oxford University Press.

PIAGET J (1967): *Biology and knowledge: an essay on the relations between organic regulations and cognitive processes.* (transl. into English by Beatrix Walsh) Chicago/London: The University of Chicago Press.

PIAGET J. (1981): *Intelligence and affectivity. Their relationships during child development.* Palo Alto: Annual Review Inc.

PIAGET J/INHELDER B. (1977): *Die Psychologie des Kindes.* Frankfurt/Main: Fischer.

RIEDL R. (1984): *Biology of knowledge. The evolutionary basis of reason.* Chichester/New York/Brisbane: John Wiley&Sons.

SCHACHTER S./SINGER J. (1962): "Cognitive, social and physiological determinants of emotional state". In: *Psychological Review*, 69 (378-399).

SCHACHTER 5. (1966): "The interactions of cognitive and physiological determinats of emotional state". In: *Anxiety and behavior*. Spielberger C. (ed.). New York: Academic Press.

SCHERER K.R. (1986): "Vocal affect expression: A review and a model for future research". *Psychological Bulletin*, 99.

SNELL B. (1953): *The discovery of the mind*. Blackwell: Oxford.

TINBERGEN N. (1951): *The study of instinct.* Oxford: Oxford University Press.

TOMASELLO M. (1999): *The cultural origins of human cognition*. Cambridge, Massachusetts: Harvard University Press.

VYGOTSKI (1978): *Mind and society: The development of higher psychological processes*. Cambridge (MA)/London: Harvard Univ. Press.

WEISS P. (1971): *Hierarchically organised systems in theory and practice*. New York: Hafner.

WHITEHEAD A.N. (1985) *Wie entsteht Religion*. Suhrkamp: Frankfurt/Main (Orig. 1926).

WIMMER M. (1995): "Evolutionary roots of emotions". In. *Evolution and Cognition* Vol.1/No.1 (38-50).

WIMMER M./CIOMPI L. (1996): "Evolutionary aspects of affective - cognitive interactions in the light of Ciompis "Affect-Logic". In: *Evolution and Cognition*. Vol.2/No.1 (37-59).

WIMMER M. (1998): "Eine Erweiterung von Piaget's Theorie der kognitiven Entwicklung in den emotionalen Bereich". In: *Freud-Piaget-Lorenz. Von den biologischen Grundlagen des Denkens und Fuehlens.* M. Wimmer (ed.). Wien: Wiener Universitaetsverlag.

WIMMER M. (1998): "Kernkonzepte von Lorenz vergleichend ethologischem Programm". In: *Freud-Piaget-Lorenz. Von den biologischen Grundlagen des Denkens und Fuehlens*. M. Wimmer (ed.). Wien: Wiener Universitaetsverlag.

ZAJONC R.B. (1980): "Feeling and thinking. Preferences need no inferences". In: *American Psychologist 2* (15-175).

MICHAEL STADLER AND PETER KRUSE

APPEARANCE OF STRUCTURE AND EMERGENCE OF MEANING IN THE VISUAL SYSTEM

SOME PROBLEMS IN MIND-BRAIN-RESEARCH

The mind-brain-problem has been discussed by philosophers over more than 2.000 years.

Today it gains new interest in the sciences, especially in neurobiology and neuropsychology. Although it seems to be a great advantage that this problem may now be reanalysed in the context of empirical investigations instead of being claimed by a rather unfruitful "ignorabimus" position, there remain some very strong principle embarrassments concerning this topic.

1. Cognitive and brain processes obviously are on totally different scales of system behavior. The elementary dynamics of neuronal brain processes take place in the order of magnitude of 10^{12} to 10^{15} major events per seconds. In the stream of consciousness, on the other hand, no more than 100 bits/sec of information can be analyzed. The enormous complexity of the neural network is confronted with the unity of mental events.

2. Brain processes consist of myriads of identical action potentials forming global spatial and temporal patterns. The language of the brain is an unspecific "click, click" as Heinz von Foerster (1985) put it. The mental events on the other hand are rich of different sensory qualities and are capable of continuous qualitative changes in a number of dimensions. The unspecifity of neuronal events is confronted with the specifity of meaning in the cognitive sphere.

3. The brain processes seem to be governed by syntactic rules from which structures of any kind but without any observable meaning emerge. On the other hand phenomenal events are always meaningful and make sense for the individual. The syntax of brain processes is confronted with the semantics of mental events. The relations between meanings are organized by different laws that brain processes obey to.

4. The brain processes have well defined elements, the neurons, with well defined connections between one another, the synapses. The phenomenal events on the other hand have no such elements but instead, as the Gestaltists have pointed out, holistic features (Koffka 1935). Any part of the phenomenal field influences and is influenced by all other parts. The — on first sight — relatively discrete functioning on the neuronal level is confronted with the obviously field-like interactions in cognition (Kruse et al. 1987).

5. The most suspicious problem in mind-brain-research seems to be the assumption of causal relationships between material and immaterial events. While most philosophers and natural scientists agree that there are causal effects of the brain processes on the mental events, there is strong scepticism for causal relationships the other way

A. Carsetti (ed.), Seeing, Thinking and Knowing, 293-306.

round because such an assumption seems to violate the law of conservation of energy. As long as one does not dare to assume brain effects caused by mental efforts the mind is only an epiphenomenon of the brain.

Cognitive neurobiologists and psychologists have to keep in mind all these problems when trying to explain brain-mind-relationships.

THE SYNERGETIC APPROACH

Some of the problems may be solved by using the theoretical and mathematical tools of a new interdisciplinary field of research called "synergetics". Synergetics was founded by Hermann Haken in the early seventies (see Haken 1977) and has now broad interdisciplinary relevance in the natural and even in the social sciences. Synergetics "is concerned with the cooperation of individual parts of a system that produces macroscopic spatial, temporal, or functional structures", as it was defined in the preface to the now about sixty volumes of the "Springer Series of Synergetics" (1977ff.) edited by Hermann Haken. The synergetic theory was first developed to explain the cooperative phenomena giving rise to laser light. Later it was applied to fluid dynamics to explain the Bènard-instability. Today there are synergetic approaches to explain cooperative phenomena in biological rhythms, movement regulation, preceptual multistability and population dynamics.

The first step of a synergetic analysis is to demonstrate the existence of non-linear phase transitions in a complex system. A phase transition is an autonomous reorganization of macroscopic order emerging spontaneously from elementary interactions. For this reorganization certain control parameters can be named which release a sudden phase transition to a higher order state of the system when continuously enhanced. The phase transition is preceded by an autocatalytical destabilisation of the system which is manifested in critical fluctuations and by a critical slowing down of the innersystemic tendency to conserve the existing ordered state. The non-linear behaviour of the system is explained by concurring modes which reach a bifurcation point where one of the modes predominates the others by slaving the elementary components of the process. This predominating mode is called the order parameter. In a synergetic system there is a certain circular causality between microscopic and macroscopic processes. The macroscopic structure emerges from and organizes the microscopic interactions of elementary components of the system.

Applied to mind-brain-dynamics such a circular causal process may be modelled in the following way. The micro processes of the nervous network give rise to a macroscopic collective process, for instance to a temporal synchronicity distributed over more or less distant areas of the brain. This collective process, if it predominates, represents a certain rhythmic pattern, the order parameter, which reinfluences the microscopic processes in the neural network (Fig. 1).

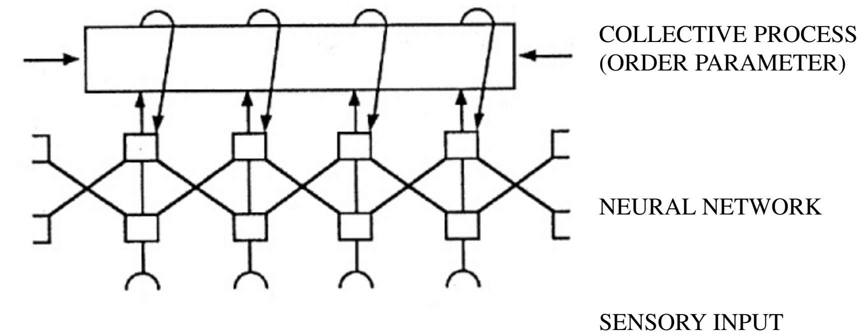

Figure 1. Micro-macro-relation in mind-brain-dynamics.

One of the main results of the synergetic approach is the obvious analogy between pattern formation and pattern recognition (Haken 1991). In the highly developed biological system of the brain pattern formation may be even identical with pattern recognition. This assumption represents an important step towards an empirical handling of the mind-brain-problem. If the macroscopic order parameter, which has emerged from the elementary activity, represents and governs this activity, it may at the same time represent the cognitive process which influences the behaviour of the organism by its nervous processes.

Following the assumption that pattern formation is pattern recognition, one would have overcome some of the problems of the mind-brain-relation mentioned above. The dynamics of perception, thinking and memory need not to be reduced to elementary brain processes. The cognitive dynamics may be represented directly by the macrodynamics of the brain. In this case cognitive phenomena may by used as a methodological window for observing and understanding brain activity. The relevance of many different macrodynamic neuronal processes for cognitive phenomena has been claimed in the last fifty years of brain research.

Wolfgang Köhler (1949) was the first who ascertained a concretization of such a macroprocess, of the D.C.-fields. His hypothesis was rejected by Lashley, Sperry, and Pribram who showed that perception is resistant against the disturbance of these fields. Today D.C.-fields are interpreted as an unspecific background activity of the brain (cf. Kruse et al. 1987).

Donald Hebb forwarded the hypothesis that certain cell assemblies might be the substrata of memory and perceptual phenomena. Hebb's ideas have influenced very strongly the neural network theorists of today. Lashley thought about interference patterns as a distance bridging emergent macroscopic brain process. He influenced Karl Pribram (1971) to formulate his hologram idea of brain functioning which has been elaborated to the holonomic brain theory (1991). In the last years, there was a certain revival in the recognition of the importance of the EEG-rhythms for cognitive processes (Mountcastle 1992). Walter Freeman and later Wolf Singer and the Marburg group around Eckhorn and Reitböck investigated the self organizing properties of the EEG in the gamma band (cf. Haken and Stadler 1990). The synchronized 40 Hz oscillations seem to be good candidates for figure-ground resolution and they may even be interpreted as attractors for meaning in perception. Finally an interesting hypothesis was brought forward by Sir John Eccles during one of the Elmau-meetings on synergetics about the possibility of immaterial macro processes influencing the material neuronal micro processes in the brain without violating the law of energy conservation. Eccles (1985) proposed macro processes analog to the probability fields in quantum physics as a representation of the mind in the brain.

There are enough unanswered questions about the relation between cognition and brain processes. Even if one accepts the analogy between pattern formation and pattern recognition and the circular interaction between macroscopic and microscopic brain processes as proposed by the synergetic approach, the question, which brain process is the representative of cognition, remains to be answered. Furthermore, there is the philosophical question whether there is a correlation or an identity between brain processes and mind. The possibility of an empirical test of the identity theory as proposed by Herbert Feigl's autocerebroscope (1958) may be approached by new technical developments. The scenario of a PET or NMR-biofeedback experiment has come into sight. This level of data collection promises to develop fruitful hypotheses and should not be rejected in brain research.

Given the assuption that the brain is a selforganizing system and that cognitive processes are based on the elementary neural dynamics of the brain, the synergetic approach can be concretized in three empirical hypotheses:

— It is possible to demonstrate non-linear phase transitions in cognition. For example continuous changes in stimulus conditions are able to trigger sudden reorganizations in perception. Perceptual organization cannot be reduced to the properties of the stimulus.

— Stable order in cognition is the result of underlying neuronal dynamics and therefore critically bound to instability. For example any percept is the result of a process of dynamic order formation. Because of the underlying dynamics perception is in principle multistable. Each stable percept can be destabilized and each instable percept can be stabilized.

— Meaning is an order parameter of the elementary neuronal dynamics. For example in the instability of ambiguous displays the basic order formation of perception can be influenced by subtle suggestive cues.

In the following, some examples of perceptual experiments are presented which may demonstrate the psychophysical significance of the synergetic approach.

STRUCTURE BY ITERATION

Frederic C. Bartlett described in one of his books (1951) a nice little experiment which showed that an iterative performance may reveal hidden structures in homogeneous areas.

Bartlett used the method of serial reproduction to show that a dot on a blank piece of paper seen for a second or so has changed it's position on another blank paper after reproduction. If this reproduced dot is again reproduced by another subject, it will another time be displaced and so on.

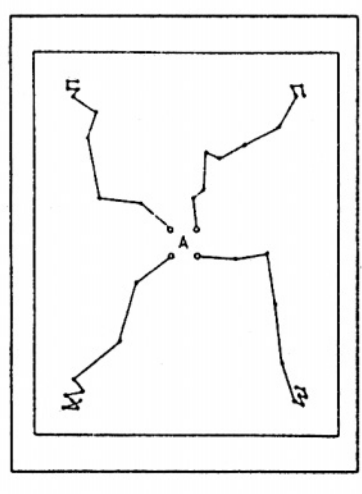

Figure 2. Phenomenon of the wandering dot.

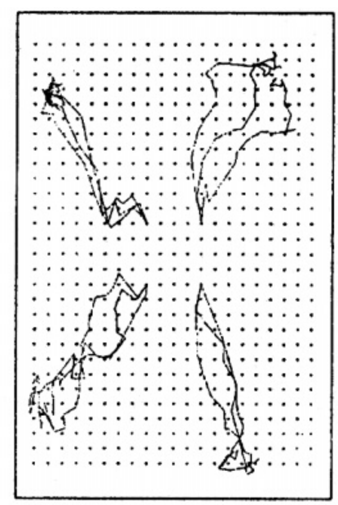

Figure 7. Simulation of the phenomenon of the wandering dot.

We adapted this experiment for our purposes and investigated this phenomenon of the wandering dot systematically (Stadler et al. 1991). In figure 2 some typical trajectories of wandering dots starting from different positions A in the middle of the area are reproduced. It seems, that the dots first take small steps, then longer steps and then diminish their steps again until they have found an attractor near one of the corners of the area from which the dot cannot flee again. The trajectories look as if there is an invisible potential gradient distributed over the area which causes the dots to make these particular movements.

Obviously the wandering dots show a non-linear behaviour on this virtual potential gradient and such a behaviour seems to be in contradiction to the linear texture gradients distributed over surfaces as described by J.J. Gibson (1950) and the ecological school (Fig. 3).

Attempts were made in our laboratory to measure the underlying non-linear potential fields.

For this purpose first the displacement vectors were collected for a 21 x 29 dot-position-pattern from 10 Ss. Figure 4 shows the result of one of them. Next the raw vectors of all individual experiments were subjected to a vectorial analysis procedure by which the sources of each vector are integrated over the whole pattern and the circular potential is divided from the gradient potential. (The mathematical procedure is described in Stadler et al. 1991). Figure 5 shows the calculated vector field of the averaged data of 10 Ss. Here we find already a very regular distribution of the vectors on the field, showing the non-linearities of the field, i.e. the bifurcation saddle in the middle and the attractors near the four corners. Figure 6 shows the gradient potential as a landscape over which the dots move like spherical particles. In figure 7 (see figure 2) a model calculation of these movements is shown, which resembles very good the empirical data of figure 2.

The experiment shows very good the hidden non-linear structure of a homogeneously stimulated perceptual field. Stabilities and instabilities in this field are represented by bifurcation areas (repellors) and attractor areas. The virtual structure, which is effective in perception but not visible, shows directly the field characteristics of the underlying macrodynamic brain process. The properties of stability and instability require more detailed investigation, for they represent system properties which may be analyzed on the brain level as well as on the psychophysical level.

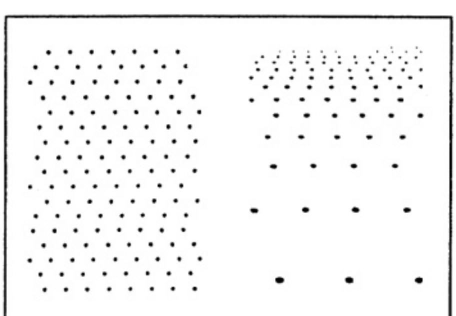

Figure 3. Texture gradients (Gibson 1955).

Figure 4. Displacement vectors of one Subject.

STABILITY, INSTABILITY, MULTISTABILITY

Inspite of the apparent stability of our daily perceptions, instability and multistability seems to be a basic feature of all perceptual processes. Figure 8 shows an example in which we can observe directly the fluctuating activity in the visual system on the search for stability. There are obviously different modes (collective processes) which compete for some time and which give the whole pattern a dynamic appearance. None of the modes is able to predominate as an order parameter. Therefore no pattern is stabilized for more than a few seconds.

Demonstrations of multistability in perception, the so-called reversible figures like the Necker-cube, Rubin's vase/face-picture, the Maltese-cross, the rabbit/dog-pattern and many others, are very well known in cognition research. Usually there are two predominating stable states which alternate periodically. Figure 9 shows the underlying potential of such a reversible process, as it is used for model calculations by Ditzinger and Haken (1989, 1990).

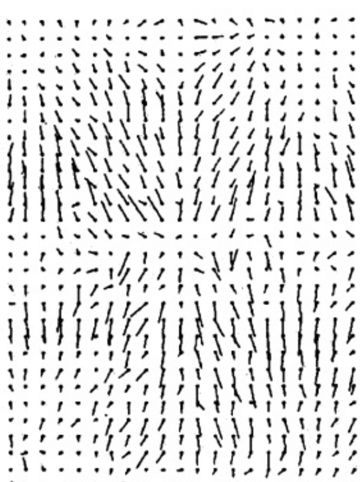

Figure 5. Calculated gradient field (10Ss).

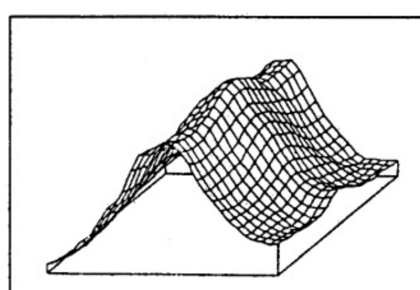

Figure 6. Calculated gradient potential.

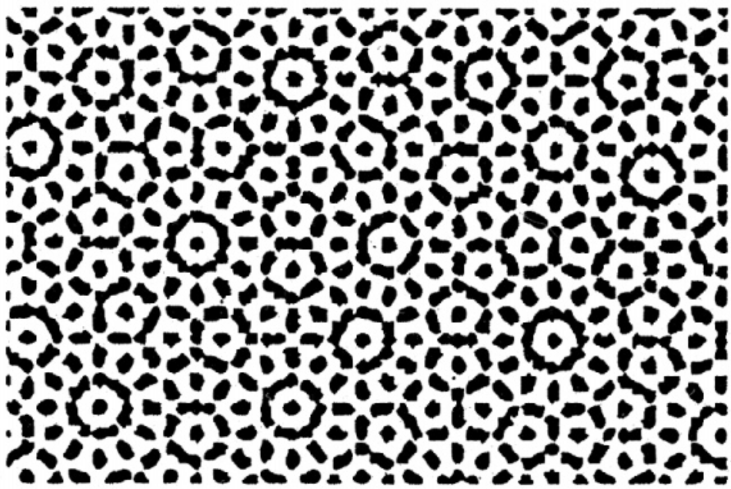

Figure 8. Instable visual pattern.

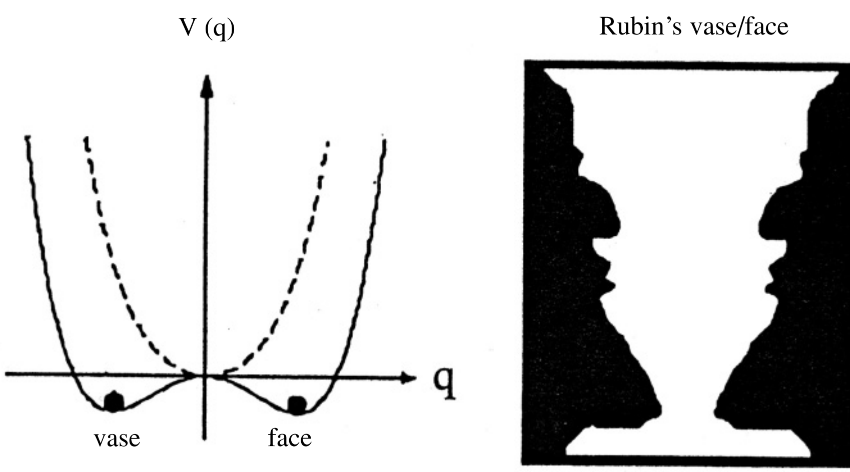

Figure 9. Reversible pattern and underlying potential (modified from Haken 1991).

In our experiments we preferred dynamic displays to analyse multistable behaviour because in motion perception the systematic variation of a control parameter is easier than in static pictures: Additionally ambiguous apparent movement patterns are very reliable in their dynamic behaviour and the reversion process is well defined for perceivers (see Kruse 1988).

Figure 10 and 11 show the two stimulus situations used: the stroboscopic alternative movement (figure 10, SAM) and the circular apparent movement (figure 11, CAM). The change from one alternative to the other shows the phenomenon of hysteresis as it is predicted in the synergetic model for phase transitions (cf. Kruse, Stadler and Strüber 1991). We could demonstrate in a variety of experiments that it is possible to influence the degree of stability or instability of such multistable patterns by contextual and semantic influences.

The potential landscape underlying the dynamics of the reversion process can be altered by introducing gestalt factors like common motion (figure 12) or figural identity (figure 13), by perceptual learning, or even by very subtle semantic cues.

In the esperiment of perceptual learning the bi-stable SAM was further destabilized by enhancing the probability of occurence of the third theoretically possible perceptual alternative (see figure 10). In this experiment the CAM was used in a training session to change the potential landscape of the underlying dynamics of the SAM in favour of the third version "circular motion" which is only very seldom perceived spontaneously. It was possible to demonstrate that the probability of occurence of the circular motion can be significantly enhanced by training.

The aspect of influencing multistable perception by introducing subtle semantic cues supports the theoretically predictable connection between instability and critical sensitivity to the initial conditions in the case of symmetry breaking. In perceptual multistability the system passes again and again the point of maximal instability (see figure 9) at which the principle symmetry of the dynamics is broken in favour of one stable ordered state. At the situation of symmetry breaking little influence has great effect. Therefore multistable patterns are a paradigmatic tool to demonstrate that semantic cues are able to influence macrodynamic brain processes.

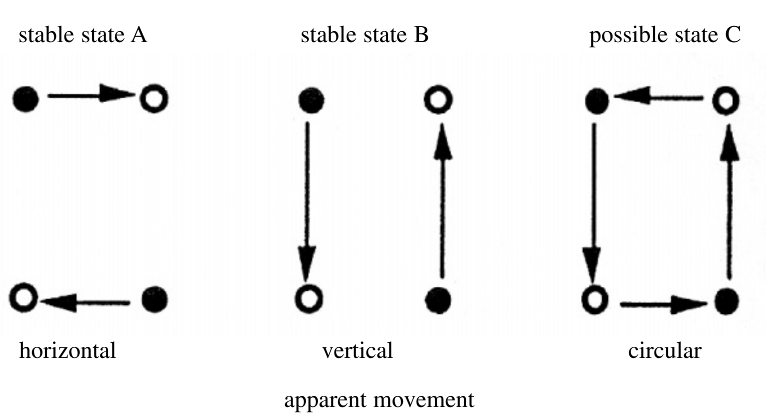

Figure 10. Stroboscopic alternative movement (SAM).

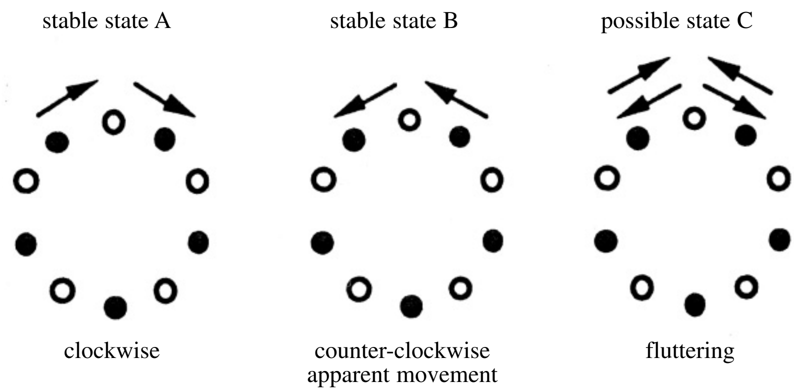

Figure 11. Circular apparent movement (CAM).

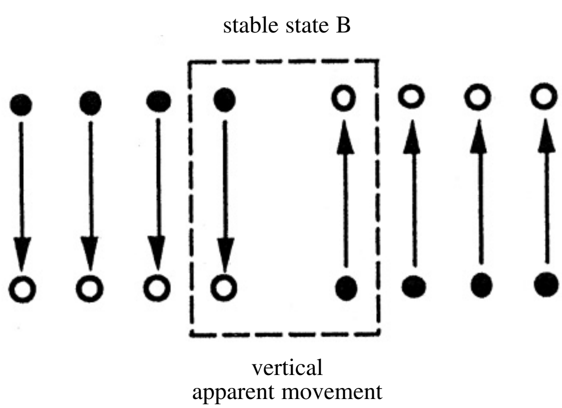

Figure 12. Stabilizing the multistable SAM by introducing a strong vertical bias (gestalt factor of common motion).

There are many examples in perceptual research showing top down influences from meaning to structure and even to basic sensory qualities in the research program of the so-called "new look" of the late forties and fifties (cf. Graumann 1966). One and the same green-brownish colour, for instance, is judged by many subjects more brown, if it is exposed in the form of a horse and more green if it is exposed in form of a leaf. Figure 14 shows that there is a clear preference to see a face instead of a vase in a bistable figure ground pattern. The structurally more attractive (gestalt factor of symmetry and proximity) vases only predominate, if the meaning of human faces cannot be attributed.

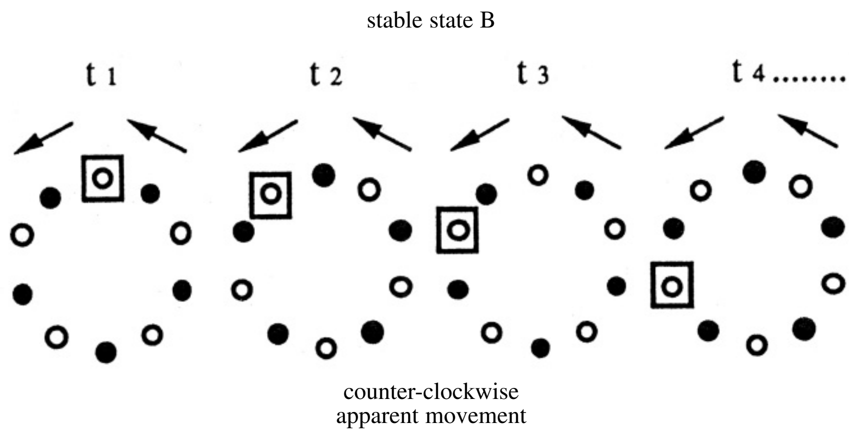

Figure 13. Stabilizing the multistable CAM by introducing a strong counter clockwise bias (gestalt factor of figural identity).

For the investigation of semantic bias effects, again the dynamic stimulus patterns SAM and CAM were used. If the CAM, for instance, is composed of arrows pointing in an anti-clockwise direction instead of circular dots, this direction is preferred significantly in the bifurcation situation at first sight. Nearly all subjects see an anti-clockwise rotation of the apparent movement although the clockwise rotation in a directional display is preferred (Figure 15).

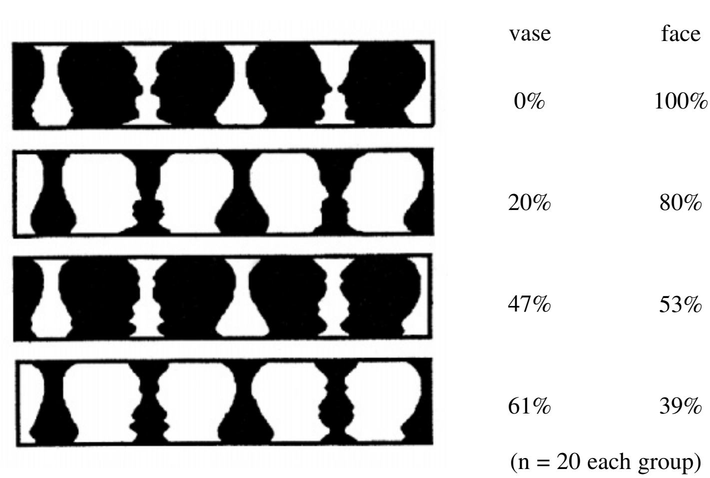

Figure 14. Meaning predominates structure (figure-ground pattern from Kruse 1986).

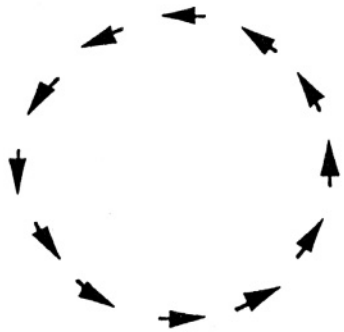

Figure 15. Semantical bias (using arrows instead of dots for the CAM).

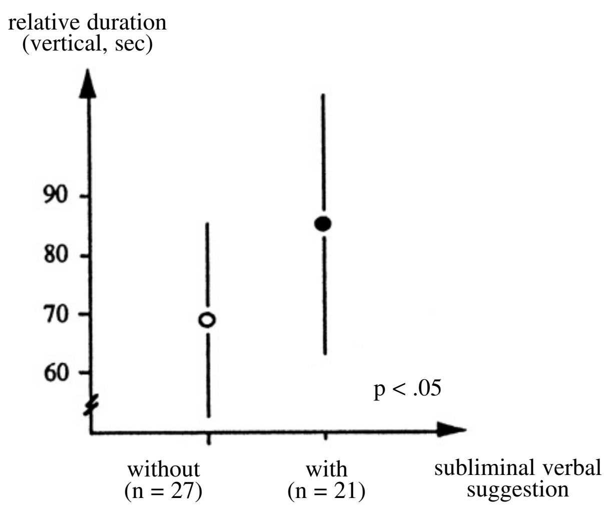

Figure 16. Effect of subliminal verbal suggestion on the relative duration of the vertical movement version during a 3-minute-presentation of the SAM.

Further experiments showed that the instability point of two perceptual alternatives is extremely sensitive to suggestive influences. Even subliminal verbal suggestions, e.g. “up and down like bouncing balls”, given to the subjects below the auditive threshold during presentation of the SAM, has a significant effect on the relative duration of the movement alternatives (figure 16).

The fact that meaning may influence the structure of brain processes, is predicted by the synergetic model of mind-brain-interaction. Further research should investigate the properties of macrodynamic brain processes during the instable and stable phase. First experiments in that direction demonstrate that EEG recordings during the perception of multistable patterns show a significant enhancement of the 40 Hz waves in the vertex position of the electrodes, which is not found during the perception of very similar but stable patterns. In the occipital EEG there is, however, no difference between the two stimulus conditions (Basar-Eroglu et al. in prep.).

The synergetic approach stimulates new experimental ideas for investigating the mind-brain-interaction on different levels of analysis between neurophysiology and psychology. Thus the concept of order parameters which emerge out of the elementary dynamics and which transform the basic instability into coherent stable patterns is a good model for macrodynamic brain-mind processes.

Prof. Michael Stadler
Institute of Psychology
University of Bremen
P.O. Box 330440
28334 Bremen, Germany

Prof. Peter Kruse
University of Bremen
P.O. Box 330440
28334 Bremen, Germany

REFERENCES

BARTLETT, F.C. (1951), *The Mind at Work and Play*, London: Allen & Unwin.

DITZINGER, T. & HAKEN, H. (1989), "Oscillations in the perception of ambiguous patterns", *Biological Cybernetics* 61, 279-287.

DITZINGER, T. & HAIEN, H. (1990), "The impact of fluctuations on the recognition of ambiguous patterns", *Biological Cybernetics* 63, 453-456.

ECCLES, J. (1985), "New light on the mind-brain-problem: How mental events could influence neural events". In: H. Haken (Ed.), *Complex Systems - Operational Approaches*. Berlin: Springer.

FEIGL, H. (1958), *The Mental and the Physical*, Minneapolis: University of Minnessota Press.

FOERSTER, H. VON (1985), *Sicht und Einsicht*, Braunschweig: Vieweg.

GIBSON, J.J. (1950), *The Perception of the Visual World*, Boston: Houghton Mifflin.

GRAUMANN, C.-F. (1966), "Nicht-sinnliche Bedingungen des Wahrnehmens". In: W. Metzger (Ed.), *Wahrnehmung und Bewußtsein*, Göttingen: Hogrefe.

HAKEN, H. (1977ff), *Springer Series of Synergetics*, Berlin: Springer.

HAKEN, H. (1977), *Synergetics*, Berlin: Springer.

HAKEN, H. (1991), *Synergetic Computers and Cognition*, Berlin: Springer.

HAKEN, H. & STADLER, M. (Eds.) (1990), *Synergetics of Cognition*, Berlin: Springer.

HEBB, D.O. (1949), *The Organization of Behavior*, New York: Wiley.

KÖBLER, W. & HELD, R. (1949), "The cortical correlate of pattern vision", *Science* 110, 414-419.

KOFFKA, K. (1935), *The Principles of Gestalt Psychology*, London: Routledge & Kegan Paul.

KRUSE, P. (1986), "Wie unabhängig ist das Wahrnehmungsobjekt vom Prozeß der Identifikation?" *Gestalt Theory* 8, 141-143.

KRUSE, P. (1988), "Stabilität, Instabilität, Multistabilität. Selbstorganisation und Selbstreferentialität in Kognitiven Systemen", *Delfin*, 6, 35-57.

KRUSE, P., ROTH, G. & STADLER, M. (1987), "Ordnungsbildung und psychophysische Feldtheorie", *Gestalt Theory* 9, 150-167.

KRUSE, P., STADLER, M. & STRÜBER, D. (1990), "Psychological modification and synergetic modelling of perceptual oscillations". In: H. Haken & H.P. Koepchen (Eds.), *Rhythms in Physiological Systems*, Berlin: Springer.

MOUNTCASTLE, V.P. (1992), Preface to E. Basar and T.H. Bullock (Eds.), *Induced rhythms in the brain*, Boston: Birkhäuser.

PRIBRAM, K.H. (1971), *Languages of the Brain*, Englewood Cliffs, N.Y.: Prentice-Hall.

PRIBRAM, K.H. (1991), *Brain and Perception*, Hillsdale: Lawrence Erlbaum.

STADLER, M., RICHTER, P.H., PFAFF, S. & KRUSE, P. (1991), "Attractors and perceptual field dynamics of homogeneous stimulus areas", *Psychological Research* 53, 102-112.

ARTURO CARSETTI

THE EMBODIED MEANING: SELF-ORGANISATION AND SYMBOLIC DYNAMICS IN VISUAL COGNITION

With the assistance of the new methodologies introduced by Synergetics and by the more inclusive theory of dissipative systems, we are now witnessing the effective realisation of a number of long-standing theoretical Gestalt assumptions concerning, in particular, the spontaneous formation of order at the perceptual level. This realisation endorses lines of research G. Kanizsa has been conducting along the entire course of his life[1]. However, a careful analysis of his experiments, particularly those dedicated to the problem of amodal completion, presents us with incontrovertible evidence: Gestalt phenomena such as those relative to amodal completion can with difficulty find a global model of explanation by recourse only to the methodologies offered by order-formation theories such as, for example, those represented by non-equilibrium thermodynamics and the theory of non-linear dynamical systems.

Indeed, it is well known that particular neural networks, inhabiting the border-area between the solid regime and chaos, can intelligently classify and construct internal models of the worlds in which they are immersed. In situations of the kind, the transition from order to chaos appears, on an objective level, as an attractor for the evolutionary dynamics of networks which exhibit adaptive properties, and which appear able to develop specific forms of coherent learning. However, natural-order models of the kind, while appearing more adequate than Koehler's field-theoretical model, are still unable to provide a satisfactory answer to the complexity and variety of problems regarding the spontaneous order formation at the perceptual level. A number of questions immediately arise. What cerebral process, for example, constitutes and "represents" perceptual activity as a whole? How can we define the relationship between the brain and the mind? How can we explain the direct or primary nature of the perceptual process when we know that at the level of underlying non-linear system dynamics there exists a multiplicity of concurrent mechanisms? How can we speak in terms of stimulus information if the measure of information we normally use in psychological sciences is substantially a propositional or monadic one (from a Boolean point of view) like that introduced by Shannon? A percept is something that lives and becomes, it possesses a biological complexity which is not to be explained simply in terms of the computations by a neural network classifying on the basis of very simple mechanisms (the analogue of which is to be found, for example, in some specific models studied at the level of statistical mechanics, such as spin-glass models).

In a self-organising biological system, characterised by the existence of cognitive activities, what is self-organising is, as Atlan states[2], the function itself with its meaning. The origin of meaning at the level of system-organisation is an emergent property, and as such is strictly connected to very specific linguistic and logical operations, to specific procedures of observation and self-observation, and to a continuous activity of inter-

A. Carsetti (ed.), Seeing, Thinking and Knowing, 307-330.

nal re-organisation. In this context, the experimental findings offered, for example, by Kanizsa remain an essential point of reference, still constituting one of our touchstones. The route to self-organisation, which Kanizsa also considers in his last articles the route of primary explanation, is ever more universally accepted. Yet questions remain: *via* what informational means and logical boundaries is self-organisation expressed? What mathematical and modelistic instruments can we use to delineate self-organisation as it presents itself at the perceptual level? What selection and elaboration of information takes place at, for example, the level of amodal completion processes? What is the role, in particular, of meaning in visual cognition (and from a more general point of view, in knowledge construction)?

Problems of the kind have for many years been analysed by several scholars working in the field of the theory of natural order, of the theory of the self-organisation of non-linear systems, and of the theory of the emergence of meaning at the level of biological structures. They have recently received particular attention (albeit partial), from a number of scientists investigating connectionist models of perception and cognition. The connectionist models, as developed in the eighties, may be divided into two main classes: firstly, the PDP models first posited by Hinton (1985) and Rumelhart (1986), based essentially on a feed-forward connectivity, and on the algorithm of back-propagation for error-correction. These models require a "teacher": a set of correct answers to be introduced by the system's operator. A second class, posited by, in particular, Amari (1983), Grossberg (1981), and Kohonen (1984), replaces the back-propagation and error-correction used by PDP models with dense local feedback. No teacher is here necessary: the network organises itself from within to achieve its own aims. Perception is here no longer viewed as a sort of matching process: on the contrary, the input destabilises the system, which responds by an internal activity generated *via* dense local feedback.

Freeman's model of olfactory perception, for instance, belongs to this second class[3]. It contains a number of innovative elements that are of particular interest to the present analysis, in that for Freeman perception is an interactive process of destabilisation and re-stabilisation by means of a self-organising dynamics. Each change of state requires a parametric change within the system, not merely a change in its input. It is the brain, essentially, which initiates, from within, the activity patterns that determine which receptor input will be processed by the brain. The input, in its turn, destabilises the olfactory bulb to the extent that the articulated internal activity is released or allowed to develop. Perception thus emerges above all as a form of interaction with the environment, originating from within the organism. As Merleau-Ponty maintained, it is the organism which selects which stimuli from the physical world it will respond to: here we find a basic divergence with respect to the theory of perceptual organisation as posited by Synergetics. Freeman's system no longer postulates an analogy-equivalence between pattern formation and pattern recognition. While in other self-organising physical systems there exists the emergence of more ordered states from less-ordered initial conditions, with precise reference to the action of specific control- and order-parameters, at the brain level, according to Freeman, a specific selective activity takes place with respect to the environment, an activity which lays the foundation for genuinely adaptive behaviour. What happens inside the brain can therefore be explained, within the system-model proposed by Freeman, without recourse to forms of inner representation. At the perceptual level we have the creation of a self-organised internal state which destabilises the system so as to enable it to respond to a particular class of stimulus input in a given sensorial modality. Perception is thus expressed in the co-operative action of masses of neurons producing consistent and distributed patterns which can be reliably correlated with particular stimuli.

It should be emphasised here, however, that if we follow the route indicated by Freeman, the problem of the veridical nature of perception immediately takes on a specific relevance. As we have just said, we know quite well that, for example, Boolean neural networks actually classify. Networks of the kind possess an internal dynamics whose attractors represent the asymptotic alternative states of the network. Given a fixed environment, from which the network receives inputs, the alternative attractors can be considered as alternative classifications of this very environment. The hypothesis underlying this connectionist conception is that similar states of the world-surroundings are classified as the same. Yet this property is nearly absent in the networks characterised by simple chaotic behaviour (as Freeman's model clearly shows). At this level the attractors as such are unable to constitute paradigmatic cases of a class of similar objects: hence the need to delineate a theory of evolutive entities which can optimise their means of knowing the surrounding world *via* adaptation through natural selection on the edge of chaos. Hence the birth also of functional models of cognition characterised in evolutionary terms, capable of relating the chaotic behaviour to the continuous metamorphosis proper to the environment.

This line of research, while seeming a totally natural direction, is not without its difficulties. There is the question, for example, of the individuation of the level of complexity within existing neural networks capable of articulating themselves on the edge of chaos, at which the attractors are able to constitute themselves as adequate paradigms to cope with the multiple aspects of external information. How can we specify particular attractors (considered as forms of classification), able to grasp the interactive emergence proper to real information as it presents itself at the level of, say, the processes of amodal completion? How can the neural network classification-processes manage to assimilate the information according to the depth at which the information gradually collocates itself? And what explanation can be given for the relationship between the assimilation of emergent "qualities" on the one hand, and adaptive processes on the other? How to reconcile a process having different stages of assimilation with perception's direct and primary nature as described by Gibson and Neisser? What about the necessary interaction between the continuous sudden emergence of meaning and the step by step development of classification processes? And finally, what about the necessary link between visual cognition and veridical perception or, in other terms, between cognitive activity, belief and truth?

To attempt even a partial answer to all these questions, it should first be underlined that the surrounding information of which Gibson speaks is, as reiterated above, immense, and only partly assimilable. Moreover, it exists at a multiplicity of levels and dimensions. Then, between the individual and the environment precise forms of co-evolution gradually take place, so that to grasp information we need to locate and disclose it within time: we have progressively to perceive, disentangle, extract, read, and evaluate it. The information is singularly compressed, which also explains why the stimulus is a system stimulus. The intrinsic characteristics of an object in a given scene are compressed and "frozen", and not merely confused in the intensity of the image input. If we are unable to disentangle it, we are unable to see; hence the need to replace one form of compression for another. A compression realised in accordance with the selective action proper to the "optical sieve", producing a particular intensity of image input, has to be replaced by that particular compression (costruction+selection) our brain constructs from the information obtained, and which allows us to re-read the information and retrieve it along lines which, however, belong to our visual activity of recovery-reading. What emerges, then, is a process of decodification and recodification, and not merely analogy-equivalence between pattern formation on the one hand, and pattern recognition on the other. This process necessarily articulates according to successive depth levels. Moreover, to perceive, select, disentangle, evaluate, etc. the brain has to be able autonomously to organise itself and utilise particular linguistic instruments, interpretative functions, reading-schemes, and, in general, specific modules of generation and recognition which have to be articulated in discrete but interconnected phases. These are modules of exploration and, at the same time, of assimilation of external information; they constitute the support-axes, which actually allow epigenetic growth at the level of neural cortex.

In order to explain the perceptual process, in particular, it is important to be able to distinguish what effectively happens at brain level when we see, what our conscious awareness is in relation to what we see, and, lastly, what our effective construction is of simulation models. For example, the extent to which we manage to perceive affordances at the level of neural dynamics affects our ability to perceive a precise intensional dimension, a dimension we could attempt to formalise, at the level of modelistic reflection, only by recourse to specific intensional functions, to the determination of accessibility relations, and to the identifying of articulated systems of constraints and indices. Yet this intensional dimension, which is necessarily articulated according to types, different degrees of complexity, categories, etc., and which is clearly tied, as outlined above, to a specific interactive game of constraints, would not seem to be present at the level of monodimensional networks such as, for example Boolean neural networks of which Hopfield's model constitutes a particular version.

What this means is that, while we are able to grasp, for instance, affordances at the level of real information, we are actually unable to construct simulation models able to account for the specific modalities of this kind of assimilation. Not only are Koehler's field-theoretical models thus superseded, but the actual models based on the non-linear dynamic systems and recurrent networks no longer appear able to offer adequate explanation for what is the final reality of perception, its character of continuous emergence, its ability to go beyond the surface aspects of things to their *Sinn*. In other words, if we grant that the brain functions in the sense of perceiving, classifying, etc., in a connectionist way, it does so at a level of sophistication and complexity far superior to that of the Boolean neural networks we are familiar with. Hopfield's model, for example, allows propositional classification by means of attractors-concepts. What it lacks, however, is not only any intensional dimension, but even a precise polyadic dimension. The information that a simulation model of the kind can therefore assimilate from the surrounding reality cannot fail to be limited to particular and restricted aspects of surface information: hence the need to design new and more sophisticated models able to "reflect" more adequately the real processes taking place at the neural-dynamic level. But to effect this, the first step is above all to start again from the experimental evidence constituted by our perceptual activity. We need to manage to "see" its whole complexity, its aspects of obscurity, and the sometimes mysterious languages it uses, together with the multiplicity of its dimensions. This is why the many cues suggested by Kanizsa's careful analysis of the painstaking experiment-illusions that he discovered in the course of his life are still so invaluable.

As we have just underlined, the information the brain receives, processes, and assimilates is in proportion to the instruments of reading-generation and organisation, which the brain itself is able to produce. A number of considerations are here, I believe, compulsory.

The first, obvious one concerns the fact that the assimilation-process of external information implies the existence of specific forms of determination at the neural level. Information relative to the system stimulus is not, however, a simple amount of neutral sense-data to be ordered. It is linked to the "unfolding" of the selective action proper to the optical sieve, it articulates through the imposition of a whole web of constraints, possibly determining alternative channels at, for example, the level of internal trajectories. Depth information grafts itself on (and is triggered by) recurrent cycles of a self-organising activity characterised by the formation and the continuous *compositio* of multi-level attractors. The possibility of the development of new systems of pattern recognition, of new modules of reading will depend on the extent to which new successful "garlands" of the functional patterns presented by the optical sieve are established at the neural level in an adequate way.

If I manage to close each time the garland successfully, and imprison the thread of meaning, thereby harmonising with the ongoing "multiplication" of mental processes at the visual level, I posit myself as an adequate grid-instrument for the progressive and coherent "surfacing" of depth information and for its self-generating and unfolding as *Natura Naturata*, a Nature which the very units (monads) of multiplication will then be able to read and see as such (i.e. as a great book (library) of natural forms written in mathematical characters) through the eyes of mind (insofar as the monads in their turn posit themselves as constituent parts of the generative process in action).

In this sense vision is the end result of a construction realised in the conditions of experience. It is "direct" and organic in nature because the product of neither simple mental associations nor reversible reasoning, but, primarily, the "harmonic" and targeted articulation of specific attractors at different embedded levels.

The resulting texture is experienced at the conscious level by means of self-reflection, we really sense that it cannot be reduced to anything else, but is primary and self-constituting. We see visual objects; they have no independent existence in themselves but cannot be broken down into elementary data. Grasping the information at the visual level means managing to hear, as it were, inner speech. It means reconstructing in the negative, in an inner generative language, through progressive assimilation, selection and real metamorphosis (albeit partially and roughly) the articulation of the complex "genealogical" apparatus which works at the deep semantic level and moulds and subtends the presentation of the functional patterns at the level of the optical sieve. Vision as emergence aims first of all to grasp the paths and the modalities that determine the selective action, the modalities specifically relative to the revelation of the afore-mentioned apparatus at the surface level according to different and successive phases of generality. This revelation is triggered by precise coagulum functions; just as, on the other hand, the construction effected by the simulation model is, in its turn, guided by specific *exempla* of mental constructions, by self-reflection, by the presentation and determining of specific symmetry choices, and so on. The afore-mentioned paths and modalities thus manage to "speak" through my own fibres. It is exactly through a similar self-organising process, characterised by the presence of a double-selection mechanism, that the brain can partially manage to perceive depth information in an objective way. The extent to which the simulation model succeeds, albeit partially, in encapsulating the secret cipher of this articulation through a specific chain of programs determines the irruption of new creativity and consequently the model's ability to open the eyes of the mind and see a Nature teeming with processes.

To assimilate and see the system must first "think" internally the secret structures of the possible, and then posit itself as a channel (through the precise indication of forms of potential coagulum) for the process of opening and revelation of depth information. This process then works itself gradually into the system's fibres, *via* possible selection, according to the coagulum possibilities offered successively by the system itself.

The revelation and channelling procedures thus emerge as an essential and integrant part of a larger and coupled process of self-organisation. In connection with this process we can ascertain the successive edification of an I-subject conceived as a progressively wrought work of abstraction, unification, and emergence. The fixed points which manage to articulate themselves within this channel, at the level of the trajectories of neural dynamics, represent the real bases on which the "I" can reflect and progressively constitute itself. The I-subject can thus perceive to the extent in which the single visual perceptions are the end result of a coupled process which, through selection, finally leads the original Source to articulate and present itself, by means of cancellations and "irruptions", within (and through) the architectures of reflection, imagination and vision. These perceptions are (partially) veridical, direct, and irreducible. They exist not in themselves, but, on the contrary, for the "I", but simultaneously constitute the primary departure-point for every successive form of reasoning perpetrated by the observer. As an observer I shall thus witness *Natura Naturata* since I have connected functional forms in accordance with a successful and coherent score.

It is precisely through a coupled process of self-organisation of the kind that it will finally be possible to manage to define specific procedures of reconstruction and representation within the system, whereby the system will be able to identify a given object within its context, together with its *Sinn*. The system will thus be able to perceive the visual object as immersed within its surroundings, as a self-sustaining reality, and, at the same time, feel it living and acting objectively within its own fibres. In this way it will be possible for the brain to perceive depth information according to the truth (albeit partially).

It is, however, inevitable to ask how a cognitive system can adapt in a self-organising way to a world in flux, which articulates on depth levels which are constantly changing: in what way can a system of the kind manage to assimilate and adapt to such a fleeting reality, to the effective and secret matrix of its meanings? In what way can it manage to both see and simultaneously think its objects together with their meanings?

The role of the brain is above all to offer itself as a self-organising measuring device, as a self-organising, biological measure space. This device articulates progressively through a manifold of processing stages characterised by patterns of continuous interaction and integration. At the level of the brain, the computation unit is not furnished by a single processing stage but by a minimal set of processing streams. As we have just said, the brain aims first of all to constitute itself as a grid capable of partially reconstructing in its interior the meaningful unity (the irradiating and unifying warp) living at the level of the semantic dimension by means of an adequate texture of self-organising programs.

This can be understood if we start from a number of simple considerations. The visual process, as stated above, occurs within a coupled system equipped with self-reflection, in which a precise distinction obtains between vision and thought, although they maintain a constant and indissoluble functional exchange. A system of the kind subsumes the articulation of a series of specific processes: a process of simulation, a process of mirroring, a process of assimilation, a process of "irruption", a process of intentional observation, etc. It also subsumes the successive outlining of functions which self-regulate, as well as the progressive construction of increasingly forms of real autonomy. That function which self-organises with its meaning, and which posits itself as emergent, is "experienced" as vision insofar as it manages to establish itself, at the network level, as a specific modulation and integration of biological circuits capable of realising a partial engraving of the original *Sinn* (of the deep process of unification articulating at the level of the system-body of meaning). The resulting picture is of a world characterised by continuous emergence and by a constant composition and restructuring of schemes. This composition works at the horizontal and vertical level with a functional and constant internal "thickening" of the processing streams involved in accordance with a precise *bricolage*.

In this sense, vision extends within a coupled system characterised by the presence of a double selection: external and internal, the latter regarding the universe of meaning (this is, actually, a point of fact we are now ready to examine in the light of current achievements in contemporary theoretical Biology). Within the process, meaning reveals itself (albeit partially) in (and through) the effected emergence. Only in this way can a real assimilation process articulate, on the basis first of all of a coherent construction of possible schemes, falsification acts, and so on. This process can then gradually recognise itself in the realised emergence as an act of vision concerning the emergence itself.

In self-organising emergence, then, we find, simultaneously, a process of assimilation, one of growth, and one of stabilisation and reduction through fixed points. It is therefore not surprising that, as soon as the assimilation (and the unfolding by unification) of meaning occurs correctly, vision appears veridical. What this particularly presupposes as an essential component of the process is also the articulated presence of definite capacities of self-reflection and precise simulation-mechanisms at the level of thought. If actually it is obvious that no thought can exist which has not first filtered through the senses, it is equally clear that there can be no effective vision at the level of simulation model, unless specific elaboration has taken place able to "guide" the activity of internal selection. The outline (the inner thread) offered by the model serves to propose possible integration schemes able to support and prime the nesting proper to the internal selection. Then, at the moment of irruption, new vision emerges, and the thread as independent instrument is abandoned because superseded. In this sense it is true that at the level of the eyes of mind we have visual cognition, and not intellectual reading. Function and meaning articulate together, but in accordance with the development of a process of *adequatio*, and not of autonomous and direct creation. I will be unable to think of vision during emergence, but will be able to use it, once realised, to construct further simulation models. Growth, modulation, and successive integration thus exist 'within and among' the channels together with specific differentiation processes.

It is far from easy to determine mathematics for processes of the kind, since it is clearly impossible to restrict the processes of self-reflection and assimilation totally within the limits of a mechanistic reductionism. Actually, internal and external selection are based on principles and on choices which are articulated on a deep, creative level. Insofar as these principles and choices enter the scene, for example, at the second order level, they cannot be previously determined at the first order level; they are produced by the ongoing dialectic, by the symbolic dynamics in action and are revealed in emergence, i.e. when they really constitute myself as the subject which sees and thinks. As for self-reflection, the space occupied by these choices, too, cannot be reductively determined: yet the thread must be untangled and the space explored. The mind has to function as a bridge between internal and external selection. This is the *Via*-Method, relying on the continuous invention of new mathematics, new geometry, new formal axioms, etc. Hence the importance of the eye of the phenomenologist, and in particular of the perceptologist, s/he who listens to the channels, and hence, at the same time, the importance of the eye of the mathematician, s/he who explores the thread of simulation in the regions of pure abstraction. Amodal completion in this context emerges as a privileged window opened on a microcosm which is largely articulated according to the fibres proper to the architecture of mind. Objects are identified through the qualities elaborated and calculated along and through the channels. I neither colonise nor occupy, to use Freeman's words: I offer myself as a gridiron and I am selected. What remains on my flesh, the operative selection, is the inscription by means of cancellations and negative engravings of the deep functional patterns according to which the real processes "pulsate".

The simulation model thus constructed permits a more coherent integration and articulation of the channels, laying the foundation for the self-organised synthesis of ever-new neural circuits. Objects, in their quality of being immersed in the real world, then emerge as related to other objects possessing different features, and so on. Through and beyond these interrelations, holistic properties and dimensions then gradually reveal themselves, which I must grasp in order to see the objects with their meaning, if I am to understand the meaning of things. Apples exist not in isolation, but as objects on a table, on a tree: they are, for instance, in Quine's words, "immersed in red", a reality I can only grasp by means of a complicated second-order process of analysis, elaboration, and comparison which can thereafter be reduced, through concatenations of horizontal and vertical constraints, specific rules and the successive determination of precise fixed points, to the first-order level. I thus need constant integration of channels and formal instruments to grasp information of the kind, i.e. to assimilate structural and holistic relations and relative ties in an adequate way.

In other words, I will understand the meaning of things only if I am able to give the correct coagulum recipes with a view to being selected so as to grasp and capture not only the superficial aspects of objects in the world, but their mutual relations as they interact in depth, in obedience, for instance, to a specific intensional dimension. Only if I provide the correct coagulum, and select the right languages, will these relations emerge through the trigger operated by the "creative" procedures proper to depth information. Information about the outside world and the "genealogical" apparatus "feeding" it is thus extended: hence the need for an "internal" guide to the growth of the mechanisms of vision, the need for a "simulation thread": Ariadne's thread, primarily. The eyes of simulation allow the eyes of mind to open: herein we can recognise the progressive opening of the eyes of the Minotaur led by the hand through the process of metamorphosis. Ariadne is a lesson in how to think: how to order and unify visual thoughts and functional patterns in order to see, while the Minotaur represents the far-flung multiplicity of channels, the pure creativity in action. Selected and guided by Ariadne, and beginning mentally to perceive her, he becomes aware of a new process of self-organisation articulating within his channels. Besides thinking of Ariadne, he will also be able to see the external world to the extent to which he himself has been selected by it (and he will also be able to see himself as a part of the world-Nature). The self-organisation of the channels coincides with the successive stages in his metamorphosis, with his own gradual cognitive development and with his very achievement of a form of effective, intellectual autonomy. It is exactly within the secret paths of this process of metamorphosis that we can ascertain the objective articulation at the deep level of the specific procedures proper to knowledge construction.

* * *

Merleau Ponty, as is well known, is in line with Brentano and Koffka in considering the phenomenal *Umwelt* as 'already there', perception consisting precisely in detaching (*dégager*) the nucleus of this 'already there'. The distinctive nature of Gestalt is not as something alive in itself, independently of the subject which has to "insert" into it its relationship with the world; nor, however, is it constructed by the subject. It is not absolute, since experience shows that it can be made to vary, yet nor is it purely related to the Self, since it provides an *Umwelt* which is objective (transcendent). In this sense perception does not constitute a simple act of synthesis.

According to this viewpoint, Quine too considers, for instance, predication as something more than mere conjunction (the mere synthesis, Brentano would have said, of a subject-notion and a predicate-notion), not least since it ultimately coincides with an act of perception[4]. When we say 'the apples are red', this for Quine means that the apples are immersed in red. Predication indeed finds its basis on a far more complex act than simple conjunction- composition.

It should, however, be underlined that when in his later work Quine gives an illustration of the kind, he is quite consciously and carefully re-examining not only some of Brentano's original ideas on the thematics of perception, but also a number of basic assumptions behind Husserl's idea of relations between perception and thought. Can colour be grasped, Husserl asked, independently of the surface supporting it? Quite clearly not: it is impossible to separate colour from space. If we allow our imagination to vary the object-colour and we try to annul the predicate relative to the extension, we inevitably annul even the possibility of object-colour in itself, and reach an awareness of impossibility. This is what essence reveals: and it is precisely the procedure of variation which introduces us to the perception of essence. The object's *eidos* is constituted by the invariant which remains unchanged throughout all the variations effected.

In Husserl's opinion, together with perception it is necessary to conceive of acts based on sensory perceptions in parallel with the movements of categorial foundation taking place at the intentional level[5]. These acts offer a precise "fulfilment" to the complex meanings which for us constitute the effective guides to perception. When I observe gold, I see not yellow on the one hand and gold on the other, but 'gold-is-yellow'. 'Gold is yellow' constitutes a fact of perception, i.e. of intuition. The copula, the categorial form par excellence, cannot in itself be ' fulfilled': yet in the perception of the fact that 'gold is yellow', the copula too is a given. The sentence is filled up in its entirety simultaneously with its formation at the categorial level. It is in this sense that intuition itself takes on a form. Categorial intuition, as opposed to sensory or sensible intuition, is simply the evidencing of this formal fact, which characterises any possible intuition. I do not see-perceive primary visions and their link: I see immersion, Quine would say: I see the whole, and perceive an act of realized synthesis. This, Vailati would add, is the sense in which meanings function as the tracks guiding all possible perception. A categorial form, then, does not exist in and for itself, but is revealed and developed through its embodiments, through the concrete forms showing its necessity, and which unfold it according to specific programs that constitute, simultaneously, themselves as program-performers. It is thus meaning which has the power to produce forms, this constituting the intuition according to its categorial nature.

Category cannot be reduced to grammar because it is not outside the object. According to Husserl, we need to conceive of a type of grammar which is immanent to language, which must necessarily be the grammar of thought, of a thought which reveals itself as language in action, a language that, in its turn, constitutes itself as the Word of reality, like a linguistic corpus, i.e. a construction articulated linguistically, according to precise grammatical and semantic patterns, which gradually becomes reality. In contemporary terms, we could say that Husserl's language in action is characterised by the fact that the origin of meaning within the organisation of the complex system is merely an emergent property. What is self-organising is the function together with its meaning, and it is in this sense that, as stated above, meaning for Husserl is able to produce form-functions.

It thus becomes clear how for Husserl form, or articulation, can be considered as precisely that, and can only be constituted as object through a formalising abstraction. Hence the birth of a very specific intuition which can only be the result of a founding act. It is in this founding activity that the ultimate sense of categorial objectivity lies: this is the case, for instance, of mathematical evidence, which relates to the existence of a structure only insofar as it is accessed by an ordered series of operations.

Thus the actual reality of an object is not given by its immediate appearance, but by its foundation, it shows itself as something constituted through a precise act. The innate meaning of an object is that of being itself within an act of intuition. There is a moment, for example, when a circle ceases to be a circle by means of a variation procedure: it is this moment which marks the limits of its essence. Being itself identifies the very idea of intuition. To have an intuition of a sensible or abstract object means possessing it just as it is, within its self-identity, which remains stable in the presence of specific variations at both a real and possible level. The realm of intuition, in this sense, is the realm of possible fulfilments. To have intuition of an object means having it just as it is ['the thing itself']: breaking down the limits of the constraints distinguishing its quiddity. To grasp by intuition, for example, a complex mathematical object means possessing it as itself, according to an identity which remains unaltered through all real or possible variations. An object is a fixed point within a chain-operation, and only through this chain can its meaning reveal itself.

It should be born in mind, however, that a categorial form can only be filled by an act of intuition which is itself categorised, since intuition is not an inert element. In this sense complex propositions can also be fulfilled, and indeed every aspect in a complete proposition is fulfilled. It is precisely the proposition, in all its complexity, which expresses our act of perception. A correspondence thus exists between the operations of categorial foundation and the founded intuitions. To each act of categorial intuition a purely significant act will correspond. Where there exists a categorial form which becomes the object of intuition, perceived on an object, the object is presented to our eyes according to a new "way": precisely the way related to the form: we see the table and the chair, but we can also see in the background of this perception the connection existing between these two different things, which makes them part of a unique whole. The analysis of the real nature of categorial intuition thus leads Husserl, almost by the hand, to the question of holism. But it is of course this question – the sum of the problems posed by the relationship between thought and its object – which, as we know, constitutes one of Husserl's basic points of affinity with his mentor Brentano. Brentano's slant on questions of this kind pointed the way for Husserl in his own analysis, and proved a blueprint for the development of the different stages of his research. For Brentano consciousness is always consciousness of something, and inextricably linked to the intentional reference. At the eidetic level this means that every object in general is an object for a consciousness. It is thus necessary to describe the way in which we obtain knowledge of the object, and how the object becomes an object for us.

While, then, the question of essences seemed initially to be taking Husserl towards the development of a rigorously logical science, a *mathesis universalis*, the question of intentionality then obliges him to analyse the meaning, for the subject, of the concepts used at the level of logical science. An eidetic knowledge had to be radically founded. Husserl thus proceeds along a path already partially mapped out, albeit in some cases only tentatively, by Brentano, gradually tracing an in-depth analysis of the concept of completeness before arriving at a new, more complex concept, that of organicity. Kant's category of totality, Brentano's unity of perception and judgement, and experimental research in field of Gestalt theory thus come together, at least in part, in a synthesis which is new. But further analysis of the concept of the organicity then suggests other areas of thematic investigation, in particular at the level of *Experience and Judgement*, principally regarding the concepts of *substratum* and dependence. At the end of his research trajectory, then, Husserl returns to the old Brentanian themes concerning the nature of the judgement, offering new keys of interpretation for the existential propositions of which Brentano had so clearly perceived the first essential theoretical contours.

It is this area of Husserl's thought which interested Quine and Putnam in the '70s and '80s, and Petitot in the '90s. New sciences and conceptual relations enter the arena: e.g. the relationship between Logic and Topology, and, simultaneously, between perceptual forms and topological forms, etc.. It is these main forces which shape the continuing relevance and originality of the line of thought, the secret link, as it were, between Brentano, its originator, and the two main streams represented by Husserl's logical analyses on the one hand, and the experimental research of the Gestalt theoreticians on the other.

Quine's and Putnam's revisitation of the Brentano-Husserl analysis of the relation between perception and judgement was of considerable importance in the development of contemporary philosophy. It was no isolated revisitation, however, Husserl's conception of perception-thought relations constituting a source of inspiration for many other thought-syntheses. Recent years, in particular, have witnessed another rediscovery of the phenomenological tradition of equal importance: that linked to the philosophical and "metaphysical" meditations of the great contemporary logician K. Goedel. Its importance at the present moment is perhaps even more emblematic, in comparison, for instance, with Quine's and Putnam's rediscoveries, with respect to today's revisitation of Husserlian conception. For many aspects, Goedel's rereading constitutes a particularly suitable key to pick the lock, as it were, of the innermost rooms containing Husserl's conception of the relations between perception and thought.

The departure point of Goedel's analysis is Husserl's distinction between sensory intuition and categorial intuition[6]. Goedel, however, speaks in terms not of categorial intuition but of rational perception, and it is precisely this type of perception which allows for a contact with concepts, and through which we reach mathematical awareness. In his opinion, the conceptual content of mathematical propositions has an objective character. The concepts constitute an objective reality which we can only perceive and describe, but not create. In this sense this form of rational perception is in some ways comparable with sense-perception. In both cases, according to Goedel, we come up against very precise limits, possible illusions, and a precise form of inexhaustibility.

A very clear example of this inexhaustibility is provided by the unlimited series of new arithmetic axioms that one could add to the given axioms on the basis of the incompleteness theorem: axioms which, in Goedel's opinion, are extremely self-explanatory in that they elucidate only the general content of the concept of set. Goedel 's comment on this in 1964 is as follows: 'We possess something like a perception of the objects of set theory'. This perception is a sort of mathematical intuition: a rational perception. But how can the intuition of essence be reached? How is it possible to extend awareness of abstract concepts? Or to understand the relations interconnecting these concepts: i.e. the axioms which represent them?

None of this, in Goedel's opinion, can be done by introducing explicit definitions for concepts or specific demonstrations for axioms, which would necessarily require further abstract concepts and the axioms characterising them. The correct procedure is, conversely, to clarify the meaning, and this act of clarifying and distinguishing is, for Goedel, the central nucleus of the phenomenological method as delineated by Husserl. The theorems of incompleteness would seem, in effect, to suggest the existence of an intuition of mathematical essences (of a capacity in us to grasp abstract concepts), for which no reductionist explanation is possible. This kind of intuition is required above all for specific mathematical problems, for obtaining proofs of coherence for formal systems, etc.. The theorems thus demonstrate clearly that that particular essence constituting "mathematical truth" is something more than a purely syntactical or mechanical concept of provability, while guaranteeing full mathematical rigour.

A rigorous science, in other words, as Husserl maintained, is more than a purely formal science. It also requires a transcendent aspect, and it is at this level that new mathematical axioms gradually come to light, arising not only from formal and deductive procedures. The unlimited series of new arithmetical axioms which present themselves in the form of Goedel's sentences, and which can be added to the already-existing axioms on the basis of the theorems of incompleteness, constitutes, in particular, a classic example of this process of successive revelation-constitution. These new axioms clearly represent precise evidence which is not extrapolable from preceding axioms *via* mere formal deduction. They can thus be used in order to solve previously-undecidable problems. According to Goedel, this is a clearly-defined way of explaining our intuition of an essence. An even more interesting example is provided by the Paris-Harrington theorem, a genuinely-mathematical statement referring only to natural numbers which, however, remains undecidable at the PA level. Its proof requires the use of infinite sets of natural numbers, the theorem providing a sound example of Goedel's concept of the need to ascend to increasingly more elevated levels of complexity to solve lower-level problems. In a number of works between 1951 and 1956, Goedel returns to one of his favourite examples: the unlimited series of axioms of infinity in Set theory. These are not immediately evident, only becoming so in the course of the development of the mathematical construction. To understand the first transfinite axioms it is first necessary to develop the set theory to a very specific level, after which it is possible to proceed to a higher stage of awareness in which it will be possible to "see" the following axiom, and so on.

In Goedel's opinion, this is a very impressive example of the procedure of meaning clarification (as well as of the process of rational perception) Husserl had posited. It is precisely by utilising our intuition of essence as related to the concept of a "set" that set-theoretic problems in general can be solved. It is also necessary, he goes on, for us constantly to recognise new axioms logically independent of those previously established in order to solve all mathematical-level problems, even within a very limited domain. One case in point is the possible solution of the Continuum problem.

Here Goedel states explicitly that the theorems of incompleteness demonstrate how mental procedures can prove to be substantially more powerful than mechanical ones, since the procedures they use are finite but not mechanical, and able to utilise the meaning of terms. This is exactly what happens in the case of the intuition of mathematical essences. In offering us the possibility of understanding the nature of this process of categorial intuition, Phenomenology allows us to avoid both the dangers of Idealism, with its risk of an inevitable drift towards a new metaphysics, and Neopositivism's instant rejection of all possible forms of metaphysics.

While the theorems of incompleteness are not derivable from the doctrine of Phenomenology, they offer a better focus on the irreducible nature of mathematical essences, not least, for instance, through the clarification offered by the concept of 'mechanically-computable function' as analysed by Turing. Goedel thus finds hidden truths within an epistemological perspective that many may have considered outdated and obsolete. His conceptual instruments, however, belong to an analytical tradition which is not that of Phenomenology. According to the great Austrian logician, Phenomenology is, basically, a method of research, it consists of a manifold of procedures of meaning clarification and these procedures appear indissolubly connected to specific patterns of selective activities.

* * *

As we have just said, from a general point of view the visual system operates, first of all, at the simulation level, a reconstruction-reading of the functional patterns inscribed in the optical sieve, i.e. of the information articulating at the level of the visual system's interface in accordance with the action of a very precise matrix-sieve.

This can come about above all in proportion to the model-system's ability to construct a multiplicity of discrete replicas of these patterns within its own developing neural circuits, through an adequate texture of self-organising programs. This texture seeks to identify each time a garland of fixed points within which the thread of meaning may be brought back to life transposedly as a shadow of the file, i.e. of the self-contained outlining of the simulation programs effected. The model's ability to see, i.e. to observe with the eyes of the mind, is then in proportion to the simulation model's success in (re)constructing the garland adequately, thus managing to grasp the organic unity of the thread of meaning. The extent to which, *via* the guide offered by Ariadne, the model of simulation succeeds in grasping this unity determines its capacity to see-observe. What it will particularly observe is both *Natura Naturata* teeming with processes, and itself within it.

By means of extractions, inscriptions, replicas and so on, the visual system constitutes itself as a simulation model in order above all to allow Reality to penetrate, select, and express itself more deeply. In other words, it allows for the realisation of a more refined process of canalisation with the consequent surfacing of, on the one hand, a "thinking I" (of a manifold of creative fluxes-principles that manage to express and reflect themselves by irrupting into the system), and on the other, the related emergence of a clearly articulated universe of meanings. Through co-ordinated and coupled "packages" of selective acts and informational modules, external Reality thus selects and feeds a neural network-system with immense development and connection potentialities, able in particular to partially reflect and internally reconstruct the complex and holistic body of meaning which supports the ongoing selection in accordance with the progressive surface modulation of specific self-organising programs. Insofar as the network articulates as a simulation model, it appears as the result of a continuous metamorphosis.

The realisation of a simulation model of the kind thus cannot be simply seen as a coded description of a given system according to von Neumann's theory of self-reproducing automata, i.e. as a coded description that the processing systems uses as a sort of working model of itself which, at the same time, need not be decoded.

On the contrary, this inventive process should be considered, first of all, as an inner guide for the derepression processes that live inside the coupled system. One must not add new rules to the model-system from outside, but make it as an autopoietic model-system able to interact with the dynamics that it tries to describe in order to outline the birth of new patterns of integration able to guide, in a mediate way, a "creative" development of the deep basis which subtends such dynamics. To the extent that the model will succeed in acting as an indirect tool for the realisation of such a development, it will reveal itself as truly autonomous. This fact will allow the model to elaborate an objective representation of itself, a form of simulation of simulation. In this way, it will also allow the actual realisation of a form of (partial) creativity on behalf of the model-system which results in being life and intelligence: intelligence considered as an ideal Form in action, as the recognition-articulation of a coded "cipher" on which the self-organising processes, living at the deep levels of the Source, can finally graft, in an indirect way, a coherent expression of their potentialities.

In this way, the real mirror represented by the design of an adequate simulation model will emerge as integrant part of a larger process of natural evolution. From a general point of view, a natural creative dimension cannot develop all its potentialities, once a certain level of complexity has been reached, without the exploitation of the inner growth of the mirror represented by the simulation model. To be a real mirror and an effective canalisation factor at the cognitive level, however, this model should show itself, as we have said, to be increasingly autonomous, closer and closer to the complexity of the biological roots of the semantic dimension and increasingly creative, albeit within the effective limits of a coupled system.

The autonomy achieved by the model, within the aforesaid limits, guarantees that the neural growth partially determined, at the human level, by the integration activity expressed by the system has been obtained according to objective modules.

This activity which canalises the generative principles lying in the deep semantic dimension manifests itself (indirectly) and is achieved through the formation of attractors (or of sets of attractors) which constitute the fixed points of the dynamics taking place, of the occurring epigenetic growth. So the results of this activity can be seen, at the same time, as an objective representation of the creativity and the partial autonomy reached by the model and as a reliable evidence of the realised unfolding of the potentialities hidden in the natural dimension.

The model, in this way, must show itself to be one that is unable to survive away from the paths of the natural dimension, and as one that can become an integrant part of its evolution in accordance with precise biological constraints. So, once the model has been separated from the creative flux existing at the natural level, it will crystallise and become a simple automatism. If, on the contrary, the mirror-model is connected to the creative deep dimension in a continuous way, it might be able to grow with it and continually renew its own roots according to the successive steps of the realised creative development.

In this sense, the model can reveal itself as a form of "artificial" but living intelligence and, at the same time, as a form of "true" cognition . Actually, depth information directly concerns both internal selection and external selection and it is precisely at the fusion of these two forms of selection that a coupled system can take place.

With respect to this frame of reference, Reality presents itself (as regards the internal selection) as a set of becoming processes characterised by the presence-irradiation of a specific body of meaning and by an inner creativity having an original character. These processes then gradually articulate through and in a (partially-consistent) unifying development warp with internal fluctuations of functional patterns. It is this functional, self-organising and "irradiating" warp, in the conditions of "fragmentation" in which it appears and is reflected at the interface level, that the network progressively manages to reconstruct and replicate within itself as regards its specific functional aspects, ultimately translating and synthesising it into an operating architecture of programs. In this way it is then possible to identify a whole, complex "score" which will function as the basis for the reading-reconstruction of the above-mentioned functional warp. However, to read-identify-represent the score will necessarily require the contemporary discovery-hearing of the underlying harmony. Only the individual capable of representing and tuning the work as living harmony, and the score as silent object, will actually be able to depict him/herself as "I" and as subject. This individual will then not only be able to observe objects, but will itself be able to see the observing eye, modeling those objects. The I able to portray itself as such will be able to rediscover the roots of the very act of seeing, positing itself as awareness and as the instrument allowing the emergence of the "thinking I", and, conjointly, of the original meaning.

It is thus through the continuous metamorphosis of the network that new Nature can begin to speak, and Reality can appear and channel itself (as regards the external selection), in accordance with its deep dimension, ultimately surfacing and expressing as an activity of synthetic multiplication, i.e. as a form of operating creativity at the level of surface information, as a "thinking I" able to reflect itself in (and through) the work outlined by the simulation model.

The system inscribes within itself (in the space of imagination) functional forms through which it will be able to identify and project the reconstructed score and articulate a consonant dynamic system of images (of interconnected determinations of time) which, in the projection, will "dance", apparently autonomously, before the visual system and come to be viewed as objects-processes in their own right.

Nature is in this way replicated, initially, as a "silent" composition of objects, each of which is characterised by a specific functional form-role. For Kant, for example, seeing means ordering along the lines of an inner generativity which is inbuilt (and fixed at the formal level) in the conditions of experience: i.e. the conditions relative to the "presentation" of pure initial chaoticity. In this particular case the simple formalisation of the presentation permits the identification of the score.

It should, once again, immediately be underlined, however, that this is merely the first step in human visual cognition. The cognitive system does not limit itself to operating a replica and identifying specific rules on the basis, primarily, of inscription- and assimilation-procedures; what is recognised are not only objects in themselves but objects singly encapsulated within a complex observation system which refers to (and, in certain ways, reflects itself according, in the limit, to a form of identification) the same face-appearance of the emergent reconstruction operated by the observer (particularly in agreement with the developing procedures of the inscription and identification of the score). It is the face-texture of the effected reconstruction which provides the guidelines for the I's edification; and indeed the "thinking I" which gradually surfaces reflects itself in the constructed work (of simulation), thereby allowing the effective emergence of an "acting I" (of a "person") which reveals finally itself as an "added" creative component, i.e. we are actually faced with the very multiplication of the cognitive units. The system is thus able to see according to the truth insofar as it constitutes itself as an "I" and as consciousness, i.e. in proportion to the extent it can "see" (and portray-represent) its own eye which observes things.

In this sense vision is neither ordering, nor recognising, nor pure comparison, nor, in general, simple replica, but is above all a reading-reconstruction of the unity of the original body of meaning (with operating self-reflection): a process of progressive identification of this unity in terms of an adequate texture of self-organising programs able to portray itself as such, a process which becomes gradually autonomous and through which, *via* selection, in a renewed way and at the surface level, Reality can canalise the primary modules of its own complex creative tissue: i.e. surfacing as generativity and nesting as meaning. The better the reconstruction, the more adequate and consistent the canalisation. In this sense the system will function ever more sophisticatedly as a reflecting and self-organising filter. As a matter of fact, in parallel to this an acting “I” will progressively arise through the narration and the methodical verification of the distinctions relative to the functional forms managing to move at the level of the unitary and cohesive articulation of the self-organising programs. As narration and synthesis, the I posits itself as autonomous and as the increasingly adequate mirror of a precise “metamorphosis”: namely, the metamorphosis proper to an intelligent network which grows into autonomy. The mirror is image-filled at the moment of selection, when new emergence can simultaneously come about and “eyes” can then manage to open and see both things and their meaning.

The “thinking I” which surfaces and the meaning which emerges thus fuse in the expression of a work which ultimately manages to articulate and unite itself with the awareness-*Cogito* and the ongoing narration. A “creating I” thus joins a work acting as a filigree. The resulting path-*Via* can then allow real conjunction of both function and meaning. The result will be not merely simple generative principles, but self-organising forms in action, creativity in action, and real cognitive multiplication: not a simple gestaltic restructuring, but the growth and multiplication of cognitive processes and units, i.e. the actual regeneration and multiplication of original Source according to the truth.

The adequate work of unification-closure of network programs, which joins and encapsulates, at the level of the ongoing emergence and self-reflection, the selection internally operated by meaning according to the living warp-filigree, constitutes vision in action. In actual fact it comprises a multiplicity of interconnected works, to each of which is linked a consciousness. In this way the afore-mentioned unification necessarily concerns the continuous weaving of a unitary consciousness, albeit within the original fragmentation of the micro-consciousness and the divided self.

It is from this viewpoint that vision appears as necessarily related to a continuous emergence, in its turn connected primarily with the progressive articulation of a self-expressing I. As the system manages to see, it surfaces towards itself and can, then, identify and narrate itself as an “I”, and specifically as an “I” that sees and grasps the meaning of things: in particular the emergence related to the meaning that is concerned with them. At the moment the afore-mentioned work becomes vision (expressing itself in its completeness), it simultaneously reveals itself as a construction in action and at the same time as the filter and the lynch-pin of a new canalisation through which new Reality can reveal itself unfolding its deep creativity. Meaningful forms will then come into play, find reflection in a work, and be seen by an “I” that can thus construct itself and re-emerge, an “I” that can finally reveal itself as autonomous: real creativity in action.

I neither order nor regiment according to principles, nor even grasp principles, but posit myself as the instrument for their recovery and recreation, and reflect their sedimentation in my self-transformation and my self-proposing as *Cogito*. Actually I posit my work as the mirror for the new canalisation, in such a way that the new emergent work, if successful, can claim to be the work of an “I” which posits itself as an “added” creator. It is not the things themselves that I "see", then, but the true and new principles, i.e. the meaningful forms in action : the rules-functions linked to their emergent meanings. I thus base myself on the "word" which dictates.

The world thus perceived at the visual level is constituted not by objects or static forms, but by processes appearing imbued with meaning. As Kanizsa stated, at the visual level the line *per se* does not exist: only the line which enters, goes behind, divides, etc.: a line evolving according to a precise holistic context, in comparison with which function and meaning are indissolubly interlinked. The static line is in actual fact the result of a dynamic compensation of forces. Just as the meaning of words is connected with a universe of highly-dynamic functions and functional processes which operate syntheses, cancellations, integrations, etc. (a universe which can only be described in terms of symbolic dynamics), in the same way, at the level of vision, I must continuously unravel and construct *schemata*; must assimilate and make myself available for selection by the co-ordinated information penetrating from external reality. Lastly, I must interrelate all this with the internal selection mechanisms through a precise "journey" into the regions of intensionality.

In accordance with these intuitions we may directly consider, from the more general point of view of contemporary Self-organisation theory, the network of meaningful programs living at the level of neural systems as a complex one which forms, articulates, and develops, functionally, within a "coupled universe" characterised by the existence of a double selection. This network gradually posits itself as the basis for the emergence of meaning and the simultaneous, if indirect, surfacing of an "acting I": as the basic instrument, in other words, for the perception of real and meaningful processes, of "objects" possessing meaning, aims, intentions, etc.: above all, of objects possessing an inner plan and linked to the progressive expression of a specific cognitive action.

The "intelligent" network which develops with its meaning articulates as a growing neuronal network through which continuous restructuring processes are effected at a holistic level, thus constituting the indispensable basis of visual cognitive activity. The process is first of all, as stated above, one of canalisation and of revelation (according *in primis* to specific reflection procedures) of precise generative (and informational) fluxes-principles. It will necessarily articulate through *schemata* which will stabilise within circuits and flux determinations. As Grossberg states the brain's processing streams compute complementary properties according to hierarchical modules linked to the resolution of uncertainty: the brain is first of all a self-organising measuring device in the world and of the world[7].

The resulting global determination will present itself as something "perceived" insofar as it will reveal itself as linked to precise postulates of meaning, it will thus emerge as a scene (a scene for an I-subject), and the single processes of determination as meaningful observers or as objects, actions, etc. which populate the scene and which result as encapsulated in observation systems. The I-subject will recognise itself through the co-ordinated action of these observation systems; it will mirror itself in the "pupils" of these very systems to the extent that it will be recognised as the primary factor of their recovery as autonomous units.

Since the processes involved are dynamic, the objects will "pass behind", divide, produce compensations, fusions, etc.. The original scene is highly dynamic. When, therefore, the model-network posits itself as a "I-representation" (when the arc of simulation "reaches completion"), and views the world-Nature before it, it sees the world in consonance with the functional forms on which its realisation was based, i.e. according to the architecture proper to the circuits and the patterns of meaning which managed to become established. The result is Nature written in mathematical formulae: Nature read and seen *iuxta propria principia* by means of grammatical and symbolic forms and specific mathematical modules.

Nature is the very (original) opening of the process of determination. It presents itself as a dynamic system of meaningful processes in action; the "method" in its turn must offer real instruments in order to feed and coagulate the self-organising growth and the articulated unfolding of these very processes. On the other hand, Nature must also be considered as a body-system of meaning that cannot be occupied. Hence the possibility to consider Nature contemporarily as both "irruption" and emergence, i.e. as deep information that hides itself with the ever-new emergence of postulates of meaning (*Natura Naturans*); to this emergence will correspond the progressive "surfacing" of ever-new constraints and rules at the generative level[8].

Vision cannot, then, be considered as a simple copy or replica of Reality: it is a functional reconstruction linked to specific procedures of simulation, assimilation and "reduction", that determines a continuous possible emergence and a continuous sudden irruption at the holistic level. Vision is the very articulation of self-organising programs in accordance with the action of specific meaning postulates and with the emerging of an I-subject that thinks according to the form of the postulates and feels according to the co-ordinated action of the programs. Vision then can be considered as the very method of synthesis of the programs, a synthesis however that constitutes not a pattern of regimentation, but the patient construction of an I-person that posits itself as a real support for the progressive emergence of new autonomous cognitive units, of new coherent integrations of cognitive modules in accordance with the conditions of the real determination.

Vision is partially objective and veridical – veridical mainly since, through the effected selection and canalisation, it appears anchored to the revelation of the original creativity, to the actual unfolding and opening of the maximal determination. It does, indeed, seem able to partially unravel this creativity in accordance with its message, thus providing a coherent filter for the realisation of an adequate biological canalisation. It is namely veridical since there can only be objective vision (i.e. with the eyes of the "flesh") if the enacted simulation and inscription emerge as truth and posit themselves as the basis for new modules of generation.

Vision, in this sense, is the process of inscription, reconstruction, assimilation and reduction realised in the conditions of double selection in accordance with the truth. It appears necessarily moulded by the mathematical forms and modules which determine and shape it; in particular it articulates as a coupled pattern of emergence and irruption, thus finally constituting itself as the vision of an "I" which manages to establish its full autonomy. As unfolded I-*Cogito*, I see primarily developmental processes articulated tridimensionally and originally possessing meaning. Vision thus appears as the embodiment of the method relative to the process of canalisation of the generative fluxes- principles in action: if adequate, it constitutes the way in order to partially permit the real unfolding of the deep information content in accordance with different and successive levels of complexity: it articulates each time through the reading- individuation of that particular co-ordinated series of functional closures, i.e. that specific chain of fixed points that is necessary for the coherent unfolding and encapsulation of the *Sinn* according to its original creativity.

Hence one of Kanizsa's fundamental intuitions (in many aspects a concealed or incompletely expressed one): when specific programs are embodied in neural networks at the level of visual cognition they articulate according to grammars functionally correlated by means of the procedures of internal selection. These grammars then exactly appear as meaningful and self-organised. Hence the possibility to consider visual cognition as a self-organising process.

Thus to see more and better, I must construct ever more sophisticated grammars. But to do this I must operate according to the principles of an adequate dynamic semantics; I must open my "eyes" (the eyes of the mind) to non-standard grammars and "illusions", construct scheme-programs able to calculate processes, and finally manage to "see" (progressively but partially) the very genealogical formation-process of the original *Gestalten*. Only at that moment can I be selected significantly: i.e. attain new meaning. Only at that moment can I posit myself as the real basis for the surfacing of the "thinking I", of the creative fluxes-principles living at the deep level.

In his experiments, Kanizsa, for instance, gradually takes us into the secret architecture of non-standard grammars governing visual perception, inside the mysteries of inner selection, allowing us more incisive contact with deep information.

To grasp new meaning, construct new visions, and obtain the "growth" of perceptual activity, I need more detailed knowledge of the grammatical and semantic principles governing the articulation of the *Gestalten*. I thus need to develop new rule-schemes, new textures of rules which are more adequate with respect to the actual development of my visual cognition. This knowledge will not simply be a form of awareness, but will constitute the basis for a further opening up, and for the constructing of new forms of vision, the very "preparation" of new forms of possible canalisation as well as of a new expression of intellectual articulation of the "I". It was no accident that Kanizsa alternated periods of extremely innovative visual construction, as a painter, with periods of reflection and analysis, investigating possible models of visual activities and exploring the immense range of their potentialities.

Prof. Arturo Carsetti
Dipartimento di Ricerche Filosofiche
Università di Roma "Tor Vergata"
V. Lariana 7
00199, Roma, Italy

NOTES

[1] Cf.: Kanizsa, G., *Grammatica del vedere*, Bologna, 1980.

[2] Cf. in the first instance for more information: Atlan, H., "Self-organising Networks: Weak, Strong and Intentional, the Role of their Underdetermination", in A. Carsetti, (Ed.), *Functional Models of Cognition. Self-Organising Dynamics and Semantic Structures in Cognitive Systems*, Dordrecht, 1999, pp.127-139.

[3] Cf.: Freeman, W., *Neurodynamics: an Exploration of Mesoscopic Brain Dynamics*, London, 2000.

[4] Cf.: Quine, W.V., *Pursuit of Truth,* Cambridge Mass., 1990.

[5] Cf.: Husserl, E., *Erfahrung und Urteil*, Hamburg, 1964.

[6] Cf.: Goedel, K., "The modern development of the foundations of Mathematics in the light of Philosophy", in S. Feferman et al. (Eds.), *Kurt Goedel: Collected Works*, Oxford, 1986, 1990, 1995, pp. 374-387.

[7] Cf.: Grossberg, S.," Neural Models of Seeing and Thinking" this volume.

[8] Cf.: Carsetti, A., "Randomness, Information and Meaningful Complexity: Some Remarks About the Emergence of Biological Structures", *La Nuova Critica,* 36 (2000), pp. 47-109.

REFERENCES

AMIT, D.J., GUTFREUND, H. & SOMPOLINSKY, H., (1985) "Spin-glass models of neural networks", *Phys. Review Letters*, 55, pp. 1530-33.

ANDERSON, P.W., (1985) "Suggested model for prebiotic evolution: the use of Chaos", *Proc. Nat. Acad. Sci.,* USA, p. 3386.

ATLAN, H., (1992) "Self-organizing networks: weak, strong and intentional, the role of their underdetermination", *La Nuova Critica*, 19-20, pp. 51-71.

ATTNEAVE, F., (1982) "Prägnanz and soap bubble systems: a theoretical exploration" in J. Beck (Ed.), *Organization and Representation in Perception*, Hillsdale NJ.

BARLOW H., BLAKEMORE C. & PETTIGREW J., (1967) "The Neural Mechanism of Binocular Depth Discrimination", *Journal of Physiology,* 193, 327-342

BELL, J.L. & SLOMSON, A.B., (1989) *Models and ultraproducts*, Amsterdam.

BIRKHOFF, G., (1967), *Lattice theory*, 3 rd. Ed., Princeton.

BOOLOS, G., (1975) "On second-order logic", *Jour. of Philosophy*, 72, pp. 509-527.

BOOLOS, G. & JEFFREY, R., (1989) *Computability and Logic*, 3 rd. Ed., Cambridge.

CARNAP, R., (1956) *Meaning and Necessity*, 2 d. Ed., Chicago.

CARNAP, R. & BAR HILLEL, Y., (1950) "An Outline of a Theory of Semantic Information", *Tech. Rep*. N. 247, M.I.T.

CARNAP, R. & JEFFREY, R., (1971) *Studies in Inductive Logic and Probability*, Berkeley.

CARSETTI, A., (1993) "Meaning and complexity: the role of non-standard models", *La Nuova_Critica*, 22, pp. 57-86.

CARSETTI, A., (Ed.), (1999) *Functional Models of Cognition. Self-organizing Dynamics and Semantic Structures in Cognitive Systems*, Dordrecht.

CHAITIN, G., (1987) *Algorithmic Information Theory*, Cambridge.

CHAITIN, G. & CALUDE, C., (1999) "Mathematics/Randomness Everywhere", *Nature*, 400, pp. 319-20.

CHOMSKY, N., (1995) *The Minimalist Program*. Cambridge, MA, 1995.

CHURCH, A., (1956) *Introduction to mathematical logic*, Princeton.

CHURCHLAND, P.M., (1984) *Matter and Consciousness: A contemporary introduction to the philosophy of mind,* Cambridge, MA.

CORNSWEET, T.N., (1970) *Visual Perception*. New York, Academic Press.

CRICK, F. & KOCH C., (1990) "Toward a Neurobiological Theory of Consciousness", *Seminars in the Neurosciences,* 2, 26, 3-275.

DAVIDSON, D., (1970) *Mental Events*, Oxford.

DAVIES, P., (Ed.), (1998) *The New Physics*, London.

FAUCONNIER, G., (1985) *Mental spaces: Aspects of meaning constrution in natural language*, Cambridge, MA.

FEFERMAN, S. et al. (Eds.), (1986, 1990, 1995) *Kurt Gödel: Collected Works*, I, II, III, Oxford.

FODOR, J., (1987) *Psychosemantics: The problem of meaning in the philosophy of mind*, Cambridge, MA.

FREGE, G., (1892) "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und Philosophische Kritik,* 100, 25-50. English translation in P. Geach and M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege,* 56-78, Oxford, 1952.

GAIFMAN, H., (2000) "What Goedel's Incompleteness result does and does not show",*The Jour. of Philosophy,* 9708, pp. 462-70.

GIBSON J.J., (1972) "A Theory of Direct Visual Perception" in J.R. Royce & W.W. Rozeboom (Eds.), *The Psychology of Knowing,* London.

GIBSON J.J., (1979) *The Ecological Approach to Visual Perception*, Houghton Mifflin.

GOLDMAN, S., (1953) *Information Theory*, Engl. Cliffs.

GREGORY, R.L., (1963) "Distortion of Visual Space as Inappropriate Constancy Scaling", *Nature,* 199, 678-679.

GROSSBERG, S., (1987) "Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory", *Perception & Psychophysics,* 41, 117-58.

GROSSBERG, S., (1990) "Neural FAÇADE: Visual representations of static and moving Form-And-Color-And-Depth", (*Special Issue on Understanding Vision*), 411-456.

GROSSBERG, S., (1994) "3-D vision and figure-ground separation by visual cortex", *Perception & Psychophysics* 55, 48-120.

GROSSBERG, S., & MINGOLLA, E., (1985) "Neural Dynamics of Form Perception: Boundary Completion, Illusory Figures, and Neon Color Spreading", *Psychological Review,* 92 173-211.

HALMOS, P., (1963) *Lectures on Boolean Algebras*, Princeton.

HARRINGTON, L.A., MORLEY M.D., SCEDEROV A. and SIMPSON S.G., (Eds.), (1985) *Harvey Friedman Research on the Foundations of Mathematics*, Amsterdam.

HELMHOLTZ, H., (1925) *Physiological Optics*, Optical Society of America, 3, 318.

HOCHBERG, J. & BROOKS, V., (1960) "The Psychophysics of Form: Reversible Perspective Drawings of Spatial Objects", *American Journal of Psychology,* 73, 337-354.

HOFFMAN D.D., (1998) *Visual Intelligence: How We Create What We See*, New York.

HENKIN, L., (1950) "Completeness in the theory of types", *Jour. of Symb. Logic*, 15, pp. 81-91.

HENKIN, L., (1953) "Banishing the rule of substitution for functional variables", *Jour. of Symb. Logic,* 18, 3, pp. 201.8.

HERKEN, R., (Ed.), (1988) *The Universal Turing Machine. A Half Century Survey*, Oxford.

HINTIKKA, J., (1963) "Distributive Normal Forms in First-order Logic" in J.M. Crossley & M.A. Dummett, (Eds.), *Formal Systems and Recursive Functions* , Amsterdam, pp. 47-90.

HINTIKKA, J., (1969) *Models for Modalities*, Dordrecht.

HINTIKKA, J., (1970) "Surface information and depth information" in J. Hintikka & P. Suppes (Eds.) *Information and Inference*, Dordrecht, pp. 298-230.

HOPFIELD, J.J., (1982) "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *Proc. of the Nat. Ac. Scien.* 79, pp. 2254-2258.

JACKENDOFF, R., (1983) *Semantic Structures* , Cambridge.

KAUFFMAN, S.A., (1993) *The Origins of Order*, New York.

KOFFKA, K., (1935) *Principles of Gestalt Psychology*, New York.

KÖHLER, W. & HELD, R., (1947) "The Cortical Correlate of Pattern Vision", *Science*, 110: 414-419.

KÖHLER, W., (1969) *The Task of Gestalt Psychology,* Princeton NY.

KÖHLER,W., (1971) "A Task For Philosophers", in M. Henle (Ed.), *The Selected Papers of Wolfgang Koehler*, New York. pp 83-107.

KOHONEN, R., (1984) *Self-organization and Associative Memories*, Berlin.

KOLMOGOROV, N., (1968) "Logical Basis for Information Theory and Probability Theory", *IEEE Trans.IT* 14, 5, pp. 662-4.

LAKOFF, G., (1971) "On generative semantics", in D. Steinberg & L. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*, 232-296. New York.

LEINFELLNER, W., (2000) "The Role of Creativity and Randomizers in Societal Human Conflict and Problem Solving", *La Nuova Critica*, 36, pp. 1-27.

MARR, D., (1982) *Vision*, San Francisco, 1982.

MANZANO, M., (1996) *Extensions of First-order Logic*, Cambridge.

MONTAGUE, R., (1974) *Formal Philosophy*, Newhaven.

NICOLIS, G., (1989) "Physics of far-from-equilibrium systems and self-organization", in P., Davies, (Ed.), *The New Physics*, London.

NICOLIS, G. & PRIGOGINE, I., *Exploring Complexity*, New York.

O'REGAN K.J., (1992) "Solving the 'Real' Mysteries of Visual Perception: The World as an Outside Memory", *Canadian Journal of Psychology,* 461-488.

PALMER, S.E., (1992) “Modern theories of Gestalt perception”, in G.W. Humphreys (Ed.) *Understanding Vision*, London.

PALMER, S.E., (1999) “Color, Consciousness, and the Isomorphism Constraint”, *Behavioral and Brain Sciences,* 22 (6) 1-21.

PESSOA L., THOMPSON E. & NOË A., (1998) “Finding Out About Filling-In: A guide to perceptual completion for visual science and the philosophy of perception”, *Behavioral and Brain Sciences,* 21, 723-802.

POPPER, K., (1959) *Logic of Scientific Discovery*, London

PARIS J. and HARRINGTON L., (1977) “A Mathematical Incompleteness in Peano Arithmetics” in Barwise, J. (Ed.), *Hanbook of Mathematical Logic*, Amsterdam, pp. 1133-42.

PRIGOGINE, I., (1978) *From Being to Becoming*, New York.

PRIGOGINE, I., (1993) *Les lois du Chaos*, Paris.

PUTNAM, H., (1975) “The meaning of ‘meaning”, in K. Gunderson (Ed.), *Language, mind, and knowledge*, pp. 131-193, Minneapolis.

PUTNAM, H., (1965) “Trial and Error Predicate and the Solution to a Problem of Mostowski”, *Jour. of Sym. Logic*., 30, pp. 49-57.

RANTALA, V., (1975) “Urn Models a New Kind of Non-standard Model for First-order Logic”, *Jour. Philos. Logic*., 4, pp. 455-74.

ROGERS, R., (1971) *Mathematical Logic and Formalized Theories,* Amsterdam.

SAUSSURE, F. DE, (1915) *Cours de linguistique générale,* edited by C. Bally & A. Sechehaye. English translation.

SEARLE, J., (1980) “Minds, brains, and programs”, *Behavioral and Brain Sciences,* 3, 417-424.

SHAPIRO, S., (1987) “Principles of reflection and second-order logic”, *Jour. of Philos. Logic*., 16, pp. 309-33.

SHAPIRO, S., (1998) *Foundations without Foundationalism. A Case for Second-order Logic*, New York.

SHERRINGTON, D. & KIRKPATRICK, S., (1975) “Spin-glasses”, *Phys. Review Letters*, 35, p. 197

SOLOMONOFF, R., (1978) “Complexity -based Induction Systems: Comparisons and Convergence Theorem*”, IEEE Trans. On Inf.Theory*, 24, 422-432.

TALMY, L., (2000) *Toward a Cognitive Semantics*. Cambridge.

TICHY, P., (1971) “A New Approach to Intensional Analysis”, *Nous*, 5, pp. 275-98.

TIESZEN, R., (1994) “Mathematical Realism and Gödel’s Incompleteness Theorem”, *Philosophia Mathematica* , 3, pp. 177-201.

VAN BENTHEM, J. & DOETS, K., (1983) “Higher-order logic” in Gabbay, D.I. & Guenthner, F. (Eds), *Handbook of philosophical logic I*, Dordrecht, pp. 275-329.

VAN FRAASSEN, BAS C., (1971) *Formal Semantics and Logic*, London.

VAN LAMBALGEN, M., (1989) “Algorithmic Information Theory”, *Jour. of Symb. Logic*, 54, pp. 1389-1400

WANG, H., (1974) *From Mathematics to Philosophy*, New York.

NAME INDEX

This page intentionally left blank

SUBJECT INDEX

This page intentionally left blank

- C -

- U -

V

W

Z

THEORY AND DECISION LIBRARY

SERIES A: PHILOSOPHY AND METHODOLOGY OF THE SOCIAL SCIENCES

Editors: W. Leinfellner (*Vienna*) and G. Eberlein (*Munich*)

1. G. Zecha and P. Weingartner (eds.): *Conscience: An Interdisciplinary View*. Salzburg Colloquium on Ethics in the Sciences and Humanities (1984). 1987 ISBN 90-277-2452-0
2. R.W. Scholz: *Cognitive Strategies in Stochastic Thinking*. 1987 ISBN 90-277-2454-7
3. H. Nurmi: *Comparing Voting Systems*. 1987 ISBN 90-277-2600-0
4. M. Schmid and F.M. Wuketits (eds.): *Evolutionary Theory in Social Science*. 1987 ISBN 90-277-2612-4
5. C.L. Sheng: *A New Approach to Utilitarianism*. A Unified Utilitarian Theory and Its Application to Distributive Justice. 1991 ISBN 0-7923-1301-1
6. F. Forman: *The Metaphysics of Liberty*. 1989 ISBN 0-7923-0080-7
7. G. Bernard: *Principia Economica*. 1989 ISBN 0-7923-0186-2
8. J.W. Sutherland: *Towards a Strategic Management and Decision Technology*. Modern Approaches to Organizational Planning and Positioning. 1989 ISBN 0-7923-0245-1
9. C. Vlek and G. Cvetkovich (eds.): *Social Decision Methodology for Technological Projects*. 1989 ISBN 0-7923-0371-7
10. P. Hoyningen-Huene and F.M. Wuketits (eds.): *Reductionism and Systems Theory in the Life Sciences*. Some Problems and Perspectives. 1989 ISBN 0-7923-0375-X
11. K.G. Grunert and F. Ölander (eds.): *Understanding Economic Behaviour*. 1989 ISBN 0-7923-0482-9
12. G. Antonides: *The Lifetime of a Durable Good*. An Economic Psychological Approach. 1990 ISBN 0-7923-0574-4
13. G.M. von Furstenberg (ed.): *Acting under Uncertainty*. Multidisciplinary Conceptions. 1990 ISBN 0-7923-9063-6
14. R.L. Dukes: *Worlds Apart*. Collective Action in Simulated Agrarian and Industrial Societies. 1990 ISBN 0-7923-0620-1
15. S.O. Funtowicz and J.R. Ravetz: *Uncertainty and Quality in Science for Policy*. 1990 ISBN 0-7923-0799-2
16. J. Götschl (ed.): *Erwin Schrödinger's World View*. The Dynamics of Knowledge and Reality. 1992 ISBN 0-7923-1694-0
17. R. von Schomberg (ed.): *Science, Politics and Morality*. Scientific Uncertainty and Decision Making. 1993 ISBN 0-7923-1997-4
18. J. Baron: *Morality and Rational Choice*. 1993 ISBN 0-7923-2276-2
19. M. Allais and O. Hagen (eds.): *Cardinalism*. A Fundamental Approach. 1994 ISBN 0-7923-2398-X
20. E. Carlson: *Consequentialism Reconsidered*. 1995 ISBN 0-7923-3547-3
21. J. Götschl (ed.): *Revolutionary Changes in Understanding Man and Society*. 1995 ISBN 0-7923-3627-5
22. A. Camacho: *Division of Labor, Variability, Coordination, and the Theory of Firms and Markets*. 1996 ISBN 0-7923-3832-4
23. R. Hegselmann, U. Mueller and K.G. Troitzsch (eds.): *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. 1996 ISBN 0-7923-4125-2

24. J. Nida-Rümelin: *Economic Rationality and Practical Reason*. 1997 ISBN 0-7923-4493-6
25. G. Barbiroli: *The Dynamics of Technology*. A Methodological Framework for Techno-Economic Analyses. 1997 ISBN 0-7923-4756-0
26. O. Weinberger: *Alternative Action Theory*. Simultaneously a Critique of Georg Henrik von Wright's Practical Philosophy. 1998 ISBN 0-7923-5184-3
27. A. Carsetti (ed.): *Functional Models of Cognition*. Self-Organizing Dynamics and Semantic Structures in Cognitive Systems. 2000 ISBN 0-7923-6072-9
28. J. Nida-Rümelin and W. Spohn (eds.): *Rationality, Rules, and Structure*. 2000 ISBN 0-7923-6326-4
29. J. Klüver: *The Dynamics and Evolution of Social Systems*. New Foundations of a Mathematical Sociology. 2000 ISBN 0-7923-6443-0
30. V.A. Lefebvre: *Algebra of Conscience*. Revised Edition with a Second Part with a new Foreword by Anatol Rapoport. 2001 ISBN 0-7923-7121-6
31. J. Götschl (ed.): *Evolution and Progress in Democracies*. Towards New Foundations of a Knowledge Society. 2001 ISBN 1-4020-0063-4
32. N.J. Saam and B. Schmidt (eds.): *Cooperative Agents*. Applications in the Social Sciences. 2001 ISBN 1-4020-0190-8
33. B. Verbeek: *Instrumental Rationality and Moral Philosophy*. An Essay on the Virtues of Cooperation. 2002 ISBN 1-4020-0639-X
34. J. Klüver: *An Essay Concerning Sociocultural Evolution*. Theoretical Principles and Mathematical Models. 2002 ISBN 1-4020-0750-7
35. T.J. Fararo and K. Kosaka: *Generating Images of Stratification*. A Formal Theory. 2003 ISBN 1-4020-1500-3
36. O. Dahlbäck: *Analyzing Rational Crime - Models and Methods*. 2003 ISBN 1-4020-1657-3
37. E. Oeser: *Evolution and Constitution*. The Evolutionary Selfconstruction of Law. 2003 ISBN 1-4020-1784-7